3F1 Random Processes Course

Nick Kingsbury

October 26, 2012

Contents

Pre-requisites

This course assumes familiarity with the following Part IB courses:

- **Linear Algebra and Numerical Analysis** (Paper 7);
- **Probability** (Paper 7);
- **Signal and Data Analysis** (Paper 6);
- **Linear Systems and Control** (Paper 6).

Booklist

- **Peebles, P.Z. Jr.**, 'Probability, Random Variables and Random Signal Principles' (McGraw Hill, 4th Edn., 2001).
  This is a good, readable text which covers most of the course material.
- **Papoulis, A.**, 'Probability and Statistics'
  (Prentice Hall, 1990).
  A good basic text which covers most of the course material.
- **Papoulis, A. and Pillai, S.U.**, 'Probability, Random Variables and Stochastic Processes' (McGraw Hill, 4th Edn., 2002).
  A more advanced but excellent text on the subject, containing a wealth of material on this course and beyond.
- **Gray, R.M. and Davisson, L.D.**, 'An Introduction to Statistical Signal Processing' (CUP, Aug 2004).
  This text explores the deeper mathematical issues underlying Random Signal Theory in a very approachable form, referring the interested reader to more advanced texts. It goes well beyond exam requirements in 3F1, but will be very useful for those interested in a broader understanding of the subject. The book is downloadable **FREE** in pdf (2.7 MB) from:

  http://ee.stanford.edu/~gray/sp.html

  A previous version of the book: 'Random Processes: A Mathematical Approach for Engineers' (Prentice Hall, 1986) may still be available in libraries etc. A new printed version has been produced by CU Press in 2005, but the free web version is still available.

Notation

Obscure mathematical notation will be avoided where possible. However, the following glossary of set-theoretic notation may be a useful reference:

$A = \{a_1, \ldots, a_M\}$ - the set A, containing M elements

$A \cup B$ - 'Union': the combined contents of sets A and B

$A \cap B$ - 'Intersection': the elements common to A and B

A^c - 'Complement': elements not in A

$\emptyset$ - The empty set: $\emptyset = \{\}$

$a \in A$ - Element a is a member of set A

$A \subset B$ - 'Subset': set A is completely contained within set B

$\{\omega : E\}$ - the set of ω's such that expression E is True

$[a, b]$ - the set of real numbers x such that $a \leq x \leq b$ (i.e. $\{x : a \leq x \leq b\}$)

(a, b) - the set of real numbers x such that $a < x < b$ (i.e. $\{x : a < x < b\}$)

$\Re$ - the real numbers: $\Re = (-\infty, +\infty)$

$\mathcal{Z}$ - the integers: $\{-\infty, \ldots, -1, 0, 1, \ldots, \infty\}$.

Acknowledgement

This course is developed from an earlier course on Random Signals and Estimation, given by Simon Godsill as part of the old E4 paper (prior to modularisation of the 3rd year). I would like to thank him very much for his help and advice, and for allowing me to base much of this course on his lecture notes and examples.

Nick Kingsbury.

1 Probability Distributions

1.1 Aims and Motivation for the Course

We aim to:

- Develop a theory which can characterize the behaviour of **real-world Random Signals and Processes**;
- Use standard **Probability Theory** for this.

Random signal theory is important for

- Analysis of signals;
- Inference of underlying system parameters from noisy observed data;
- Design of optimal systems (digital and analogue signal recovery, signal classification, estimation...);
- Predicting system performance (error-rates, signal-to-noise ratios, ...).

Examples:

- **Speech signals** (see fig. 1.1) - use probability theory to characterize that some sequences of vowels and consonants are more likely than others, some waveforms more likely than others for a given vowel or consonant.
  Use this to achieve: speech recognition, speech coding, speech enhancement, ...
- **Digital communications** (see fig. 1.2) – characterise the properties of the digital data source (mobile phone, digital television transmitter, ...), characterize the noise/distortions present in the transmission channel.
  Use this to achieve: accurate regeneration of the digital signal at the receiver, analysis of the channel characteristics...

Probability theory is used to give a mathematical description of the behaviour of real-world systems which involve elements of **randomness**. Such a system might be as simple as a coin-flipping experiment, in which we are interested in whether 'Heads' or 'Tails' is the outcome, or it might be more complex, as in the study of random errors in a coded digital data stream (e.g. a digital TV signal or mobile phone signal).

The basics of probability theory should be familiar from the IB Probability and Statistics course. Here we summarize the main results from that course and develop them into a framework that can encompass random signals and processes.

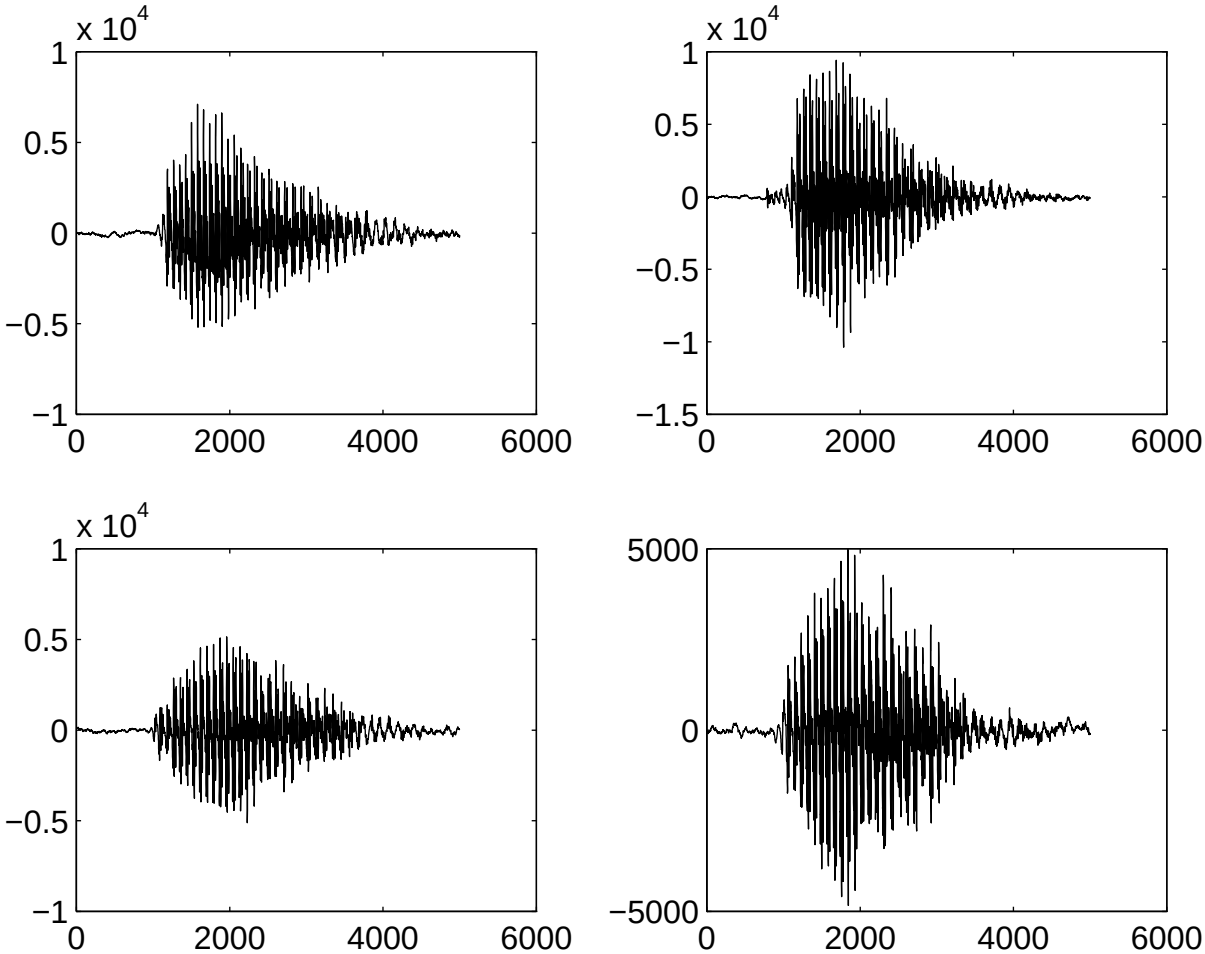


Fig. 1.1: Four utterances of the vowel sound 'Aah'.

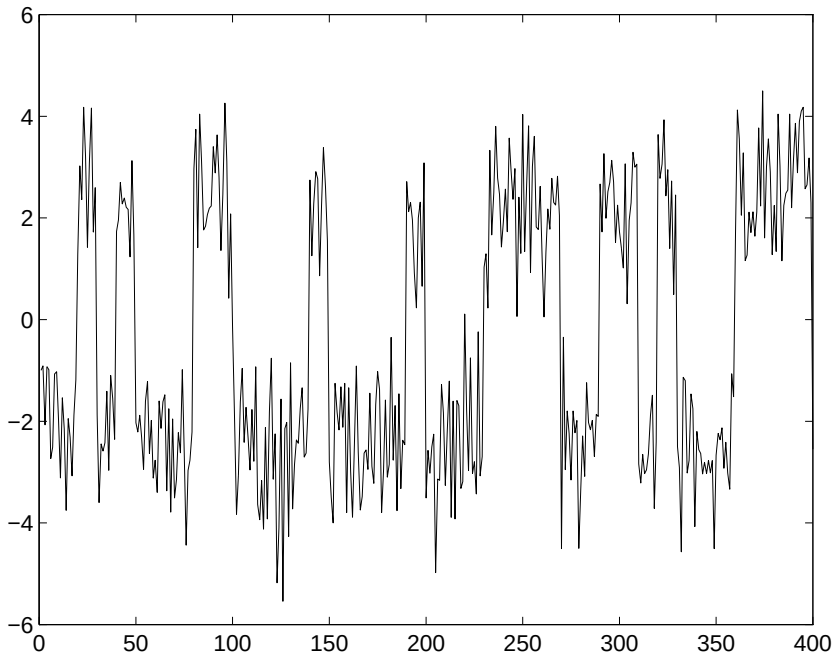


Fig. 1.2: Digital data stream from a noisy communications Channel.

1.2 Probability Distributions

The distribution P_X of a random variable X is simply a probability measure which assigns probabilities to events on the real line. The distribution P_X answers questions of the form:

What is the probability that X lies in some subset F of the real line?

In practice we summarize P_X by its **Probability Mass Function – pmf** (for discrete variables only), **Probability Density Function – pdf** (mainly for continuous variables), or **Cumulative Distribution Function – cdf** (for either discrete or continuous variables).

1.2.1 Probability Mass Function (pmf)

Suppose the discrete random variable X can take a set of M real values $\{x_1, \ldots x_M\}$, then the **pmf** is defined as:

$$p_X(x_i) = \Pr\{X = x_i\} = P_X(\{x_i\}), \quad \text{where} \quad \sum_{i=1}^{M} p_X(x_i) = 1 \tag{1.1}$$

e.g. For a normal 6-sided die, $M = 6$ and $p_X(x_i) = \frac{1}{6}$. For a pair of dice being thrown, $M = 11$ and the pmf is as shown in fig. 1.3(a).

1.2.2 Cumulative Distribution Function (cdf)

The **cdf** can describe discrete, continuous or mixed distributions of X and is defined as:

$$F_X(x) = \Pr\{\, X \leq x \,\} = P_X(\,(-\infty, x]\,) \tag{1.2}$$

For discrete X:

$$F_X(x) = \sum\{p_X(x_i) : x_i \leq x\} \tag{1.3}$$

giving step-like cdfs as in the example of fig. 1.3(b).

Properties follow directly from the Axioms of Probability:

1. $0 \leq F_X(x) \leq 1$
2. $F_X(-\infty) = 0, \qquad F_X(\infty) = 1$
3. $F_X(x)$ is non-decreasing as x increases
4. $\Pr\{x_1 < X \leq x_2\} = F_X(x_2) - F_X(x_1)$
5. $\Pr\{X > x\} = 1 - F_X(x)$

Where there is no ambiguity we will often drop the subscript X and refer to the cdf as $F(x)$.

1.2.3 Probability Density Function (pdf)

The **pdf** of X is defined as the derivative of the cdf:

$$f_X(x) = \frac{d}{dx} F_X(x) \tag{1.4}$$

The pdf can also be interpreted in derivative form as:

$$f_X(x)\ \delta x = \Pr\{x < X \leq x + \delta x\} = F_X(x + \delta x) - F_X(x) \quad \text{as} \quad \delta x \to 0 \tag{1.5}$$

For a discrete random variable with pmf given by $p_X(x_i)$:

$$f_X(x) = \sum_{i=1}^{M} p_X(x_i)\ \delta(x - x_i) \tag{1.6}$$

An example of the pdf of the 2-dice discrete random process is shown in fig. 1.3(c). (Strictly the delta functions should extend vertically to infinity, but we show them only reaching the values of their areas, $p_X(x_i)$.)

The pdf and cdf of a continuous distribution (in this case the **normal** or **Gaussian** distribution) are shown in fig. 1.3(d,e). Note that the cdf is the integral of the pdf and should always go from zero to unity for a valid probabilty distribution.

Properties of pdfs:

1. $f_X(x) \geq 0$
2. $\int_{-\infty}^{\infty} f_X(x)\ dx = 1$
3. $F_X(x) = \int_{-\infty}^{x} f_X(\beta)\ d\beta$
4. $Pr\{x_1 < X \leq x_2\} = \int_{x_1}^{x_2} f_X(\beta)\ d\beta$

As for the cdf, we will often drop the subscript X and refer simply to $f(x)$ when no confusion can arise.

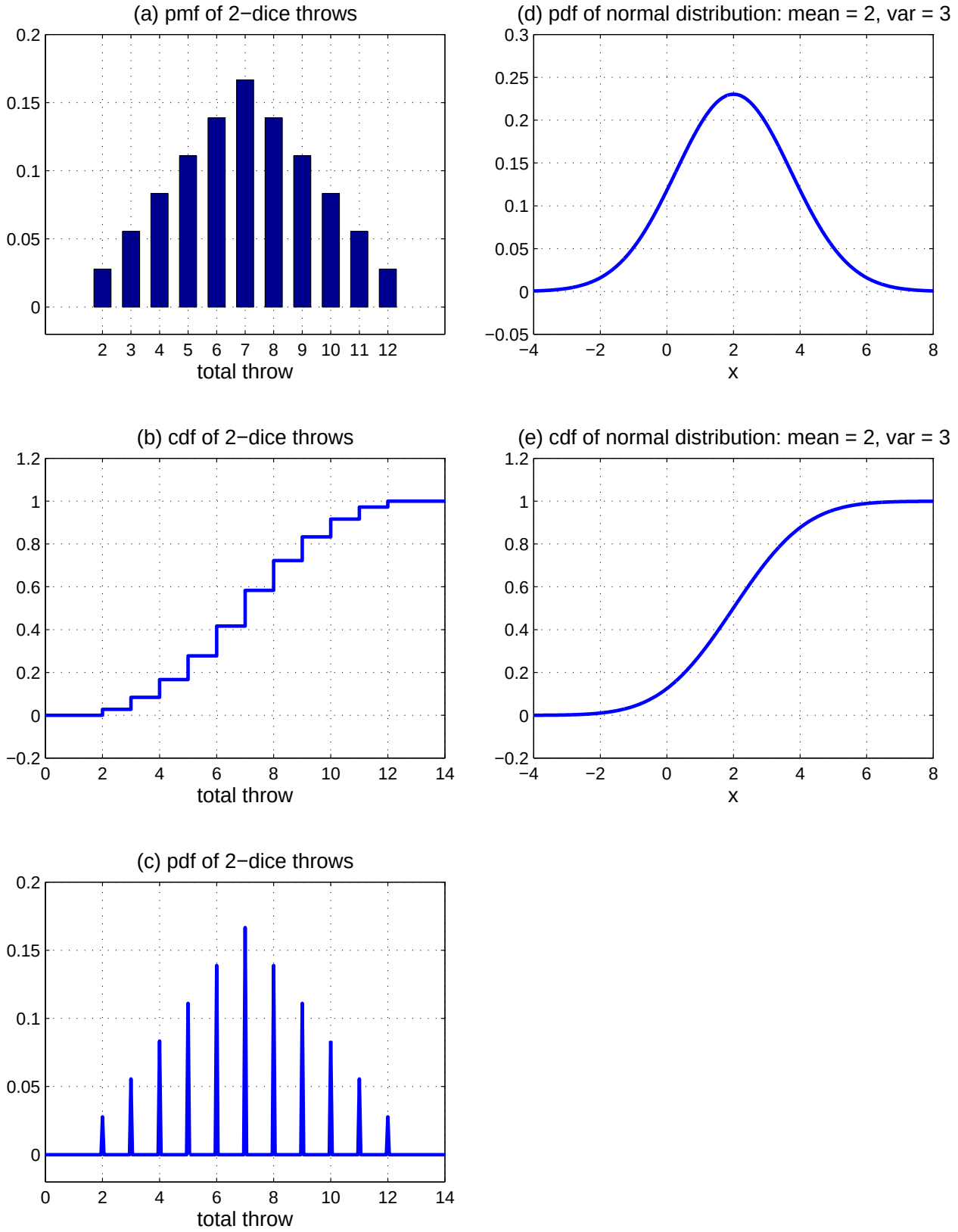


Fig. 1.3: Examples of pmfs, cdfs and pdfs: (a) to (c) for a discrete process, the sum of two dice; (d) and (e) for a continuous process with a normal or Gaussian distribution, whose mean = 2 and variance = 3.

1.3 Conditional probabilities and Bayes' Rule

If A and B are two separate but possibly dependent random events, then:

1. The **joint probability of A and B** occurring together $= \Pr\{A, B\}$
2. The **conditional probability of A**, given that B occurs $= \Pr\{A|B\}$
3. The **conditional probability of B**, given that A occurs $= \Pr\{B|A\}$

From elementary rules of probability (Venn diagrams):

$$\Pr\{A, B\} = \Pr\{A|B\}\ \Pr\{B\} = \Pr\{B|A\}\ \Pr\{A\} \tag{1.7}$$

Dividing the right-hand pair of expressions by $\Pr\{B\}$ gives Bayes' rule:

$$\Pr\{A|B\} = \frac{\Pr\{B|A\}\ \Pr\{A\}}{\Pr\{B\}} \tag{1.8}$$

In problems of probabilistic inference, we are often trying to estimate the most probable underlying model for a random process, based on some observed data or evidence. If A represents a given set of model parameters, and B represents the set of observed data values, then the terms in equation (1.8) are given the following terminology:

- $\Pr\{A\}$ is the **prior** probability of the model being A (in the absence of any evidence);
- $\Pr\{B\}$ is the probability of the **evidence** being B;
- $\Pr\{B|A\}$ is the **likelihood** that the evidence would be B, given that the model was A;
- $\Pr\{A|B\}$ is the **posterior** probability of the model being A, given that the evidence is B.

Quite often, we try to find the model A which maximises the posterior $\Pr\{A|B\}$. This is known as **maximum a posteriori** or **MAP** model selection.

The following example illustrates the concepts of Bayesian model selection.

1.3.1 Example – Loaded Dice

Problem:

Given a tub containing 100 six-sided dice, in which one die is known to be loaded towards the six to a specified extent, derive an expression for the probability that, after a given set of throws, an arbitrarily chosen die is the loaded one? Assume the other 99 dice are all fair (not loaded in any way). The loaded die is known to have the following pmf:

$$p_L(1) = 0.05; \quad p_L(2) \ldots p_L(5) = 0.15; \quad p_L(6) = 0.35.$$

Hence derive a good serial testing strategy for finding the loaded die from the tub.

Solution:

The pmfs of the fair dice may be assumed to be: $p_F(i) = \frac{1}{6}$ for $i = 1 \ldots 6$.

Let each die have one of two states, $S = L$ if it is loaded and $S = F$ if it is fair. These are our two possible **models** for the random process and they have underlying pmfs given by $\{p_L(1), \ldots p_L(6)\}$ and $\{p_F(1), \ldots p_F(6)\}$ respectively.

After N throws of the chosen die, let the sequence of throws be $\Theta_N = \{\theta_1, \ldots \theta_N\}$, where each $\theta_i \in \{1, \ldots 6\}$. This is our **evidence**.

We shall now calculate the probability that this die is the loaded one. We therefore wish to find the **posterior** $\Pr\{S = L | \Theta_N\}$.

We cannot evaluate this directly, but we can evaluate the **likelihoods**, $\Pr\{\Theta_N | S = L\}$ and $\Pr\{\Theta_N | S = F\}$, since we know the expected pmfs in each case. We also know the **prior** probabilities $\Pr\{S = L\}$ and $\Pr\{S = F\}$ before we have carried out any throws, and these are $\{0.01, 0.99\}$ since only one die in the tub of 100 is loaded. Hence we can use Bayes' rule:

$$\Pr\{S = L | \Theta_N\} = \frac{\Pr\{\Theta_N | S = L\} \Pr\{S = L\}}{\Pr\{\Theta_N\}} \tag{1.9}$$

Similarly:

$$\Pr\{S = F | \Theta_N\} = \frac{\Pr\{\Theta_N | S = F\} \Pr\{S = F\}}{\Pr\{\Theta_N\}} \tag{1.10}$$

The denominator term $\Pr\{\Theta_N\}$ ensures that $\Pr\{S = L | \Theta_N\}$ and $\Pr\{S = F | \Theta_N\}$ sum to unity (as they must). It can most easily be calculated from the numerators of the above two expressions:

$$\begin{aligned} \Pr\{\Theta_N\} &= \Pr\{\Theta_N, S = L\} + \Pr\{\Theta_N, S = F\} \\ &= \Pr\{\Theta_N | S = L\} \Pr\{S = L\} + \Pr\{\Theta_N | S = F\} \Pr\{S = F\} \end{aligned} \tag{1.11}$$

so that

$$\begin{aligned} \Pr\{S = L | \Theta_N\} &= \frac{\Pr\{\Theta_N | S = L\} \Pr\{S = L\}}{\Pr\{\Theta_N | S = L\} \Pr\{S = L\} + \Pr\{\Theta_N | S = F\} \Pr\{S = F\}} \\ &= \frac{1}{1 + R_N} \end{aligned} \tag{1.12}$$

where

$$R_N = \frac{\Pr\{\Theta_N|\, S=F\}\ \Pr\{S=F\}}{\Pr\{\Theta_N|\, S=L\}\ \Pr\{S=L\}} \tag{1.13}$$

To calculate the likelihoods, $\Pr\{\Theta_N|\, S=L\}$ and $\Pr\{\Theta_N|\, S=F\}$, we simply take the product of the probabilities of each throw occurring in the sequence of throws Θ_N, given each of the two models respectively (since each new throw is independent of all previous throws, given the model). So, after N throws, these likelihoods will be given by:

$$\Pr\{\Theta_N|\, S=L\} = \prod_{i=1}^{N} p_L(\theta_i) \quad \text{and} \quad \Pr\{\Theta_N|\, S=F\} = \prod_{i=1}^{N} p_F(\theta_i) \tag{1.14}$$

We can now substitute these probabilities into the above expression for R_N and include $\Pr\{S=L\} = 0.01$ and $\Pr\{S=F\} = 0.99$ (see previous page) to get the desired **a posteriori** probability $\Pr\{S=L|\,\Theta_N\}$ after N throws using equ. (1.12).

We may calculate this iteratively by noting that

$$\begin{aligned} \Pr\{\Theta_N|\, S=L\} &= \Pr\{\Theta_{N-1}|\, S=L\} \times p_L(\theta_N) \\ \text{and} \quad \Pr\{\Theta_N|\, S=F\} &= \Pr\{\Theta_{N-1}|\, S=F\} \times p_F(\theta_N) \end{aligned} \tag{1.15}$$

so that

$$R_N = R_{N-1}\,\frac{p_F(\theta_N)}{p_L(\theta_N)} \quad \text{where} \quad R_0 = \frac{\Pr\{S=F\}}{\Pr\{S=L\}} = 99 \tag{1.16}$$

If we calculate this after every throw of the current die being tested (i.e. as N increases), then we can either move on to test the next die from the tub if $\Pr\{S=L|\,\Theta_N\}$ becomes sufficiently small (say $< 10^{-4}$) or continue to test the current die. We finally accept the current die as the loaded one when $\Pr\{S=L|\,\Theta_N\}$ becomes large enough (say > 0.995). (These thresholds correspond approximately to $R_N > 10^4$ and $R_N < 5\,.\,10^{-3}$ respectively.)

The choice of these thresholds for $\Pr\{S=L|\,\Theta_N\}$ is a function of the desired tradeoff between speed of searching versus the probability of failure to find the loaded die, either by moving on to the next die even when the current die is the loaded one, or by incorrectly selecting a fair die as the loaded one.

The lower threshold, $p_1 = 10^{-4}$, is the more critical, because it affects how long we spend before discarding each fair die. The probability of correctly detecting all the fair dice before the loaded die is reached is $(1-p_1)^n \approx 1 - np_1$, where $n \approx 50$ is the expected number of fair dice tested before the loaded one is found. So the failure probability due to incorrectly assuming the loaded die to be fair $\approx np_1 \approx 0.005$.

The upper threshold, $p_2 = 0.995$, is much less critical on search speed, since the 'loaded' result occurs only once, so it is a good idea to set it very close to unity. The failure probability caused by selecting a fair die to be the loaded one is is just $1-p_2 = 0.005$. Hence, for the given thresholds, the overall failure probability $= 0.005 + 0.005 = 0.01$.

We could make the failure probability 0.001 by choosing $p_1 = 10^{-5}$ and $p_2 = 0.9995$, but this would slow down the search (by about 30% - try to explain this factor!).

Note on computation:

In problems with significant amounts of evidence (e.g. large N), the evidence probability and the likelihoods can both get very very small, sufficient to cause floating-point underflow on many computers if equations such as (1.14) are computed directly. However the ratio of likelihood to evidence probability still remains a reasonable size and is an important quantity which must be calculated correctly.

One solution to this problem is to compute only the ratio of likelihoods, as in equ. (1.16). A more generally useful solution is to compute log(likelihoods) instead. The product operations in the expressions for the likelihoods then become sums of logarithms. Even the calculation of likelihood ratios such as R_N and comparison with appropriate thresholds can be done in the log domain. After this, it is OK to return to the linear domain if necessary since R_N should be a reasonable value as it is the **ratio** of very small quantities.

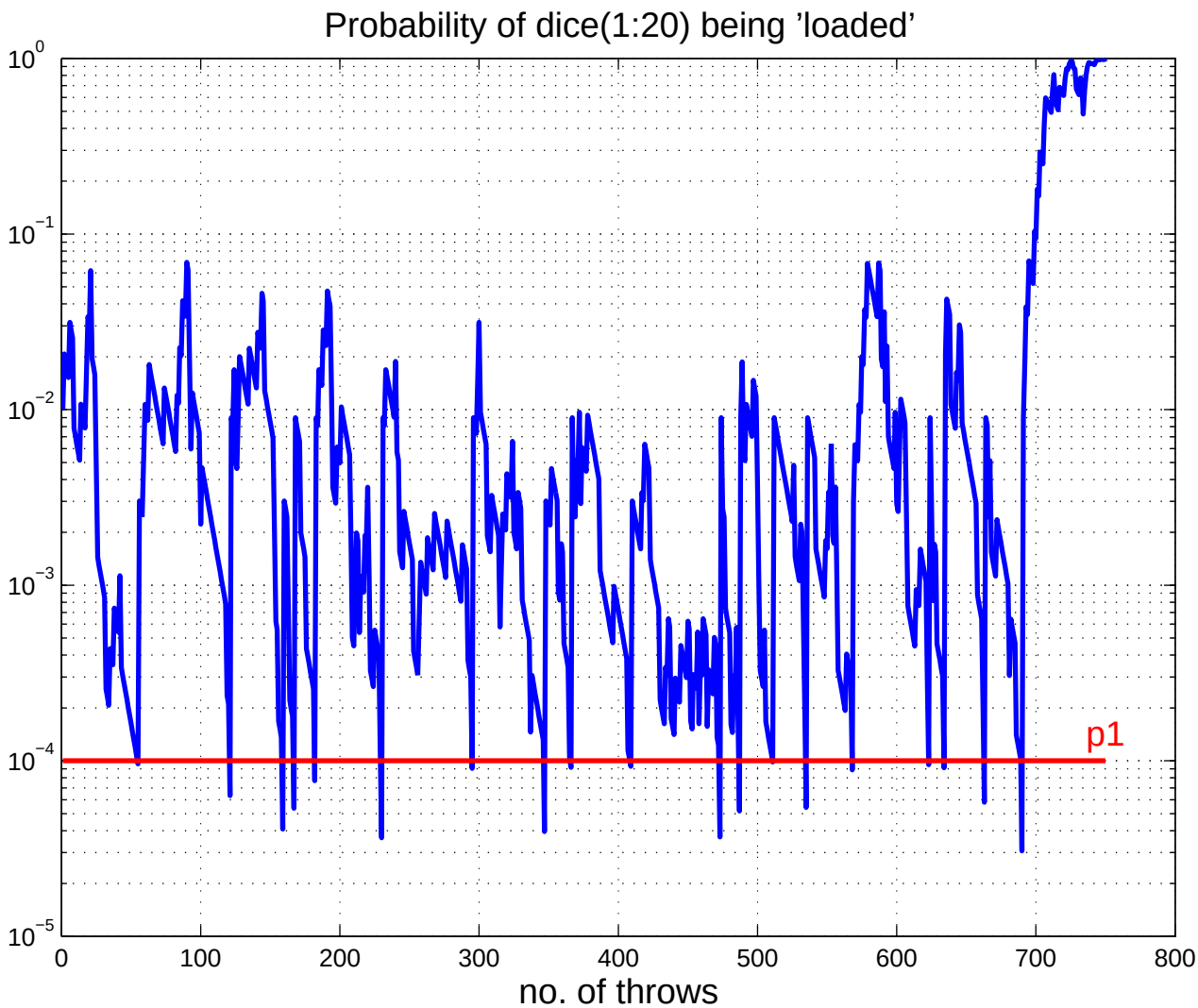


Fig. 1.4: Probability of the current die being 'loaded' as the throws progress (in this case die 20 is the loaded one). A new die is selected whenever the probability falls below p_1.

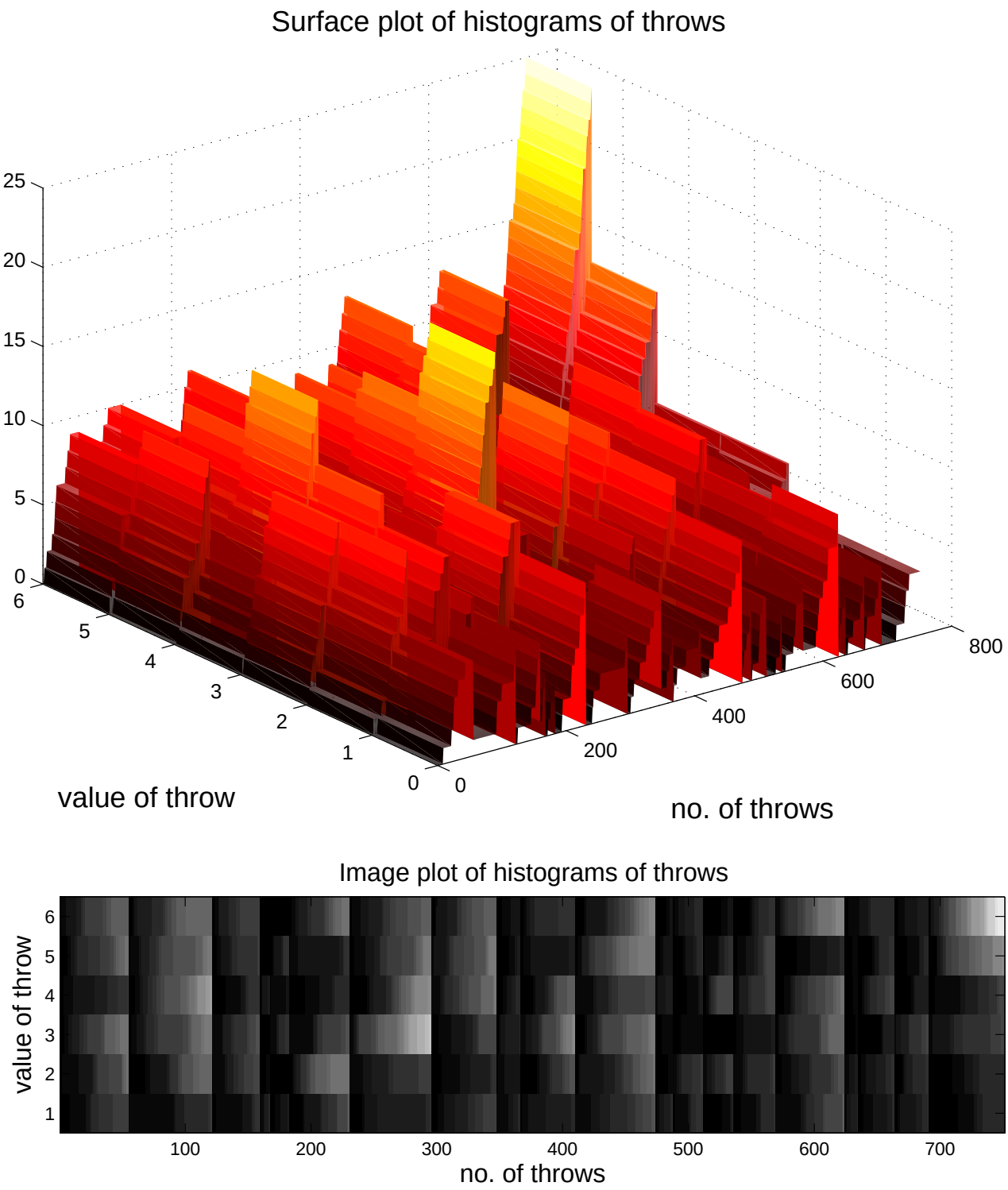


Fig. 1.5: Histograms of the dice throws as the throws progress. Histograms are reset to zero when each new die is selected.

1.4 Joint and Conditional cdfs and pdfs

1.4.1 Cumulative distribution functions

We define the **joint** cdf to be

$$F_{X,Y}(x,y) = \Pr\{ (X \leq x) \text{ AND } (Y \leq y)\} \tag{1.17}$$

and **conditional** cdf to be

$$F_{X|Y}(x|y) = \Pr\{ (X \leq x)|(Y \leq y)\} \tag{1.18}$$

Hence we get the following rules:

Conditional probability (cdf):

$$F_{X|Y}(x|y) = \Pr\{ (X \leq x)|(Y \leq y)\} = \frac{F_{X,Y}(x,y)}{F_Y(y)} \tag{1.19}$$

Bayes Rule (cdf):

$$F_{X|Y}(x|y) = \frac{F_{Y|X}(y|x)F_X(x)}{F_Y(y)} \tag{1.20}$$

Total probability (cdf):

$$F_{X,Y}(x,+\infty) = F_X(x) \qquad \text{since} \quad F_Y(+\infty) = 1 \tag{1.21}$$

Conditional cdf's have similar properties to standard cdf's, i.e.

$$F_{X|Y}(-\infty|y) = 0, \quad F_{X|Y}(+\infty|y) = 1 \quad \text{etc.} \tag{1.22}$$

1.4.2 Probability density functions

We define joint and conditional pdfs in terms of corresponding cdfs. The **joint** pdf is defined to be

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} \tag{1.23}$$

and the **conditional** pdf is defined to be

$$f_{X|Y}(x|y) = \tfrac{\partial}{\partial x} F'(x \,|\, Y=y) \quad \text{where} \quad F'(x \,|\, Y=y) = \Pr\{X \le x \,|\, Y=y\} \tag{1.24}$$

Note that $F'(x \,|\, Y=y)$ is different from the conditional cdf $F(x \,|\, y) = \Pr\{X \le x \,|\, Y \le y\}$, defined in equ. (1.18).

However there is a slight problem. The event, $Y=y$, has zero probability for continuous random variables, hence probability conditional on $Y=y$ is not directly defined and $F'(x \,|\, Y=y)$ cannot be found by direct application of event-based probability. However all is OK if we consider it as a limiting case, as is usual in calculus:

$$\begin{aligned} F'(x|Y=y) &= \lim_{\delta y \to 0} \Pr\{X \le x \mid y < Y \le y+\delta y\} \\ &= \lim_{\delta y \to 0} \frac{F_{X,Y}(x,y+\delta y) - F_{X,Y}(x,y)}{F_Y(y+\delta y) - F_Y(y)} = \frac{\frac{\partial}{\partial y} F_{X,Y}(x,y)}{f_Y(y)} \end{aligned} \tag{1.25}$$

Joint and conditional pdfs have similar properties and interpretation to ordinary pdfs:

$$f(x,y) \ge 0, \quad \int\int f(x,y)\,dx\,dy = 1, \quad f(x|y) \ge 0, \quad \int f(x|y)\,dx = 1 \tag{1.26}$$

[N.B. - from now on interpret $\int$ as $\int_{-\infty}^{\infty}$ unless otherwise stated.]

For pdfs we get the following rules:

Conditional pdf: (from equations (1.23) to (1.25))

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \tag{1.27}$$

Bayes Rule (pdf):

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)\, f_X(x)}{f_Y(y)} \tag{1.28}$$

Total Probability (pdf):

$$\int f_{Y|X}(y|x) f_X(x)\,dx = \int f_{Y,X}(y,x)\,dx = f_Y(y) \int f_{X|Y}(x|y)\,dx = f_Y(y) \tag{1.29}$$

This final result is often referred to as the **Marginalisation Integral** and $f_Y(y)$ as the **Marginal Probability**.

2 Random Vectors, Signals and Functions

2.1 Random Vectors

Random Vectors are simply groups of random variables, arranged as vectors. E.g.:

$$\mathbf{X} = [X_1 \dots X_n]^T \tag{2.1}$$

where $X_1, \dots X_n$ are n separate random variables.

In general, all of the previous results can be applied to random vectors as well as to random scalars, but vectors allow some interesting new results too.

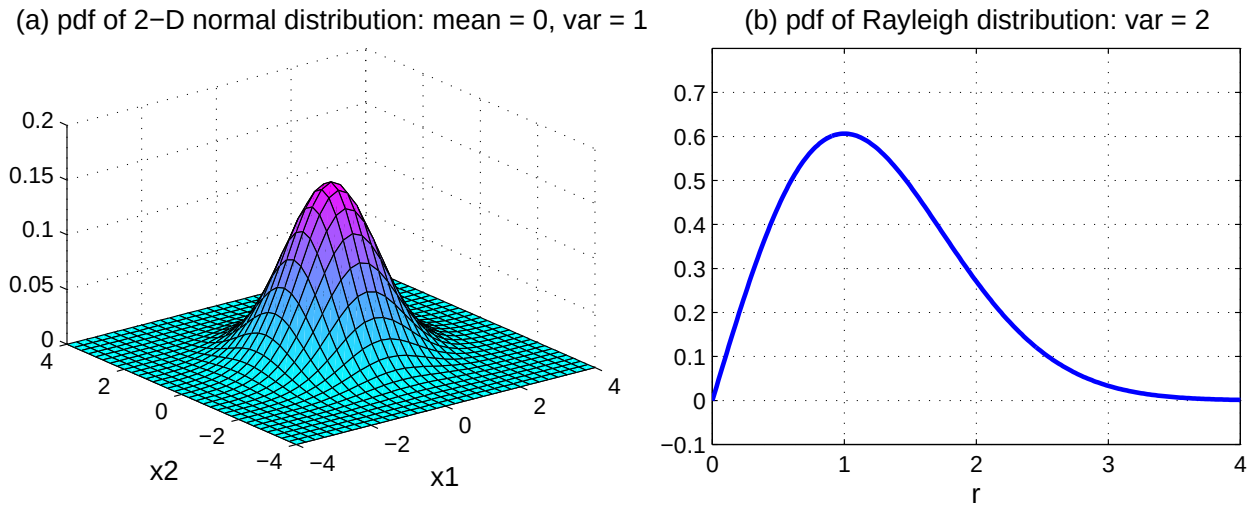


Fig. 2.1: pdfs of (a) a 2-D normal distribution, and (b) a Rayleigh distribution, corresponding to the magnitude of the 2-D random vectors in (a).

2.1.1 Example – Arrows on a target

Suppose that arrows are shot at a target and land at random distances from the target centre. The horizontal and vertical components of these distances are formed into a 2-D random error vector. If each component of this error vector is an independent variable with zero-mean Gaussian pdf of variance σ^2, calculate the pdf's of the radial magnitude and the phase angle of the error vector.

Let the error vector be

$$\mathbf{X} = [X_1 \quad X_2]^T \tag{2.2}$$

X_1 and X_2 each have a zero-mean Gaussian pdf given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-x^2/2\sigma^2} \tag{2.3}$$

Since X_1 and X_2 are independent, the 2-D pdf of $\mathbf{X}$ is

$$f_{\mathbf{X}}(x_1, x_2) = f(x_1)\, f(x_2) = \frac{1}{2\pi\sigma^2}\, e^{-(x_1^2+x_2^2)/2\sigma^2} \tag{2.4}$$

In polar coordinates

$$x_1 = r\cos\theta \quad \text{and} \quad x_2 = r\sin\theta \tag{2.5}$$

Hence

$$f_{\mathbf{X}}(x_1, x_2) = \tfrac{1}{2\pi\sigma^2}\ e^{-(x_1^2+x_2^2)/2\sigma^2} = \tfrac{1}{2\pi\sigma^2}\ e^{-r^2/2\sigma^2} \tag{2.6}$$

To calculate the radial pdf, we first calculate the probability of the arrows landing in a narrow ring of radius r_0 and width δr:

$$\begin{aligned}
\Pr\{r_0 < r < r_0 + \delta r\} &= \int_{r=r_0}^{r_0+\delta r}\int_{\theta=-\pi}^{\pi} f_{\mathbf{X}}(x_1, x_2)\ r\,d\theta\ dr \\
&= \int_{r=r_0}^{r_0+\delta r}\int_{\theta=-\pi}^{\pi} \tfrac{1}{2\pi\sigma^2}\ e^{-r^2/2\sigma^2}\ r\,d\theta\ dr \\
&= \int_{r=r_0}^{r_0+\delta r} \tfrac{2\pi}{2\pi\sigma^2}\ e^{-r^2/2\sigma^2}\ r\,dr \\
&\simeq \tfrac{1}{\sigma^2}\ r_0\ e^{-r_0^2/2\sigma^2}\ \delta r
\end{aligned}$$

This probability is the radial pdf times the ring width, $f_R(r_0)\ \delta r$. Therefore, taking the limit as $\delta r \to 0$, the radial pdf of the error vector is:

$$f_R(r) = \tfrac{1}{\sigma^2}\ r\ e^{-r^2/2\sigma^2} \qquad \text{where} \quad r \geq 0 \tag{2.7}$$

This is known as a **Rayleigh distribution** and is shown in fig. 2.1(b).

The 2-D pdf of $\mathbf{X}$ in (2.6) depends only on r and not on θ, so the angular pdf of the error vector is constant over any 2π interval and is therefore

$$f_\Theta(\theta) = \tfrac{1}{2\pi} \quad \text{so that} \quad \int_{-\pi}^{\pi} f_\Theta(\theta)\ d\theta = 1 \tag{2.8}$$

2.2 Random Signals

Random signals are random variables which evolve, often with time (e.g. audio noise), but also with distance (e.g. intensity in an image of a random texture), or sometimes another parameter.

They can be described as usual by their cdf and either their pmf (if the amplitude is discrete, as in a digitised signal) or their pdf (if the amplitude is continuous, as in most analogue signals).

However a very important additional property is how rapidly a random signal fluctuates. Clearly a slowly varying signal such as the waves in an ocean is very different from a more rapidly varying signal such as vibrations in a vehicle. We will see later in section 4 how to deal with these frequency dependent characteristics of randomness.

For the moment we shall assume that random signals are sampled at regular intervals and that each signal is equivalent to a sequence of samples of a given random process, as in the following example.

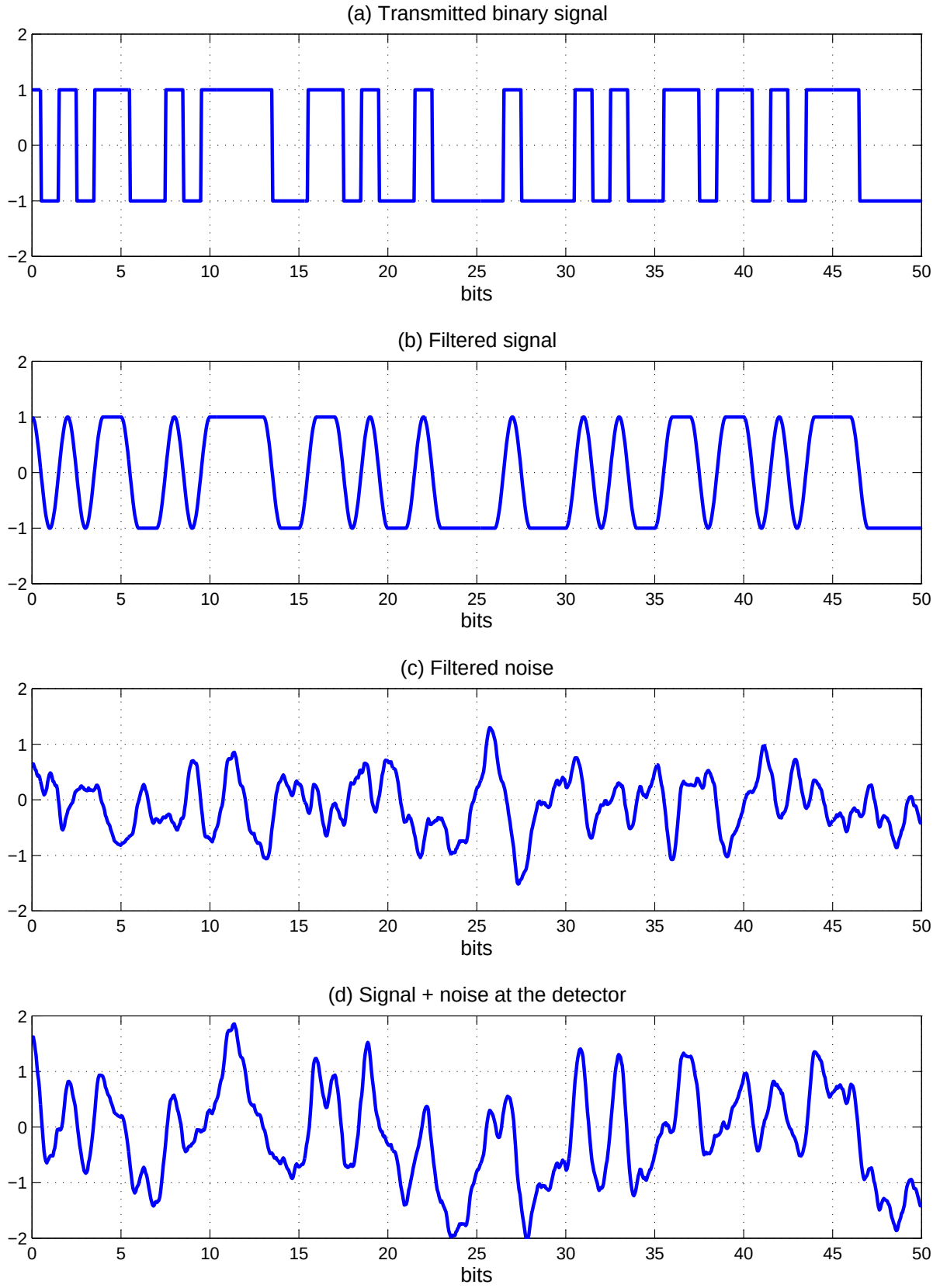


Fig. 2.2: Detection of signals in noise: (a) the transmitted binary signal; (b) the binary signal after filtering with a half-sine receiver filter; (c) the channel noise after filtering with the same filter; (d) the filtered signal plus noise at the detector in the receiver.

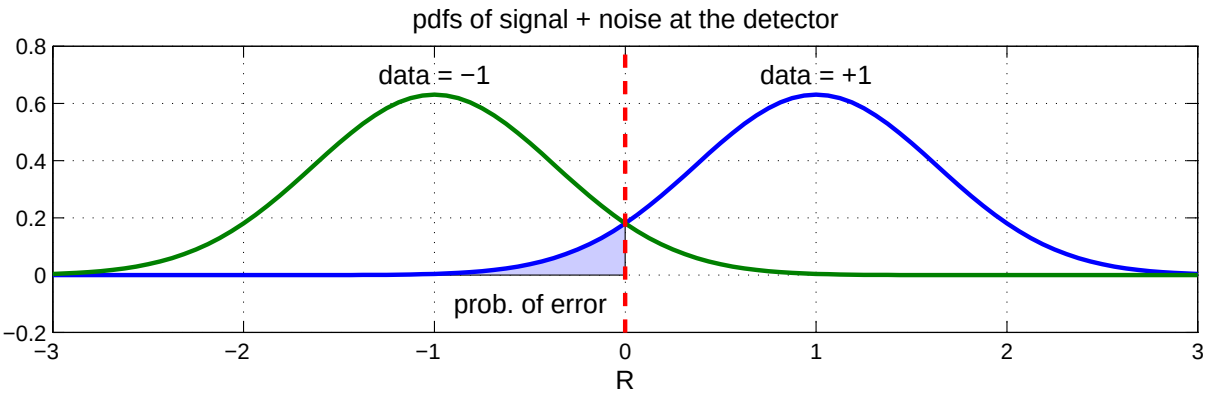


Fig. 2.3: The pdfs of the signal plus noise at the detector for the two data states, ±1. The vertical dashed line is the detector threshold and the shaded area to the left of the origin represents the probability of error when data = +1.

2.2.1 Example – Detection of a binary signal in noise

We now consider the example of detecting a binary signal after it has passed through a channel which adds noise. The transmitted signal is typically as shown in fig. 2.2(a).

In order to reduce the channel noise, the receiver will include a lowpass filter. The aim of the filter is to reduce the noise as much as possible without reducing the peak values of the signal significantly. A good filter for this has a half-sine impulse response of the form:

$$h(t) = \begin{cases} \dfrac{\pi}{2T_b} \sin\left(\dfrac{\pi t}{T_b}\right) & \text{if } \; 0 \le t \le T_b \quad (T_b = \text{bit period}) \\ 0 & \text{elsewhere} \end{cases} \tag{2.9}$$

This filter will convert the rectangular data bits into sinusiodally shaped pulses as shown in fig. 2.2(b) and it will also convert wide bandwidth channel noise into the form shown in fig. 2.2(c). Bandlimited noise of this form will usually have an approximately Gaussian pdf.

Because this filter has an impulse response limited to just one bit period and has unit gain at zero frequency (the area under $h(t)$ is unity), the signal values at the centre of each bit period at the detector will still be ±1. If we choose to sample each bit at the detector at this optimal mid point, the pdfs of the signal plus noise at the detector will be as shown in fig. 2.3.

Let the filtered data signal be $D(t)$ and the filtered noise be $V(t)$, then the detector signal is

$$R(t) = D(t) + V(t) \tag{2.10}$$

If we assume that +1 and −1 bits are equiprobable and the noise is a symmetric zero-mean process, the optimum detector threshold is clearly midway between these two states, i.e. at zero. The probability of error when the data = +1 is then given by:

$$\Pr\{\text{error}|D=+1\} = \Pr\{R(t) < 0 | D=+1\} = F_V(-1) = \int_{-\infty}^{-1} f_V(v)\, dv \tag{2.11}$$

where F_V and f_V are the cdf and pdf of the noise voltage V. This is the shaded area in fig. 2.3.

Similarly the probability of error when the data $= -1$ is then given by:

$$\Pr\{\text{error}|D=-1\} = \Pr\{R(t)>0|D=-1\} = 1 - F_V(+1) = \int_1^\infty f_V(v)\,dv \tag{2.12}$$

Hence the overall probability of error is:

$$\begin{aligned}
\Pr\{\text{error}\} &= \Pr\{\text{error}|D=+1\}\ \Pr\{D=+1\} + \Pr\{\text{error}|D=-1\}\ \Pr\{D=-1\} \\
&= \int_{-\infty}^{-1} f_V(v)\,dv\ \ \Pr\{D=+1\} + \int_1^\infty f_V(v)\,dv\ \ \ \Pr\{D=-1\} \\
&= \int_1^\infty f_V(v)\,dv\ (\Pr\{D=+1\} + \Pr\{D=-1\}) \\
&\quad (\text{since } f_V \text{ is symmetric about zero}) \\
&= \int_1^\infty f_V(v)\,dv
\end{aligned} \tag{2.13}$$

To be a little more general and to account for signal attenuation over the channel, we shall assume that the signal values at the detector are $\pm v_0$ (rather than ± 1) and that the filtered noise at the detector has a zero-mean Gaussian pdf with variance σ^2.

Hence:

$$f_V(v) = \tfrac{1}{\sqrt{2\pi\sigma^2}}\ e^{-v^2/2\sigma^2} \tag{2.14}$$

and so, letting $v = \sigma u$ where u is a unit variance Gaussian process:

$$\begin{aligned}
\Pr\{\text{error}\} &= \int_{v_0}^\infty f_V(v)\,dv \ = \ \int_{v_0/\sigma}^\infty f_V(\sigma u)\,\sigma\,du \\
&= \tfrac{1}{\sqrt{2\pi}} \int_{v_0/\sigma}^\infty e^{-u^2/2}\,du \ = \ Q\left(\frac{v_0}{\sigma}\right)
\end{aligned} \tag{2.15}$$

where

$$Q(x) = \tfrac{1}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2}\,du \tag{2.16}$$

This integral has no analytic solution, but a good (and reasonably simple) approximation to it exists and is discussed in some detail in section 2.4 and plotted in fig. 2.6.

From equ. (2.15) we may obtain the probability of error in the binary detector, which is often expressed as the **bit error rate** or **BER**. For example, if $\Pr\{\text{error}\} = 2\ .\ 10^{-3}$, this would often be expressed as a bit error rate of $2\ .\ 10^{-3}$, or alternatively as 1 error in 500 bits (on average).

The argument (v_0/σ) in (2.15) is the signal-to-noise voltage ratio (SNR) at the detector, and the BER rapidly diminishes with increasing SNR (see fig. 2.6).

2.3 Functions of Random Variables

Problem:

If the random variable Y is a monotonic increasing function of random variable X such that

$$Y = g(X) \quad \text{and} \quad X = g^{-1}(Y) \tag{2.17}$$

(inversion of $g(.)$ requires it to be monotonic so that there is only one value of X for every valid value of Y) then, given the cdf $F_X(x)$ and the function $g(.)$, what are $F_Y(y)$ and $f_Y(y)$, the cdf and pdf of Y?

Solution:

If $g(.)$ is monotonic increasing, the cdf of Y is given by

$$F_Y(y) = \Pr\{Y \le y\} = \Pr\{g(X) \le g(x)\} = \Pr\{X \le x\} = F_X(x) \tag{2.18}$$

since $Y = g(X)$ and $y = g(x)$.

The pdf of Y may be found as follows:

$$f_Y(y) = \tfrac{d}{dy} F_Y(y) = \tfrac{d}{dy} F_X(x) = \left(\tfrac{d}{dx} F_X(x)\right) \tfrac{dx}{dy} = f_X(x) \, \tfrac{dx}{dy} \tag{2.19}$$

$$\text{Defining} \quad \tfrac{dy}{dx} = g'(x), \qquad f_Y(y) = \frac{f_X(x)}{g'(x)} \tag{2.20}$$

This relation is illustrated in fig. 2.4, using a geometric construction to relate f_Y to f_X via $Y = g(X)$. The area under each of the pdfs between a given pair of dashed lines must be the same, because the probability of being in a given range of X must be the same as the probability of being in the equivalent range of Y.

If $g(.)$ is monotonic decreasing (instead of increasing), then equ. (2.18) becomes

$$F_Y(y) = \Pr\{g(X) \le g(x)\} = \Pr\{X \ge x\} = 1 - F_X(x) \tag{2.21}$$

and by a similar argument we find that

$$f_Y(y) = \frac{f_X(x)}{-g'(x)} \tag{2.22}$$

In principle, any **non-monotonic** function $g(.)$ can be split into a finite number of monotonic sections and in that case the pdf result can be generalised to

$$f_Y(y) = \sum_i \left[\frac{f_X(x)}{|g'(x)|}\right]_{x=x_i} \tag{2.23}$$

where the x_i are all the solutions of $g(x) = y$ at any given y. However care is needed in this case, because if $g(x)$ is smooth then $g'(x)$ will become zero at the section boundaries and so $f_Y(y)$ will tend to infinity at these points.

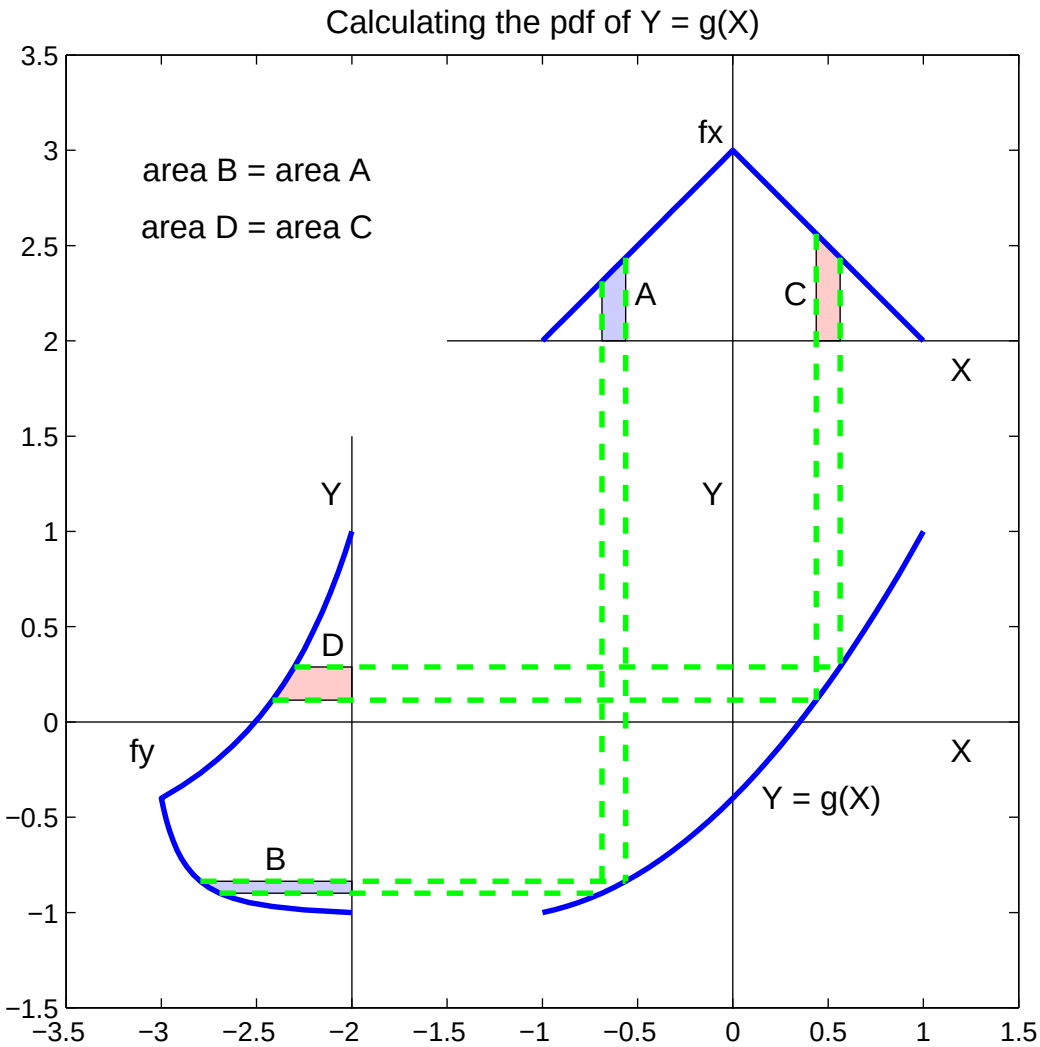


Fig.2.4: Illustration of monotonic mapping of pdfs by plotting $f_Y(y)$ rotated by 90°. The non-linearity in this case is $g(X) = 0.4X^2 + X - 0.4$, which is monotonic for $-1 \leq X \leq 1$.

2.3.1 Example – Generation of a Gaussian pdf from a uniform pdf

If X has a uniform pdf from 0 to 1 (and zero elsewhere), and we wish to generate Y using $Y = g(X)$ such that Y has a Gaussian (normal) pdf of unit variance and zero mean, what is the required function $g(.)$?

(This function is often needed in computers, because standard random number generators tend to have uniform pdfs, while simulation of noise from the real world requires Gaussian pdfs.)

For these pdfs:

$$f_X(x) = \begin{cases} 1 & \text{for } 0 \le x \le 1 \\ 0 & \text{elsewhere} \end{cases} \tag{2.24}$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \tag{2.25}$$

The corresponding cdfs are

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } 0 \le x \le 1 \\ 1 & \text{for } x > 1 \end{cases} \tag{2.26}$$

$$F_Y(y) = \int_{-\infty}^{y} f_Y(u)\, du \tag{2.27}$$

From our previous analysis

$$F_Y(y) = F_X(g^{-1}(y)) = g^{-1}(y) \quad \text{if } 0 \le g^{-1}(y) \le 1 \tag{2.28}$$

$$\therefore\ g^{-1}(y) = \int_{-\infty}^{y} f_Y(u)\, du \tag{2.29}$$

This integral has no analytic solution, so we cannot easily invert this result to get $g(.)$. However a numerical (or graphical) solution is shown in fig. 2.5(a).

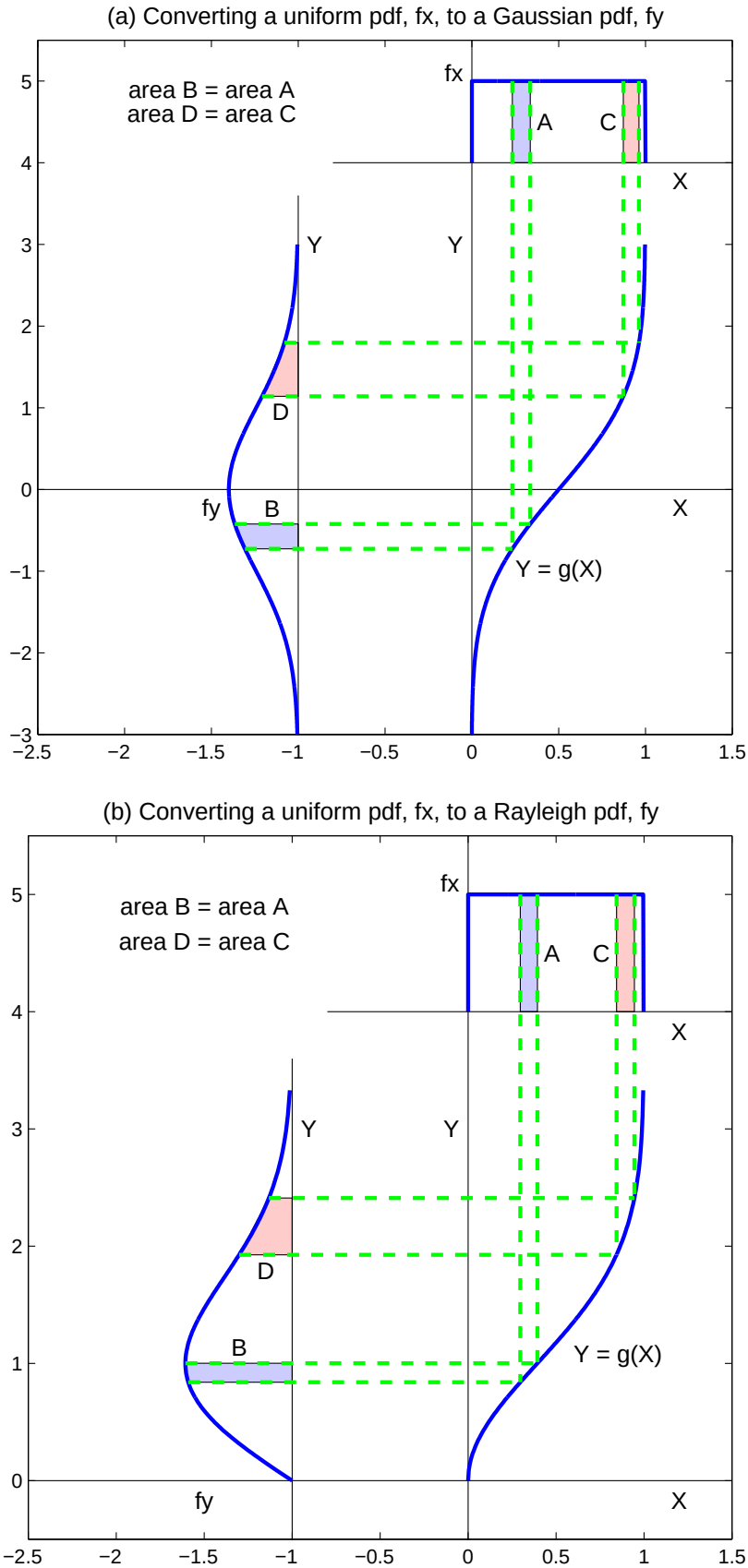


Fig.2.5: Conversion of a uniform pdf to (a) a Gaussian pdf and (b) a Rayleigh pdf. A numerical solution for $g(X)$ was required for (a) in order to invert equ. (2.29), whereas (b) uses the analytic solution for $g(X)$, given in equ. (2.32).

We can get an analytic solution to this problem as follows. If we generate a 2-D Gaussian from polar coordinates, we need to use two random variables to generate r and θ with the correct distributions. In particular (see section 2.1.1), r requires a Rayleigh distribution which can be integrated analytically and hence gives a relatively simple analytic solution for $g(.)$.

Assuming that we start with a uniform pdf from 0 to 1 as before, generating θ is easy as we just scale the variable by 2π to get random phases uniformly distributed from 0 to 2π.

Once we have $\{r, \theta\}$, we can convert to cartesian components $\{x_1, x_2\}$ to obtain two variables, both with Gaussian pdfs as desired.

To generate r correctly, we need a Rayleigh pdf with variance = 2 (to generate 2 unit-variance processes $\{x_1, x_2\}$).

Hence:

$$f_R(r) = \begin{cases} r\ e^{-r^2/2} & \text{for } r \geq 0 \\ 0 & \text{elsewhere} \end{cases} \tag{2.30}$$

and from equ. (2.29):

$$g^{-1}(y) = \int_{-\infty}^{y} f_R(r)\ dr = \int_0^y r\ e^{-r^2/2}\ dr = \left[-e^{-r^2/2}\right]_0^y = 1 - e^{-y^2/2} \tag{2.31}$$

To get $y = g(x)$, we just invert the formula for $x = g^{-1}(y)$. Hence

$$\begin{aligned} x &= 1 - e^{-y^2/2} \\ -y^2/2 &= \ln(1-x) \\ \therefore\ y = g(x) &= \sqrt{-2\ln(1-x)} \qquad \text{(only for } 0 \leq x < 1\text{)} \end{aligned} \tag{2.32}$$

This conversion is illustrated in fig. 2.5(b).

Summarising the complete algorithm to convert uniform pdfs to Gaussian pdfs:

1. Generate a 2-D random vector $\mathbf{x} = [x_1, x_2]^T$ with uniform pdfs from $[0, 0]^T$ to $[1, 1]^T$, by two calls to a standard random number generator function (e.g. rand() in Matlab; although this whole procedure is unnecessary in Matlab as there is already a built-in Gaussian random generator, randn(), which most probably uses this algorithm!).

2. Convert x_1 into r with Rayleigh pdf using
$$r = g(x_1) = \sqrt{-2\ln(1-x_1)} \tag{2.33}$$

3. Convert x_2 into θ with uniform pdf from 0 to 2π using
$$\theta = 2\pi x_2 \tag{2.34}$$

4. Generate two independent random variables with Gaussian pdfs of unit variance and zero mean using
$$y_1 = r\cos\theta \quad \text{and} \quad y_2 = r\sin\theta \tag{2.35}$$

5. Repeat steps 1 to 4 for each new pair of Gaussian variables required.

Note y_1 and y_2 may be scaled by σ to adjust their variance to be σ^2, and an offset μ may be added to them in order to produce a non-zero mean equal to μ.

2.4 Approximation Formulae for the Gaussian Error Integral, Q(x)

A Gaussian pdf with unit variance is given by:

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2} \tag{2.36}$$

The probability that a signal with a pdf given by $f(x)$ lies above a given threshold x is given by the Gaussian Error Integral or Q function:

$$Q(x) = \int_x^\infty f(u)\, du \tag{2.37}$$

There is no analytical solution to this integral, but it has a simple relationship to the error function, $\mathrm{erf}(x)$, or its complement, $\mathrm{erfc}(x)$, which are tabulated in many books of mathematical tables.

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2}\, du \qquad \text{and} \qquad \mathrm{erfc}(x) = 1 - \mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-u^2}\, du \tag{2.38}$$

$$\therefore\ Q(x) = \tfrac{1}{2}\, \mathrm{erfc}\left(\frac{x}{\sqrt{2}}\right) = \tfrac{1}{2}\left[1 - \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right] \tag{2.39}$$

Note that $\mathrm{erf}(0) = 0$ and $\mathrm{erf}(\infty) = 1$, and therefore $Q(0) = 0.5$ and $Q(x) \to 0$ very rapidly as x becomes large.

It is useful to derive simple approximations to $Q(x)$ which can be used on a calculator and avoid the need for tables.

Let $v = u - x$, then:

$$Q(x) = \int_0^\infty f(v+x)\, dv = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-(v^2+2vx+x^2)/2}\, dv = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \int_0^\infty e^{-vx}\, e^{-v^2/2}\, dv \tag{2.40}$$

Now if $x \gg 1$, we may obtain an approximate solution by replacing the $e^{-v^2/2}$ term in the integral by unity, since it will initially decay much slower than the e^{-vx} term.

$$\therefore\ Q(x) < \frac{e^{-x^2/2}}{\sqrt{2\pi}} \int_0^\infty e^{-vx}\, dv = \frac{e^{-x^2/2}}{\sqrt{2\pi}\ x} \tag{2.41}$$

This approximation is an upper bound since $e^{-v^2/2} \leq 1$, but its ratio to the true value of $Q(x)$ becomes less than 1.1 only when $x > 3$, as shown in fig. 2.6. Empirically we obtain a much better approximation to $Q(x)$ by altering the denominator above from $(\sqrt{2\pi}\ x)$ to $(1.64x + \sqrt{0.76x^2 + 4})$ to give:

$$Q(x) \approx \frac{e^{-x^2/2}}{1.64x + \sqrt{0.76x^2 + 4}} \tag{2.42}$$

This improved approximation (developed originally by Borjesson and Sundberg, IEEE Trans. on Communications, March 1979, p 639) gives a curve indistinguishable from $Q(x)$ in fig. 2.6 and its ratio to the true $Q(x)$ is now within $\pm 0.3\%$ of unity for all $x \geq 0$ as shown in fig. 2.7. This accuracy is sufficient for nearly all practical problems.

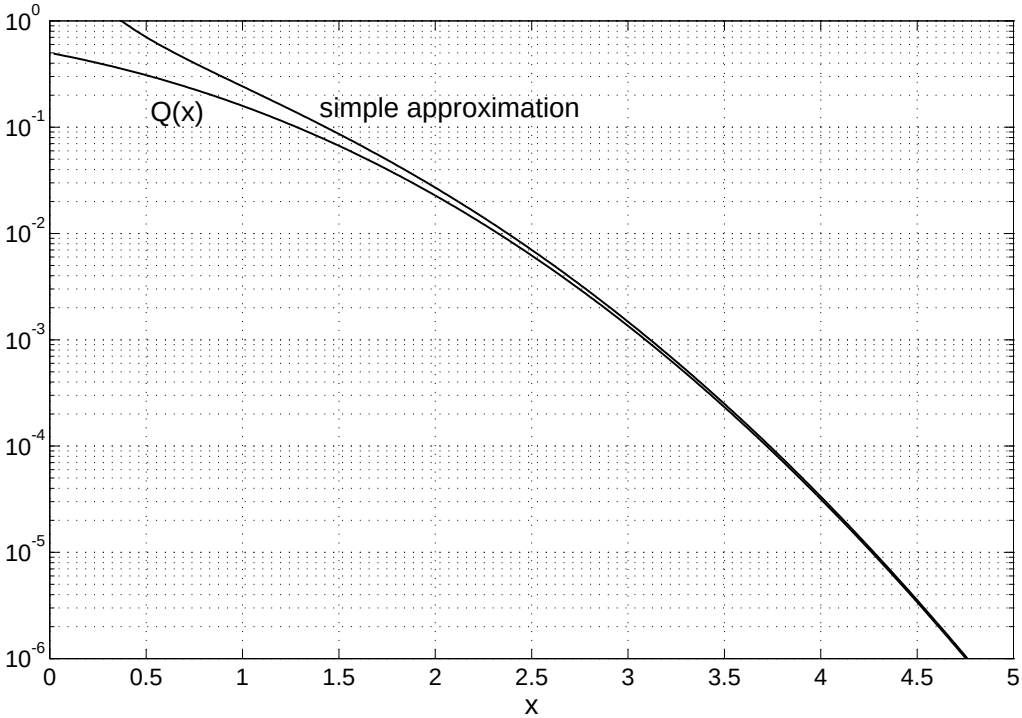


Fig. 2.6: $Q(x)$ and the simple approximation of equ. (2.41).

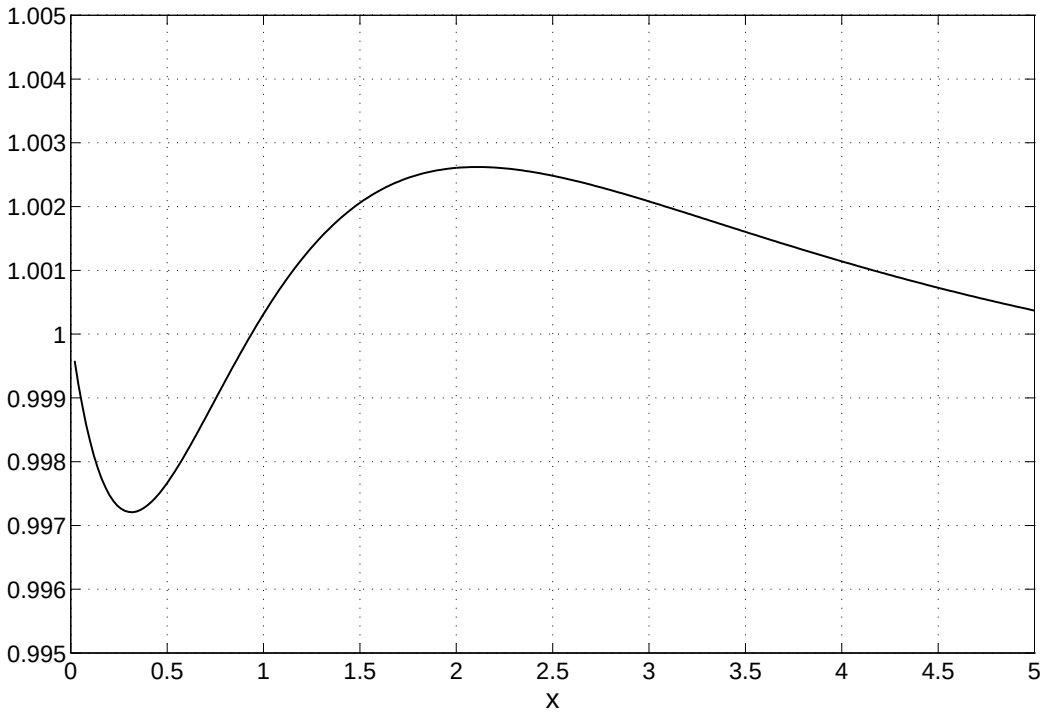


Fig. 2.7: The ratio of the improved approximation of $Q(x)$ in equ. (2.42) to the true value, obtained by numerical integration.

3 Expectations, Moments and Characteristic Functions

3.1 Expectation

Expectations form a fundamental part of random signal theory. In simple terms

the **Expectation Operator** calculates the **mean** of a random quantity

although the concept turns out to be much more general and useful than just this.

If X has pdf $f_X(x)$ (correctly normalised so that $\int_{-\infty}^{+\infty} f_X(x)\ dx = 1$), its expectation is given by:

$$E[X] = \int_{-\infty}^{+\infty} x\ f_X(x)\ dx = \overline{X} \tag{3.1}$$

For discrete processes, we substitute equ. (1.6) in here to get

$$E[X] = \int_{-\infty}^{+\infty} x\ \sum_{i=1}^{M} p_X(x_i)\ \delta(x - x_i)\ dx = \sum_{i=1}^{M} x_i\ p_X(x_i) = \overline{X} \tag{3.2}$$

Now, what is the mean value of some function, $Y = g(X)$?

Using the result of equ. (2.19) for pdfs of related processes Y and X:

$$f_Y(y)\ dy = f_X(x)\ dx \tag{3.3}$$

Hence (again assuming infinite integral limits unless stated otherwise)

$$E[g(X)] = E[Y] = \int y\, f_Y(y)\, dy = \int g(x)\, f_X(x)\, dx \tag{3.4}$$

This is an important result which allows us to use the Expectation Operator for many purposes including the calculation of moments and other related parameters of a random process.

Note, expectation is a **Linear Operator**:

$$E[a\, g_1(X) + b\, g_2(X)] = a\, E[g_1(X)] + b\, E[g_2(X)] \tag{3.5}$$

3.2 Important examples of Expectation

We get **Moments** of a pdf by setting $g(X) = X^n$ in equ. (3.4):

$$n\text{th order moment:} \quad E[X^n] = \int x^n \, f_X(x) \, dx \tag{3.6}$$

$n = 1$**:** 1st order moment, $E(X)$ = Mean value

$n = 2$**:** 2nd order moment, $E[X^2]$ = Mean-squared value (Power or energy)

$n > 2$**:** Higher order moments, $E[X^n]$, give more detail about $f_X(x)$.

3.2.1 Central Moments:

Central moments are moments about the **centre** or mean of a distribution

$$n\text{th order central moment:} \quad E[(X - \overline{X})^n] = \int (x - \overline{X})^n \, f_X(x) \, dx \tag{3.7}$$

Some important parameters from central moments of a pdf are:

- Variance, $n = 2$:

$$\begin{aligned} \sigma^2 &= E[(X - \overline{X})^2] = \int (x - \overline{X})^2 \, f_X(x) \, dx \\ &= \int x^2 \, f_X(x) \, dx - 2\overline{X} \int x \, f_X(x) \, dx + \overline{X}^2 \int f_X(x) \, dx \\ &= E[X^2] - 2\overline{X}^2 + \overline{X}^2 = E[X^2] - \overline{X}^2 \end{aligned} \tag{3.8}$$

- Standard deviation, $\sigma = \sqrt{\text{variance}}$.

- Skewness, $n = 3$:

$$\gamma = \frac{E[(X - \overline{X})^3]}{\sigma^3} \tag{3.9}$$

  $\gamma = 0$ if the pdf of X is symmetric about $\overline{X}$, and becomes more positive if the tail of the distribution is heavier when $X > \overline{X}$ than when $X < \overline{X}$.

- Kurtosis (or excess), $n = 4$:

$$\kappa = \frac{E[(X - \overline{X})^4]}{\sigma^4} - 3 \tag{3.10}$$

  $\kappa = 0$ for a Gaussian pdf and becomes more positive for distributions with heavier tails.

Note that skewness and kurtosis are normalised by dividing the central moments by appropriate powers of σ to make them dimensionless. Kurtosis is usually offset by -3 (as above) to make it zero for Gaussian pdfs.

3.2.2 Example: Central Moments of a Normal Distribution

The normal (or Gaussian) pdf with zero mean is given by:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-x^2/2\sigma^2} \tag{3.11}$$

What is the nth order central moment for the Gaussian?

Since the mean is zero, the nth order central moment is given by

$$E[X^n] = \int x^n\, f_X(x)\, dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int x^n\, e^{-x^2/2\sigma^2}\, dx \tag{3.12}$$

$f_X(x)$ is a function of x^2 and therefore is symmetric about zero. So all the odd-order moments will integrate to zero (including the 1st-order moment, giving zero mean). The even-order moments are then given by:

$$E[X^n] = \frac{2}{\sqrt{2\pi\sigma^2}} \int_0^\infty x^n\, e^{-x^2/2\sigma^2}\, dx \quad \text{for } n \text{ even} \tag{3.13}$$

The integral is calculated by substituting $u = x^2/2\sigma^2$, and hence $du = x/\sigma^2\, dx$, to give:

$$\int_0^\infty x^n\, e^{-x^2/2\sigma^2}\, dx = \tfrac{1}{2}(2\sigma^2)^{\frac{n+1}{2}} \int_0^\infty u^{\frac{n-1}{2}}\, e^{-u}\, du = \tfrac{1}{2}(2\sigma^2)^{\frac{n+1}{2}}\, \Gamma(\tfrac{n+1}{2}) \tag{3.14}$$

Here $\Gamma(z)$ is the Gamma function, which is defined as an integral for all real $z > 0$ and is like the factorial function but generalised to allow non-integer arguments. Values of the Gamma function can be found in mathematical tables. It is defined as follows:

$$\Gamma(z) = \int_0^\infty u^{z-1} e^{-u}\, du \tag{3.15}$$

Using integration by parts, $\Gamma(z)$ has the important factorial-like property that

$$\Gamma(z+1) = z\Gamma(z) \quad \text{for all } z \neq 0 \tag{3.16}$$

$$= z! \quad \text{if } z \in \mathcal{Z} \text{ and } z > 0 \tag{3.17}$$

The following results hold for the Gamma function (see below for a way to evaluate $\Gamma(\frac{1}{2})$ etc.):

$$\Gamma(\tfrac{1}{2}) = \sqrt{\pi}, \quad \Gamma(1) = 1, \quad \text{and hence } \Gamma(\tfrac{3}{2}) = \tfrac{\sqrt{\pi}}{2}, \quad \Gamma(2) = 1, \ \ldots \tag{3.18}$$

Hence

$$E[X^n] = \begin{cases} 0 & n \text{ odd} \\ \frac{1}{\sqrt{\pi}}(2\sigma^2)^{\frac{n}{2}}\, \Gamma(\frac{n+1}{2}) & n \text{ even} \end{cases} \tag{3.19}$$

Valid pdf, $n = 0$:

$$E[X^0] = \tfrac{1}{\sqrt{\pi}}\, \Gamma(\tfrac{1}{2}) = 1 \quad \text{as required for a valid pdf.} \tag{3.20}$$

Note that the normalisation factor $1/\sqrt{2\pi\sigma^2}$ in the expression for the pdf of a unit variance Gaussian (e.g. equ. (3.11)) arises directly from the above result.

Mean, $n = 1$:

$$E[X] = 0 \quad \text{so the mean is zero.} \tag{3.21}$$

Variance, $n = 2$:

$$E[(X - \overline{X})^2] = E[X^2] = \tfrac{1}{\sqrt{\pi}}(2\sigma^2)\ \Gamma(\tfrac{3}{2}) = \tfrac{1}{\sqrt{\pi}}(2\sigma^2)\ \tfrac{\sqrt{\pi}}{2} = \sigma^2 \tag{3.22}$$

Therefore standard deviation $= \sqrt{\text{variance}} = \sigma$.

Skewness, $n = 3$:

$$E[X^3] = 0 \quad \text{so the skewness is zero.} \tag{3.23}$$

Kurtosis, $n = 4$:

$$E[(X - \overline{X})^4] = E[X^4] = \tfrac{1}{\sqrt{\pi}}(2\sigma^2)^2\ \Gamma(\tfrac{5}{2}) = \tfrac{1}{\sqrt{\pi}}(2\sigma^2)^2\ \tfrac{3\sqrt{\pi}}{4} = 3\sigma^4 \tag{3.24}$$

Hence

$$\kappa = \frac{E[(X - \overline{X})^4]}{\sigma^4} - 3 = 3 - 3 = 0 \tag{3.25}$$

3.2.3 Evaluation of $\Gamma(\frac{1}{2}), \Gamma(1)$ etc.

From the definition of Γ and substituting $u = x^2$:

$$\begin{aligned} \Gamma(\tfrac{1}{2}) &= \int_0^\infty u^{-1/2}\, e^{-u}\, du = \int_0^\infty x^{-1}\, e^{-x^2}\, 2x\, dx \\ &= 2\int_0^\infty e^{-x^2}\, dx = \int_{-\infty}^{\infty} e^{-x^2}\, dx \end{aligned} \tag{3.26}$$

Using the following squaring trick to convert this to a 2-D integral in polar coordinates:

$$\begin{aligned} \Gamma^2(\tfrac{1}{2}) &= \int_{-\infty}^{\infty} e^{-x^2}\, dx \int_{-\infty}^{\infty} e^{-y^2}\, dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\, dx\, dy \\ &= \int_0^\infty \int_{-\pi}^{\pi} e^{-r^2}\, r\, d\theta\, dr = \Big[-\tfrac{1}{2}\, 2\pi\, e^{-r^2}\Big]_0^\infty = \pi \end{aligned} \tag{3.27}$$

and so (ignoring the negative square root since $e^{-x^2} \geq 0$ everywhere):

$$\Gamma(\tfrac{1}{2}) = \sqrt{\pi} \simeq 1.7725 \tag{3.28}$$

Hence, using $\Gamma(z+1) = z\Gamma(z)$:

$$\Gamma(\{\tfrac{3}{2}, \tfrac{5}{2}, \tfrac{7}{2}, \tfrac{9}{2}, \ldots\}) = \{\tfrac{1}{2}\sqrt{\pi}, \tfrac{3}{4}\sqrt{\pi}, \tfrac{15}{8}\sqrt{\pi}, \tfrac{105}{16}\sqrt{\pi}, \ldots\} \tag{3.29}$$

The case for $z = 1$ is straightforward:

$$\Gamma(1) = \int_0^\infty u^0\, e^{-u}\, du = \Big[-e^{-u}\Big]_0^\infty = 1 \tag{3.30}$$

so

$$\Gamma(\{2, 3, 4, 5, \ldots\}) = \{1, 2, 6, 24, \ldots\} \tag{3.31}$$

3.3 Sums of random variables

Consider the random variable Y formed as the sum of two independent random variables X_1 and X_2:

$$Y = X_1 + X_2 \tag{3.32}$$

where X_1 has pdf $f_1(x_1)$ and X_2 has pdf $f_2(x_2)$.

We can write the joint pdf for y and x_1 by rewriting the conditional probability formula:

$$f(y, x_1) = f(y|x_1)\, f_1(x_1) \tag{3.33}$$

It is clear that the event 'Y takes the value y conditional upon $X_1 = x_1$' is equivalent to X_2 taking a value $y - x_1$ (since $X_2 = Y - X_1$). Hence

$$f(y|x_1) = f_2(y - x_1) \tag{3.34}$$

Now $f(y)$ may be obtained using the **Marginal Probability** formula (equ. (1.29) from section 1.4.2). Hence

$$\begin{aligned} f(y) &= \int f(y|x_1)\, f_1(x_1)\, dx_1 \\ &= \int f_2(y - x_1)\, f_1(x_1)\, dx_1 \\ &= f_2 * f_1 \quad \text{(Convolution)} \end{aligned} \tag{3.35}$$

This result may be extended to sums of three or more random variables by repeated application of the above arguments for each new variable in turn. Since convolution is a commutative operation, for n independent variables we get:

$$f(y) = f_n * (f_{n-1} * \ldots * f_2 * f_1) = f_n * f_{n-1} * \ldots * f_2 * f_1 \tag{3.36}$$

An example of this effect occurs when multiple dice are thrown and the scores are added together. In the 2-dice example of fig. 1.3(a,b,c) we saw how the pmf approximated a triangular shape. This is just the convolution of two uniform 6-point pmfs for each of the two dice.

Similarly if two variables with Gaussian pdfs are added together, we shall show in section 3.4.2 that this produces another Gaussian pdf whose variance is the sum of the two input variances.

3.4 Characteristic Functions

You have already encountered the **Moment Generating Function** of a pdf in the Part IB probability course. This function was closely related to the **Laplace Transform** of the pdf.

Now we introduce the **Characteristic Function** for a random variable, which is closely related to the **Fourier Transform** of the pdf.

In the same way that Fourier Transforms allow easy manipulation of signals when they are convolved with linear system impulse responses, Characteristic Functions allow easy manipulation of convolved pdfs when they represent sums of random processes.

The **Characteristic Function** of a pdf is defined as:

$$\Phi_X(u) = E[e^{juX}] = \int_{-\infty}^{\infty} e^{jux} f_X(x)\, dx = \mathcal{F}(-u) \tag{3.37}$$

where $\mathcal{F}(u)$ is the Fourier Transform of the pdf.

Note that whenever f_X is a valid pdf, $\Phi(0) = \int f_X(x)\, dx = 1$

Properties of Fourier Transforms apply with $-u$ substituted for ω. In particular:

1. **Convolution - (sums of independent rv's)**

$$\begin{aligned} Y &= \sum_{i=1}^{N} X_i \\ \Rightarrow f_Y &= f_{X_1} * f_{X_2} \ldots * f_{X_N} \\ \Rightarrow \Phi_Y(u) &= \prod_{i=1}^{N} \Phi_{X_i}(u) \end{aligned} \tag{3.38}$$

2. **Inversion**

$$f_X(x) = \tfrac{1}{2\pi} \int e^{-jux}\ \Phi_X(u)\ du \tag{3.39}$$

3. **Moments**

$$\begin{aligned} \frac{d^n \Phi_X(u)}{du^n} &= \int (jx)^n\ e^{jux}\ f_X(x)\ dx \\ \Rightarrow E[X^n] &= \int x^n f_X(x)\ dx = \frac{1}{j^n} \left.\frac{d^n \Phi_X(u)}{du^n}\right|_{u=0} \end{aligned} \tag{3.40}$$

4. **Scaling**

$$\begin{aligned} \text{If } Y = aX, \quad f_Y(y) &= \frac{f_X(x)}{a} \quad \text{from equ. (2.20)} \\ \text{then} \quad \Phi_Y(u) &= \int e^{juy} f_Y(y)\, dy \\ &= \int e^{juax} f_X(x)\, dx \\ &= \Phi_X(au) \end{aligned} \tag{3.41}$$

3.4.1 Characteristic Function of a Gaussian pdf

The Gaussian or normal distribution is very important, largely because of the **Central Limit Theorem** which we shall prove below. Because of this (and as part of the proof of this theorem) we shall show here that **a Gaussian pdf has a Gaussian characteristic function** too.

A Gaussian distribution with mean μ and variance σ^2 has pdf:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2\,/\,2\sigma^2} \tag{3.42}$$

Its characteristic function is obtained as follows, using a trick known as **completing the square of the exponent**:

$$\begin{aligned}
\Phi_X(u) &= E[e^{juX}] = \int f_X(x)\, e^{jux}\, dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int e^{-(x^2-2\mu x+\mu^2-2\sigma^2 jux)\,/\,2\sigma^2}\, dx \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \int e^{-(x-(\mu+ju\sigma^2))^2\,/\,2\sigma^2}\, dx\right)\; e^{(2j\mu u\sigma^2-u^2\sigma^4)\,/\,2\sigma^2} \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \int e^{-y^2\,/\,2\sigma^2}\, dy\right)\; e^{(2j\mu u-u^2\sigma^2)\,/\,2} \quad \text{where} \quad y = x-(\mu+ju\sigma^2) \\
&= e^{j\mu u}\; e^{-\sigma^2u^2\,/\,2}
\end{aligned} \tag{3.43}$$

since the integral in brackets is equivalent to that of a Gaussian pdf and integrates to unity.

Thus the characteristic function of a Gaussian pdf is also Gaussian in magnitude, $e^{-\sigma^2u^2\,/\,2}$, with standard deviation $1/\sigma$, and with a linear phase rotation term, $e^{j\mu u}$, whose rate of rotation equals the mean μ of the pdf. This coincides with standard results from Fourier analysis of Gaussian waveforms and their spectra (e.g. Fourier transform of a Gaussian waveform with time shift).

3.4.2 Summation of two or more Gaussian random variables

If two variables, X_1 and X_2, with Gaussian pdfs are summed to produce X, their characteristic functions will be multiplied together (equivalent to convolving their pdfs) to give

$$\Phi_X(u) = \Phi_{X_1}(u)\;\Phi_{X_2}(u) = e^{ju(\mu_1+\mu_2)}\; e^{-u^2(\sigma_1^2+\sigma_2^2)\,/\,2} \tag{3.44}$$

This is the characteristic function of a Gaussian pdf with mean $(\mu_1+\mu_2)$ and variance $(\sigma_1^2+\sigma_2^2)$.

Further Gaussian variables can be added and the pdf will remain Gaussian with further terms added to the above expressions for the combined mean and variance.

3.4.3 Central Limit Theorem

The central limit theorem states broadly that if a large number N of independent random variables of arbitrary pdf, but with equal variance σ^2 and zero mean, are summed together and scaled by $1/\sqrt{N}$ to keep the total energy independent of N, then the pdf of the resulting variable will tend to a zero-mean Gaussian with variance σ^2 as N tends to infinity.

This result is obvious from the previous result **if the input pdfs are also Gaussian**, but it is the fact that it applies **for arbitrary input pdfs** that is remarkable, and is the reason for the importance of the Gaussian (or normal) pdf. Noise generated in nature is nearly always the result of summing many tiny random processes (e.g. noise from electron energy transitions in a resistor or transistor, or from distant worldwide thunder storms at a radio antenna) and hence tends to a Gaussian pdf.

Although for simplicity, we shall prove the result only for the case when all the summed processes have the **same** variance and pdfs, the central limit result is more general than this and applies in many cases even when the variances and pdfs are not all the same.

Proof:

Let X_i ($i = 1$ to N) be the N independent random processes, each with zero mean and variance σ^2, which are combined to give

$$X = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i \tag{3.45}$$

Then, if the characteristic function of each input process before scaling is $\Phi(u)$ and we use equ. (3.41) to include the scaling by $1/\sqrt{N}$, the characteristic function of X is

$$\Phi_X(u) = \prod_{i=1}^{N} \Phi_{X_i}\left(\frac{u}{\sqrt{N}}\right) = \Phi^N\left(\frac{u}{\sqrt{N}}\right) \tag{3.46}$$

Taking logs:

$$\ln(\Phi_X(u)) = N \ \ln\left(\Phi\left(\frac{u}{\sqrt{N}}\right)\right) \tag{3.47}$$

Using Taylor's theorem to expand $\Phi\left(\frac{u}{\sqrt{N}}\right)$ in terms of its derivatives at $u = 0$ (and hence its moments) gives

$$\begin{aligned} \Phi\left(\frac{u}{\sqrt{N}}\right) \ &= \ \Phi(0) + \frac{u}{\sqrt{N}}\Phi'(0) + \frac{1}{2}\left(\frac{u}{\sqrt{N}}\right)^2 \Phi''(0) \\ &\quad + \frac{1}{6}\left(\frac{u}{\sqrt{N}}\right)^3 \Phi'''(0) + \frac{1}{24}\left(\frac{u}{\sqrt{N}}\right)^4 \Phi''''(0) + \ldots \end{aligned} \tag{3.48}$$

From the **Moments** property of characteristic functions with zero mean:

$$\begin{aligned}
\Phi(0) &= E[X_i^0] = 1 \quad \text{valid pdf} \\
\Phi'(0) &= jE[X_i] = 0 \quad \text{zero mean} \\
\Phi''(0) &= j^2E[X_i^2] = -\sigma^2 \quad \text{variance} \\
\Phi'''(0) &= j^3E[X_i^3] = -j\gamma\sigma^3 \quad \text{scaled skewness} \\
\Phi''''(0) &= j^4E[X_i^4] = (\kappa+3)\sigma^4 \quad \text{scaled kurtosis} \\
&\text{etc.}
\end{aligned} \tag{3.49}$$

These are all constants, independent of N, and dependent only on the shape of the pdfs f_{X_i}.

Substituting these moments into equations (3.47) and (3.48) and using the series expansion, $\ln(1+x) = x +$ (terms of order x^2 or smaller), gives

$$\begin{aligned}
\ln(\Phi_X(u)) &= N \ \ln\Big(\Phi\Big(\tfrac{u}{\sqrt{N}}\Big)\Big) \\
&= N \ \ln\Big(1 - \tfrac{u^2}{2N}\ \sigma^2 + (\text{terms of order } N^{-3/2} \text{ or smaller})\Big) \\
&= N\ \Big(-\tfrac{u^2\sigma^2}{2N} + (\text{terms of order } N^{-3/2} \text{ or smaller})\Big) \\
&= -\tfrac{u^2\sigma^2}{2} + (\text{terms of order } N^{-1/2} \text{ or smaller}) \\
&\to -\tfrac{u^2\sigma^2}{2} \quad \text{as } N \to \infty
\end{aligned} \tag{3.50}$$

$$\therefore\ \Phi_X(u) \to e^{-u^2\sigma^2/2} \quad \text{as } N \to \infty \tag{3.51}$$

Note that, if the input pdfs are symmetric, the skewness will be zero and the error terms will decay as N^{-1} rather than $N^{-1/2}$; and so convergence to a Gaussian characteristic function will be more rapid.

Hence we may now infer from equations (3.42), (3.43) and (3.51) that the pdf of X as $N \to \infty$ will be given by

$$f_X(x) = \tfrac{1}{\sqrt{2\pi\sigma^2}}\ e^{-x^2/2\sigma^2} \tag{3.52}$$

Thus we have proved the required central limit result.

Fig. 3.1(a) shows an example of convergence when the input pdfs are uniform, and N is gradually increased from 1 to 50. By $N = 12$, convergence is good, and this is how some 'Gaussian' random generator functions operate – by summing typically 12 uncorrelated random numbers with uniform pdfs.

For some less smooth or more skewed pdfs, convergence can be slower, as shown for a highly skewed triangular pdf in fig. 3.1(b); and pdfs of discrete processes are particularly problematic in this respect, as illustrated in fig. 3.1(c).

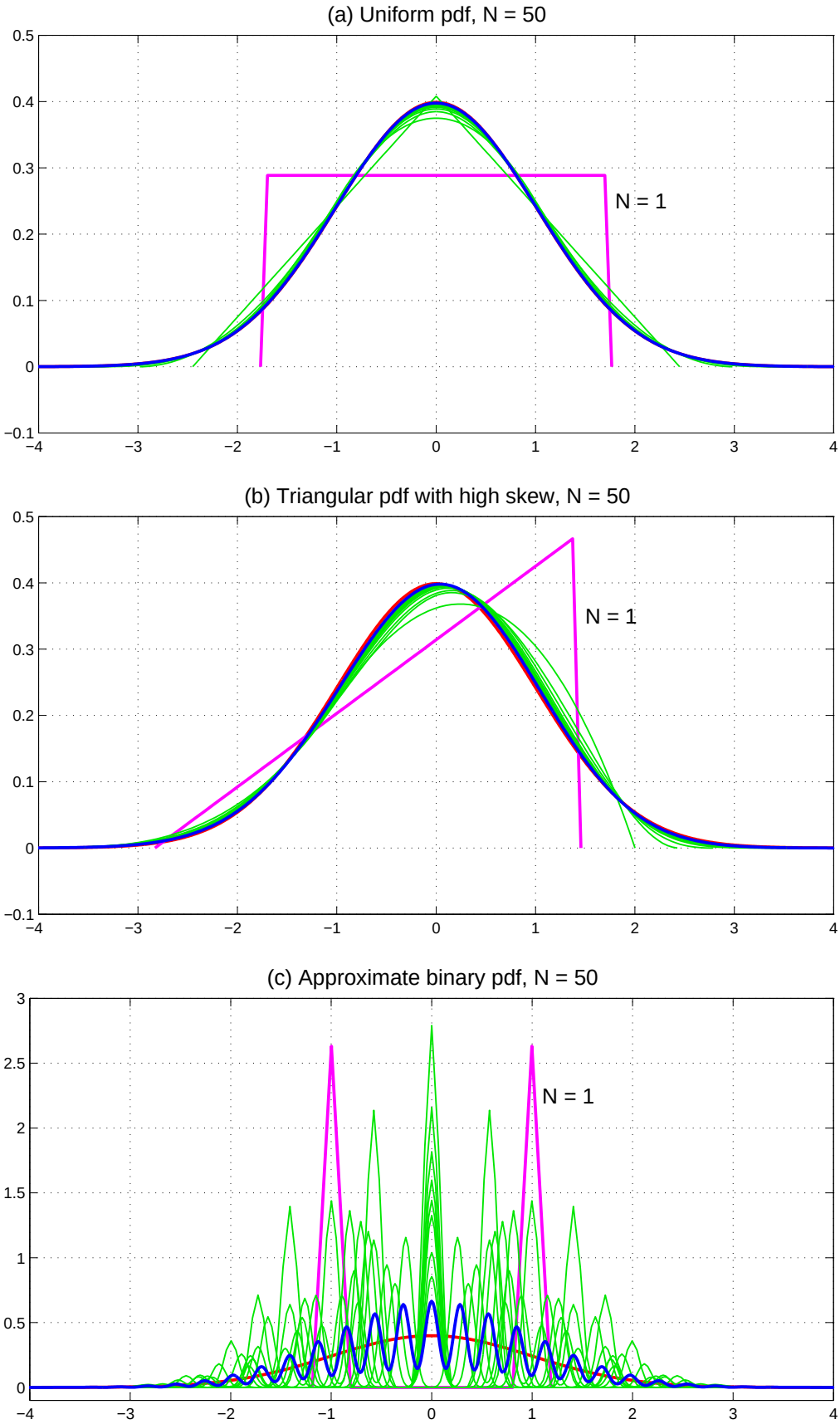


Fig.3.1: Convergence towards a Gaussian pdf (Central Limit Theorem) for 3 different input pdfs for $N = 1$ to 50. Note that the uniform pdf (a) with smallest higher-order moments converges fastest. Curves are shown for $N = \{1, 2, 3, 4, 6, 8, 10, 12, 15, 20, 30, 50\}$.

4 Correlation Functions and Power Spectra

4.1 Random Processes

We met **Random Signals** briefly in section 2.2 and now we return to consider them in detail. We shall assume that they evolve continuously with time t, although they may equally well evolve with distance (e.g. a random texture in image processing) or some other parameter.

We can imagine a generalization of our previous ideas about random experiments so that the outcome of an experiment can be a 'Random Object', an example of which is a signal waveform chosen at random from a set of possible signal waveforms, which we term an **Ensemble**. This ensemble of random signals is known as a **Random Process**.

Two examples of a Random Process $X(t, \alpha)$ are shown in fig. 4.1, where t is time and α is an index to the various members of the ensemble. The first example (a) shows several samples from a process whose statistics do not vary with time (a stationary process), while the second example (b) shows a time-dependent process such as different instances of similar speech sounds.

- t is assumed to belong to some set $\mathcal{T}$ (the time axis).
- α is assumed to belong to some set $\mathcal{A}$ (the sample space).
- If $\mathcal{T}$ is a continuous set, such as $\Re$ or $[0, \infty)$, then the process is termed a **Continuous Time** random process.
- If $\mathcal{T}$ is a discrete set of time values, such as the integers $\mathcal{Z}$, the process is termed a **Discrete Time Process** or **Time Series**.
- The members of the ensemble can be the result of different random events, such as different instances of the sound 'ah' during the course of this lecture. In this case α is discrete.
- Alternatively the ensemble members are often just different portions of a single random signal. If the signal is a continuous waveform, then α may also be a continuous variable, indicating the starting point of each ensemble waveform.

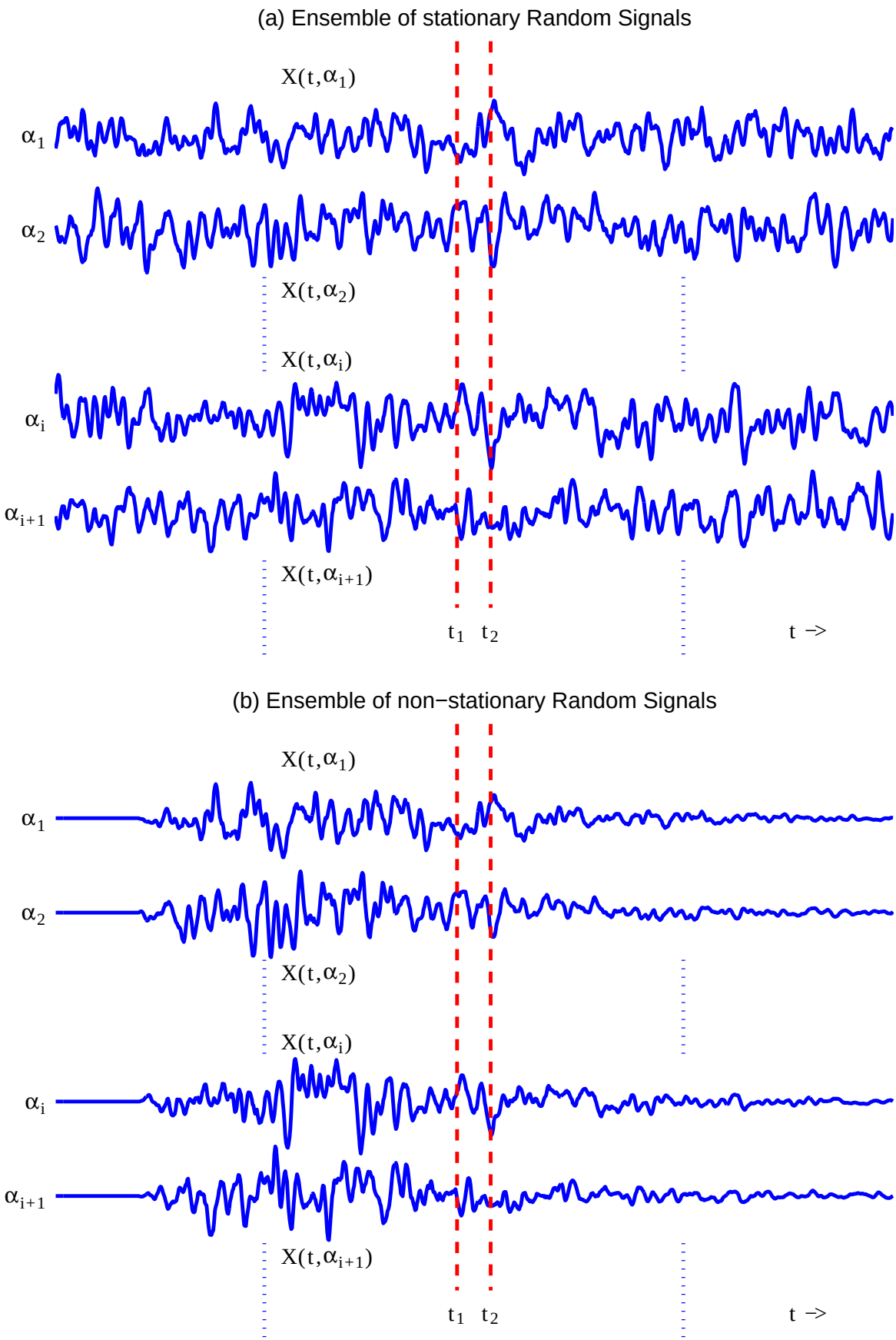
We will often drop the explicit dependence on α for notational convenience, referring simply to **random process** $\{X(t)\}$.

If we consider the process $\{X(t)\}$ at one particular time $t = t_1$, then we have a **random variable** $X(t_1)$.

If we consider the process $\{X(t)\}$ at N time instants $\{t_1, t_2, \ldots, t_N\}$, then we have a **random vector**:

$$\mathbf{X} = [X(t_1), X(t_2), \ldots X(t_N)]^T$$

We can study the properties of a random process by considering the behaviour of random variables and random vectors extracted from the process, using the probability theory derived earlier in this course.


(a) Ensemble of stationary Random Signals
X(t,α1)
α1
α2
X(t,α2)
X(t,αi)
αi
αi+1
X(t,αi+1)
t1 t2
t ->
(b) Ensemble of non−stationary Random Signals
X(t,α1)
α1
α2
X(t,α2)
X(t,αi)
αi
αi+1
X(t,αi+1)
t1 t2
t ->
Fig. 4.1: Ensemble representations of a random process:
(a) with stationary properties; (b) with time-varying properties.

4.2 Correlation and Covariance

Correlation and covariance are techniques for measuring the similarity of one set of signal samples to another. For a random process $X(t,\alpha)$ they are defined as follows.

- **Auto-correlation function:**

$$r_{XX}(t_1,t_2) = E[X(t_1,\alpha)\,X(t_2,\alpha)] = \int\int x_1\,x_2\,f(x_1,x_2)\,dx_1\,dx_2 \qquad (4.1)$$

where the expectation is performed over all $\alpha \in \mathcal{A}$ (i.e. the whole ensemble), and $f(x_1,x_2)$ is the joint pdf when x_1 and x_2 are samples taken at times t_1 and t_2 from the **same random event** α of the random process X.

- **Auto-covariance function:**

$$\begin{aligned} c_{XX}(t_1,t_2) &= E[(X(t_1,\alpha)-\overline{X(t_1)})(X(t_2,\alpha)-\overline{X(t_2)})] \\ &= \int_{x_1}\int_{x_2}(x_1-\overline{X(t_1)})(x_2-\overline{X(t_2)})\,f(x_1,x_2)\,dx_1\,dx_2 \\ \text{using equ (1.29):} \quad &= r_{XX}(t_1,t_2) - \overline{X(t_1)}\int x_2\,f_2(x_2)\,dx_2 \\ &\quad - \overline{X(t_2)}\int x_1\,f_1(x_1)\,dx_1 + \overline{X(t_1)}\;\overline{X(t_2)} \\ &= r_{XX}(t_1,t_2) - \overline{X(t_1)}\;\overline{X(t_2)} \qquad (4.2) \end{aligned}$$

where the same conditions apply as for auto-correlation and the means $\overline{X(t_1)}$ and $\overline{X(t_2)}$ are taken over all $\alpha \in \mathcal{A}$. Covariances are equivalent to correlations except that the **means are removed from the data first**.

- **Cross-correlation function:**

If we have two different processes, $X(t,\alpha)$ and $Y(t,\alpha)$, both arising as a result of the same random event α, then cross-correlation is defined as

$$r_{XY}(t_1,t_2) = E[X(t_1,\alpha)\,Y(t_2,\alpha)] = \int\int x_1\,y_2\,f(x_1,y_2)\,dx_1\,dy_2 \qquad (4.3)$$

where $f(x_1,y_2)$ is the joint pdf when x_1 and y_2 are samples of X and Y taken at times t_1 and t_2 as a result of the **same random event** α. Again the expectation is performed over all $\alpha \in \mathcal{A}$.

- **Cross-covariance function:**

$$\begin{aligned} c_{XY}(t_1,t_2) &= E[(X(t_1,\alpha)-\overline{X(t_1)})(Y(t_2,\alpha)-\overline{Y(t_2)})] \\ &= \int_{x_1}\int_{y_2}(x_1-\overline{X(t_1)})(y_2-\overline{Y(t_2)})\,f(x_1,y_2)\,dx_1\,dy_2 \\ &= r_{XY}(t_1,t_2) - \overline{X(t_1)}\;\overline{Y(t_2)} \qquad (4.4) \end{aligned}$$

For Deterministic Random Processes in which X (and Y) depend deterministically on the random variable α (or some function of it), we can simplify the above integrals by expressing the joint pdf in that space. E.g. for auto-correlation:

$$r_{XX}(t_1,t_2) = E[X(t_1,\alpha)\,X(t_2,\alpha)] = \int_{\mathcal{A}} x(t_1,\alpha)\,x(t_2,\alpha)\,f(\alpha)\,d\alpha \qquad (4.5)$$

4.3 Stationarity

Stationarity in a Random Process implies that its statistical characteristics **do not change with time**. Put another way, if one were to observe a stationary random process at some time t it would be impossible to distinguish the **statistical characteristics** at that time from those at some other time t'. Fig. 4.1(a) shows a stationary process.

4.3.1 Strict Sense Stationarity (SSS)

Choose a Random Vector of length N from a Random Process:

$$\mathbf{X} = [X(t_1),\, X(t_2),\, \ldots,\, X(t_N)]^T \tag{4.6}$$

Its Nth order cdf is

$$F_{X(t_1),\, \ldots\, X(t_N)}(x_1,\ \ldots\ x_N) = \Pr\{X(t_1) \leq x_1,\ \ldots\ X(t_N) \leq x_N\} \tag{4.7}$$

$X(t)$ is defined to be **Strict Sense Stationary** iff:

$$F_{X(t_1),\, \ldots\, X(t_N)}(x_1,\ \ldots\ x_N) = F_{X(t_1+c),\, \ldots\, X(t_N+c)}(x_1,\ \ldots\ x_N) \tag{4.8}$$

for all time shifts c, all finite N and all sets of time points $\{t_1,\ \ldots\ t_N\}$.

4.3.2 Wide Sense (Weak) Stationarity (WSS)

If we are only interested in the properties of moments up to 2nd order (mean, autocorrelation, covariance, ...), which is the case for many practical applications, a weaker form of stationarity can be useful:

$X(t)$ is defined to be **Wide Sense Stationary** (or Weakly Stationary) iff:

1. The mean value is independent of t

$$E[X(t)] = \mu \quad \text{for all } t \tag{4.9}$$

2. Autocorrelation depends only upon $\tau = t_2 - t_1$

$$E[X(t_1)\, X(t_2)] = E[X(t_1)\, X(t_1+\tau)] = r_{XX}(\tau) \quad \text{for all } t_1 \tag{4.10}$$

Note that, since 2nd-order moments are defined in terms of 2nd-order probability distributions, strict sense stationary processes are **always** wide-sense stationary, but not necessarily **vice versa**.

4.4 Ergodicity

Many stationary random processes are also **Ergodic**. For an Ergodic Random Process we can exchange **Ensemble Averages** for **Time Averages**. This is equivalent to assuming that our ensemble of random signals is just composed of all possible time shifts of a single signal $X(t)$. Fig. 4.1(a) could be an Ergodic process, while Fig. 4.1(b) could not (it is non-stationary).

Recall from equ. (3.4) that the expectation of a function of a random variable is given by

$$E[g(X)] = \int g(x)\, f_X(x)\, dx \tag{4.11}$$

This result also applies if we have a **random function** $g(.)$ of a **deterministic variable** such as time t. Hence

$$E[g(t)] = \int g(t)\, f_T(t)\, dt \tag{4.12}$$

Because t is linearly increasing, the pdf $f_T(t)$ is uniform over our measurement interval, say $-T$ to T, and will be $1\,/\,2T$ to make the pdf valid (integral = 1). Hence

$$E[g(t)] = \int_{-T}^{T} g(t)\, \tfrac{1}{2T}\, dt = \tfrac{1}{2T} \int_{-T}^{T} g(t)\, dt \tag{4.13}$$

If we wish to measure over all time, then we take the limit as $T \to \infty$.

This leads to the following results for Ergodic WSS random processes:

- **Mean Ergodic:**

$$E[X(t)] = \int_{-\infty}^{\infty} x\, f_{X(t)}(x)\, dx = \lim_{T\to\infty} \tfrac{1}{2T} \int_{-T}^{T} X(t)\, dt \tag{4.14}$$

- **Correlation Ergodic:**

$$\begin{aligned} r_{XX}(\tau) &= E[X(t)X(t+\tau)] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1\, x_2\, f_{X(t),X(t+\tau)}(x_1, x_2)\, dx_1\, dx_2 \\ &= \lim_{T\to\infty} \tfrac{1}{2T} \int_{-T}^{T} X(t)\, X(t+\tau)\, dt \end{aligned} \tag{4.15}$$

  and similarly for other correlation or covariance functions.

Ergodicity greatly simplifies the measurement of WSS processes and it is often assumed when estimating moments (or correlations) for such processes.

In almost all practical situations, processes are stationary **only over some limited time interval** (say T_1 to T_2) rather than over all time. In that case we deliberately keep the limits of the integral finite and adjust $f_{X(t)}$ accordingly. For example the autocorrelation function is then measured using

$$r_{XX}(\tau) = \tfrac{1}{T_2 - T_1} \int_{T_1}^{T_2} X(t)\, X(t+\tau)\, dt \tag{4.16}$$

This avoids including samples of X which have incorrect statistics, but it can suffer from errors due to limited sample size.

4.5 Spectral Properties of Random Signals

4.5.1 Relation of Power Spectral Density to ACF

The autocorrelation function (ACF) of an ergodic random signal tells us how correlated the signal is with itself as a function of time shift τ. In particular, for $\tau = 0$

$$r_{XX}(0) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X^2(t)\, dt = \text{ mean power of } X(t) \tag{4.17}$$

Note that if $T \to \infty$,

$$r_{XX}(\tau) = r_{XX}(-\tau) \leq r_{XX}(0) \quad \text{for all } \tau \tag{4.18}$$

As τ becomes large, $X(t)$ and $X(t+\tau)$ will usually become decorrelated and, as long as X is zero mean, r_{XX} will tend to zero.

Hence the ACF will have its **maximum at $\tau = 0$** and decay symmetrically to zero (or to μ^2, if the mean, $\mu \neq 0$) as $|\tau|$ increases.

The width of the ACF (to say its half-power points) tells us how slowly X is fluctuating or how bandlimited it is. Fig. 4.2(b) shows how the ACF of a rapidly fluctuating (wide-band) random signal, as in fig. 4.2(a) upper plot, decays quickly to zero as $|\tau|$ increases, whereas, for a slowly fluctuating signal, as in fig. 4.2(a) lower plot, the ACF decays much more slowly.

The ACF measures an entirely different aspect of randomness from amplitude distributions such as pdf and cdf.

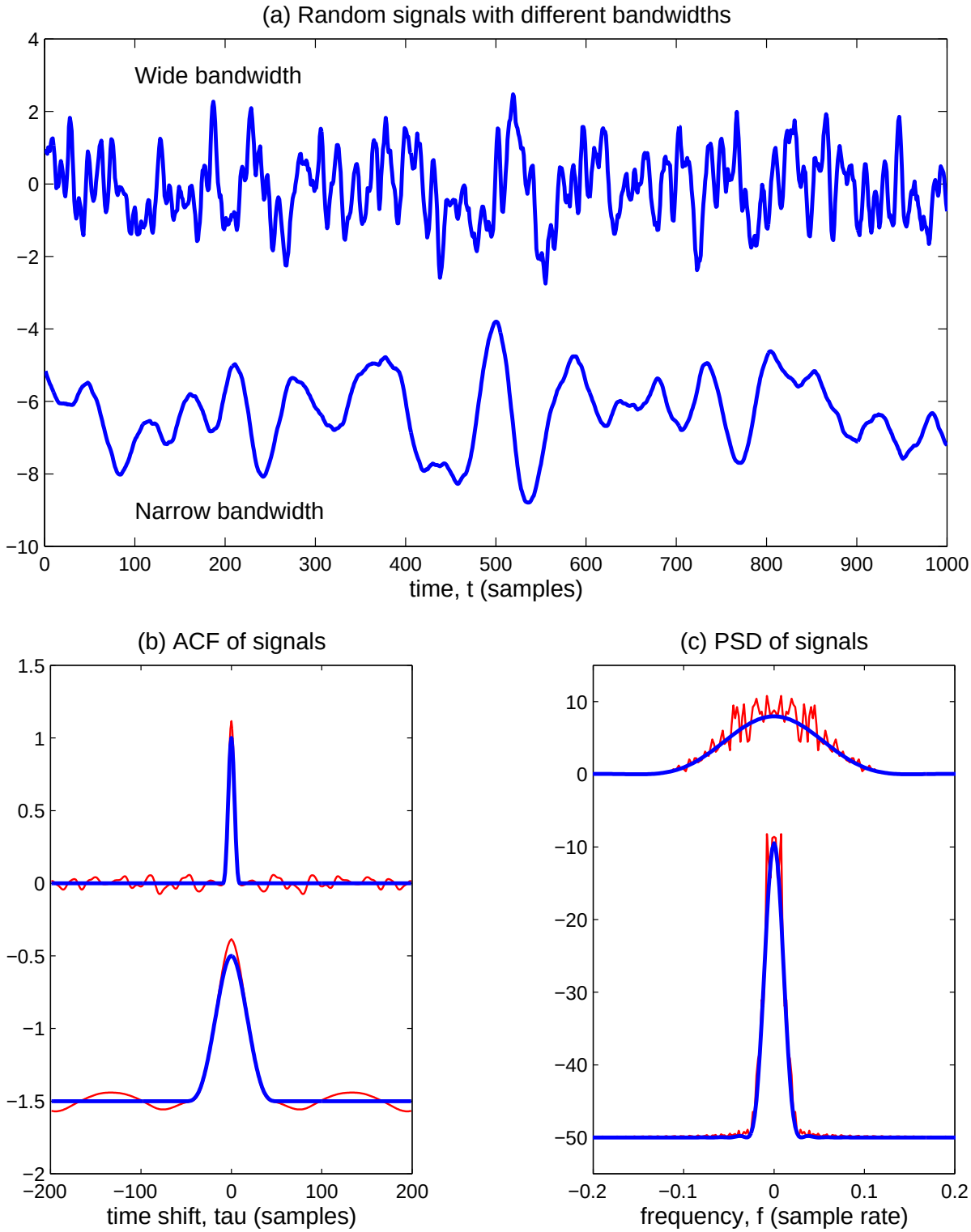


Fig. 4.2: Illustration of the different properties of wide band (upper) and narrow band (lower) random signals: (a) the signal waveforms with unit variance; (b) their autocorrelation functions (ACFs); and (c) their power spectral densities (PSDs). In (b) and (c), the thin fluctuating curves show the actual values measured from 4000 samples of the random waveforms while the thick smooth curves show the limits of the ACF and PSD as the lengths of the waveforms tend to infinity.

As with deterministic signals, we may formalise our ideas of rates of fluctuation by transforming to the **Frequency (Spectral) Domain** using the **Fourier Transform**:

$$\mathcal{F}_u(\omega) = \mathrm{FT}\{u(t)\} = \int u(t)\, e^{-j\omega t}\, dt \tag{4.19}$$

The **Power Spectral Density** (PSD) of a random process X is defined to be the Fourier Transform of its ACF:

$$\mathcal{S}_X(\omega) = \mathrm{FT}\{r_{XX}(\tau)\} = \int r_{XX}(\tau)\, e^{-j\omega\tau}\, d\tau \tag{4.20}$$

$$r_{XX}(\tau) = \mathrm{FT}^{-1}\{\mathcal{S}_X(\omega))\} = \tfrac{1}{2\pi}\int \mathcal{S}_X(\omega)\, e^{j\omega\tau}\, d\omega \tag{4.21}$$

N.B. $\{X(t)\}$ must be **at least** Wide Sense Stationary (WSS) for this to be valid.

From equations (4.17) and (4.21) we see that the mean signal power is given by:

$$r_{XX}(0) = \tfrac{1}{2\pi}\int \mathcal{S}_X(\omega)\, d\omega = \int \mathcal{S}_X(2\pi f)\, df \tag{4.22}$$

Hence $\mathcal{S}_X$ has units of power per Hertz. Note that we must integrate over **all** frequencies, both positive and negative, to get the correct total power.

Fig. 4.2(c) shows how the PSDs of the signals relate to the ACFs in fig. 4.2(b).

Properties of PSDs for real-valued $X(t)$:

1. $\mathcal{S}_X(\omega) = \mathcal{S}_X(-\omega)$
2. $\mathcal{S}_X(\omega)$ is Real-valued
3. $\mathcal{S}_X(\omega) \geq 0$

Properties 1 and 2 are because ACFs are real and symmetric about $\tau = 0$; and 3 is because $\mathcal{S}_X$ represents **power** density.

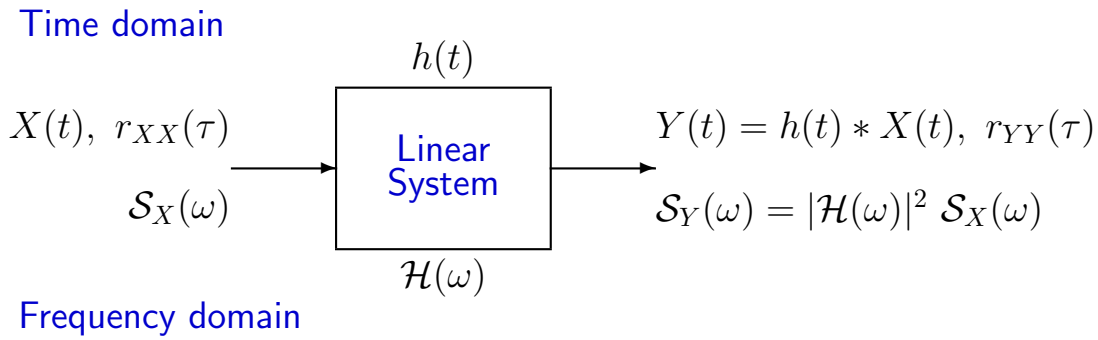


Fig. 4.3: Block diagram of a linear system with a random input signal, $X(t)$.

4.5.2 Linear system (filter) with WSS input

Let the linear system with input $X(t)$ and output $Y(t)$ have an impulse response $h(t)$, so

$$Y(t) = h(t) * X(t) = \int h(\beta)\, X(t-\beta)\, d\beta \tag{4.23}$$

Then the ACF of Y is

$$\begin{aligned} r_{YY}(t_1,t_2) &= E\left[Y(t_1)\; Y(t_2)\right] \\ &= E\left[\left(\int h(\beta_1)X(t_1-\beta_1)\,d\beta_1\right)\left(\int h(\beta_2)X(t_2-\beta_2)\,d\beta_2\right)\right] \\ &= E\left[\int\int h(\beta_1)\,h(\beta_2)\,X(t_1-\beta_1)\,X(t_2-\beta_2)\,d\beta_1\,d\beta_2\right] \\ &= \int\int h(\beta_1)\,h(\beta_2)\,E\left[X(t_1-\beta_1)\,X(t_2-\beta_2)\right]\,d\beta_1\,d\beta_2 \\ &= \int\int h(\beta_1)\,h(\beta_2)\,r_{XX}(t_1-\beta_1,t_2-\beta_2)\,d\beta_1\,d\beta_2 \end{aligned} \tag{4.24}$$

If X is WSS, then we substitute $\tau = t_2 - t_1$ and $t = t_1$ to get

$$\begin{aligned} r_{YY}(\tau) &= E\left[Y(t)\; Y(t+\tau)\right] \\ &= \int\int h(\beta_1)\,h(\beta_2)\,r_{XX}(\tau+\beta_1-\beta_2)\,d\beta_1\,d\beta_2 \\ &= r_{XX}(\tau) * h(-\tau) * h(\tau) \end{aligned} \tag{4.25}$$

Taking Fourier transforms:

$$\begin{aligned} \mathcal{S}_Y(\omega) &= \mathrm{FT}\{r_{YY}(\tau)\} \\ &= \int\left(\int\int h(\beta_1)\,h(\beta_2)\,r_{XX}(\tau+\beta_1-\beta_2)\,d\beta_1\,d\beta_2\right)\,e^{-j\omega\tau}\,d\tau \\ &= \int\int h(\beta_1)\,h(\beta_2)\left(\int r_{XX}(\tau+\beta_1-\beta_2)\,e^{-j\omega\tau}\,d\tau\right)\,d\beta_1\,d\beta_2 \\ &= \int\int h(\beta_1)\,h(\beta_2)\left(\int r_{XX}(\lambda)\,e^{-j\omega(\lambda-\beta_1+\beta_2)}\,d\lambda\right)\,d\beta_1\,d\beta_2 \\ &= \left(\int h(\beta_1)\,e^{j\omega\beta_1}\,d\beta_1\right)\left(\int h(\beta_2)\,e^{-j\omega\beta_2}\,d\beta_2\right)\left(\int r_{XX}(\lambda)\,e^{-j\omega\lambda}\,d\lambda\right) \\ &= \mathcal{H}^*(\omega)\;\mathcal{H}(\omega)\;\mathcal{S}_X(\omega) \qquad \text{where} \quad \mathcal{H}(\omega) = \mathrm{FT}\{h(t)\} \end{aligned} \tag{4.26}$$

$$\text{i.e:} \quad \mathcal{S}_Y(\omega) \;=\; |\mathcal{H}(\omega)|^2 \; \mathcal{S}_X(\omega) \tag{4.27}$$

Hence the PSD of Y = the PSD of X $\times$ the power gain $|\mathcal{H}|^2$ of the system at frequency ω.

Thus if a large and important system is subject to random perturbations (e.g. a power plant subject to random load fluctuations), we may measure $r_{XX}(\tau)$ and $r_{YY}(\tau)$, transform these to $\mathcal{S}_X(\omega)$ and $\mathcal{S}_Y(\omega)$, and hence obtain

$$|\mathcal{H}(\omega)| = \sqrt{\frac{\mathcal{S}_Y(\omega)}{\mathcal{S}_X(\omega)}} \tag{4.28}$$

Hence we may measure the system frequency response **without taking the plant off line**.

But this does not give any information about the **phase** of $\mathcal{H}(\omega)$.

To get the phase of $\mathcal{H}(\omega)$, we instead measure the **Cross-Correlation Function** (CCF) between X and Y:

$$\begin{aligned} r_{XY}(t_1, t_2) &= E\left[X(t_1)\; Y(t_2)\right] \\ &= E\left[X(t_1)\left(\int h(\beta)\, X(t_2-\beta)\, d\beta\right)\right] \\ &= E\left[\int h(\beta)\, X(t_1)\, X(t_2-\beta)\, d\beta\right] \\ &= \int h(\beta)\, E\left[X(t_1)\, X(t_2-\beta)\right]\, d\beta \\ &= \int h(\beta)\, r_{XX}(t_1, t_2-\beta)\, d\beta \end{aligned} \tag{4.29}$$

If $X(t)$, and hence $Y(t)$, are WSS:

$$r_{XY}(\tau) = E\left[X(t)\; Y(t+\tau)\right] = \int h(\beta)\, r_{XX}(\tau-\beta)\, d\beta = h(\tau) * r_{XX}(\tau) \tag{4.30}$$

and taking Fourier transforms:

$$\mathcal{S}_{XY}(\omega) = \mathrm{FT}\{r_{XY}(\tau)\} = \mathcal{H}(\omega)\; \mathcal{S}_X(\omega) \tag{4.31}$$

where $\mathcal{S}_{XY}(\omega)$ is known as the **Cross Spectral Density** between X and Y.

$$\therefore \;\; \mathcal{H}(\omega) = \frac{\mathcal{S}_{XY}(\omega)}{\mathcal{S}_X(\omega)} \tag{4.32}$$

Hence we obtain the **amplitude and phase** of $\mathcal{H}(\omega)$. As before, this is achieved without taking the plant off line.

Note that for WSS processes, $r_{XY}(\tau) = r_{YX}(-\tau)$, but that (unlike r_{XX} and r_{YY}) these need not be symmetric about $\tau = 0$, since in general $r_{XY}(\tau) \neq r_{XY}(-\tau)$. Hence the cross spectral density $\mathcal{S}_{XY}(\omega)$ need not be purely real (unlike $\mathcal{S}_X(\omega)$), and the phase of $\mathcal{S}_{XY}(\omega)$ gives the phase of $\mathcal{H}(\omega)$.

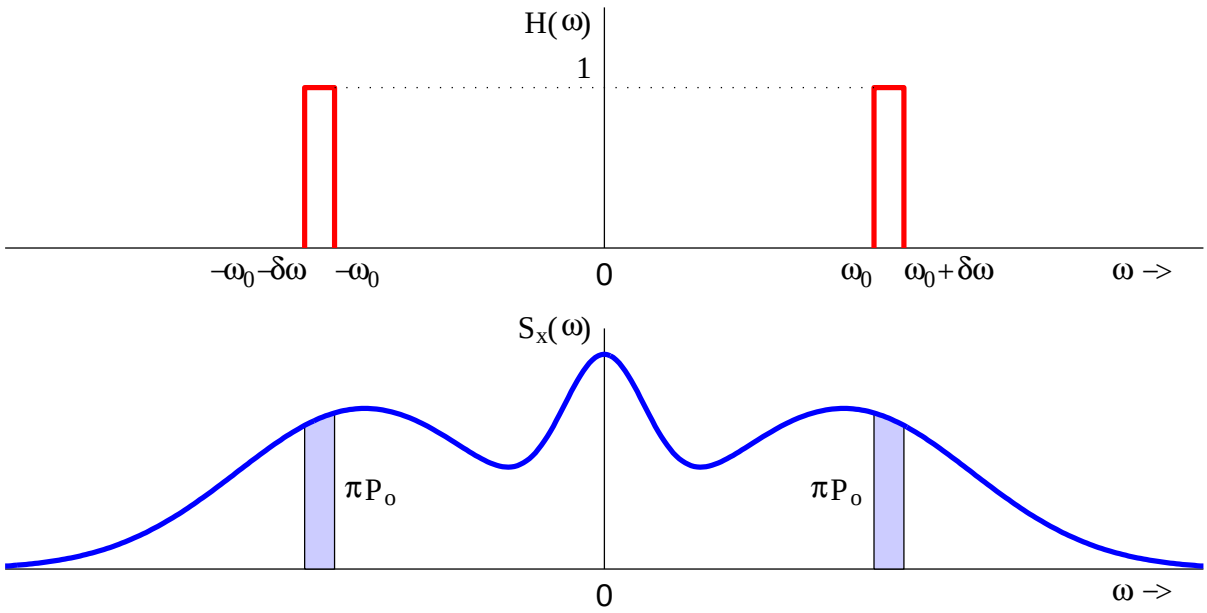


Fig. 4.4: Narrowband filter frequency response and PSD of filter input and output.

4.5.3 Physical Interpretation of Power Spectral Density

Let us pass $X(t)$ through a narrow-band filter of bandwidth $\delta\omega = 2\pi\,\delta f$, as shown in fig. 4.4:

$$\mathcal{H}(\omega) = \begin{cases} 1 & \text{for } \omega_0 < |\omega| \le \omega_0 + \delta\omega \\ 0 & \text{otherwise} \end{cases} \tag{4.33}$$

Now we find the average power at the filter output (shaded area in fig. 4.4, divided by 2π):

$$\begin{aligned} P_o &= r_{YY}(0) = \tfrac{1}{2\pi}\int_{-\infty}^{\infty} \mathcal{S}_Y(\omega)\,d\omega \\ &= \tfrac{1}{2\pi}\int_{-\infty}^{\infty} \mathcal{S}_X(\omega)\,|\mathcal{H}(\omega)|^2\,d\omega \\ &= \tfrac{1}{2\pi}\left(\int_{-(\omega_0+\delta\omega)}^{-\omega_0} \mathcal{S}_X(\omega)\,d\omega + \int_{\omega_0}^{\omega_0+\delta\omega} \mathcal{S}_X(\omega)\,d\omega\right) \\ &\simeq \tfrac{1}{2\pi}\left(\mathcal{S}_X(-\omega_0) + \mathcal{S}_X(\omega_0)\right)\,\delta\omega \;=\; 2\,\mathcal{S}_X(\omega_0)\,\frac{\delta\omega}{2\pi} \end{aligned} \tag{4.34}$$

since $\mathcal{S}_X(-\omega) = \mathcal{S}_X(\omega)$ if X is purely real.

Expressed in terms of frequencies f_0 and δf in Hz:

$$P_o \;\simeq\; 2\;\mathcal{S}_X(2\pi f_0)\;\delta f \tag{4.35}$$

The factor of 2 appears because our filter responds to both negative and positive frequency components of X.

Hence $\mathcal{S}_X$ is indeed a **Power Spectral Density** with units V^2/Hz (assuming unit impedance).

4.6 White and Coloured Processes

4.6.1 White Noise

If we have a zero-mean Wide Sense Stationary process X, it is a **White Noise Process** if its ACF is a delta function at $\tau = 0$, i.e. it is of the form:

$$r_{XX}(\tau) = P_X \ \delta(\tau) \tag{4.36}$$

where P_X is a constant.

The PSD of X is then given by

$$\mathcal{S}_X(\omega) = \int P_X \, \delta(\tau) \, e^{-j\omega\tau} \, d\tau \ = P_X \ e^{-j\omega 0} = P_X \tag{4.37}$$

Hence X is **white**, since it contains equal power at **all** frequencies, as in **white light**.

P_X is the PSD of X at all frequencies.

But:

$$\text{Power of } X = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{S}_X(\omega) \, d\omega = \infty \tag{4.38}$$

so the White Noise Process is unrealisable in practice, because of its infinite bandwidth.

However, it is very useful as a conceptual entity and as an approximation to realisable 'nearly white' processes which have finite bandwidth, but which are 'white' over all frequencies of practical interest. For 'nearly white' processes, $r_{XX}(\tau)$ is a narrow pulse of non-zero width, and $\mathcal{S}_X(\omega)$ is flat from zero up to some relatively high cutoff frequency, above which it then decays to zero.

4.6.2 Strict Whiteness and i.i.d. Processes

Usually the above concept of whiteness is sufficient, but a much stronger definition is as follows:

Pick a set of times $\{t_1, t_2, \ldots, t_N\}$ to sample $X(t)$.

If, for **any choice** of $\{t_1, t_2, \ldots, t_N\}$ with N finite, the random variables $X(t_1), X(t_2), \ldots X(t_N)$ are **jointly independent**, i.e. their joint pdf is given by

$$f_{X(t_1),X(t_2),\ldots X(t_N)}(x_1, x_2, \ldots x_N) = \prod_{i=1}^{N} f_{X(t_i)}(x_i) \tag{4.39}$$

and the marginal pdfs are identical, i.e.

$$f_{X(t_1)} = f_{X(t_2)} = \ldots = f_{X(t_N)} = f_X \tag{4.40}$$

then the process is termed **Independent and Identically Distributed (i.i.d.)**.

If, in addition, f_X is a pdf with zero mean, we have a **Strictly White Noise Process**.

An i.i.d. process is 'white' because the variables $X(t_i)$ and $X(t_j)$ are jointly independent, even when separated by an infinitesimally small interval between t_i and t_j .

4.6.3 Additive White Gaussian Noise (AWGN)

In many systems the concept of **Additive White Gaussian Noise (AWGN)** is used. This simply means a process which has a Gaussian pdf, a white PSD, and is linearly added to whatever signal we are analysing.

Note that although 'white' and 'Gaussian' often go together, this is **not necessary** (especially for 'nearly white' processes).

E.g. a very high speed random bit stream has an ACF which is approximately a delta function, and hence is a nearly white process, but its pdf is clearly not Gaussian – it is a pair of delta functions at $+V$ and $-V$, the two voltage levels of the bit stream.

Conversely a nearly white gaussian process which has been passed through a lowpass filter (see next section) will still have a Gaussian pdf (as it is a summation of Gaussians) but will no longer be white.

4.6.4 Coloured Processes

A random process whose PSD is not white or nearly white, is often known as a **coloured noise** process.

We may obtain coloured noise $Y(t)$ with PSD $\mathcal{S}_Y(\omega)$ simply by passing white (or nearly white) noise $X(t)$ with PSD P_X through a filter with frequency response $\mathcal{H}(\omega)$, such that from equ. (4.27)

$$\mathcal{S}_Y(\omega) \;=\; \mathcal{S}_X(\omega)\;|\mathcal{H}(\omega)|^2 = P_X\;|\mathcal{H}(\omega)|^2 \tag{4.41}$$

Hence if we design the filter such that

$$|\mathcal{H}(\omega)| = \sqrt{\frac{\mathcal{S}_Y(\omega)}{P_X}} \tag{4.42}$$

then $Y(t)$ will have the required coloured PSD.

For this to work, $\mathcal{S}_X(\omega)$ need only be constant (white) over the passband of the filter, so a **nearly white** process which satisfies this criterion is quite satisfactory and realisable.

Using equations (4.25) and (4.36), the ACF of the coloured noise is given by

$$\begin{aligned} r_{YY}(\tau) &= r_{XX}(\tau) * h(-\tau) * h(\tau) \\ &= P_X\;\delta(\tau) * h(-\tau) * h(\tau) \\ &= P_X\,h(-\tau) * h(\tau) \end{aligned} \tag{4.43}$$

where $h(\tau)$ is the impulse response of the filter.

Fig. 4.2 shows two examples of coloured noise, although the upper waveform is more 'nearly white' than the lower one, as can be seen in fig. 4.2(c) in which the upper PSD is flatter than the lower PSD. In these cases, the coloured waveforms were produced by passing uncorrelated random noise samples (white up to half the sampling frequency) through half-sine filters (as in equ. (2.9)) of length $T_b = 10$ and 50 samples respectively.

5 Summary

In this course we have covered the following key topics:

Probability Distributions (pmf, cdf and pdf): to define the amplitude distribution of a random process.

Conditional probabilities and Bayes' rule: to deal with dependent random processes and model selection.

Random vectors and signals: more complicated random processes which are multi-dimensional and/or functions of time.

Functions of random processes: to convert one random process into another with a different distribution.

Expectations and moments: to parameterise pdfs and cdfs.

Sums of random processes: produce convolution of pdfs.

Characteristic functions: to turn convolutions into products and allow easy calculation of moments.

Correlation functions of random processes: to define time-dependent behaviour (e.g. fast or slow variation).

Stationarity and ergodicity: constraints on the time-dependent behaviour of random signals.

Spectral properties and linear system responses: frequency-dependent behaviour of random signals and the effects of linear systems on this. Analysis of linear systems using random (online) excitation waveforms.

White and coloured processes: random signals with important spectral properties.

Significant further material on discrete-time random signals will be presented in module 3F3.