

Appendix 2

Probability, statistics & time-series: basic principles¹

A 4.1 Probability distributions & random variables

A 2.1.2 Random variables

A *sample point* is a point that is associated with each outcome of an experiment.

A *sample space* S is the totality of points $s_i (i=1,2,\dots,n)$ that correspond to all possible outcomes of an experiment.

Finite sample spaces contain a countable number of sample points.

A *countably, or denumerable, infinite sample space* contains an infinite number of sample points that can be placed in one-to-one correspondence with the set of natural numbers \mathbb{N} . ($\mathbb{N} = \{ 1, 2, 3 \dots \}$)

A *discrete sample space* is a space that contains a finite or countably-infinite number of sample points.

A *continuous sample space* is a space that contains sample points that form a continuum.

A *random variable* is a function $x(s)$ whose value is defined at each sample point $s_1, s_2, s_3, \dots, s_n$ in a sample space S .

A 2.1.3 Probability density functions

If x represents a continuous random variable in a sample space S , the probability density function (or simply a probability density), $p(x)$ is a function that satisfies the following conditions:

- (1) $p(x) \geq 0$, for all $x \in S$. [The probability of every sample in S is ≥ 0 .]
- (2) $\int_S P(x) dx = 1$ [These probabilities all sum to 1.]
- (3) For any $x_1 \leq x_2$ in S , the probability of x lying in the range $x_1 \leq x \leq x_2$ is

¹ While no originality is claimed over these brief notes, no identifiable individual sources are acknowledged. They are informally compiled here for the benefit of those reading Chapter 5 who may not be familiar with the underlying statistical principles. One of the difficulties in reading mathematics is the pronunciation of unfamiliar symbols. The intention behind the inclusion of the Greek alphabet in the footer is to assist the reader in that process.

A2-2 Appendix 2

$$\int_{x_1}^{x_2} p(x) dx = P(x_1 \leq x \leq x_2).$$

Note that $P(x_1 \leq x \leq x_2)$ represents the area under the graph $y = p(x)$ in the xy -plane between two arbitrary points x_1 and x_2 as illustrated in Figure A2.1.

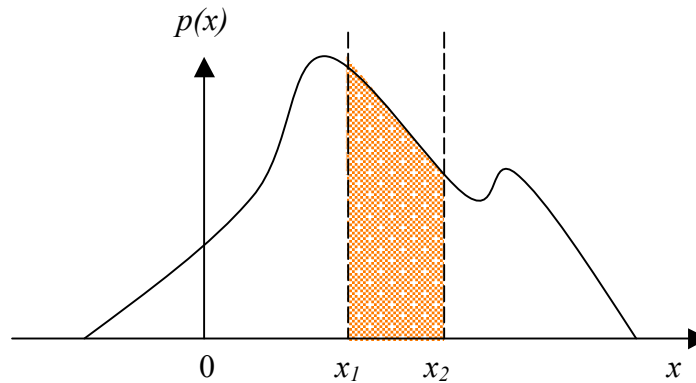


Figure A2.1. The probability of x being between x_1 and x_2 is the area under the curve of the probability density function $p(x)$.

The cumulative probability density function, or ‘probability distribution function, or just the ‘distribution function’ (CDF for short), is a function which completely describes of a real-values random variable, a function When the random variable takes values in the set of real numbers, the probability distribution is completely described by the cumulative distribution function, whose value at each real x is the probability that the random variable is smaller than or equal to x .

A 2.2 Central tendency

The term *central tendency* refers to the way values in a distribution cluster. three common measures of central tendency are the

- **mode**, or the value occurring most often. Distributions with two or more different such points are called *bimodal* or *multimodal*.
- **median**, or middle value in an ordered list of them, and
- **average** or **arithmetic mean**, is the sum of all the values divided by how many of them there are. The arithmetic mean, \bar{x} , of a data distribution of n scores is defined as

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Alpha	Beta	Gamma	Delta	Epsilon	Zeta	Eta	Theta	Iota	Kappa	Lambda	Mu
A α	B β	Γ γ	Δ δ	E ε	Z ζ	H η	Θ θ	I ι	K κ	Λ λ	M μ

A 2.3 Moments of dispersion about the statistical mean

The statistical moments of a distribution about this mean, also called *central moments*, are all defined by

$$\mu'_n = \sum_{i=1}^n (x_i - \alpha)^n P(x)$$

where n is the order of the moment, α is the value around which the moment is taken and $P(x)$ is the probability function. Moments can be thought of as forces. Torque, the twisting force, for example, is described as the turning-moment. Other physical examples are the moment of inertia, angular momentum and magnetic moment, which is a measure of the strength and direction of a magnetic source. The moments about the mean of a probability density function (see § later) are called central moments and they describe “forces” that shape the function, independent of translation—shifting its location in the x-y plane, for example.

A 2.3.1 The 1st central moment ($\mu_1 : 0$)

The first moment about the mean is the sum of the deviations of the scores from the mean, divided by the number of number of scores. The mean is the “balancing point” of the sample scores so the first moment is always 0.

$$\mu_1 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})$$

Note that the convention is to use the expression $(x_i - \bar{x})$ not $(\bar{x} - x_i)$, so that all positive resultants are above the mean.

A 2.3.2 The 2nd central moment (μ_2, s^2 or σ^2 : the variance)

The second moment about the mean is the sum of the *squares* of the deviations of the scores from the mean, divided by the number of number of scores.

$$\mu_2 = s^2 = \sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 .$$

The variance measures the spread of the distribution by averaging the effect of large and small deviations from the mean. Only the absolute value of these deviations are

<i>Nu</i>	<i>Xi</i>	<i>Omicron</i>	<i>Pi</i>	<i>Rho</i>	<i>Sigma</i>	<i>Tu</i>	<i>Upsilon</i>	<i>Phi</i>	<i>Chi</i>	<i>Psi</i>	<i>Omega</i>
N v	Ξ ξ	Ο ο	Π π	Ρ ρ	Σ σ	Τ τ	Υ υ	Φ φ	Χ χ	Ψ ψ	Ω ω

A2-4 Appendix 2

captured. That is, because these deviations $(x_i - \bar{x})$ are squared, the side of the mean on which they occur is lost. The relative location of the location of the mean in the range of the distribution is addressed.

The square-root of the variance, $\sqrt{\mu_2} = \sigma$ or s , is known as the *standard deviation*. Statistical analyses will often report the standard deviation because its value is more intuitively grasped: it represents the average deviation between the mean and the observed values.

A 2.3.3 The 3rd central moment (μ_3 : skewness)

The third moment about the mean is *skewness*, sometimes abbreviated to *skew*, is a measure of the degree of asymmetry of a distribution. The tapering sides of a distribution are called *tails*. Skewness is defined as

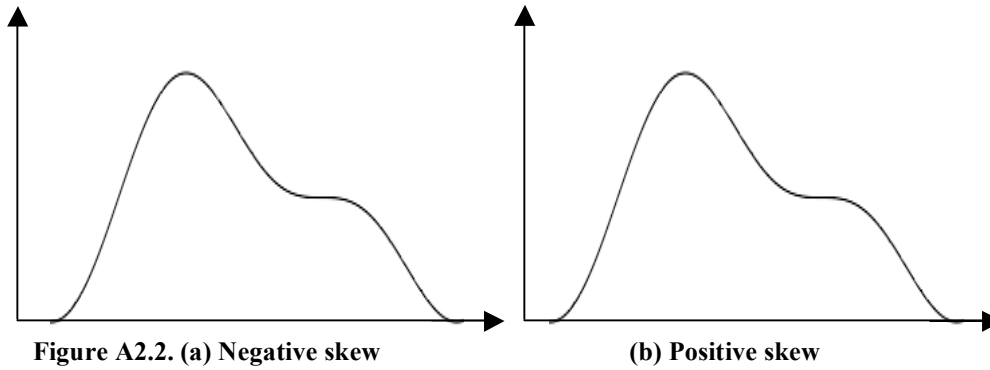
$$\mu_3 = \frac{\frac{1}{n} \sum_{i=1}^n (x - \bar{x})^3}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x - \bar{x})^2}} = \frac{\frac{1}{n} \sum_{i=1}^n (x - \bar{x})^3}{\sqrt{\mu_2}}$$

where μ_2 is the second moment, the variance. In Figure A2.2 the curves on the right side of the distributions taper differently to those on the left. The skew of

- (a) is **negative**: The left tail is longer than the right and the mass of the distribution is concentrated on the right of the figure. It has a few relatively low values. The distribution is said to be left-skewed. In such a distribution, the mean is lower than median, which is lower than the mode (i.e.; mean < median < mode). In this case the *skewness coefficient* is lower than zero.
- (b) is **positive**: The right tail is longer than the left; the greater mass of the distribution is concentrated on the left of the figure. It has a few relatively high values. The distribution is said to be right-skewed. In such a distribution, the mean is greater than median, which is greater than the mode (i.e.; mean > median > mode); in which case the skewness coefficient is greater than zero.

If there is no skewness or the distribution is symmetric, as the bell-shaped normal curve, then the mean = median = mode.

Alpha	Beta	Gamma	Delta	Epsilon	Zeta	Eta	Theta	Iota	Kappa	Lambda	Mu
A α	B β	Γ γ	Δ δ	E ε	Z ζ	H η	Θ θ	I ι	K κ	Λ λ	M μ



In a skewed (lopsided) distribution, the mean is farther out in the long tail than is the median. Skewness is the third standardised moment, symbolised as γ_1 and defined as

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

where μ_3 is the third moment about the mean and σ is the standard deviation. For a sample of n values, the sample skewness, g_1 , is

$$g_1 = \frac{m_3}{m_2^{3/2}} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^{3/2}}$$

where x_i is the i^{th} value, \bar{x} is the sample mean, m_3 is the sample third central moment, and m_2 is the sample variance.

Karl Pearson suggested two simpler calculations of the approximate measure of skewness:

1. $\frac{(\text{mean} - \text{mode})}{\text{standard deviation}}$ and 2. $\frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$

A 2.3.4 The 4th central moment (μ_4 : kurtosis)

The kurtosis is the statistic that indicates whether the samples are clustered closely around the mean the mean or spread out over a wide range with many scores at the extremes. The kurtosis is defined as

$$\mu_4 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4$$

A distribution that had a disproportionate number of scores at the extremes of their range is known as *fat-tailed* or *leptokurtotic*, as illustrated in Figure A2.3.

<i>Nu</i>	<i>Xi</i>	<i>Omicron</i>	<i>Pi</i>	<i>Rho</i>	<i>Sigma</i>	<i>Tu</i>	<i>Upsilon</i>	<i>Phi</i>	<i>Chi</i>	<i>Psi</i>	<i>Omega</i>
N v	Ξ ξ	O o	Π π	Ρ ρ	Σ σ	Τ τ	Υ υ	Φ φ	Χ χ	Ψ ψ	Ω ω

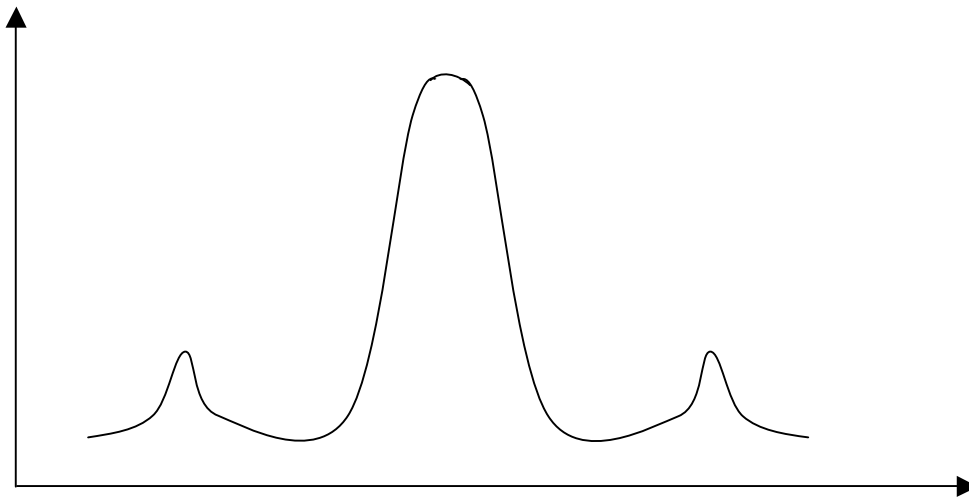


Figure A2.3. Graph of a leptokurtotic function with a small positive skew.

A 2.4 Stationary processes and time series analysis

A stationary process is a stochastic process whose joint probability distribution does not change when shifted in time or space. As a result, parameters such as the mean and variance, if they exist, also do not change over time or position.

Stationarity is used as a tool in time series analysis, where the raw data are often transformed to become stationary. For example, economic data are often seasonal and/or dependent on the price level. Processes are described as *trend stationary* if they are a linear combination of a stationary process and one or more processes exhibiting a trend. Transforming this data to leave a stationary data set for analysis is referred to as de-trending.

A uniform distribution, also known as *white noise*, is stationary. However, the sound of a cymbal crashing is not stationary because the acoustic power of the crash, and hence its variance, diminishes with time. An example of a discrete-time stationary process is a Bernoulli scheme, where the sample space is discrete, so that the random variable may take one of N possible values. Other examples of a discrete-time stationary process with continuous sample space include autoregressive and moving average processes—both of which are subsets of the autoregressive moving average model.

Alpha	Beta	Gamma	Delta	Epsilon	Zeta	Eta	Theta	Iota	Kappa	Lambda	Mu
A α	B β	Γ γ	Δ δ	E ε	Z ζ	H η	Θ θ	I ι	K κ	Λ λ	M μ

A 2.5 Independent random variables

The previous section describes independence of events. this section considers the independence of random variables. If X is a real-valued random variable and a is a number, then the set of outcomes that corresponds to the value of X being less-than-or-equal-to the value of event a is written $\{X \leq a\}$.

Since such a set of outcomes have probabilities, it makes sense to refer to them as events that are independent of other events of this sort. Two random variables X and Y are independent if-and-only-if, for any numbers a and b , the events $\{X \leq a\}$ (the set of outcomes where values of X are less-than-or-equal-to a) and $\{Y \leq b\}$ are independent events as described above. Similarly, an arbitrary collection of random variables—possibly more than just two of them—is independent precisely if, for any finite collection X_1, \dots, X_n and any finite set of numbers a_1, \dots, a_n , the events $\{X_1 \leq a_1\}, \dots, \{X_n \leq a_n\}$ are independent events as defined above.

A 2.6 Correlation

In statistical usage, the strength and direction of the linear relationship between two random variables is known as their correlation, and the measure of it is called their *correlation coefficient*. Essentially, the correlation coefficient of two random variables is a measure of the extent to which they are not independent. The correlation of two random variables is defined only if both of their standard deviations are finite and nonzero. Correlation strength varies from 0 (no correlation) to 1 (identical) and the *direction* of the correlation varies from 1 (same direction) to -1 (opposite direction). The *value* of the correlation coefficient varies from -1 (negatively correlated) to 0 (no correlation) to 1 (highly correlated). The closer the correlation coefficient of two random variables is to -1 or 1, the stronger is the correlation between them.

It is important to note that the *causes* underlying a correlation may be unknown, and establishing a strong correlation between two random variables is not a sufficient condition to establish a causal relationship, positive or negative.

Cross-correlation is a term used in measuring the similarity of two time series. In signal processing, its value represents the measure of similarity of two waveforms as a function of a time-lag that is applied to one of them. Because the technique

<i>Nu</i>	<i>Xi</i>	<i>Omicron</i>	<i>Pi</i>	<i>Rho</i>	<i>Sigma</i>	<i>Tu</i>	<i>Upsilon</i>	<i>Phi</i>	<i>Chi</i>	<i>Psi</i>	<i>Omega</i>
N ν	Ξ ξ	Ο ο	Π π	Ρ ρ	Σ σ	Τ τ	Υ υ	Φ φ	Χ χ	Ψ ψ	Ω ω

A2-8 Appendix 2

employed to calculate cross-correlation involves the taking of successive inner-products of the two time signals, it is also known as a *sliding dot product* or *sliding inner-product*. Cross-correlation techniques are commonly used to search for a short feature in a longer duration signal, such as a motive or phrase in a birdsong extemporisation, for example.

Autocorrelation is a technique for finding repeated and/or underlying patterns in a time-series; the cross-correlation of a signal with itself. Autocorrelation is frequently used, for example, to find the ‘missing fundamental’ frequency as implied by the presence of its harmonics.

Decorrelation is a term for any process that is used to reduce autocorrelation within a signal while preserving other aspects of the signal such as its statistical distribution characteristics. An example of an uncorrelated time series is Brownian motion; an independent random walk in which the size and direction of each move is not dependant on the size or direction of any previous moves.

A statistical analysis of time series data is concerned with the distribution of values without the *sequence* of values in a time series being taken into account and so it completely disregards any spectral (correlation) information. Thus, Bachelier’s hypothesis that market prices are a random walk was an hypothesis, recently shown by Mandelbrot and others not to be true, that there is no autocorrelation (such as trends) in market prices.

A simple method for decorrelating a time-series while preserving its statistical properties is to randomly shuffle the order of its samples.

<i>Alpha</i>	<i>Beta</i>	<i>Gamma</i>	<i>Delta</i>	<i>Epsilon</i>	<i>Zeta</i>	<i>Eta</i>	<i>Theta</i>	<i>Iota</i>	<i>Kappa</i>	<i>Lambda</i>	<i>Mu</i>
A α	B β	Γ γ	Δ δ	E ε	Z ζ	H η	Θ θ	I ι	K κ	Λ λ	M μ