Chapter 10 - RANDOM VARIABLES AND PROBABILITY DENSITY FUNCTIONS

Introduction

In this chapter we introduce probability density functions for single random variables, and extend them to multiple, jointly-distributed variables. Particular emphasis is placed on conditional probabilities and density functions, which play a key role in Bayesian detection theory.

10.1 Random variables and probability density functions

10.1.1 Random variables, events and probability

A *random variable* is a number assigned to every outcome of an experiment. For example, the result $1 \leq n \leq 6$ of rolling a die once, or a temperature measurement x from a patient are random variables. If the experiment is repeated, the random variable may take a different value on each trial. A random variable can take either discrete values (within the set of integers) as in the die example, or continuous values (any real number) as in the temperature example. In many cases, one is interested in whether a random variable lies within a particular range of values, for example, that the result of rolling a die is a 3, or that it is between 2 and 5. Such sets of values of random variables are called *events*. When an experiment is repeated, the *probability* of an event is defined as the limit of the frequency of occurrence of that event when the number of trials becomes large:

$$Pr(\varepsilon) \triangleq \lim_{N_t \to \infty} \frac{N_\varepsilon}{N_t}, \tag{10.1}$$

where N_ε is the the number of trials in which the event occurs, and N_t is the total number of trials. It is clear from this definition that the probability of an event must always be between 0 and 1:

$$0 \leq Pr(\varepsilon) \leq 1 \tag{10.2}$$

Two events are said to be disjoint if the sets of values of the random variable that they represent are nonoverlapping. For example, in the die experiment, the events $n = 2$ and $n > 3$ are disjoint. If ε_1 and ε_2 are two disjoint events, then the probability of their union is:

$$Pr(\varepsilon_1 \cup \varepsilon_2) = Pr(\varepsilon_1) + Pr(\varepsilon_2) \qquad \text{if } \varepsilon_1 \cap \varepsilon_2 = \emptyset \tag{10.3}$$

If ε_1, $\varepsilon_2, \ldots \varepsilon_N$ are disjoint events and their union is the certain event (the set of all possible values of a random variable), then their probabilities sum to 1:

$$\sum_{i=1}^{N} Pr(\varepsilon_i) \;=\; 1 \tag{10.4}$$

For example, in the die experiment, the events $n = 1$, $n = 2, \ldots, n = 6$ are disjoint and represent all possible values of n, so their probabilities sum to 1. In the simplest case of a fair die, all 6 probabilities are equal to 1/6.

10.1.2 Probability density function

Given a continuous random variable x, the probability of any event can be derived from the *probability density function (pdf).* The pdf evaluated at X is a limit of the normalized probability that x lies in the small interval $[X,\ X + \Delta X]$:

$$f_x(X) \stackrel{\triangle}{=} \lim_{\Delta X \to 0} \frac{1}{\Delta X} Pr(X \leq x \leq X + \Delta) \tag{10.5}$$

In this expression, the lower-case x denotes the random variable, while the upper-case X refers to a particular value of this variable. The probability of an arbitrary event ε defined by a random variable x is obtained by integrating the pdf over the set of values of x defining the event:

$$Pr(\varepsilon) = \int_{\varepsilon} f_x(X)dX \tag{10.6}$$

For example, the probability that x is between two numbers X_1 and X_2 is:

$$Pr(X_1 \leq x \leq X_2) = \int_{X_1}^{X_2} f_x(X)dX \tag{10.7}$$

Specializing (10.7) to $X_1 = X$ and $X_2 = X + \Delta X$, and taking the limit when ΔX becomes small gives the definition of the pdf (10.5), as expected.

For continuous-valued random variables, the pdf is usually (but not always) a continuous function of X. However, for a discrete-valued random variable n, the pdf is a weighted sum of impulses located at integer values of X:

$$f_n(X) \;=\; \sum_{i=-\infty}^{\infty} P_i\ \delta(X - i), \quad \text{with } P_i \stackrel{\triangle}{=} Pr(n = i) \tag{10.8a}$$

In that case, the integral in (10.7) can be replaced by a sum:

$$Pr(N_1 \leq n \leq N_2) = \sum_{i=N_1}^{N_2} P_i \tag{10.8b}$$

For both discrete and continuous-valued random variables, the pdf must have the following properties:

$$f_x(X) \;\geq\; 0 \quad \text{for all } X \tag{10.9a}$$

$$\int_{-\infty}^{\infty} f_x(X)dX = 1 \tag{10.9b}$$

Examples of pdfs are given in Chapter 11. Two important ones are:

1. The *Gaussian pdf,* which is continuous-valued:

$$f_x(X) \;=\; \frac{1}{\sqrt{2\pi}\sigma}\;\exp\left(\;-\;\frac{(X\;-\;\mu)^2}{2\sigma^2}\right) \tag{10.10}$$

2. The *Bernoulli pdf,* which is discrete-valued:

$$Pr(n=i) = \begin{cases} p & \text{if } i=1 \\ 1-p & \text{if } i=0 \end{cases} \tag{10.11a}$$

The random variable n can only take the two values 0 or 1. The Bernoulli pdf can be conveniently rewritten as:

$$f_n(X) \;=\; p^X\;(1-p)^{(1-X)}, \qquad \text{with } X \;=\; 0 \text{ or } 1. \tag{10.11b}$$

10.1.3 Expected value

The expected value, or *mean* of a random variable is defined by the integral:

$$E(x) \;=\; \mu_x \;\triangleq\; \int_{-\infty}^{\infty}\; X\; f_x(X)\; dX \tag{10.12}$$

It can be shown that, if an experiment is repeated many times, and the value X_i of the random variable x is measured on each trial i, then the mean is the limit of the average value of the X_i when the number of trials becomes large:

$$E(x) \;=\; \lim_{N\to\infty}\;\frac{1}{N}\;\sum_{i=1}^{N}\; X_i \tag{10.13}$$

This is a direct consequence of the definition of probability (10.1).

The pdf can be used to obtain the means of arbitrary functions of random variables:

$$E\left(g(x)\right) \;=\; \int_{-\infty}^{\infty}\; g(X)\; f_x(X)\; dX \tag{10.14}$$

In particular, the *variance* of a random variable is defined by:

$$\sigma_x^2 \;\triangleq E\left((x-\mu_x)^2\right) \;=\; \int_{-\infty}^{\infty}\;(x-\mu_x)^2\; f_x(X)\; dX \;=\; E(x^2)\;-\;\mu_x^2 \tag{10.15}$$

As the notation indicates, the mean of a Gaussian random variable (10.10) is μ and its variance σ^2. The mean of a Bernoulli variable (10.10) if p and its variance $p\;(1-p)$.

10.2 Multiple random variables and joint probability density functions

Often, the outcome of an experiment is not a single random variable, but two or more random variables. For example, one can simultaneously roll two dice, or measure both the temperature and the heart rate of a patient. Here, we only consider the case of two random variables x and y; generalization to an arbitrary number of variables is trivial except for cumbersome notation.

Together, two random variables define joint events representing the outcome that x and y occupy a specific region $R(X, Y)$ of the $X - Y$ plane. For example, a simple event is defined by the rectangular region:

$$\varepsilon \triangleq [X_1 \leq x \leq X_2 \;\text{ and }\; Y_1 \leq y \leq X_2] \tag{10.16}$$

In general, it is not possible to derive the probabilities of joint events on x and y from the first-order pdfs $f_x(X)$ and $f_y(Y)$ because the two variables may not be independent. Instead, we need to know the *joint probability density function* $f_{xy}(X, Y)$. The joint pdf evaluated at (X, Y) is the limit of the normalized probability that x and y lie in a small rectangle of area $\Delta X \Delta Y$ located at the coordinates (X, Y):

$$f_{xy}(X, Y) \;\triangleq\; \lim_{\Delta X \to 0, \Delta Y \to 0} \frac{1}{\Delta X \Delta Y}\; Pr(X \leq x \leq X + \Delta X \;\text{ and }\; Y \leq y \leq Y + \Delta Y) \tag{10.17}$$

As the pdf for a single random variable, the joint pdf is non negative

$$f_{xy}(X) \;\geq\; 0 \qquad \text{for all } X,\; Y \tag{10.18a}$$

and integrates to unity

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{xy}(X)\; dX\; dY \;=\; 1 \tag{10.18b}$$

Moreover, the first-order or "marginal" p.d.f's for x and y can are generated from the joint pdf by integrating over the other variable:

$$f_x(X) \;=\; \int_{-\infty}^{\infty} f_{xy}(X,\; Y)\; dY \tag{10.19a}$$

$$f_y(Y) \;=\; \int_{-\infty}^{\infty} f_{xy}(X,\; Y)\; dX \tag{10.19b}$$

More generally, the probability that x and y lie in an arbitrary region R of the $X - Y$ plane is obtained by integrating the joint pdf over that region:

$$Pr([x, y] \in R) = \int \int_R f_{xy}(X, Y)\; dX\; dY \tag{10.20}$$

For example,

$$Pr(X_1 \leq x \leq X_2 \;\text{ and }\; Y_1 \leq y \leq X_2) \;=\; \int_{X_1}^{X_2} \int_{Y_1}^{Y_2} f_{xy}(X, Y)\; dX\; dY \tag{10.21}$$

10.2.1 Mean of a function of two random variables

The joint pdf is used for computing expressions of the form $E\left(g(x, y)\right)$, where $g(X,\; Y)$ is an arbitrary function of two variables:

$$E\left(g(x,\; y)\right) \;=\; \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(X, Y)\; f_{xy}(X, Y)\; dX\; dY \tag{10.22}$$

For example, the *correlation* r_{xy} between two random variables is the expected value of their product. It can be obtained by specializing (10.22) to the function $g(X,\ Y)\ =\ XY$:

$$r_{xy} \triangleq E(xy) \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} XY\ f_{xy}(X,\ Y)\ dX\ dY \tag{10.23}$$

Similarly, the *covariance* c_{xy} is the correlation of the variables with the means subtracted out:

$$c_{xy} \triangleq E\left((x-\mu_x)\ (y-\mu_y)\right) \;=\; r_{xy} \;-\; \mu_x\mu_y \tag{10.24}$$

Two random variables are said to be *uncorrelated* if their correlation is the product of their means, i.e. if

$$E(xy) = E(x)E(y) \tag{10.25}$$

An equivalent statement is that the covariance c_{xy} is zero.

10.3 Conditional probability density functions

10.3.1 Conditional probability and statistical independence

The *conditional probability* of an event ε_A given another event ε_B is defined as the ratio of the joint probability of the two events to the probability of ε_B:

$$Pr(\varepsilon_A|\varepsilon_B) \triangleq \frac{Pr(\varepsilon_A \text{ and } \varepsilon_B)}{Pr(\varepsilon_B)} \tag{10.26}$$

For example, assume we have an unbiased die in which the probabilities of all 6 faces showing are equal $Pr(n=i)\ =\ 1/6,\ 1\ \leq\ i\ \leq\ 6$. We will evaluate the conditional probability that $n=5$ given that $n>3$. ¿From the definition (10.26), this is

$$Pr(n=5|n>3) \;=\; \frac{Pr(n=5 \text{ and } n>3)}{Pr(n>3)} \;=\; \frac{Pr(n=5)}{Pr(n>3)} \;=\; \frac{1/6}{1/2} \;=\; 1/3 \tag{10.27}$$

Thus, in the absence of any knowledge about the value of n, the probability $Pr(n=5)$ is $1/6$. If, however, we are told that $n>3$, then $Pr(n=5)$ increases to $1/3$.

Two events ε_A and ε_B are said to be *statistically independent* if the conditional probability of one event given the other one is equal to its unconditioned probability:

$$Pr(\varepsilon_A|\varepsilon_B) \;=\; Pr(\varepsilon_A) \tag{10.28a}$$

This definition means that knowledge that ε_B has occurred does not alter the probability of ε_A. This corresponds well with the intuitive notion of independent events.

¿From the definition of the conditional probability (10.26), an alternative statement of statistical independence is that the joint probability of the two events is the product of the probabilities of the two separate events:

$$Pr(\varepsilon_A \text{ and } \varepsilon_B) \;=\; Pr(\varepsilon_A)\ Pr(\varepsilon_B) \tag{10.28b}$$

In the above die example, the two events $n = 5$ and $n > 3$ were obviously not independent. However, if we roll two dice and measure the numbers n and m showing on each die, then it makes sense that the two events $n = 5$ and $m = 3$ are independent. In fact, any pairs of events defined on n and m respectively are independent:

$$Pr(n = i \text{ and } m = j) \;=\; Pr(n = i)\; Pr(n = j) \qquad \text{for all } i,\ j \tag{10.29}$$

The two random variables n and m are said to be independent.

10.3.2 Conditional density function

The notion of conditional probability is easily extended to probability density functions. Specifically, given two random variables x and y, the conditional pdf $f_{x|y}(X|Y)$ is defined as a limit:

$$f_{x|y}(X|Y) \;\triangleq\; \lim_{\Delta X \to 0, \Delta Y \to 0} \frac{1}{\Delta X} \frac{Pr(X \leq x \leq X + \Delta X \text{ and } Y \leq y \leq Y + \Delta Y)}{Pr(Y \leq y \leq Y + \Delta Y)} \tag{10.30}$$

Applying the definitions of $f_{xy}(X, Y)$ and $f_y(Y)$, this becomes:

$$f_{x|y}(X|Y) \;=\; \frac{f_{xy}(X, Y)}{f_y(Y)} \tag{10.31}$$

Because the denominator $f_y(Y)$ can be obtained by integrating the joint pdf $f_{xy}(X, Y)$ over X, the conditional pdf is entirely specified by the joint pdf. In other words, a conditional pdf gives no additional information over the joint pdf. In many cases, however, it is much easier to work with conditional pdf's rather than joint pdf's, which can be hard to specify.

The conditional pdf of x given Y behaves in every respect like an ordinary pdf. It is nonnegative:

$$f_{x|y}(X|Y) \;\geq\; 0 \qquad \text{for all } X,\ Y \tag{10.32a}$$

It integrates to unity:

$$\int_{-\infty}^{\infty} f_{x|y}(X|Y)\; dX \;=\; \frac{\int_{-\infty}^{\infty} f_{xy}(X, Y)\; dX}{f_y(Y)} \;=\; \frac{f_y(Y)}{f_y(Y)} \;=\; 1 \tag{10.32b}$$

It generates the conditional probabilities of all events defined by the random variable x given that $y = Y$. For example:

$$Pr(X_1 \;\leq\; x \;\leq\; X_2 \mid y = Y) \;=\; \int_{X_1}^{X_2} f_{x|y}(X|Y)\; dX \tag{10.32c}$$

Moreover, the unconditioned pdf $f_x(X)$ can be obtained from the conditional pdf and the marginal pdf for the conditioning variable y:

$$f_x(X) \;=\; \int_{-\infty}^{\infty} f_{xy}(X, Y)\; dY \;=\; \int_{-\infty}^{\infty} f_{x|y}(X|Y)\; f_y(Y)\; dY \tag{10.33}$$

10.3.3 Independent random variables

Two random variables x and y are said to be *statistically independent* if every event defined by x is independent of every event defined by y. It is easily shown that this will be the case if and only if:

$$f_{xy}(X,Y) \;=\; f_x(X)\; f_y(Y) \qquad \text{for all } X,\; Y \tag{10.34a}$$

This independence condition can also be stated in terms of the conditional pdf:

$$f_{x|y}(X|Y) \;=\; f_x(X) \qquad \text{for all } X,\; Y \tag{10.34b}$$

Again, this fits with the intuitive notion of independence that knowing the value of y gives no information about the value of x.

If two random variables are independent, they are also uncorrelated:

$$\begin{aligned} E(xy) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} XY\; f_{xy}(X,Y)\; dX\; dY \\ &= \left(\int_{-\infty}^{\infty} X\; f_x(X)\; dX\right)\left(\int_{-\infty}^{\infty} Y\; f_x(Y)\; dY\right) \\ &= E(x)\; E(y) \end{aligned} \tag{10.35}$$

However, the converse is not true: Statistical independence is a much stronger condition on random variables than being uncorrelated. As a counter example, if x has a symmetric pdf with respect to $x=0$, then x and x^2 are uncorrelated random variables, but they are obviously not independent. We will see however that, if x and y are uncorrelated Gaussian random variables, then they are also independent.

10.3.4 Bayes' rule

Consider two jointly-distributed random variables x and y. ¿From the definition of the conditional pdf (10.31), we have:

$$f_{xy}(X,Y) \;=\; f_{x|y}(X|Y)\; f_y(Y) \tag{10.36a}$$

By exchanging the roles of x and y, we can also write:

$$f_{xy}(X,Y) \;=\; f_{y|x}(Y|X)\; f_x(X) \tag{10.36b}$$

Since the two expressions are equal, we get:

$$f_{y|x}(Y|X) \;=\; \frac{f_y(Y)\; f_{x|y}(X|Y)}{f_x(X)} \tag{10.37a}$$

Further making use of (10.33), we can express $f_{y|x}(Y|X)$ in terms of $f_{x|y}(X|Y)$ and $f_y(Y)$:

$$f_{y|x}(Y|X) \;=\; \frac{f_y(Y)\; f_{x|y}(X|Y)}{\int_{-\infty}^{\infty} f_y(Y)\; f_{x|y}(X|Y)\; dY} \tag{10.37b}$$

Either form of (10.37) is known as *Bayes' rule.*

While, at first sight it might seem somewhat arbitrary, Bayes' rule is of fundamental importance in statistical applications such as pattern classification, hypothesis testing, signal detection and estimation. Specifically, assume that the values of the random variable y represent different states of nature, while x is the result of a measurement intended to inform us about the state of nature. Prior to any measurement, we may have some information about which states of nature are the most likely. This information is represented by the *a priori* density $f_y(Y)$. We also have a probabilistic model relating the states of nature Y to our measurement x. This measurement model is represented by the conditional pdf $f_{x|y}(X|Y)$. After making a measurement, and finding that $x = X$, our knowledge of the state of nature is (hopefully) improved. This knowledge is represented by the *a posteriori* density $f_{y|x}(Y|X)$. For example, we might want to pick the value of Y that maximizes the a posteriori probability to estimate the state of nature given the observation that $x = X$. Bayes' rule allows us to compute the a posteriori probability in terms of the two known distributions $f_y(Y)$ and $f_{x|y}(X|Y)$. This is useful because a posteriori probabilities are often exceedingly difficult to directly estimate.

10.3.5 Probability density function for the sum of two random variables

To further illustrate the usefulness of conditional pdf's we will derive the pdf for the sum s of two random variables x and y. We temporarily assume that y takes a particular value Y. The conditional pdf for s given $y = Y$ can expressed as a function of the conditional pdf for x:

$$f_{s|y}(S|Y) \;=\; f_{x|y}(S - Y|Y) \tag{10.38}$$

This just states that, if $x + y = S$, and $y = Y$, then x must be equal to $S - Y$. Using (10.33), the unconditioned pdf $f_s(S)$ can then be obtained by summing the conditional pdf over all Y:

$$f_s(S) \;=\; \int_{-\infty}^{\infty} f_{s|y}(S|Y)\, f_y(Y)\, dY \;=\; \int_{-\infty}^{\infty} f_{x|y}(S - Y|Y)\, f_y(Y)\, dY \;=\; \int_{-\infty}^{\infty} f_{xy}(S - Y,\, Y)\, dY \tag{10.39}$$

This is the desired result in the general case.

In the special case when x and y are statistically independent, the pdf of s takes on a particularly simple form. Specifically, if x and y are independent, then

$$f_{x|y}(X|Y) \;=\; f_x(X), \tag{10.40}$$

so that

$$f_s(S) \;=\; \int_{-\infty}^{\infty} f_{x|y}(S - Y|Y)\, f_y(Y)\, dY \;=\; \int_{-\infty}^{\infty} f_x(S - Y)\, f_y(Y)\, dY \;=\; f_x(S) * f_y(S) \tag{10.41}$$

Thus, the pdf for the sum of two statistically independent random variables is the convolution of the pdf's of the two random variables.

This result is the basis for the *central limit theorem,* which states that the pdf for the sum of a large number of independent random variables approaches a Gaussian pdf. This theorem follows from the fact that the convolution of many functions approaches a Gaussian regardless of the shapes of the convolved functions.

10.3.6 Conditional mean and variance

We have seen that the conditional pdf $f_{x|y}(X|Y)$ behaves in every respect like an ordinary pdf. Therefore, using is (10.12), it can be used to obtain the mean of x given that $y = Y$:

$$\mu_{x|y}(Y) \triangleq \int_{-\infty}^{\infty} X \ f_{x|y}(X|Y) \ dX \tag{10.42}$$

Thus quantity is known as the *conditional mean.* As the notation indicates, it depends on the value Y of the conditioning variable y. The conditional mean plays an important role in nonlinear least-squares estimation.

More generally, given an arbitrary function $g(X)$, the conditional mean of $g(x)$ given $y = Y$ can also be derived from the conditional pdf:

$$E\left(g(x)|y = Y\right) = \int_{-\infty}^{\infty} g(X) \ f_{x|y}(X|Y) \ dX \tag{10.43}$$

For example, the *conditional variance* is obtained by specializing (10.43) to $g(x) = (x - \mu_{x|y})^2$:

$$\sigma^2_{x|y}(Y) = \int_{-\infty}^{\infty} \left(X - \mu_{x|y}(Y)\right)^2 \ f_{x|y}(X|Y) \ dX \tag{10.44}$$

Because the conditional mean $\mu_{x|y}$ depends on the value Y of the random variable y, it is a random variable, specifically one that is solely a function of y. Therefore, its expected value can be obtained as a special case of (10.14):

$$E(\mu_{x|y}) = \int_{-\infty}^{\infty} \mu_{x|y}(Y) \ f_y(Y) \ dY \tag{10.45a}$$

Inserting the definition (10.43) of the conditional mean, we get:

$$E(\mu_{x|y}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X \ f_{x|y}(X|Y) \ f_y(Y) \ dX \ dY \tag{10.45b}$$

Further making use of the definition of the conditional pdf (10.31), this becomes:

$$E(\mu_{x|y}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X \ f_{xy}(X,Y) \ dX \ dY = \int_{-\infty}^{\infty} X \ f_x(X) \ dX = \mu_x \tag{10.46}$$

Thus, the mean of the conditional mean $\mu_{x|y}$ over all values of the conditioning variable y is the unconditioned mean μ_x.

10.4 Gaussian random variables

10.4.1 Joint pdf of a Gaussian random vector

Given a large number of random variables, it is rarely possible to specify their joint pdf, except in the trivial case of independent variables. An important exception is that of *Gaussian random* *variables* which are entirely specified from their means and covariances. Specifically, let

$$\mathbf{x} \triangleq \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \tag{10.47}$$

be an N-dimensional vector of random variables, a.k.a. a *random vector.* This random vector is said to be Gaussian if its joint pdf is of the form:

$$f_x(\mathbf{X}) \triangleq \left((2\pi)^N|\mathbf{C}|\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\,(\mathbf{X} - \mathbf{M})^{\mathbf{T}}\,\mathbf{C^{-1}}\,(\mathbf{X} - \mathbf{M})\right), \tag{10.48a}$$

where $\mathbf{X}$ and $\mathbf{M}$ are N-dimensional column vectors defined by

$$\mathbf{X} \triangleq \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} \quad \text{and} \quad \mathbf{M} \triangleq E(\mathbf{x}) = \begin{bmatrix} \mu_{x_1} \\ \mu_{x_2} \\ \vdots \\ \mu_{x_N} \end{bmatrix}, \tag{10.48b}$$

and

$$\mathbf{C} \triangleq E\left((\mathbf{x} - \mathbf{M})\,(\mathbf{x} - \mathbf{M})^T\right) \tag{10.48c}$$

is the *covariance matrix.* $\mathbf{C}$ is an $N \times N$ symmetric, positive-definite matrix whose elements c_{ij} are the covariances of the vector elements:

$$c_{ij} \triangleq = E\left((x_i - \mu_{x_i})\,(x_j - \mu_{x_j})\right) = c_{ji}, \qquad 1 \le i,\ j \le N \tag{10.48d}$$

$|\mathbf{C}|$ denotes the determinant of $\mathbf{C}$, and $\mathbf{C}^{-1}$ its inverse.

10.4.2 Uncorrelated Gaussian variables

In the special case when the random variables x_i are uncorrelated, the expression for the joint pdf is greatly simplified because the covariance matrix is diagonal:

$$c_{ij} = \sigma_{x_i}^2\,\delta_{ij}, \qquad 1 \le i,\ j \le N \tag{10.49}$$

The inverse covariance matrix $\mathbf{C}^{-1}$ is also diagonal, and the determinant $|\mathbf{C}|$ is the product of the variances. Thus, the joint pdf can be factored out into:

$$f_x(\mathbf{X}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_{x_i}} \exp\left(-\frac{(X_i - \mu_{x_i})^2}{2\sigma_{x_i}^2}\right) \tag{10.50}$$

This is the product of the N first-order Gaussian pdf's for each of the component variables x_i:

$$f_x(\mathbf{X}) = \prod_{i=1}^{N} f_{x_i}(X_i) \tag{10.51}$$

This shows that the N random variables x_i are independent. Thus, we have shown that uncorrelated Gaussian random variables are independent. Key to this result is that the joint pdf of a Gaussian vector depends only on the mean vector and covariance matrix.

10.4.3 Joint pdf for $N = 2$

Most properties of the joint pdf of Gaussian random vector can be understood from the two-dimensional case ($N = 2$). Specifically, let x and y be two jointly Gaussian random variables. Their covariance matrix can be written as:

$$\mathbf{C} \;=\; \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}, \qquad (10.52a)$$

where $\rho \overset{\triangle}{=} c_{xy/}\sigma_x\sigma_y$ is the *correlation coefficient,* which is always between -1 and 1 (Cauchy-Schwarz inequality). The determinant $|\mathbf{C}|$ is $\sigma_x^2\sigma_y^2\ (1-\rho^2)$, and the inverse matrix is

$$\mathbf{C^{-1}} = \frac{1}{1\ -\ \rho^2} \begin{bmatrix} \frac{1}{\sigma_x^2} & \frac{-\rho}{\sigma_x\sigma_y} \\ \frac{-\rho}{\sigma_x\sigma_y} & \frac{1}{\sigma_y^2} \end{bmatrix} \qquad (10.52b)$$

Inserting these values into (10.48a), we obtain

$$f_{xy}(X,Y) \;=\; \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$$
$$\exp\left(-\ \frac{1}{2\ (1-\rho^2)}\left[\frac{(X-\mu_x)^2}{\sigma_x^2}\ -\ 2\rho\frac{(X-\mu_x)(Y-\mu_y)}{\sigma_x\sigma_y}\ +\ \frac{(Y-\mu_y)^2}{\sigma_y^2}\right]\right) \qquad (10.53)$$

Figure 10.1 shows second-order Gaussian pdf's for four different values of ρ. For $\rho = 0$, iso-probability contours are ellipses centered at $[\mu_x, \mu_y]$, with principal axes parallel to the coordinates. If, in addition, $\sigma_x = \sigma_y$, then the ellipses become circles. As ρ increases from 0 to 1, the iso-probability contours become increasingly elongated ellipses along the line defined by the equation $(X-\mu_x)/\sigma_x = (Y-\mu_y)/\sigma_y$.

It is easily shown that the conditional pdf for x given $y = Y$ is also Gaussian with mean

$$\mu_{x|y}(Y) \;=\; \mu_x \;+\; \rho\ \sigma_x\ \frac{Y-\mu_y}{\sigma_y} \qquad (10.54a)$$

and variance

$$\sigma_{x|y}^2 \;=\; (1-\rho)^2\ \sigma_x^2 \qquad (10.54b)$$

Thus, the conditional pdf can be expressed as

$$f_{x|y}(X) \;=\; \frac{1}{\sqrt{2\pi(1-\rho)^2}\sigma_x}\ \exp\left(-\ \frac{[X-\mu_x-\rho\sigma_x(Y-\mu_y)/\sigma_y]^2}{2\ (1-\rho)^2\ \sigma_x^2}\right) \qquad (10.54c)$$

Note that, the larger $|\rho|$, the smaller the conditioned variance $\sigma_{x|y}^2$, so that x becomes increasingly concentrated near its conditional mean $\mu_{x|y}(Y)$. This makes sense because, if x and y are highly correlated, and we know that $y = Y$, then X can only take a narrow range of values defined by Y.

10.4.4 Properties of Gaussian random vectors

In Chapter 11, we will discuss the importance of Gaussian random variables as models for many physical and biological processes. In addition, Gaussian variables play a fundamental role in probability theory because their properties make them particularly simple to analyze:

1. The result of an arbitrary linear operation on a Gaussian vector is a Gaussian random variable. In fact, this property can be used to define Gaussian vectors.

2. Higher-order moments of Gaussian random variables can be computed from the mean and covariance matrix by means of the *moment factorization theorem.* This theorem states that, if x_1, x_2, x_3 and x_4 are jointly Gaussian random variables, then the mean of their product is formed from the following formula:

$$\begin{aligned} E(x_1 x_2 x_3 x_4) \;\; &= \;\; E(x_1 x_2) E(x_3 x_4) \; + \; E(x_1 x_3) E(x_2 x_4) \; + \; E(x_1 x_4) E(x_2 x_3) \\ & \quad - \; 2 E(x_1) E(x_2) E(x_3) E(x_4) \end{aligned} \tag{10.55}$$

   Thus, for example:

$$E(x^4) \; = \; 3 \; E(x^2)^2 \; - \; 2 \; \mu_x^4 \tag{10.56a}$$

$$E(x^2 \; y^2) \; = \; E(x^2) \; E(y^2) \; + \; 2 \; E(xy)^2 \; - \; 2 \; \mu_x^2 \; \mu_y^2 \tag{10.56b}$$

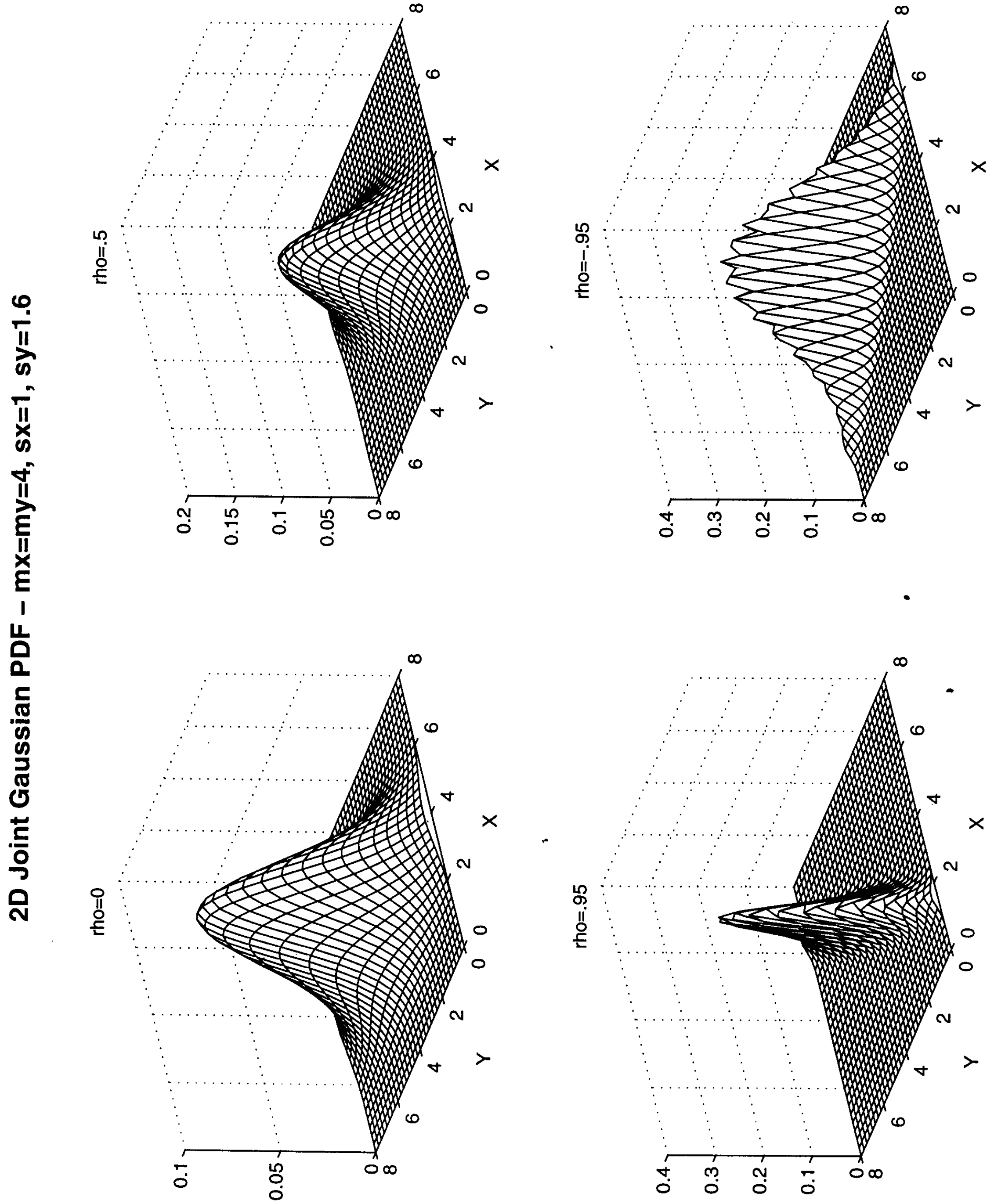

Figure 10.1: 2D Joint Gaussian PDF - $m_x = m_y = 4$, $s_x = 1$, $s_y = 1.6$