Review of Probability

Image Processing

Prof. Vidya Manian
Dept. of Electrical and Comptuer Engineering

A ***random experiment*** is an experiment in which it is not possible to predict the outcome.

Example: Coin toss:

Let n denote the total number of tosses, n_H the number of heads that turn up, and n_T the number of tails. Clearly,

$$n_H + n_T = n.$$

Relative frequency

Dividing both sides by *n* gives

$$\boxed{\frac{n_H}{n} + \frac{n_T}{n} = 1.}$$

The term n_H/n is called the ***relative frequency*** of the event we have denoted by *H*, and similarly for n_T/n.

P(event) : after several tosses

Probability

The first important property of *P* is that, for an event *A*,

$$0 \le P(A) \le 1.$$

That is, the probability of an event is a positive number bounded by 0 and 1. For the certain event, *S*,

$$P(S) = 1.$$

Conditional Probability

The relative frequency of event *A* occurring, ***given that*** event *B* has occurred, is given by

This ***conditional probability*** is denoted by *P*(*A*/*B*), *P*(*A*/*B*) as the ***probability of*** *A* ***given*** *B*.

Bayes' theorem

A little manipulation of the preceding results yields the following important relationships

$$\boxed{P(A/B) = \frac{P(B/A)P(A)}{P(B)}}$$

and

$$\boxed{P(AB) = P(A)P(B/A) = P(B)P(A/B).}$$

The second expression may be written as

$$\boxed{P(B/A) = \frac{P(A/B)P(B)}{P(A)}}$$

which is known as ***Bayes' theorem***, so named after the 18th century mathematician Thomas Bayes.

Example: Suppose that we want to extend the expression

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

to three variables, *A*, *B*, and *C*. Recalling that *AB* is the same as A $\cap$ B, we replace *B* by *B* $\cup$ *C* in the preceding equation to obtain

$$P(A \cup B \cup C) = P(A) + P(B \cup C) - P(A \cap [B \cup C]).$$

The second term in the right can be written as

$$P(B \cup C) = P(B) + P(C) - P(BC).$$

From the Table discussed earlier, we know that

$$A \cap [B \cup C] = (A \cap B) \cup (A \cap C)$$

so,

$$\begin{aligned} P(A \cap [B \cup C]) &= P([A \cap B] \cup [A \cap C]) \\ &= P(AB \cup AC) \\ &= P(AB) + P(AC) - P(ABC). \end{aligned}$$

Collecting terms gives us the final result

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC).$$

Proceeding in a similar fashion gives

$$P(ABC) = P(A)P(B/A)P(C/AB).$$

The preceding approach can be used to generalize these expressions to *N* events.

Statistical Independence

If A and B are ***statistically independent***, then $P(B/A) = P(B)$ and it follows that

$$\boxed{\begin{aligned} P(A/B) &= P(A) \\ P(B/A) &= P(B) \end{aligned}}$$

and

$$\boxed{P(AB) = P(A)P(B).}$$

It was stated earlier that if sets (events) A and B are ***mutually exclusive***, then $A \cap B = \emptyset$ from which it follows that $P(AB) = P(A \cap B) = 0$. As was just shown, the two sets are statistically independent if $P(AB)=P(A)P(B)$, which we assume to be nonzero in general. ***Thus, we conclude that for two events to be statistically independent, they cannot be mutually exclusive***.

For three events *A*, *B*, and *C* to be independent, it must be true that

$$\boxed{\begin{aligned} P(AB) &= P(A)P(B) \\ P(AC) &= P(A)P(C) \\ P(BC) &= P(B)P(C) \end{aligned}}$$

and

$$\boxed{P(ABC) = P(A)P(B)P(C).}$$

In general, for *N* events to be statistically independent, it must be true that, for all combinations $1 \leq i \leq j \leq k \leq \ . \ . \ . \leq N$

$$\boxed{\begin{aligned} P(A_iA_j) &= P(A_i)P(A_j) \\ P(A_iA_jA_k) &= P(A_i)P(A_j)P(A_k) \\ &\vdots \\ P(A_1A_2\cdots A_N) &= P(A_1)P(A_2)\cdots P(A_N). \end{aligned}}$$

Example

(a) An experiment consists of throwing a single die twice. The probability of any of the six faces, 1 through 6, coming up in either experiment is 1/6. Suppose that we want to find the probability that a 2 comes up, followed by a 4. These two events are statistically independent (the second event does not depend on the outcome of the first). Thus, letting *A* represent a 2 and *B* a 4,

$$P(AB) = P(A)P(B) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}.$$

We would have arrived at the same result by defining "2 followed by 4" to be a single event, say *C*. The sample set of all possible outcomes of two throws of a die is 36. Then, *P*(*C*)=1/36.

Example (Con't): (b) Consider now an experiment in which we draw one card from a standard card deck of 52 cards. Let *A* denote the event that a king is drawn, *B* denote the event that a queen or jack is drawn, and *C* the event that a diamond-face card is drawn. A brief review of the previous discussion on relative frequencies would show that

$$P(A) = \frac{4}{52},$$
$$P(B) = \frac{8}{52},$$

and

$$P(C) = \frac{13}{52}.$$

Example (Con't): Furthermore,

$$P(AC) = P(A \cap C) = P(A)P(C) = \frac{1}{52}$$

and

$$P(BC) = P(B \cap C) = P(B)P(C) = \frac{2}{52}.$$

Events *A* and *B* are mutually exclusive (we are drawing only one card, so it would be impossible to draw a king and a queen or jack simultaneously). Thus, it follows from the preceding discussion that $P(AB) = P(A \cap B) = 0$ [and also that $P(AB) \neq P(A)P(B)$].

Example (Con't): (c) As a final experiment, consider the deck of 52 cards again, and let A_1, A_2, A_3, and A_4 represent the events of drawing an ace in each of four successive draws. If we replace the card drawn before drawing the next card, then the events are statistically independent and it follows that

$$\begin{aligned} P(A_1A_2A_3A_4) &= P(A_1)P(A_2)P(A_3)P(A_4) \\ &= \left[\frac{4}{52}\right]^4 \approx 3.5 \times 10^{-5}. \end{aligned}$$

Example (Con't): Suppose now that we do not replace the cards that are drawn. The events then are no longer statistically independent. With reference to the results in the previous example, we write

$$\begin{aligned}
P(A_1A_2A_3A_4) &= P(A_1)P(A_2A_3A_4/A_1) \\
&= P(A_1)P(A_2/A_1)P(A_3A_4/A_1A_2) \\
&= P(A_1)P(A_2/A_1)P(A_3/A_1A_2)P(A_4/A_1A_2A_3) \\
&= \frac{4}{52}\cdot\frac{3}{51}\cdot\frac{2}{50}\cdot\frac{1}{49} \approx 3.7\times 10^{-6}.
\end{aligned}$$

Thus we see that not replacing the drawn card reduced our chances of drawing fours successive aces by a factor of close to 10. This significant difference is perhaps larger than might be expected from intuition.

Random Variables

A ***random variable***, *x*, is a real-valued function ***defined*** on the events of the sample space, *S*.

In words, for each event in *S*, there is a real number that is the corresponding value of the random variable.

a random variable maps each event in *S* onto the real line.

Example: Consider again the experiment of drawing a single card from a standard deck of 52 cards. Suppose that we define the following events. *A*: a heart; *B*: a spade; *C*: a club; and *D*: a diamond, so that *S* = {*A*, *B*, *C*, *D*}. A random variable is easily defined by letting *x* = 1 represent event *A*, *x* = 2 represent event *B*, and so on.

consider the experiment of throwing a single die and observing the value of the up-face. We can define a random variable as the numerical outcome of the experiment (i.e., 1 through 6), but there are many other possibilities. For example, a binary random variable could be defined simply by letting *x* = 0 represent the event that the outcome of throw is an even number and *x* = 1 otherwise.

To handle ***continuous random variables*** we need some additional tools.
For example, given a continuous function we know that the area of the function between two limits *a* and *b* is the integral from *a* to *b* of the function.

However, the area ***at a point*** is zero because the integral from say, *a* to *a* is zero.

We are dealing with the same concept in the case of continuous random variables.

Continuous random variables

Discrete random variables takes discrete values between 0 and 1.

Continuous random variables lies in a specified ***range***. In particular, we are interested in the probability that the random variable is less than or equal to (or, similarly, greater than or equal to) a specified constant *a*. We write this as

$$\boxed{F(a) = P(x \leq a).}$$

If this function is given for all values of *a* (i.e., – ∞ < *a* < ∞), then the values of random variable *x* have been defined. Function *F* is called the ***cumulative probability distribution function*** or simply the ***cumulative distribution function*** (cdf) or ***distribution function***

Cumulative distribution function

$$F_X(x) = P(X \leq x).$$

Cdf F(x) of a random variable.

Properties of cdf’s

1. $F(-\infty) = 0$
2. $F(\infty) = 1$
3. $0 \leq F(x) \leq 1$
4. $F(x_1) \leq F(x_2) \quad \text{if} \quad x_1 < x_2$
5. $P(x_1 < x \leq x_2) = F(x_2) - F(x_1)$
6. $F(x^+) = F(x),$

where $x^+ = x + \varepsilon$, with ε being a positive, infinitesimally small number.

Pdf properties

The ***probability density function*** (pdf) of random variable *x* is defined as the derivative of the cdf:

$$\boxed{p(x) = \frac{dF(x)}{dx}.}$$

The pdf satisfies the following properties:

1. $p(x) \geq 0$ for all x
2. $\int_{-\infty}^{\infty} p(x)dx = 1$
3. $F(x) = \int_{-\infty}^{x} p(\alpha)d\alpha$, where α is a dummy variable
4. $P(x_1 < x \leq x_2) = \int_{x_1}^{x_2} p(x)dx.$

The preceding concepts are applicable to discrete random variables. In this case, there is a finite no. of events and we talk about ***probabilities***, rather than probability density functions.

Integrals are replaced by summations and, sometimes, the random variables are subscripted.

For example, in the case of a discrete variable with N possible values we would denote the probabilities by $P(x_i)$, i=1, 2,…, N.

In Sec. 3.3 of the book we used the notation $p(r_k)$, $k = 0,1,\ldots, L - 1$, to denote the ***histogram*** of an image with L possible gray levels, r_k, $k = 0,1,\ldots, L - 1$, where $p(r_k)$ is the probability of the kth gray level (random event) occurring.

The discrete random variables in this case are gray levels.

Uppercase letters (e.g., P) are frequently used to distinguish between probabilities and probability density functions (e.g., p) when they are used together in the same discussion.

Random variable transformation

If a random variable *x* is ***transformed*** by a monotonic transformation function *T*(*x*) to produce a new random variable *y*, the probability density function of *y* can be obtained from knowledge of *T*(*x*) and the probability density function of *x*, as follows:

$$\boxed{p_y(y) = p_x(x)\left|\frac{dx}{dy}\right|}$$

where the subscripts on the *p*'s are used to denote the fact that they are different functions, and the vertical bars signify the absolute value.

Expected value

The ***expected value*** of a function *g*(*x*) of a ***continuous*** random variable is defined as

$$E[g(x)] \; = \int_{-\infty}^{\infty} g(x)p(x)dx.$$

If the random variable is ***discrete*** the definition becomes

$$E[g(x)] \; = \sum_{i=1}^{N} g(x_i)P(x_i).$$

The expected value is one of the operations used most frequently when working with random variables. For example, the expected value of random variable *x* is obtained by letting *g*(*x*) = *x*:

$$E[x] = \overline{x} = m = \int_{-\infty}^{\infty} xp(x)dx$$

when *x* is continuous and

$$E[x] = \overline{x} = m = \sum_{i=1}^{N} x_i P(x_i)$$

when *x* is discrete. The expected value of *x* is equal to its ***average*** (or ***mean***) ***value***, hence the use of the equivalent notation $\overline{x}$ and *m*.

The ***variance*** of a random variable, denoted by σ^2, is obtained by letting $g(x) = x^2$ which gives

$$\boxed{\sigma^2 = E[x^2] = \int_{-\infty}^{\infty} x^2 p(x)dx}$$

for continuous random variables and

$$\boxed{\sigma^2 = E[x^2] = \sum_{i=1}^{N} x_i^2 P(x_i)}$$

for discrete variables.

Of particular importance is the variance of random variables that have been ***normalized*** by subtracting their mean. In this case, the variance is

$$\sigma^2 = E[(x-m)^2] = \int_{-\infty}^{\infty}(x-m)^2 p(x)dx$$

and

$$\boxed{\sigma^2 = E[(x-m)^2] = \sum_{i=1}^{N}(x_i - m)^2 P(x_i)}$$

for continuous and discrete random variables, respectively. The square root of the variance is called the ***standard deviation***, and is denoted by σ.

*n*th ***central moment*** of a continuous random variable

$g(x) = (x - m)^n$:

$$\mu_n = E[(x-m)^n] = \int_{-\infty}^{\infty} (x-m)^n p(x)dx$$

$$\mu_n = E[(x-m)^n] = \sum_{i=1}^{N} (x_i - m)^n P(x_i)$$

for discrete variables, where we assume that $n \geq 0$. Clearly, $\mu_0=1$, $\mu_1=0$, and $\mu_2=\sigma^2$. The term ***central*** when referring to moments indicates that the mean of the random variables has been subtracted out.

The moments defined above in which the mean is not subtracted out sometimes are called ***moments about the origin***.

In image processing, moments are used for a variety of purposes, including histogram processing, segmentation, and description.

In general, moments are used to characterize the probability density function of a random variable.

The second, third, and fourth central moments are intimately related to the ***shape*** of the probability density function of a random variable.
The second central moment (the centralized variance) is a measure of ***spread*** of values of a random variable about its mean value,
the third central moment is a measure of ***skewness*** (bias to the left or right) of the values of *x* about the mean value, and the fourth moment is a relative measure of ***flatness***. In general, knowing all the moments of a density specifies that density.

Example: Consider an experiment consisting of repeatedly firing a rifle at a target, and suppose that we wish to characterize the behavior of bullet impacts on the target in terms of whether we are shooting high or low.. We divide the target into an upper and lower region by passing a horizontal line through the bull's-eye. The events of interest are the vertical distances from the center of an impact hole to the horizontal line just described. Distances above the line are considered positive and distances below the line are considered negative. The distance is zero when a bullet hits the line.

In this case, we define a random variable directly as the value of the distances in our sample set. Computing the mean of the random variable indicates whether, ***on average***, we are shooting high or low. If the mean is zero, we know that the average of our shots are on the line. However, the mean does not tell us how far our shots deviated from the horizontal. The variance (or standard deviation) will give us an idea of the ***spread of the shots***. A small variance indicates a tight grouping (with respect to the mean, and in the vertical position); a large variance indicates the opposite. Finally, a third moment of zero would tell us that the spread of the shots is symmetric about the mean value, a positive third moment would indicate a high bias, and a negative third moment would tell us that we are shooting low more than we are shooting high with respect to the mean location.

Gaussian probability density function

A random variable is called ***Gaussian*** if it has a probability density of the form

$$\boxed{p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-m)^2/\sigma^2}}$$

where *m* and σ are as defined in the previous section. The term ***normal*** also is used to refer to the Gaussian density.

A plot and properties of this density function are given in Section 5.2.2 of the book.

The cumulative distribution function corresponding to the Gaussian density is

$$\boxed{\begin{aligned} F(x) &= \int_{-\infty}^{x} p(x)dx \\ &= \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x} e^{-(x-m)^2/\sigma^2}dx. \end{aligned}}$$

which, as before, we interpret as the probability that the random variable lies between minus infinite and an arbitrary value *x*. This integral has no known closed-form solution, and it must be solved by numerical or other approximation methods. Extensive tables exist for the Gaussian cdf.

Several random variables

In general, we consider in this section the case of *n* random variables, which we denote by x_1, x_2,…, x_n (the use of *n* here is not related to our use of the same symbol to denote the *n*th moment of a random variable).

It is convenient to use vector notation when dealing with several random variables. Thus, we represent a ***vector random variable*** **x** as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Then, for example, the cumulative distribution function introduced earlier becomes

$$\begin{aligned} F(\mathbf{a}) &= F(a_1, a_2, \ldots, a_n) \\ &= P\{x_1 \leq a_1, x_2 \leq a_2, \ldots, x_n \leq a_n\} \end{aligned}$$

the ***cdf of a random variable vector*** often is written simply as *F*(**x**).

As in the single variable case, the ***probability density function of a random variable vector*** is defined in terms of derivatives of the cdf; that is,

$$
\begin{aligned}
p(\mathbf{x}) &= p(x_1, x_2, \dots, x_n) \\
&= \frac{\partial^n F(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n}.
\end{aligned}
$$

The ***expected value*** of a function of **x** is defined basically as before:

$$\begin{aligned} E[g(\mathbf{x})] &= E[g(x_1, x_2, \ldots, x_n)] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) p(x_1, x_2, \ldots, x_n) dx_1 dx_2 \cdots dx_n. \end{aligned}$$

Cases dealing with expectation operations involving pairs of elements of **x** are particularly important. For example, the joint moment (about the origin) of order *kq* between variables x_i and x_j

$$\eta_{kq}(i,j) = E[x_i^k x_j^q] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_i^k x_j^q p(x_i, x_j) dx_i dx_j.$$

When working with any two random variables (any two elements of **x**) it is common practice to simplify the notation by using *x* and *y* to denote the random variables. In this case the joint moment just defined becomes

$$\boxed{\eta_{kq} = E[x^k y^q] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^q p(x,y)dxdy.}$$

It is easy to see that η_{k0} is the *k*th moment of *x* and η_{0q} is the *q*th moment of *y*, as defined earlier.

The moment $\eta_{11} = E[xy]$ is called the ***correlation*** of x and y. As we will see in Chapter 4, correlation is an important concept in image processing. In fact, it is important in most areas of signal processing, where typically it is given a special symbol, such as R_{xy}:

$$\boxed{R_{xy} = \eta_{11} = E[xy] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyp(x,y)dxdy.}$$

If the condition

$$\boxed{R_{xy} = E[x]E[y]}$$

holds, then the two random variables are said to be ***uncorrelated***. From our earlier discussion, we know that if *x* and *y* are ***statistically independent***, then *p*(*x*, *y*) = *p*(*x*)*p*(*y*), in which case we write

$$\boxed{R_{xy} = \int_{-\infty}^{\infty} xp(x)dx \int_{-\infty}^{\infty} yp(y)dy = E[x]E[y].}$$

Thus, we see that ***if two random variables are statistically independent then they are also uncorrelated***. The converse of this statement is ***not*** true in general.

The joint central moment of order *kq* involving random variables *x* and *y* is defined as

$$\boxed{\begin{aligned}\mu_{kq} &= E[(x-m_x)^k(y-m_y)^q] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-m_x)^k(y-m_y)^q p(x,y)dxdy\end{aligned}}$$

where $m_x = E[x]$ and $m_y = E[y]$ are the means of *x* and *y*, as defined earlier. We note that

$$\boxed{\mu_{20} = E[(x-m_x)^2]} \quad \text{and} \quad \boxed{\mu_{02} = E[(y-m_y)^2]}$$

are the variances of *x* and *y*, respectively.

The moment μ_{11}

$$\boxed{\begin{aligned}\mu_{11} &= E[(x-m_x)(y-m_y)] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-m_x)(y-m_y)p(x,y)dxdy\end{aligned}}$$

is called the ***covariance*** of *x* and *y*. As in the case of correlation, the covariance is an important concept, usually given a special symbol such as C_{xy}.

By direct expansion of the terms inside the expected value brackets, and recalling the $m_x = E[x]$ and $m_y = E[y]$, it is straightforward to show that

$$\boxed{\begin{aligned} C_{xy} &= E[xy] - m_y E[x] - m_x E[y] + m_x my \\ &= E[xy] - E[x]E[y] \\ &= R_{xy} - E[x]E[y]. \end{aligned}}$$

From our discussion on correlation, we see that the covariance is zero if the random variables are either uncorrelated ***or*** statistically independent.

This is an important result worth remembering.

If we divide the covariance by the square root of the product of the variances we obtain

$$\boxed{\begin{aligned}\gamma &= \frac{\mu_{11}}{\sqrt{\mu_{20}\mu_{02}}}\\ &= \frac{C_{xy}}{\sigma_x\sigma_y}\\ &= E\left[\frac{(x-m_x)}{\sigma_x}\,\frac{(y-m_y)}{\sigma_y}\right].\end{aligned}}$$

The quantity γ is called the ***correlation coefficient*** of random variables *x* and *y*. It can be shown that γ is in the range $-1 \leq \gamma \leq 1$

Multivariate Gaussian density

As an illustration of a probability density function of more than one random variable, we consider the ***multivariate Gaussian probability density function***, defined as

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{C}|^{1/2}} e^{-\frac{1}{2}\left[(\mathbf{x}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})\right]}$$

where *n* is the ***dimensionality*** (number of components) of the random vector **x**, **C** is the ***covariance matrix*** (to be defined below), |**C**| is the determinant of matrix **C**, **m** is the ***mean vector*** (also to be defined below) and *T* indicates transposition

The ***mean vector*** is defined as

$$\boxed{\mathbf{m} = E[\mathbf{x}] = \begin{bmatrix} E[x_1] \\ E[x_2] \\ \vdots \\ E[x_n] \end{bmatrix}}$$

and the ***covariance matrix*** is defined as

$$\boxed{\mathbf{C} = E[(\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^T].}$$

The element of **C** are the covariances of the elements of **x**, such that

$$\boxed{c_{ij} = C_{x_i x_j} = E[(x_i - m_i)(x_j - m_j)]}$$

where, for example, x_i is the *i*th component of **x** and m_i is the *i*th component of **m**.

Covariance matrices are ***real*** and ***symmetric***.

The elements along the main diagonal of **C** are the variances of the elements **x**, such that $c_{ii}= \sigma_{x_i}^{2}$.

When all the elements of **x** are uncorrelated or statistically independent, $c_{ij} = 0$, and the covariance matrix becomes a ***diagonal matrix***.

If all the variances are equal, then the covariance matrix becomes proportional to the ***identity matrix***, with the constant of proportionality being the variance of the elements of **x**.

Example: Consider the following ***bivariate*** (*n* = 2) Gaussian probability density function

$$\boxed{p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{C}|^{1/2}} e^{-\frac{1}{2}\left[(\mathbf{x}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})\right]}}$$

with

$$\mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}$$

and

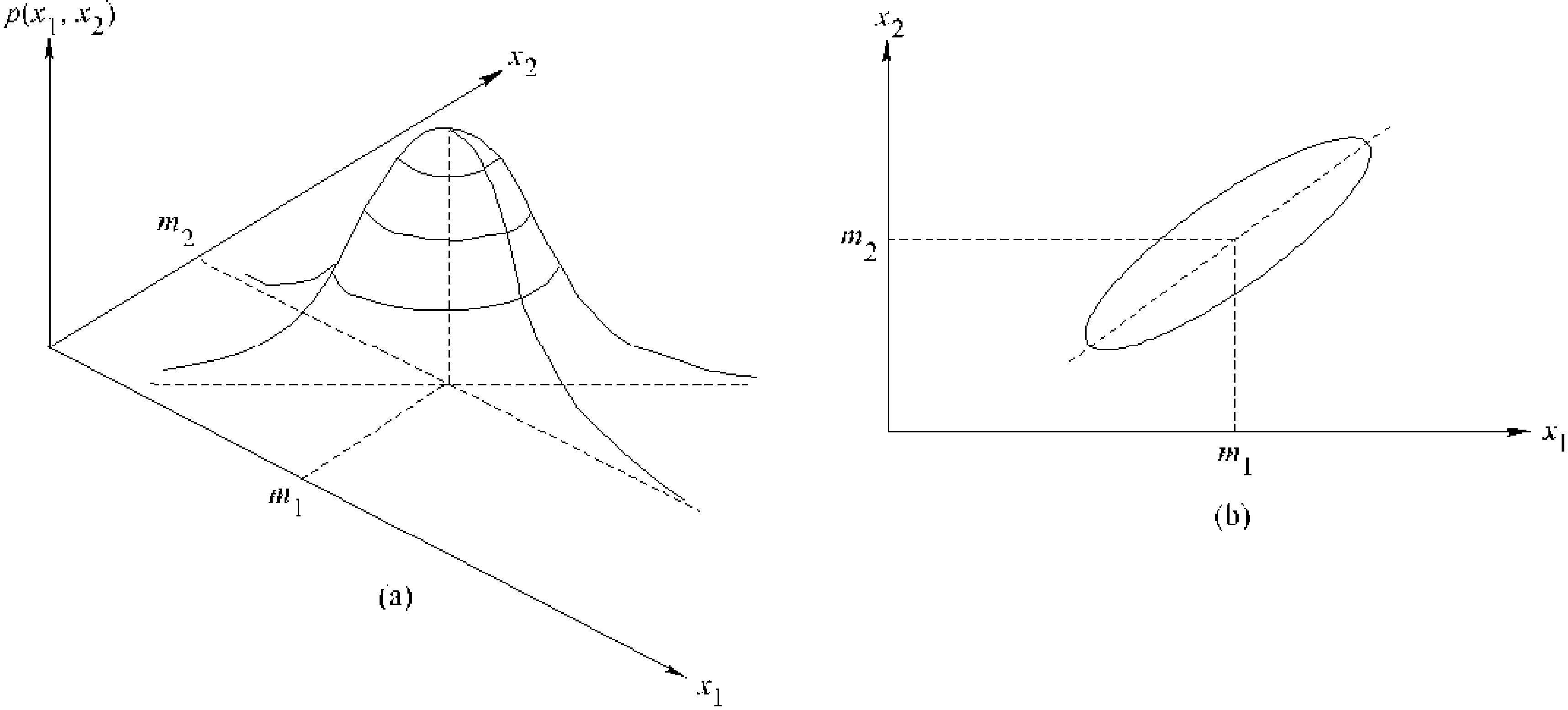
$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

where, because **C** is known to be symmetric, $c_{12} = c_{21}$.
A schematic diagram of this density is shown in Part (a) of the following figure.

Part (b) is a horizontal slice of Part (a)

The main directions of data spread are in the directions of the eigenvectors of **C**. Furthermore, if the variables are uncorrelated or statistically independent, the covariance matrix will be diagonal and the eigenvectors will be in the same direction as the coordinate axes x_1 and x_2 (and the ellipse shown would be oriented along the x_1 - and x_2-axis).
If, the variances along the main diagonal are equal, the density would be symmetrical in all directions (in the form of a bell) and Part (b) would be a circle. Note in Parts (a) and (b) that the density is centered at the mean values (m_1,m_2).

Multivariate Gaussian density



(a)

(b)

Linear transformation of random variables

A linear transformation of a vector **x** to produce a vector **y** is of the form **y** = **Ax**. Of particular importance in our work is the case when the rows of **A** are the eigenvectors of the covariance matrix.

Because **C** is real and symmetric, it is always possible to find *n* orthonormal eigenvectors from which to form **A**.