Multivariate Distributions

CIVL 7012/8012

Multivariate Distributions

- Engineers often are interested in more than one measurement from a single item.

- Multivariate distributions describe the probability of events defined in terms of multiple random variables

Joint Probability Distributions

- Some random variables are not independent of each other, i.e., they tend to be related.
  - Urban atmospheric ozone and airborne particulate matter tend to vary together.
  - Urban vehicle speeds and fuel consumption rates tend to vary inversely.

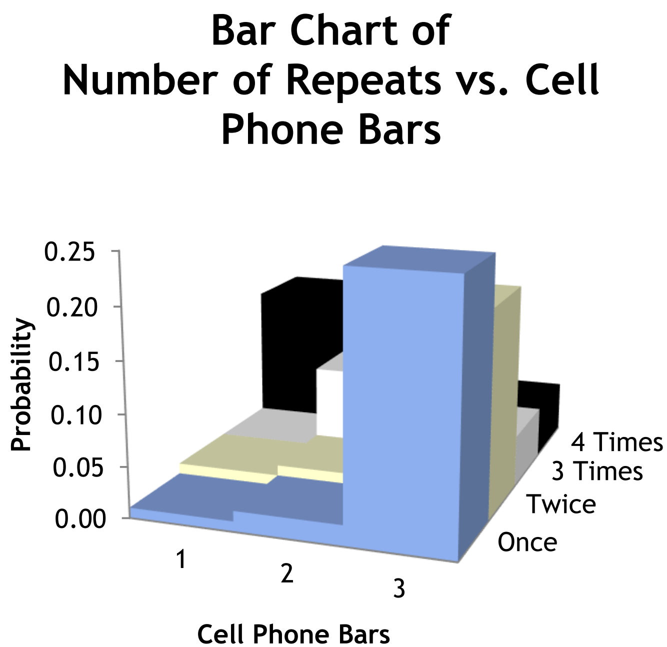
- A joint probability distribution will describe the behavior of several random variables, say, X and Y. The graph of the distribution is 3-dimensional: x, y, and $f(x,y)$.

You use your cell phone to check your airline reservation. The airline system requires that you speak the name of your departure city to the voice recognition system.

- Let Y denote the number of times that you have to state your departure city.
- Let X denote the number of bars of signal strength on you cell phone.

y = number of times city name is stated	x = number of bars of signal strength		
	1	2	3
1	0.01	0.02	0.25
2	0.02	0.03	0.20
3	0.02	0.10	0.05
4	0.15	0.10	0.05

Figure 5-1 Joint probability distribution of X and Y. The table cells are the probabilities. Observe that more bars relate to less repeating.



Joint Probability distributions

The joint probability mass function of the discrete random variables X and Y, denoted as $f_{XY}(x, y)$, satifies:

(1) $f_{XY}(x, y) \geq 0$ All probabilities are non-negative

(2) $\sum_x \sum_y f_{XY}(x, y) = 1$ The sum of all probabilities is 1

(3) $f_{XY}(x, y) = P(X = x, Y = y)$ (5-1)

Joint Probability Density Function Defined

The joint probability density function for the continuous random variables X and Y, denotes as $f_{XY}(x,y)$, satisfies the following properties:

$$(1)\ \ f_{XY}(x,y) \geq 0 \text{ for all } x,\ y$$

$$(2)\ \ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dxdy = 1$$

$$(3)\ \ P\big((X,Y)\subset R\big) = \iint_R f_{XY}(x,y)\,dxdy \qquad (5\text{-}2)$$

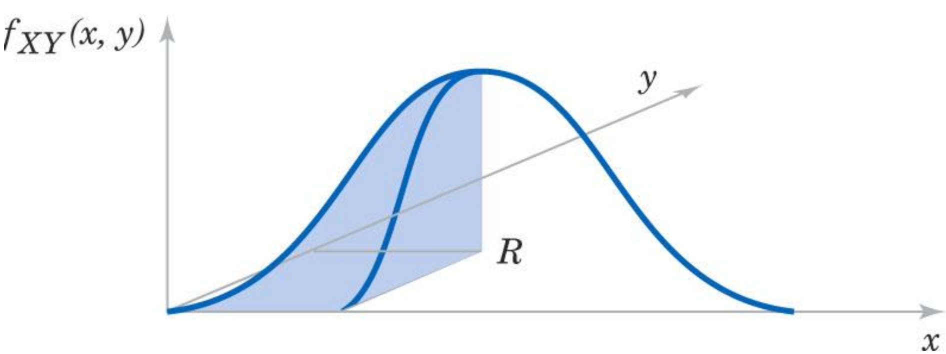


Figure 5-2 Joint probability density function for the random variables X and Y. Probability that (X, Y) is in the region R is determined by the volume of $f_{XY}(x,y)$ over the region R.

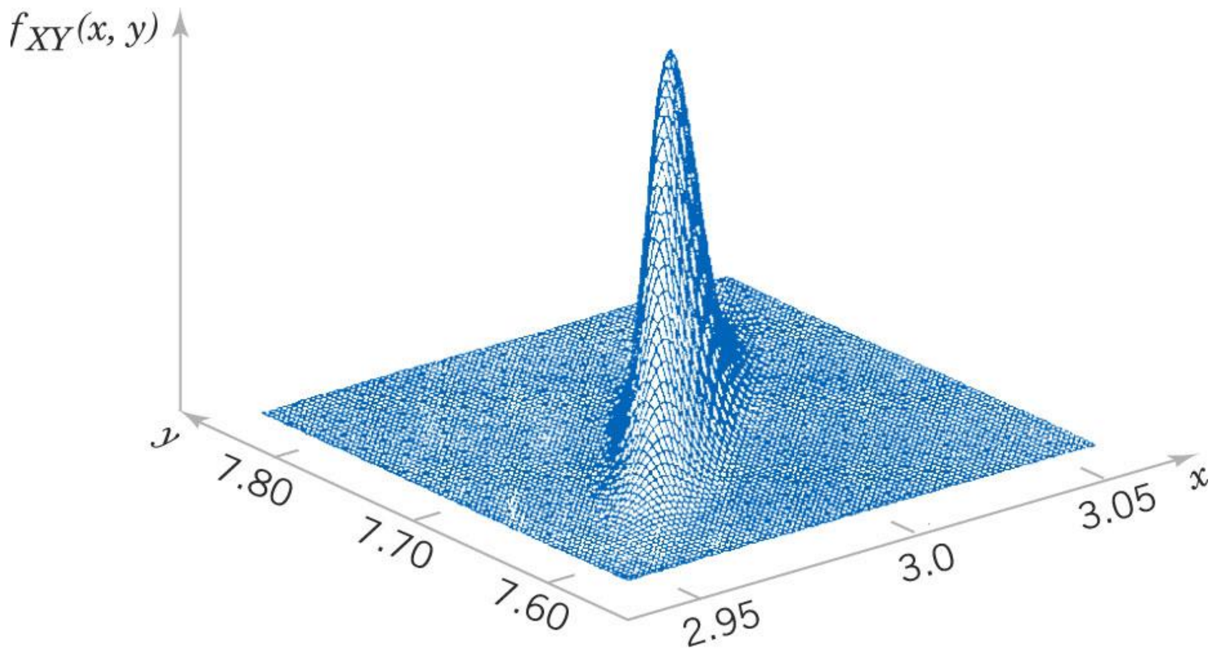


Figure 5-3 Joint probability density function for the continuous random variables *X* and *Y* of different dimensions of an injection-molded part. Note the asymmetric, narrow ridge shape of the PDF – indicating that small values in the *X* dimension are more likely to occur when small values in the *Y* dimension occur.

Marginal Probability Distributions

- The individual probability distribution of a random variable is referred to as its **marginal probability distribution**.
- In general, the marginal probability distribution of X can be determined from the joint probability distribution of X and other random variables. For example, to determine $P(X = x)$, we sum $P(X = x, Y = y)$ over all points in the range of (X, Y) for which $X = x$. Subscripts on the probability mass functions distinguish between the random variables.

Marginal Probability Distributions (discrete)

For a discrete joint PMF, there are marginal distributions for each random variable, formed by summing the joint PMF over the other variable.

$$f_X(x) = \sum_y f(xy)$$

$$f_Y(y) = \sum_x f(xy)$$

y = number of times city name is stated	x = number of bars of signal strength			
	1	2	3	$f(y)$ =
1	0.01	0.02	0.25	0.28
2	0.02	0.03	0.20	0.25
3	0.02	0.10	0.05	0.17
4	0.15	0.10	0.05	0.30
$f(x)$ =	0.20	0.25	0.55	1.00

Figure 5-6 From the prior example, the joint PMF is shown in green while the two marginal PMFs are shown in blue.

Marginal Probability Distributions (continuous)

- Rather than summing, like for a discrete joint PMF, we integrate a continuous joint PDF.
- The marginal PDFs are used to make probability statements about one variable.
- If the joint probability density function of random variables X and Y is $f_{XY}(x,y)$, the marginal probability density functions of X and Y are:

$$f_X(x) = \int_y f_{XY}(x,y)\,dy$$

$$f_Y(y) = \int_x f_{XY}(x,y)\,dx \qquad (5\text{-}3)$$

Example

Twenty chips of the same size are placed in a bowl. Each chip has a pair (x, y) of numbers written on it. Assume that there are 4 chips with $(1, 1)$ written on them, 3 with $(2, 1)$, 1 with $(3, 1)$, 2 with $(1, 2)$, 4 with $(2, 2)$, and 6 with $(3, 2)$. In other words, the pairs of numbers are distributed among the 20 chips in accordance with Table 2.6-1.

If a chip is drawn at random, the probability of obtaining a chip with, for example, $(2, 1)$, is 3/20 under the assumption that each chip has the same probability of being picked. Also, the probability of drawing a chip with $(1, 2)$ is 2/20. Let us denote the coordinates of the chip that is to be selected at random by (X, Y), so that (X, Y) is a pair of discrete random variables. With this convention, the probability of obtaining a chip with coordinates $(2, 1)$ can be written

$$P(X = 2, Y = 1) = \frac{3}{20},$$

Table 2.6-1 Classification of 20 Chips

	x			
y	1	2	3	Total
1	4	3	1	8
2	2	4	6	12
Total	6	7	7	20

Example

Moreover, we can ask for probabilities about X and Y alone. For example,

$$P(X = 2) = P(X = 2, Y = 1) + P(X = 2, Y = 2) = \frac{3}{20} + \frac{4}{20} = \frac{7}{20},$$

because the event $X = 2$ is the union of the mutually exclusive events $(X = 2, Y = 1)$ and $(X = 2, Y = 2)$.

Table 2.6-2 $f(x, y) = P(X = x, Y = y)$

	x			
y	1	2	3	$P(Y = y)$
1	4/20	3/20	1/20	8/20
2	2/20	4/20	6/20	12/20
$P(X = x)$	6/20	7/20	7/20	

Mean & Variance of a Marginal Distribution

Means E(X) and E(Y) are calculated from the discrete and continuous marginal distributions.

	Discrete	Continuous
	$E(X)=\sum_R x\cdot f_X(x)$	$=\int_R x\cdot f_X(x)\,dx=\mu_X$
	$E(Y)=\sum_R y\cdot f_Y(y)$	$=\int_R y\cdot f_Y(y)\,dy=\mu_Y$
	$V(X)=\sum_R x^2\cdot f_X(x)-\mu_X^2$	$=\int_R x^2\cdot f_X(x)\,dx-\mu_X^2$
	$V(Y)=\sum_R y^2\cdot f_Y(y)-\mu_Y^2$	$=\int_R y^2\cdot f_Y(y)\,dy-\mu_Y^2$

Example

y = number of times city name is stated	x = number of bars of signal strength					
	1	2	3	$f(y)$ =	$y*f(y)$ =	$y^2*f(y)$ =
1	0.01	0.02	0.25	0.28	0.28	0.28
2	0.02	0.03	0.20	0.25	0.50	1.00
3	0.02	0.10	0.05	0.17	0.51	1.53
4	0.15	0.10	0.05	0.30	1.20	4.80
$f(x)$ =	0.20	0.25	0.55	1.00	2.49	7.61
$x*f(x)$ =	0.20	0.50	1.65	2.35		
$x^2*f(x)$ =	0.20	1.00	4.95	6.15		

$E(X) = 2.35$ $V(X) = 6.15 – 2.35^2 = 6.15 – 5.52 = 0.6275$

$E(Y) = 2.49$ $V(Y) = 7.61 – 2.49^2 = 7.61 – 16.20 = 1.4099$

Conditional Probability Distributions

Given continuous random variables X and Y with

joint probability density function $f_{XY}(x,y)$,

the conditional probability densiy function of Y given X =x is

$$f_{Y|x}(y)=\frac{f_{XY}(x,y)}{f_X(x)} \quad \text{for} \quad f_X(x)>0 \qquad (5\text{-}4)$$

which satifies the following properties:

(1) $f_{Y|x}(y)\geq 0$

(2) $\int f_{Y|x}(y)\,dy=1$

(3) $P(Y\subset B|X=x)=\int_B f_{Y|x}(y)\,dy$ for any set B in the range of Y

Conditional discrete PMFs can be shown as tables.

y = number of times city name is stated	x = number of bars of signal strength				f(x|y) for y =			Sum of
	1	2	3	$f(y)$ =	1	2	3	f(x|y) =
1	0.01	0.02	0.25	0.28	0.036	0.071	0.893	1.000
2	0.02	0.03	0.20	0.25	0.080	0.120	0.800	1.000
3	0.02	0.10	0.05	0.17	0.118	0.588	0.294	1.000
4	0.15	0.10	0.05	0.30	0.500	0.333	0.167	1.000
$f(x)$ =	0.20	0.25	0.55					
1	0.050	0.080	0.455					
2	0.100	0.120	0.364					
3	0.100	0.400	0.091					
4	0.750	0.400	0.091					
Sum of f(y|x) =	1.000	1.000	1.000					

Conditional Probability Distributions

For the example in Table 2.6-2, we find

Table 2.6-2 $f(x, y) = P(X = x, Y = y)$

	x			
y	1	2	3	$P(Y = y)$
1	4/20	3/20	1/20	8/20
2	2/20	4/20	6/20	12/20
$P(X = x)$	6/20	7/20	7/20	

$$g(y|x = 1) = \begin{cases} \dfrac{4/20}{6/20} = \dfrac{2}{3}, & y = 1 \\[2ex] \dfrac{2/20}{6/20} = \dfrac{1}{3}, & y = 2 \end{cases}$$

$$(y|x = 2) = \begin{cases} \dfrac{3/20}{7/20} = \dfrac{3}{7}, & y = 1 \\[2ex] \dfrac{4/20}{7/20} = \dfrac{4}{7}, & y = 2 \end{cases}$$

$$g(y|x = 3) = \begin{cases} \dfrac{1/20}{7/20} = \dfrac{1}{7}, & y = 1 \\[2ex] \dfrac{6/20}{7/20} = \dfrac{6}{7}, & y = 2 \end{cases}$$

Covariance and Correlation Coefficient

The covariance between the random variables X and Y, denoted as $\operatorname{cov}(X,Y)$ or σ_{XY} is

$$\sigma_{XY} = E\left[\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right] = E\left(XY\right)-\mu_X\mu_Y \qquad (5\text{-}14)$$

The units of σ_{XY} are units of X times units of Y.

For example, if the units of X are feet
and the units of Y are pounds,
the units of the covariance are foot-pounds.

Unlike the range of variance, $-\infty < \sigma_{XY} < \infty$.

Covariance and Correlation Coefficient

y = number of times city name is stated	x = number of bars of signal strength		
	1	2	3
1	0.01	0.02	0.25
2	0.02	0.03	0.20
3	0.02	0.10	0.05
4	0.15	0.10	0.05

The probability distribution of Example 5-1 is shown.

By inspection, note that the larger probabilities occur as *X* and *Y* move in opposite directions. This indicates a negative covariance.

Covariance and Correlation Coefficient

The correlation between random variables X and Y, denoted as ρ_{XY}, is

$$\rho_{XY} = \frac{\operatorname{cov}(X,Y)}{\sqrt{V(X) \cdot V(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \qquad (5\text{-}15)$$

Since $\sigma_X > 0$ and $\sigma_Y > 0$,

ρ_{XY} and $\operatorname{cov}(X,Y)$ have the same sign.

We say that ρ_{XY} is normalized, so $-1 \le \rho_{XY} \le 1$ (5-16)

Note that ρ_{XY} is dimensionless.

Variables with non-zero correlation are correlated.

Example

The joint PMF of precipitation, X (in.) and runoff, Y (cfs) (discretized here for simplicity) due to storms at a given location is shown in the table below.

a. What is the probability that the next storm will bring a precipitation of 2 or more inches and a runoff of more than 20 cfs?
b. After a storm, the rain gauge indicates a precipitation of 2 in. What is the probability that the runoff in this storm is 20 cfs or more?
c. Are X and Y statistically independent?
d. Determine and plot the marignal PMF of runoff.
e. Determine and plot the PMF of runoff for a storm whose precipitation is 2 in.
f. Determine the correlation coefficient between precipitation and runoff.

	X = 1	X = 2	X = 3
Y = 10	0.05	0.15	0.0
Y = 20	0.10	0.25	0.25
Y = 30	0.0	0.10	0.10

Example

The daily water levels (normalized to the respective full condition) of two reservoirs A and B are denoted by two random variables x and y having the following joint PDF.

a. Determine the marginal density function of the daily water level of Reservoir A.
b. If a reservoir A is half full on a given day, what is the probability that the water level will be more than half full in reservoir B?
c. Is there any statistical correlation between the water levels in the two reservoirs?

$$f(x,y)=\frac{6}{5}(x+y^2) \quad 0<x<1; 0<y<1$$