Communications Engineering

Probability Theory and Random Variables

Probability Theory

- *Probability theory* is concerned with phenomena that can be modelled by an experiment that is subject to *chance*.
- Such a chance experiment is referred to as a *random experiment*, e.g. tossing a fair coin whose possible outcomes are “heads” or “tails”.

Probability Theory

- To be more precise in the description of a random experiment, we require *three* features:
  1. The experiment is repeatable under identical conditions.
  2. On any trial of the experiment, the outcome is unpredictable.
  3. For a large number of trials of the experiment, the outcomes exhibit *statistical regularity*, i.e. a definite *average* pattern of outcomes is observed if the experiment is repeated over a large number of times.

Relative Frequency

- Let *event* A denote one of the possible outcomes of a random experiment, e.g. A could denote "heads" in a coin-tossing experiment.
- Suppose that in n trials of the experiment, event A occurs $N_n(A)$ times.
- Then the ratio

$$0 \leq \frac{N_n(A)}{n} \leq 1$$

- denotes the *relative frequency* of event A.

Relative Frequency and Probability

- The experiment exhibits *statistical regularity* if for any sequence of n trials the relative frequency converges to the same limit as n becomes large.
- Thus, one may define the *probability* of event A as

$$P(A) = \lim_{n \to \infty} \left(\frac{N_n(A)}{n} \right)$$

- The probability of an event is intended to represent the likelihood that a trial of the experiment will result in the occurrence of that event.

Probability Theory

- A more formal definition of *probability* is the following.
- Let A be an event, or *outcome*, of a random experiment, and $P(A)$ be a real number called the probability of A. Then the following three axioms serve to define probability:
  - $0 < P(A) < 1$
  - $P(\mathbf{S}) = 1$ where $\mathbf{S}$ is the sample space or totality of all possible outcomes (i.e. the sure event).
  - If $A + B$ is the union of two mutually exclusive events, then $P(A+B) = P(A) + P(B)$.

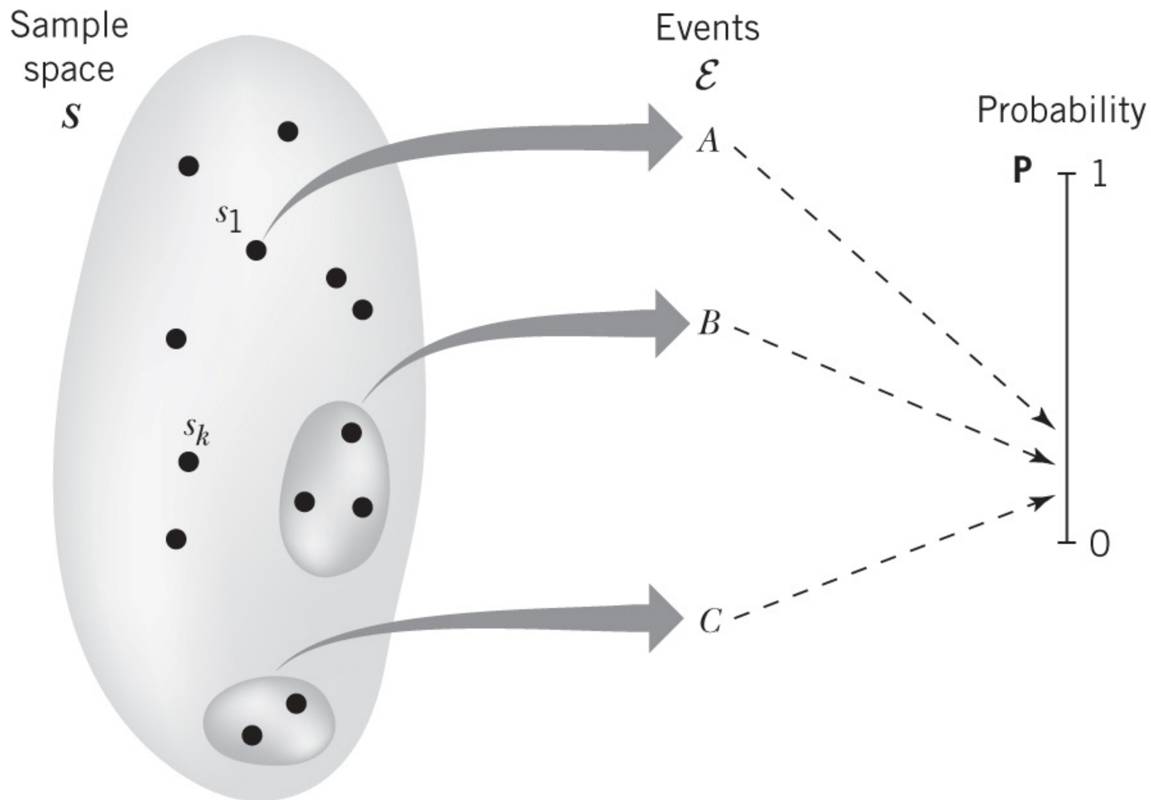


Illustration of the relationship between
sample space, events, and probability

Properties of Probability

- The *complement* of event *A*, denoted *A*, is given by

$$P(A) = 1 - P(A)$$

- This property is used to investigate the *non-occurrence* of an event.
- If M mutually exclusive events $A_1, A_2, \ldots, A_M$ have the exhaustive property that $A_1+A_2+\ldots+A_M = \boldsymbol{S}$, then $P(A_1)+P(A_2)+\ldots+P(A_M) = 1$.

Properties of Probability

- When events A and B are <u>not</u> mutually exclusive, then the probability of the *union* event "A OR B" equals

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

- where $P(AB)$ is called the probability of the *joint* event "A AND B".
- The probability $P(AB)$ is called a *joint probability*, and represents the probability of events A and B occurring simultaneously.

Conditional Probability

- Suppose we perform an experiment that involves a pair of events A and B. Let $P(B|A)$ denote the probability of event B <u>given</u> that event A has occurred.
- The probability $P(B|A)$ is called the *conditional probability* of B given A.
- Assuming that A has a nonzero probability, the conditional probability $P(B|A)$ is defined by

$$P(B \mid A) = \frac{P(AB)}{P(A)}$$

Conditional Probability

- Similarly

$$P(A \mid B) = \frac{P(AB)}{P(B)}$$

- In other words, the joint probability may be expressed in terms of the conditional probabilities as follows:

$$\begin{aligned} P(AB) &= P(B \mid A)P(A) \\ &= P(A \mid B)P(B) \end{aligned}$$

Conditional Probability

- This is an important result known as *Bayes' Rule* and it defines the relationship between the conditional and joint probabilities.
- If the events A and B are statistically independent, i.e. the occurrence of event A does not influence the occurrence of B, then

$$P(AB) = P(A)P(B)$$

Conditional Probability

- Then the conditional probabilities simply reduce to the elementary probabilities as follows:

$$P(B \mid A) = P(B)$$

$$P(A \mid B) = P(A)$$

Notation

- It is important to be aware of the different notation that is sometimes used. For example:
- AB represents the *intersection* of events A and B, and is sometimes denoted $A \cap B$ (by mathematicians). It is analogous to the AND operation in digital logic.
- $A+B$ represents the *union* of events A and B, and is sometimes denoted $A \cup B$ (by mathematicians). It is analogous to the OR operation in digital logic.

Random Variables

- A *random variable* is a rule, or a functional relationship, that assigns a real number to each possible outcome of a random experiment. This is convenient from the standpoint of analysis.
- A standard notation is to denote random variables by capital letters (X, Y etc.) and the values they take on by the corresponding lower case letters (x,y etc.).
- Random variables may be discrete, continuous, or mixed depending on whether they take on countable (discrete) or uncountable (continuous) number of values or both.

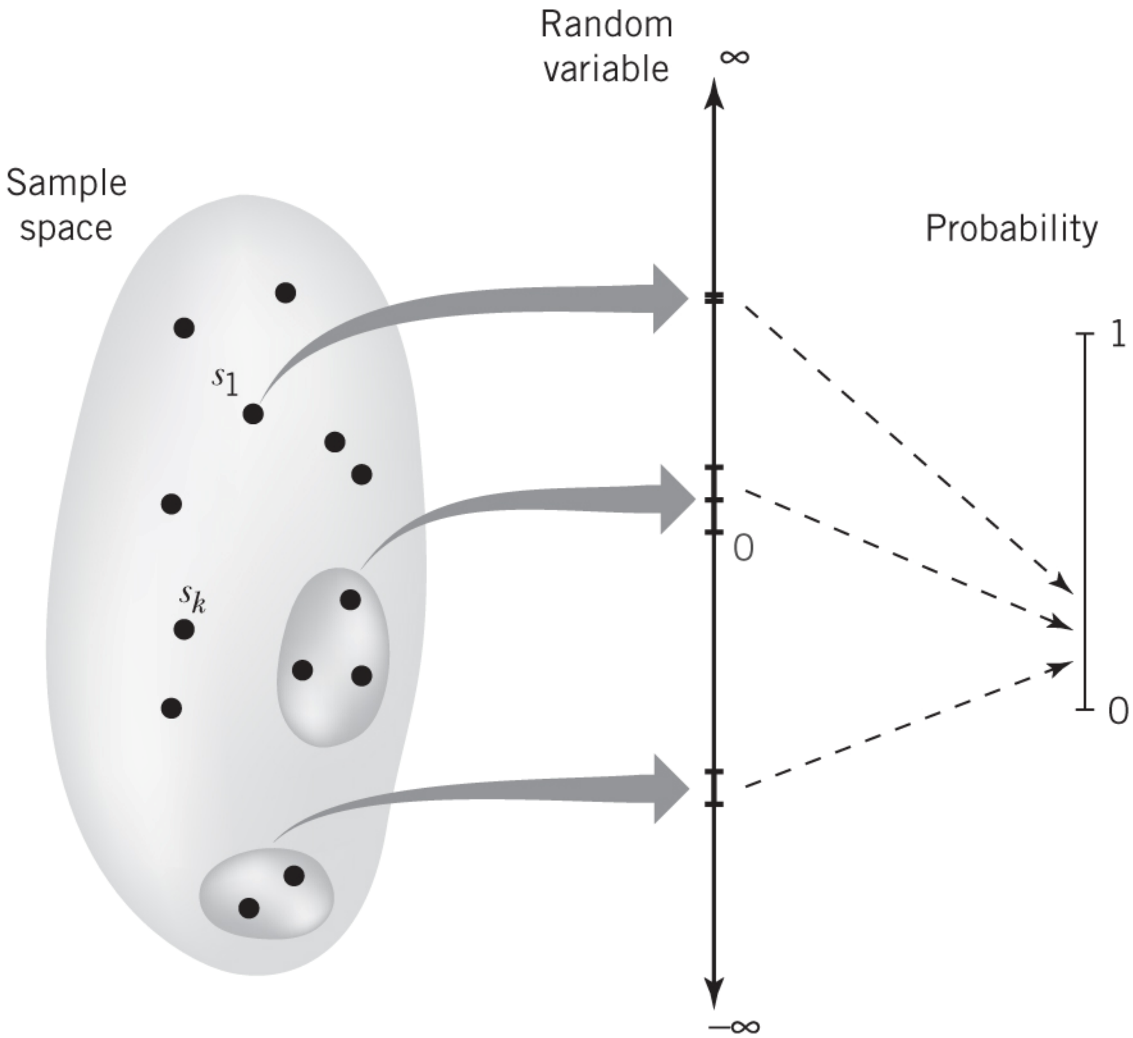


Illustration of the relationship between sample space, random variable, and probability

Random Variables

- For example, consider that the sample space S denotes the outcomes of the throw of a die and consists of a set of six sample points that may be taken to be the integers $1,2,\ldots,6$.
- Then, if we identify the sample point k with the event that k dots show when the die is cast, the function $X(k) = k$ is random variable such that $X(k)$ equals the numbers of dots showing when the die is cast.
- Here $X(k)$ is a *discrete random variable* or more precisely the random variable $X(k)$ can only take on a finite number of values in any finite interval.

Random Variables

- If however, the random variable X can take on any value in a whole observation interval, then X is called a *continuous random variable.*
- An example of this might be where X represents the amplitude of a noise voltage at a particular instant of time and thus can take on any value between plus and minus infinity.
- Clearly, it is necessary to develop a probabilistic description of random variables.

Distribution Function

- Consider the random variable X and the probability of the event $X \leq x$ which is denoted by $P(X \leq x)$.
- Clearly, this probability is a function of the dummy variable *x*, where

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

- The function $F_X(x)$ is called the *cumulative distribution function (cdf)* or simply the *distribution function* of the random variable X.

Distribution Function

- The distribution function $F_X(x)$ has the following properties:

$$0 \le F_X(x) \le 1$$

$$F_X(x_1) \le F_X(x_2) \qquad if \qquad x_1 \le x_2$$

$$F_X(-\infty) = 0$$

$$F_X(+\infty) = 1$$

Probability Density Function

- An alternative description of the probability of the random variable X is the *probability density function (pdf)* denoted $p_X(x)$ where,

$$p_X(x) = \frac{dF_X(x)}{dx}$$

- Note that as with the distribution function, the *pdf* is a function of the real variable x.

Probability Density Function

- The name "density function" arises from the fact that the probability of the event $x_1 \le X \le x_2$ equals

$$P(x_1 \le X \le x_2) = P(X \le x_2) - P(X \le x_1)$$
$$= F_X(x_2) - F_X(x_1)$$
$$= \int_{x_1}^{x_2} p_X(x)dx$$

- The probability of an interval is therefore the area under the *pdf* in that interval.

Probability Density Function

- The probability density function (*pdf*) has the following properties

$$p_X(x) \geq 0$$

$$\int_{-\infty}^{\infty} p_X(x)dx = F_X(+\infty) - F_X(-\infty) = 1$$

- The probability density function (*pdf*) is always a nonnegative function with a total area of one.

Statistical Averages

- One of the primary uses of probability theory is to evaluate the *average* value of a random variable (which represents some physical phenomenon) or to evaluate the average value of some function of the random variable.
- The *mean* value m_X or *expected* value of a random variable X is defined by

$$m_X = E[X] = \int_{-\infty}^{\infty} x p_X(x) dx$$

Expectation Operator

- Here, $\mathrm{E}[X]$ is the *expectation operator* and it is used to obtain the mean value for both discrete and continuous random variables.
- Furthermore, it is worth noting the expectation operator $\mathrm{E}[X]$ is linear.
- Considering the more general case where X is a random variable and $g(x)$ is a real valued function, if the argument of $g(x)$ is the random variable X where

$$Y = g(X)$$

Expectation Operator

- Then, the mean or expected value m_Y of Y is

$$m_Y = E[Y]$$

$$= \int_{-\infty}^{\infty} y p_Y(y)\, dy$$

$$= \int_{-\infty}^{\infty} g(x)\, p_X(x)\, dx$$

- In other words

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)\, p_X(x)\, dx$$

Moments

- For the special case of $g(X) = X^n$, we obtain the n^{th} *moment* of the probability distribution of the random variable X, i.e.

$$E[X^n] = \int_{-\infty}^{\infty} x^n p_X(x)dx$$

- For the purposes of communications systems analysis, the most important moments are the first two moments. Setting $n = 1$ gives the *mean* of the random variable whereas setting $n = 2$ gives the *mean square value*.

Central Moments

- One can also define *central moments*, which are simply the moments of the differences between a random variable X and its mean m_X.
- Thus, the n^{th} central moment is

$$E[(X - m_X)^n] = \int_{-\infty}^{\infty} (x - m_X)^n p_X(x) dx$$

- For $n = 1$, the central moment is zero. However, for $n = 2$ the central moment is referred to as the *variance* of the random variable X.

Variance

- The variance *var(X)* of the random variable X is defined by

$$\mathrm{var}(X) = E[(X - m_X)^2]$$

$$= \int_{-\infty}^{\infty} (x - m_X)^2 p_X(x) dx$$

- The variance of a random variable X is commonly denoted σ_X^2 where the square of the variance, namely σ_X, is called the *standard deviation* of X.

Chebyshev's Inequality

- The variance σ_X^2 is a measure of the "randomness" of the random variable X.
- By specifying the variance σ_X^2 we are constraining the width of probability density $p_X(x)$ of the random variable X function about its mean m_X.
- A precise statement of this constraint is given by *Chebyshev's Inequality* which states that for any positive number ε, we have

$$P\left(|X - m_X| \geq \varepsilon\right) \leq \frac{\sigma_X^2}{\varepsilon^2}$$

Variance and Mean-Square Value

- The variance σ_X^2 and the mean-square value $E[X^2]$ are related by

$$\begin{aligned}\sigma_X^2 &= E\left[X^2 - 2m_X X + m_X^2\right] \\ &= E\left[X^2\right] - 2m_X E\left[X\right] + m_X^2 \\ &= E\left[X^2\right] - m_X^2\end{aligned}$$

- In other words, the variance is equal to the difference between the mean-square value and the square of the mean.

Joint Moments

- If X and Y are two random variable with a joint (or bivariate) probability density function $p_{XY}(x,y)$, then the *joint moments* of $p_{XY}(x,y)$ are defined by

$$E[X^m Y^n] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^m y^n p_{XY}(x,y)\,dx\,dy$$

- A special case is when $m = n = 1$ where the resulting joint moment is known as the *correlation*.

Correlation

- The *correlation* of two random variables X and Y is defined by

$$E[XY] = \int_{-\infty}^{\infty} xyp_{XY}(x, y)dxdy$$

- The two random variables X and Y are said to be *orthogonal* if and only if their correlation is zero.

$$E[XY] = 0$$

Joint Central Moment

- The *joint central moments* of $p_{XY}(x,y)$ are defined by

$$E[(X-m_X)^m(Y-m_Y)^n] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-m_X)^m(y-m_Y)^n\, p_{XY}(x,y)\,dx\,dy$$

- In this case, when $m = n = 1$, the joint central moment is called the *covariance* of X and Y.

$$\begin{aligned}\operatorname{cov}[XY] &= E[(X-m_X)(Y-m_Y)] \\ &= E[(X-E[X])(Y-E[Y])] \\ &= E[XY]-E[X]E[Y]\end{aligned}$$

Covariance

- Since the calculation of the covariance involves first subtracting the mean parts of X and Y before correlating the resulting zero mean variable, the covariance refers to the correlation of the *varying* parts of X and Y .
- If X and Y are zero mean random variables, then the correlation and the covariance are the same.

Correlation Coefficient

- If the covariance of X and Y is normalised with respect to their variances σ_X^2 and σ_Y^2 respectively, then the *correlation coefficient* ρ results where

$$\rho = \frac{\operatorname{cov}[XY]}{\sigma_X^2 \sigma_Y^2}$$

$$\rho = 0$$

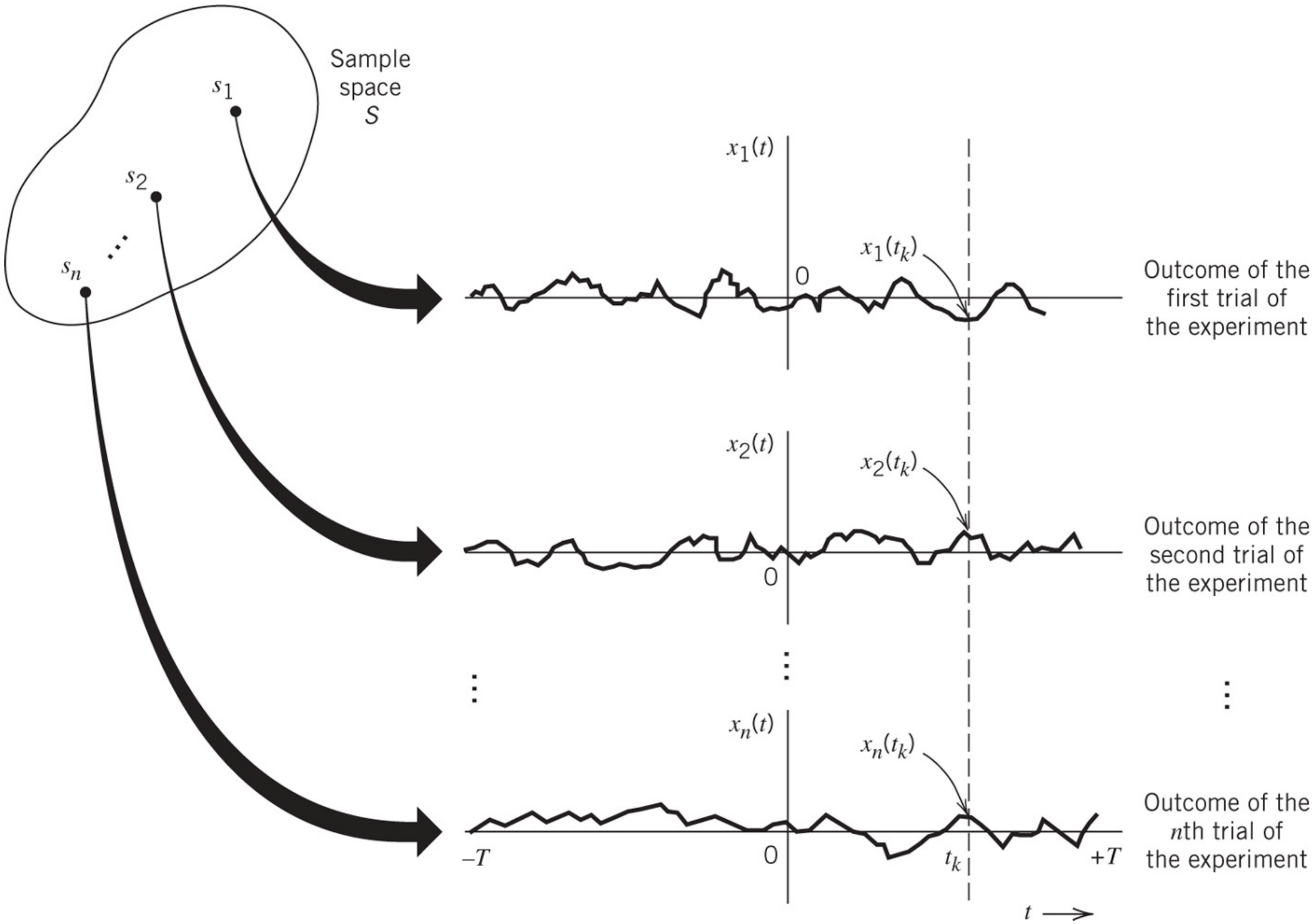
- The two random variable X and Y are said to be *uncorrelated* if and only if their correlation coefficient is zero.

Statistical Independence

- It is intuitively obvious that *statistically independent* random variables (i.e. random variables arising from physically separate processes) must be uncorrelated.
- However, uncorrelated random variables are <u>not</u> necessarily statistically independent.
- Consider two signals described by $\cos\omega t$ and $\sin\omega t$ which although they will have a zero correlation, are not necessarily independent since they differ in terms of a phase shift only.
- It follows then that independence is a stronger (i.e. more restrictive) statistical condition than uncorrelatedness.

Random Processes

- A *random process*, $X(A,t)$, can be viewed as a function of two variables, an event A and time, and is strictly defined in terms of an *ensemble* (or collection) of time functions.
- For a specific time t_k, $X(A,t)$, is a random variable $X(t_k)$ whose value simply depends on the event.
- For a specific event A_j and a specific time t_k, $X(A_j,t_k)$ is simply a number.
- By convention, a random process is usually denoted $X(t)$ where the dependence on A is



An ensemble of sample functions

Random Processes

- A random process whose distribution functions are continuous can be described statistically with a probability density function (*pdf*).
- In general, the form of the *pdf* of a random process will be different for different times.
- In most cases it is not practical to determine empirically the probability distribution of a random process.
- However, a partial description based upon the *mean* and the *autocorrelation function* is often adequate for analysing communications systems.

Random Processes

- The *mean* of a random process $X(t)$ is defined by

$$E[X(t_k)] = \int_{-\infty}^{\infty} x p_{X_k}(x) dx = m_X(t_k)$$

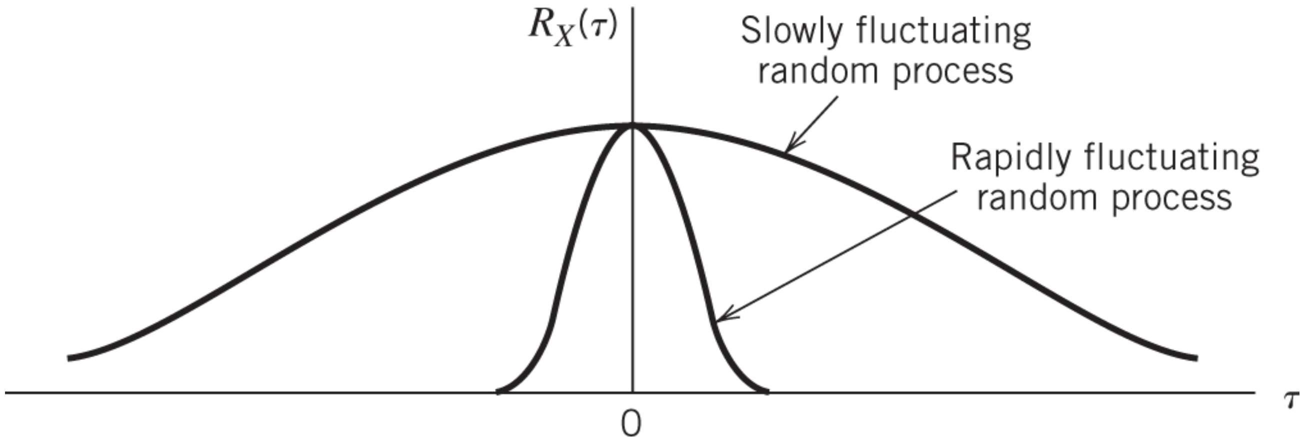
- where $X(t_k)$ is the random variable obtained by observing the random process at a time t_k and $p_{Xk}(x)$ is the *pdf* of $X(t_k)$ over the *ensemble* of events at time t_k.

Random Processes

- The *autocorrelation function* of the random process $X(t)$ is a function of two variables t_1 and t_2 as follows

$$R_X(t_1, t_2) = E[X(t_1)X(t_2)]$$

- where $X(t_1)$ and $X(t_2)$ are the random variables obtained by observing $X(t)$ at time t_1 and t_2 respectively.
- The autocorrelation function is a measure of the degree to which two time samples of the same random process are related.



Illustrating the autocorrelation functions of slowly and rapidly fluctuating random process

Stationarity

- A random process $X(t)$ is said to be *stationary* in the *strict sense* if none of its statistics are affected by a shift in the time origin.
- A random process is said to be *wide-sense stationary (WSS)* if two of its statistics, its mean and autocorrelation function, do not vary with a shift in the time origin.
- Strict sense stationary implies wide sense stationary, but the reverse does not hold.

Stationarity

- A process is said to be wide-sense stationary if

$$E[X(t)] = m_X = constant$$

and

$$R_X(t_1,t_2) = R_X(t_1-t_2)$$

- Most of the analysis of communications systems is predicated on information signals and noise being wide-sense stationary.
- From a practical point of view, it is not necessary for a random process to be stationary for all time, but only for the observation interval of interest.

Autocorrelation Function

- Just as the variance provides a measure of randomness for random variables, the autocorrelation function provides a similar measure for random processes.
- For a wide-sense stationary process, the autocorrelation function is only a function of the time difference $\tau = t_1 - t_2$, i.e.

$$R_X(\tau) = E[X(t)X(t+\tau)] \qquad for -\infty < \tau < \infty$$

Autocorrelation Function

- For a zero mean wide-sense stationary process, $R_X(\tau)$ indicates the extent to which the random values of the process separated by τ seconds in time are statistically correlated.
- The properties of a real-valued wide-sense stationary process are:

$R_X(\tau) = - R_X(\tau)$ *symmetrical about zero*

$R_X(\tau) < R_X(0)$ for all τ *max value occurs at the origin*

$R_X(\tau) \Leftrightarrow G_X(f)$ *Wiener-Khintchine theorem*

$R_X(0) = E[X^2(t)]$ *Value at origin is equal to the average power*

Ergodicity and Time Averaging

- In order to compute m_X and $R_X(\tau)$ using ensemble averaging, one would average across all the sample functions of the process.
- When a random process belongs to a special class, known as *ergodic processes*, its time averages equal its ensemble averages and the statistical properties of the process can be determined by *time averaging* over a single sample function.
- For a random process to be *ergodic*, it must be stationary in the strict sense (the reverse does not hold).

Ergodicity and Time Averaging

- Ergodicity relates to the equivalence of ensemble and time averages whereby the mean may be calculates using

$$m_X = \lim_{T \to \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} X(t)\,dt$$

- Similarly, for the autocorrelation function

$$R_X(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} X(t)\,X(t+\tau)\,dt$$

Ergodicity and Time Averaging

- Testing for the ergodicity of a random process is usually quite difficult.
- In practice, one has to make an intuitive judgement as to whether it is reasonable to interchange time and ensemble averages.
- A reasonable assumption in the analysis of most communication signals is that the random waveforms are ergodic in the mean and autocorrelation function.