# Contents

1 On the predictability of seizures 3

1.1 Predictability - Terminology made clear 7

1.2 How to Estimate LRD 12

1.3 Seizure Frequency Dataset 24

1.4 Analysis - Estimation of $\alpha$ 26

1.5 Memory and Predictability of Seizures 33

1.6 Conclusions 36
Chapter 1

On the predictability of seizures

“The farther backward you can look, the farther forward you are likely to see.”

- Sir Winston Churchill, (1874-1965)

Prediction is a difficult task - to be able to forecast the future an in-depth knowledge of both past and present is required. The more complex a system is, the more information that is necessary to predict the same distance into the future. Although Winston Churchill’s words were probably spoken in a more philosophical sense they hold true also for numerical analysis - how far back in time do we have to look in order to be able to tell anything about the future?

Perhaps a good example that points out the difficulties of this problem is a type of prediction well known to us all: the weather. We all turn our TVs and expect that there is truth in what is being told to us about what the weather will be like tomorrow, the day after, or even next week. But how are these predictions made? The most comprehensive and versatile solution is to create a physical model of the earth’s atmosphere that accounts for all factors that affect the weather at any point on this earth. Humidity, topography, air pressure, and even how many people decide to drive to work today all need to be considered. This is of course an incredibly complex system for which even our most modern modelling tools may not be sophisticated enough.

“When the number of factors coming into play in a phenomenological complex is too large, scientific method in most cases fails us. One need only think of the weather, in which case prediction even for a few days ahead is impossible. Nevertheless no one doubts that we are confronted with a causal connection whose causal components are in
the main known to us. Occurrences in this domain are beyond the reach of exact prediction because of the variety of factors in operation, not because of any lack of order in nature.”

- *Science, Philosophy and Religion, Albert Einstein, (1879-1955)*

A physical model could be used to explain what the mechanisms that generate known patterns are, but they are often too complex to realise even with appropriate simplifications. In such cases an alternate approach is to base predictions on the assumption of a *stochastic* model. Future events are inferred through the analysis of statistics gathered over time - the more information that is collected, so long as this information is still valid, the more accurate the forecasts are likely to be. The statistics do not themselves explain anything about the processes involved (even though physical processes may be inferred through the study of these statistics) and are therefore known as *black box* methods. Weather prediction is an example of this, possible only because monitoring of weather patterns has been in place for many decades. Although a physical model is less limited in what it can tell us, for such a complex system statistical analysis is today the only practical solution. Future developments may change this.

The same observations hold true for seizure prediction - epilepsy is undoubtedly the result of one of the more complex systems that humans are trying to understand. Current “predictors” of seizures rely on the assumption of stochasticity simply because our models of the brain, presented in the last chapter, are not developed enough to be usable for this task.

In any case whether physical or black box models are used the fundamental question “are seizures predictable?” is rarely asked. They have been assumed predictable because up to 50% of people suffering from epilepsy are able to predict their own seizures, usually with warning symptoms such as headaches or mood alterations that appear well before the clinical onset [8, 14]. In some cases people are able to learn how to use these warning symptoms (also known as *auras*) to prevent their incidence [1]. However, short term auras are accepted as the point at which a seizure has already begun without yet impairing consciousness. Long term auras have been linked to the build-up of epileptic activity. Auras are therefore *detectors* rather than predictors of epilepsy\(^1\). A systematic study that determines if seizures are indeed predictable does not to the best of our knowledge exist.

Given that auras often occur well before clinical onset the question “is there even a need for seizure prediction?” is also valid. The purpose of seizure prediction

---

\(^1\)Another argument for the existence of a pre-seizure state is the ability of some dogs to “predict” seizures - but once again since these dogs also predict pseudo-seizures they are detectors rather than predictors, most likely through motor-based symptoms [6, 11].
algorithms is to provide sufficient warning to allow for some form of intervention. This is predominantly important for patients that cannot be surgically or pharmacologically treated - by knowing when a seizure is imminent they can retain some level of control over their own lives. It is also useful for the more timely delivery of fast acting drugs or electrical stimulation, in this way reducing the side effects associated with traditional pre-emptive treatments. But if the development of seizures is slow enough so that the electrographic onset (sometimes manifest as auras) is present some time before impairing consciousness, then the same can be achieved through the detection rather than prediction of these symptoms. It was shown in Chapter ?? that this is already feasible for intra-cranial records with detection algorithms available today. Some research groups ([?]) decided that this direction is the most logical, but although preliminary results are promising the question as to whether the seizure can be aborted once it has started has not yet been adequately answered [12].

Thus detection based treatment has not yet made the need for the prediction of seizures redundant. Prediction can also be considered an alternative form of treatment that targets the prevention of lapses into seizure states. For example, there are cases in which epileptic people feel the build-up of an oncoming seizure over hours, days or even months in which they crave for a seizure to occur so that they may feel better. This supports the hypothesis of seizures as reset mechanisms - the solution to a problem rather than the cause[10]. If prediction were possible then this reset could be performed externally, thereby abating the need for a seizure - a preventative rather than reactive measure.

On a more practical note prediction would also provide longer warning time in which action could be taken. This is important seeing as breaching the blood-brain barrier in the delivery of fast acting drugs, as well as the use of electrical stimulation, are both novel methods for which much research is still necessary. If longer horizons are provided then traditional drugs that take minutes rather than seconds to react could be used today with a successful predictor.

The need for prediction may not yet void, but predictors today are far from clinically applicable. The foundation of even the most sophisticated prediction algorithms is based on many assumptions: they assume the existence of a dynamical model - even though this model is not known ; they assume that the measured EEG is appropriate in the representation of this model - this may not be true (see Chapter ??) ; they assume that relatively short amounts of data can be used to infer statistical invariants such as lyapunov exponents and synchronicity - this has been disproved in Chapter ?? . Although it is not impossible that dynamical characteristic can be tracked in this way, it seems unlikely that these

---

2Although it is also possible that this desire for the seizure is a psychological need by the patient to associate some positivity to the situation.
methods are reliable considering the volatility observed in EEG records. Their performance creates further doubt as to what these measures are really detecting in the first place. They work only in very specific cases and have for the most part been invalidated in recent publications when tested on larger and independent datasets. In some cases the methods were shown to perform no better than random (see [12] for details). Furthermore nothing is known about the nuances of the data being tested: What do the EEG records leading up to the seizure look like? What do the inter-seizure events look like? Are they stereotyped? Is it possible that these methods are acting as a form or signature recognition, and are therefore yet another form of detector rather than predictor? Without knowing exact details of the data under test (beyond the typical information such as sampling rate, number of channels, etc) then these questions are unanswerable.

Methods based on signature recognition exploit only phenomenological information - they do not attempt to understand anything about the generation of the seizures. They cannot themselves be used to infer anything generally about the dynamics of epilepsy, and seizures that are not typical are likely missed. It has been argued, and it could be true, that these statistics are indeed measuring a slow change in the EEG that is representative of the dynamics. Patients for whom this does not work can be explained by a recruitment that is too fast to notice - it all makes sense under this paradigm. But just as valid is the argument that what is being detected is a stereotyped pattern leading to a seizure - we have to accept the fact that this could also be true. Until continuous, standardised data sets that include a blind validation phase are available prediction (as opposed to detection) of epilepsy cannot be verified.

Fortunately measures have been put in place making a more unified effort possible. The International Workshop of Seizure Prediction is a regular meeting scheduled every two years promoting collaboration and providing common datasets for testing. The third such meeting was held in Freiberg, Germany, in 2007. The standardisation of datasets as well as the the introduction of validation tests that compare algorithms against random predictors [16, 17, 18] provide a more rigorous platform if not a solution to the problem of seizure prediction.

Successful or otherwise ; predictors or detectors ; dynamical indicators or stochastic fluctuations - current prediction have not addressed the fundamental question posed previously - how predictable are seizures? The work in this chapter is an attempt to answer this, in part, with particular interest placed in the identification of memory in the epileptic brain. If no memory exists and seizures are caused by a sudden, abrupt transition then they are unpredictable and therapy should concentrate on their detection. If, on the other hand, memory exists then arguably seizures can be predicted. How long this memory is and what type or
The methods used to identify the existence of memory involve the detection of scale-invariance in these datasets. This is a phenomenon that indicates that behaviour at different time-scales is similar to each other. Scale-invariance does not itself indicate memory unless it is of a specific type. Because the words scale-invariance, power-laws, long-range memory and many other terms that are all relevant to this study are frequently used, confused and abused in literature, the next section is dedicated to defining and discussing what exactly this terminology means. A description of the datasets is then given in Section 1.3, and the analysis and discussion of results are presented in Section 1.4 and Section 1.5 respectively. As usual, boxed comments and summaries throughout give a conceptual overview of key points.

\section*{1.1 Predictability - Terminology made clear}

In the context of predictability, the notions of self-similarity, scale-invariance, self-organisation, criticality, power-laws and long-range dependency are used in literature liberally and often interchangeably without much explanation of their meaning, their assumptions, and their relevance. For the person unfamiliar with the theory this may become quite confusing, so this section is here to clarify the
Let’s start with the most general concept - self-similarity. An arbitrary entity (e.g. a picture, an object, a time-series) is said to be self-similar if the properties of the entity are the same when it is looked at as a whole or in parts. A common example is that of fractals, the most famous of which is the Mandelbrot set shown in Figure 1.1. (a) shows the Mandelbrot set in its entirety. Enlarging a part of the set, as in (b), results in an image that shows patterns that are very similar to the whole. This continues indefinitely at finer scales. The properties at different scales do not look exactly the same as each other, but similar type of structures exist. Self-similarity exists in the real-world - coastlines, snow-flakes and even the flows and eddies in Leonardo da Vinci’s drawings are considered self-similar [20].

Scale-invariance is an instance of self-similarity. The Mandelbrot set in Figure 1.1 is spatially scale invariant because similar properties are present at all spatial scales of the picture. More relevant to this chapter is temporal scale-invariance, which for a stochastic process implies that the statistical properties at different time scales (e.g. days versus hours versus minutes) effectively remain the same. Scale-invariance has been studied for earthquake frequency, internet traffic and economic data [13], and has also been identified in many biological systems including the timing of ion-channel opening in neurons, auditory nerve fibre action potentials, and human heartbeats [2].

Self-organisation is the ability of a system (of interest here - the brain) to organise itself into a state of increased complexity without the need for external interference - the system can change thanks to properties of the network and its dynamics rather than through the need of external input. Self-organised criticality (SOC) is then the ability of this system to evolve toward a self-similar or scale-invariant state. These concepts are explained very well in [4]. The identification of SOC in brain function is an active area of research because it is believed that understanding why it exists will reveal important mechanisms in brain function. For example in [5] a study of spontaneous activity in in-vitro brain slices revealed that SOC exists in the ‘avalanches’ of activity that occur - a trend that may be important for optimal information transfer and stability in cortical networks. In [15] evidence of self-organised criticality is found in the focus of some types of epilepsies, where the seizure itself is the self-similar state. These observations are important in the development of models, which must be capable of replicating this behaviour. Conversely these observations can help validate models, such as the Integrate-fire ones described in Chapter ?? which have been shown to exhibit SOC-like behaviour [19].

So, regardless of the presence of SOC, systems can be scale-invariant (and therefore self-similar) in both the spatial and temporal domain. How can the scale-
1.1. PREDICTABILITY - TERMINOLOGY MADE CLEAR

Temporal scale-invariance is studied to this end because of its relevance to later sections. The concepts presented here are applicable also to statistics that can be extracted from spatial information.

Implied by temporal scale-invariance is that short/small events occur frequently and long/large events occur infrequently but with non-negligible probability. If an arbitrary function $f(t)$ describes a temporal scale-invariant process, then the log-log plot of $f(t)$ versus time $t$ is a straight line. That is,

$$
\log(f(t)) = -\beta \log(t) + c
$$

or alternatively

$$
f(t) = \exp(c)t^{-\beta} = Ct^{-\beta}
$$

This is known as a power-law with positive scaling exponent $\beta$ and positive $c = \log C$. $\beta$ is the negative of the gradient of the straight line formed in the log-log domain. $f(t)$ displays scale invariance because dilating or contracting time $t$ by a constant $k$ (that is, changing the temporal scale) simply multiplies $f(t)$ by $k^\beta$ - also a constant. The power-law becomes

$$
\log(f(kt)) = -\beta \log(kt) + c = -\beta \log(t) - \beta \log(k) + c
$$

so the dilation results in an additive factor of $-\beta \log k$ in the log-log domain - the gradient/exponent of the power-law remains the same.

The word “power-law” thus describes a particular relationship that may be present in an arbitrary function. By itself it does not mean much - events with random distributions that obey a power-law are easily generated. However under certain conditions if power-laws are observed in the statistics of a system then information about the underlying process, such as its predictability, may be inferred.

\footnote{Here the negative sign in front of $\beta$ is there to demonstrate a decay or negative gradient in the relationship. This is convenient for the analysis in this chapter but not necessary in the general sense of a power-law.}
Scale-invariance is an example of self-similarity, in which the statistical properties at large scales are the same as those at small scales. Scale-invariance may be temporal or spatial. Self-organised criticality is the ability of a system to, without external input, make its dynamics scale-invariant when a critical stage is reached.

If the log-log plot of a function \( f(t) \) versus time \( t \) is a straight line with gradient \( \beta \), then the process is said to obey a power-law. Power-laws are scale-invariant processes.

Let’s stray a little now and introduce the concept of short-range versus long-range dependence. The autocorrelation of \( f(t) \) (described by Equation ?? in Chapter ??) describes how dependent events at different times are on each other. If the auto-correlation decreases very fast, then current events can say very little about future events. This is known as short-range dependence (SRD). The most trivial example of an SRD process is an independently drawn random variable, where the auto-correlation for all values other than \( t = 0 \) should be negligible. This is the shortest type of dependence - i.e. none at all.

However, if the auto-correlation decreases more slowly then there is some information about the future in the present event. If the decay is particularly slow then the dependence spans far into the future - this is known as long-range dependence (LRD). Processes that exhibit LRD are thus said to have long memory. Formally, the area under the auto-correlation of an SRD process is finite, and the decay near zero is very fast, whereas the area under the auto-correlation of an LRD process is infinite.

Determining whether a process is SRD or LRD can be done by examining the power-laws that exist in the statistics of collected data, so long as it is stationary. The amount of dependence in a process can be characterised by an exponent \( \alpha \), with values \( 0 \leq \alpha < 1 \). This exponent is calculated from the gradient \( \beta \) extracted from a power-law relationship\(^4\). For the trivial (random) SRD process, \( \alpha = 0 \). The closer \( \alpha \) gets to 1 the longer the dependence and therefore the longer the memory in the process. Higher values of \( \alpha \) represent smoother trends and less volatility.

The values of \( \alpha \) have been restricted to the range \( 0 \leq \alpha < 1 \) so that they only test for LRD. Furthermore, this restriction implies that if a process is scale-invariant then it is also LRD, and vice versa [13]. This is not true in the more general sense. Values outside this range from observed data are indicative of non-stationarity, noise, or the existence of more complex dynamics not describable by LRD.

\(^4\)The relationship between \( \alpha \) and \( \beta \) depends on the method of extracting the power law. More on this in Section 1.2
The exponent $\alpha$ is often re-expressed in terms of the *hurst exponent* $H$ - the “index of dependence” which can in turn be directly related to fractal dimensions such as those described in Chapter ??.

So as not to confuse matters, the relationship between $H$ and $\alpha$ is assumed to be\footnote{Some sources define this as $\alpha = 2H + 1$. The alternate definition is used in this text to restrict LRD values to $0 \leq \alpha < 1$} $\alpha = 2H - 1$, so that $H = 0.5$ indicates a random distribution (i.e. no memory), and values of $0.5 < H < 1$ are indicative of LRD. Since $\alpha$ and $H$ can in any case always be expressed in terms of each other only $\alpha$ is used from now on.

Is LRD a desirable property? There are two ways in which the existence of LRD may be interpreted: matter:

1. LRD is bad news - the presence of long memory creates many problems in analysis, especially for short data sequences. The structure of the underlying system is complex, and difficult to analyse because of the power-law nature of the probability distributions of such processes - these requires many data points to calculate accurate statistics. The observation time must be long enough so that at least some of the infrequent events occur.

2. LRD is good news because at least there is some structure in the system. Consider measurements that look random, but are in fact LRD. For a completely random system the acquisition of more data does not make the system more predictable because no information is gained about the future from one more observation. However, because memory exists in an LRD process the longer the period of observation the more predictable that it becomes \[13\]. In theory it is possible to reduce the prediction error to an arbitrarily low level by increasing the observation time accordingly. In practice, of course, this is often not feasible.

This information can be used to determine the predictability of seizures. Data which consists of only the times at which epileptic events occur is collected from different patients spanning different amounts of time. These events look random, but may in fact be correlated - determining whether this system has memory, that is, identifying the presence of long-range dependence could mean that these events are predictable. If no memory exists then this type of data cannot be used to predict, although other forms of data may still be usable.

Robust methods for estimating $\alpha$ are described next in Section 1.2.
In summary, if a power-law exists in the statistics of a function $f(t)$ then it displays scale-invariance. The gradient of the power law, $\beta$, can then be used to determine if there is memory in the system or not by computing the scaling exponent $\alpha$. The relationship between $\alpha$ and $\beta$ is dependent on the method of computing the power law.

If $\alpha = 0$, then there is no memory in the system. If $0 < \alpha < 1$, then long-range dependency or memory exists in the system. Memory or LRD can mean that a system becomes more predictable with longer observation times. This information could be important in discerning how predictable epileptic seizures are.

### 1.2 How to Estimate LRD

The clinical data used for analysis in later sections (described in detail in Section 1.3) is in the form of discrete time points at which epileptic events occur. The strength or duration of the events is not known. This is called a point process, which can be equally expressed as the discrete times at which events occur or as the inter-event times. The data can be converted to resemble continuous time by drawing a time-line in which a value of 0 is assigned at times when no event occurs, and 1 at the time points at which events occur. The inter-event times of these datasets look random in nature.

This section illustrates the process of robust estimation of long range dependence and/or scale-invariance through the use of well-known generators of both LRD and random (trivial SRD) processes. These methods can then be applied to the clinical data. To generate point processes samples are drawn from well known probability distributions, with or without memory.

For a random process, the most widely used random generator is the Gaussian or Normal distribution, with probability density function (PDF)

$$P_{GAUSS}(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{(x - \mu)^2}{2\sigma^2} \right)$$  \hspace{1cm} (1.4)

where $P_{GAUSS}(x)$ is the probability that event with value $x$ has of occurring. For a sufficiently large number of events the sample mean approaches $\mu$, and the variance approaches $\sigma^2$. Point processes such as the ones in the clinical datasets can only take positive values. Because negative values of $x$ in the above distribution are possible suitable clipping or rejection of negative events is applied.
1.2. HOW TO ESTIMATE LRD

to the time series. Results are not affected by these transformations. These positive values are then representative of the inter-event times. An example time series drawn from $P_{GAUSS}(x)$ with mean $\mu = 3.5$ and variance $\sigma^2 = 1$ (both in arbitrary units) is shown in Figure 1.2(a). Provided that the samples are drawn independently then there is no correlation and thus no memory in this time series - this is the example used for the trivial SRD process with expected $\alpha = 0$.

A stochastic LRD process is more difficult to simulate. A well accepted example is that Fractional Gaussian Noise (fGN), derived from Fractional Brownian Motion (fBM) (also known as a random walk process). fBM is a random discrete time series $B_\alpha[n]$ with $n = 1, 2, \ldots, N$. It is constructed as

$$B_\alpha[n] = \sum_{i=0}^{n} X_i, \quad n = 1, 2, 3, \ldots, N$$

where $X_i$ are randomly generated samples. Brownian motion has the property that

$$B_\alpha[N] \sim N^{\alpha+1} X_i$$

where $\sim$ denotes an equality in distribution, that is, equality in PDF including mean and variance. Therefore the properties at different scales (i.e. different $N$) scale according to a scaling exponent $\alpha$. If $X_i$ are independently drawn variables from Gaussian distribution described by Equation 1.4 then fGN is defined as

$$F_\alpha[n] = B[n] - B[n - 1], \quad n = 2, 3, 4, \ldots, N$$

The distribution of $F_\alpha$ follows a that of a Gaussian, with variance proportional to the delay between samples (in this case 1) to the power of $\alpha + 1$. When $\alpha = 0$ then the elements in $F_{\alpha=0}$ are independent, and the random walk $B_{\alpha=0}$ is truly random. With $0 < \alpha < 1$ the system experiences memory in that if an increment in a particular direction occurs, then it is likely that the motion in $B_\alpha$ continues in this direction. The larger the $\alpha$, the greater memory this process exhibits.

Better probability distributions that generate discrete and positive point processes exist (e.g. the Poisson distribution), but no analogous distribution exists for an LRD process. This makes comparisons between findings more tedious. In any case, all observations made about the random case in this section are the same regardless of whether samples are drawn from a Gaussian or a Poisson distribution.
fGN looks random, even though slightly different than the Gaussian process in (a). fBM, on the other hand, displays more structure. In (d) the PDF of each of the 3 cases is shown to follow the expected distributions. Notice the much greater variability in the case of fBM, although at face value the distributions of the Gaussian and fGN cases look very similar. Scale-invariance is not evident at face value and further processing is necessary.

The remainder of this section shows how $\alpha$ can be estimated from sequences obtained using these distributions. The random Gaussian distribution or trivial SRD case is referred to from now on to have probability density function $P_{\text{SRD}}(x)$. The fGN is referred to have PDF $P_{\text{LRD}}(x)$ with $\alpha = 0.8$.

The first step in determining LRD is to check whether scale-invariance is likely, and the easiest way to do this is to derive inter-event probability histogram (IPH), which gives an idea of the PDF of the process by calculating the probability that an inter-event time occurs [2]. For data from an unknown distribution this requires knowledge about the maximum resolution. In the case of the processes drawn from $P_{\text{SRD}}(x)$ and $P_{\text{LRD}}(x)$ the maximum resolution is known.

The distribution is verified for examples of different lengths in Figure 1.3(a) and (b), where the latter is now the PDF of the fBM and not the fGN. The histograms are presented in a log-log plot to demonstrate that the probability of large events for $P_{\text{SRD}}(x)$ is very low - a very fast (exponential) decay occurs at large events. (b) in contrast shows that large events occur at non-trivial probability for $P_{\text{LRD}}(x)$. This is regardless of the number of samples, $N$, used. An exponential curve is also shown in these plots for comparison. This slow decay is sometimes referred to as a heavy tail (although more formal definitions exist), which have a very wide range of possible values with long/large events occurring at non-negligible probabilities [13].

From the purely mathematical point of view there is no reason why a heavy tailed distribution has to be LRD, but in processes derived from dynamical systems heavy tails generally lend themselves to the existence of long memory [13]. Although a heavy tail does not necessarily indicate the existence of LRD, it cannot exist without it. This first step is a simple way to determine whether it is worth continuing with analysis - for example, the process generated by $P_{\text{SRD}}(x)$ can be rejected, whereas the process generated by $P_{\text{LRD}}(x)$ requires further analysis and validation.

Traditional methods of estimating the fractal exponent $\alpha$ include the identification of power-laws, constructed in a variety of different ways, from the same data set. These methods are described in great detail in references such as [20], [7] and [13], and typically involve higher order statistics (e.g. variance) because first order statistics (e.g. mean) do not reveal the differences between random and LRD processes [13]. Variance is the most often used second order statistic.
1.2. HOW TO ESTIMATE LRD

![Simulated Random (Gaussian) Inter-event Times](image1)
(a)

![Simulated Fractional Gaussian Noise (fGN) Inter-event Times](image2)
(b)

![Simulated Fractional Brownian Motion (fBM)](image3)
(c)

![Simulated Gaussian PDF](image4)
(d)

Figure 1.2: Simulated time series for (a) Random (gaussian) process, (b) Fractional Gaussian Noise (fGN) with $\alpha = 0.8$, and (c) the consequent Fractional Brownian Motion (fBM). Upon visual inspection both the (a) and (b) look predominantly random even though some structure exists in fGN. Similar probability density functions (PDFs) are apparent for both these processes in (d). However the consequent fBM shows a lot more structure and its PDF is volatile and different from that of a random gaussian process. This is due to the introduction of the memory parameter $\alpha$.

Other typical methods include:
Figure 1.3: Inter-event probability histograms (IPH) drawn on log-log plots for both the random ($\alpha = 0$) and the LRD ($\alpha = 0.8$) simulated time series. (a) shows that the IPH of random events falls exponentially so that the probability of large events is negligible. (b) shows that the IPH for an LRD process has large events that occur at non-trivial probability. This is known as a heavy tail - a property that is necessary for memory to exist.

- **Auto-correlation**: (defined in Section ?? in Chapter ??) The auto-correlation not only gives an idea of how far in time there is non-trivial coupling, but can also be used directly to estimate $\alpha$. In the log-log plot of absolute auto-correlation magnitude versus delay if a power-law is formed then its gradient $\beta = 1 - \alpha$ [2].

- **Power Spectral Density (PSD)**: Section ?? in Chapter ?? Since the auto-correlation can be related to the PSD, it follows that $\alpha$ can also be estimated by analysing the frequency content of the signal. Long memory relates to low frequencies (and slow time scales). The power-law generated at low frequencies in the PSD scales as $\beta = \alpha$.

- **Fano Factor**: This is an estimation of the variance in the number of events observed in a time period $T$. Each $T$ is representative of a different time scale. The gradient of the fano factor scales as $\beta = \alpha + 1$ for increasing $T$. In the random case, the fano factor does not depend on the duration of the observation, resulting in an $\beta = 1$ ($\alpha = 0$) as expected. Any variation from $\beta = 1$ indicates a complexity or richness of information not present in a process with no memory [2].

The above examples are all simple ways of estimating the scaling exponent $\alpha$, related to $\beta$ in different ways depending on the way that the power-law is drawn\(^7\). However although these methods work well under optimal conditions they are

\(^7\)Note that the estimated value of $\alpha$ may not be the same as the true scaling exponent. It is the aim of this work to find out to what confidence the value of $\alpha$ may be representative of the true value.
1.2. HOW TO ESTIMATE LRD

Each wrought with their own idiosyncrasies for data series that are short, noisy, and not stationary - results are difficult to interpret.

These days wavelet theory (described in detail in Chapter ??) are used as an alternative. Wavelets work well because they are themselves scale-invariant processes - a property that is not true for other analysis tools [13]. They have proved effective in detecting the existence of many different types of long memory. Under optimal conditions they are not better at estimating $\alpha$ than previously mentioned methods [2], their major advantage being their ability to do so under non-stationarity, provided that the changes are smooth enough [2, 7, 13].

Wavelets have been described in detail in Chapter ?? . They isolate activity at different frequencies and time-scales, and as such, they are intuitively equivalent to the estimation of $\alpha$ through PSD based methods - larger scales correspond to lower frequencies. If an orthogonal mother wavelet is used and only dyadic sampling is allowed (that is, let dilation factor $a = 2^k$ for $k = 1, 2, 3...$) then the wavelet coefficients between scales are only weakly correlated. Thus the analysis at each scale is largely de-coupled from all other scales [13].

The variance $s_k$ of the wavelet coefficients $d_k[n]$ at scale $k$ is defined as

$$s_k = \frac{1}{L_k} \sum_{l=1}^{L_k} |d[k, l]|^2$$

(1.8)

where there are $L_k$ coefficients at each scale $k$. This is a second order statistic that can then be used to estimate $\alpha$. Because each value of $k$ represents an exponential increase in bandwidth as well as centre frequency, plotting $\log_2(s_k)$ versus $k$ is the equivalent of a log-log plot (also known as a scalogram) used to identify the power-law. In reality a small bias $g_k$ needs to be introduced because of the non-linear nature of the logarithm, so that the plot versus $k$ becomes

$$y_k = \log_2(s_k) - g_k$$

(1.9)

A good approximation of $g_k$ when $L_k$ is large enough and under the simplifying assumption that coefficients at different $k$ are completely de-correlated is given by [7]

---

8In fact, if the DB-2 (also known as the Haar wavelet) is used then the analysis can be equivalent to the Fano Factor calculation because they both measure variance at different time scales in similar ways. The difference is that the Fano Factor does not handle the correlation present between scales very well, whereas the wavelet methods do, provided that sampling is dyadic [13].
The detection of scaling using the plot of $y_k$ versus $k$ then results from the identification of a region of alignment in this plot. Not all scales need to be involved in the long memory process, but if a sufficient number of scales line up then a power-law, which may then be tested for LRD, is evident. In practice the region of alignment must consist of at least 4 scales. This number is arbitrary, seeing as memory could exist in as little as 2 scales, but it is entirely too easy for up to 3 scales to line up randomly. A minimum of 4 scales is necessary before results are taken seriously [13]. The gradient at the region of alignment can be equated to the scaling exponent, that is, $\beta = \alpha$.

In reality the points are unlikely to line up perfectly. However the quantity $y_k$ is only accurate to a certain degree of confidence dependent on the number of available coefficients. Since $L_k$ halves at each iteration $k$ then the errors expected at each scale increase monotonically with larger scales. An estimate of the expected variance $\sigma^2_k$ in $y_k$ can also be used to estimate error bars, an approximation of which is given by

$$\sigma^2_k \approx \frac{2}{L_k \ln^2 2}$$

Error bars can then be drawn on $y_k$ as multiples of $\sigma_k$ for different confidence levels. Notice that more data not only increases the confidence in the error bars (i.e. decreases $\sigma_k$), but also makes more scales available for analysis - results are more conclusive. A small dataset can lead to problems that could imply that (a) the observation time may not be sufficiently long for LRD to be detectable, or (b) error bars become large enough so that LRD cannot be distinguished from random - that is, the method is incapable of rejecting either possibility. Longer time series do not remove but do reduce the chance of these limitations becoming significant.

To apply the above theory on our distributions $P_{t\text{SRD}}(x)$ and $P_{t\text{LRD}}(x)$, the following steps are undertaken:

1. **Generate inter-event sequences**: This is done using Equations 1.4 and 1.7 respectively for a random (trivial SRD) and LRD process. This includes ensuring that all values are positive. Three sequences of length $N = 500$, $N = 1000$, and $N = 10000$ were generated for each case.

2. **Create time-series**: A continuous-time process is approximated by creating a time-series in which a zero indicates no event, and a 1 indicates an event
1.2. HOW TO ESTIMATE LRD

9. In this case, values were arbitrarily assigned a resolution of the minimum inter-event time observed in generated sequences. Using a finer resolution does not affect the estimates (see later analysis).

3. **Use wavelet tools:** Calculate $y_k$ and $\sigma_k$ for as many scales as are available\textsuperscript{10}. Although the number of scales are dependent on the resolution of the time series created in step 2, the number of scales available for $N = 500$ is roughly 1 less than for $N = 1000$ and 4-5 less than for $N = 10000$.

4. **Calculate $\alpha$:** Select the interval over which, within the confidence limits, points align in the scalogram of $y_k$ versus $k$. Compute $\alpha$ using a line of best fit. A $\chi^2$ goodness of fit test is computed to determine how well the data fits this line - visual judgement is not sufficient. A statistic $Q$ is provided representative of this test, with a value greater than 0.05 deemed an acceptable fit. For the readers not familiar with $\chi^2$ goodness of fit tests, this is explained in Appendix ??.

5. **Distinguish LRD:** Determine if the value of $\alpha$ falls within values indicative of LRD. If not, reject as noise.

Figure 1.4(a) and (b) show estimates for the trivial SRD and LRD process respectively. The expected $\alpha = 0$ for SRD and $\alpha = 0.8$ for LRD are also plotted for comparison. Notice that in both cases, even for relatively low number of events the limiting value of the gradient of $y_k$ versus $k$ approaches these expected values. Note also that the error bars increase with increasing number of scales. All calculations were performed using a DB-3 wavelet.

How robust are the estimations of LRD? Several factors could affect our calculations: choice of resolution in step 2, noise in the data and choice of wavelet type and order in step 3. Let’s see how each of these affect the estimates.

First, in Figure 1.5 step 3 is applied to the LRD process for different resolutions, ranging from low (smaller than the smallest inter-event time) to high (larger than the smallest inter-event time) in arbitrary units. The coarser the resolution (higher value) the fewer scales that are available for analysis. The values of $y_k$ themselves are affected by the choice of resolution, but it is clear that the

\textsuperscript{9}This step is performed because inter-event times are inherently discrete-time processes, whereas the theory presented is relevant to continuous time only. Note that a way to filter the discrete-time process so that methods can be applied directly to this data exist. Details on this can be found in [7] and [13], but results in this chapter were not significantly affected by the choice of method.

\textsuperscript{10}The base code used in this analysis is freely available for download on http://www.cubinlab.ee.unimelb.edu.au/~darryl/secondorder_code.html for non-commercial purposes. The authors, Patrice Abry and Darryl Veitch, retain copyright of this code.
CHAPTER 1. ON THE PREDICTABILITY OF SEIZURES

Figure 1.4: Wavelet estimation tools are shown to reliably estimate \( \alpha \) at large scales for both the random and the LRD processes, even for simulations involving relatively low number of events. The figures show the power-law relationships formed between \( y_k \) and \( k \). The expected value of \( \alpha = 0 \) for the random case and \( \alpha = 0.8 \) in the LRD case is included in each plot for comparison.

Figure 1.5: Changing the resolution at which \( \alpha \) is estimated does not affect the gradient at large scales, even though the raw values of \( y_k \) do change. Thus although it is not valid to use a resolution higher than that used in the recording process, so long as the selected resolution is reasonable analysis can continue with little thought of its impact on estimates. This is in stark contrast to traditional estimators of \( \alpha \) that are very volatile under such design choices.

Another type of noise that is likely present in the clinical data is that the recording process is prone to human error and events may be missing. A suitable question to ask is then how is the analysis affected by removal of events. To answer this, step 3 was computed for both random and LRD processes with events randomly removed at different rates [9]. The results are in Figure 1.6, and they show that even for large removal rate of 0.7 the trends at the larger scales remain relatively
1.2. HOW TO ESTIMATE LRD

Figure 1.6: The estimation of $\alpha$ in an LRD process is robust when events are missing from datasets, even at large probabilities. However, the targeted removal of large events destroys this structure because LRD is dependent on their inclusion. This is demonstrated for both the random and the LRD process above. Neither the random nor the selective removal of long events changes the estimation of $\alpha = 0$ for the random process, as seen in (a). On the other hand the structure at large scales is shown to suffer for the LRD process in (b) when long events are not present.

unaffected. Of course such large removal rates result in fewer scales available for analysis and there is greater uncertainty in the results, but the trends are still observable.

If we are selective in the types of events that are remove, on the other hand, by targeting only the long events, the values of $y_k$ at larger scales $k$ are affected. This is also shown in Figure 1.6 for both LRD and random processes, and it occurs because the properties of scale-invariance for larger $k$ are dependent on the longer events even though they occur at lower probabilities. Fortunately the noise in the clinical data is not expected to be selective toward long events only.

Non-stationarities in the data may also affect the estimation of $\alpha$. If there are reasons to believe that the data is stationary, as is the case for our simulated sequences, then the estimations of $\alpha$ should be taken seriously [13]. However, in a physical process such as the frequency of epileptic events the generated data may be affected by internal or external mechanisms that change the mean frequency or the variance of events. These changes could come about from changes in drug dosages, the worsening or bettering of the epilepsy, or in the short-term normal
rhythms such as day and night. Traditional methods of estimation of LRD could not be used to analyse these cases because they cannot cope with changes in stationarity, but wavelet tools can - provided these changes are smooth enough. This is shown in Figure 1.7. In (a) the mean of the events is changed (smoothly) by modulating the original time series by a polynomial trend. It is seen that although the values of $y_k$ are affected (as compared to the original data), the $\alpha$ estimated from the gradient at large $k$ is not. Changes in variance of events are abruptly introduced in the second half of the sequence in (b), and again the estimate of $\alpha$ is unchanged.

These results are not unexpected, as the theory supports the observations. The order $O$ of the Daubechies wavelet is capable of removing polynomial trends of order up to $O - 1$ [2, 13]. For example the DB-3 ($O = 3$) wavelet, which has been used in all analysis so far, is capable of removing quadratic trends. Increasing the order increases the capabilities of coping with less stationarity. The choice of wavelet family in this type of analysis is therefore not as important as the choice of wavelet order, so long as the wavelets used are orthogonal. Since wavelets are naturally equipped with the ability to change order quickly and easily, once again they are shown superior to traditional estimation methods.

In practice, the selection of the wavelet order should be dependent on the data, because the smoothness of the non-stationarities are unknown. The order of the wavelet is systematically increased until stable results are observed, and the lowest order wavelet should be used at that point. Low order wavelets are desired because less data that is necessary - the tradeoff is between the number of scales available for analysis versus higher order wavelets which can cope with more complicated forms of non-stationarity. Figure 1.7(a) shows that the introduced trend is smooth enough so that selection of higher order wavelets does not change results.

Finally, because for real data it is not known whether the non-stationarity is smooth enough changes over time should be checked for less obvious volatility. An accepted method is to break up the data into adjacent segments to determine if the values of $\alpha$ are consistent over time. This should be done for any estimation method used, not just wavelet based tools. This step is not shown here for the simulated sequences as it is known that these are stationary.
1.2. HOW TO ESTIMATE LRD

Figure 1.7: Wavelet based tools for the computation of $\alpha$ are robust under smooth changes in stationarity. In (a) the mean of an LRD process with $\alpha = 0.8$ is modulated with a smooth polynomial. Even though the actual values of $y_k$ change when compared to the original, the estimated gradient $\alpha$ at large scales does not, even for relatively low wavelet orders. Wavelets are also shown to be robust under changes in the variance of data, as shown in (b) where a sharp transition is observed half way through the sequence. The ability of wavelets to cope with such changes in stationarity make them more suitable for data such as the one described in Section 1.3 because it is expected to contain such transitions.

Wavelet tools described in Section ?? in Chapter ?? have been shown effective in the estimation of $\alpha$ by using the variance of the wavelet coefficients at each scale $k$. These methods are more robust than traditional ones because they cope better with:

1. **Resolution/Noise:** The choice of resolution, and thus the allowed tolerance for noise, does not affect the estimation of $\alpha$, only the number of scales available for analysis.

2. **Missing Data:** Even with a relatively large number of missing events (which may have gone unrecorded) estimates of $\alpha$ using wavelet methods are not affected.

3. **Non-Stationarity:** A wavelet of order $O$ is not affected by non-stationarity of polynomial trends of order less than or equal to $O - 1$. The ability for wavelets to cope with non-stationary data as well as the flexibility in the parameter choice $O$ that allows different levels of non-stationarity makes wavelets much more powerful than traditional methods.
Now that practical aspects of the detection of the presence of long memory have been discussed they can be applied to real data. This is done in Section 1.4 and discussed in Section 1.5, but first a detailed description of the clinical data is provided in the next section.

1.3 Seizure Frequency Dataset

A summary of the epilepsy data available for analysis is given in Table 1.1. Each of the 6 datasets belong to different subjects, and comprise of the times at which epileptic events occur. From this inter-event times can be extracted, and time-series created with a maximum resolution shown in Table 1.1. This refers to the best resolution that the events were reliably recorded at by the patient or at which the events were extracted from an EEG time-series. The resolution at which analysis was performed may differ from this value - specifics are given in Section 1.4, but should never be better than the number given here. The time series are plotted in Figure 1.8.

There are two types of data. Datasets 1-4 are records of epileptic events maintained by the patient him/herself over a period of 1-25 years. These records are prone to much noise - faulty memory may lead to erroneous event times; lapses in discipline may mean some events are missing; only clinical epileptic events are present since sub-clinical seizures are not detectable by the patient; drug dosages usually change in the span of many years, perhaps affecting the stationarity of event frequency. Nevertheless this is the best data of this nature that we are ever likely to have, seeing as EEG monitoring at these time-scales is not feasible. Analysis of such data may be helped by using coarser resolution than that shown in Table 1.1 and removing time periods in which lapses in the record-keeping occur (these usually have been annotated by the patient). Furthermore, one can make the perhaps unjustified assumption that drug dosages change the mean frequency or variance of events but not the intrinsic relationship between events. The use of wavelet analysis that can remove the effect of such changes if they are smooth enough is then justified. (Note that a few more datasets than the ones presented here exist, but have been removed because they do not add any further information. Similar conclusions can be extracted from these).

Datasets 5-6 are shorter-term sequences extracted from EEG recordings. Dataset 5 belongs to a patient who experienced an unusually large number of epileptiform discharges in the EEG. These events are not necessarily clinical. That there are many events makes this 4-day EEG monitoring suitable for analysis of this type. Finally, Dataset 6 belongs to a 3 hour EEG record taken from a long-evens rat recorded in the Shanghai Institute of Brain Functional Genomics (East China Normal University, Shanghai, China). During a surgical procedure in
### 1.3. SEIZURE FREQUENCY DATASET

#### Table 1.1: Seizure frequency data

<table>
<thead>
<tr>
<th>Dataset No.</th>
<th>Data Source</th>
<th>Study Duration</th>
<th>No. of Sz.</th>
<th>Resolution</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Human</td>
<td>25 years (1982-2006)</td>
<td>911</td>
<td>1 day</td>
<td>Many drug dosage changes</td>
</tr>
<tr>
<td>2</td>
<td>Human</td>
<td>10 years (1997-2006)</td>
<td>1050</td>
<td>1 day</td>
<td>Seizures occur in clusters (665 clusters)</td>
</tr>
<tr>
<td>3</td>
<td>Human</td>
<td>5 years (2001-2006)</td>
<td>412</td>
<td>0.5 days</td>
<td>Drug changes unavailable</td>
</tr>
<tr>
<td>4</td>
<td>Human</td>
<td>1 year (2005-2006)</td>
<td>397</td>
<td>15 mins</td>
<td>Drug changes unavailable</td>
</tr>
<tr>
<td>5</td>
<td>Human (EEG)</td>
<td>4 days</td>
<td>2722</td>
<td>1 secs</td>
<td>Subject experienced an unusually large number of epileptiform discharges</td>
</tr>
<tr>
<td>6</td>
<td>Rat (EEG)</td>
<td>3 hours</td>
<td>1429</td>
<td>0.5 secs</td>
<td>Spontaneous epilepsy</td>
</tr>
</tbody>
</table>

preparation for a different experiment the rat developed frequent spontaneous epilepsy. The records for the rat were taken in 1 hour periods. No drug changes occurred throughout the entire recording time. Furthermore, in Datasets 5-6 it was possible to extract the duration of the events as well as their times, so that the strength of the events may be used as well.

The short duration of both these studies, as well as the extraction of events directly from EEG records, makes their integrity greater than Datasets 1-4 because events are known to at most be very rarely missed. The stationarity in the epileptic discharges is also more likely. Non-stationarity may be experienced in Dataset 5 because of the day-night changes, although this is shown not to affect results in the next section.

One could argue that a typical epileptic patient is unlikely to experience this number of epileptiform discharges, or that the analysis on rat EEG data may not translate to human epilepsy. However the purpose of this chapter is to detect the presence of memory in the epilepsy, thus although the analysis may be relevant only to specific cases it could have greater implications in our understanding of epilepsy as a whole.

In any case, all datasets are relatively short in the number of observed events. The computation of memory is therefore limited by fairly large error bars, and care in the interpretation of these results is necessary. The analysis of this data, and their relevance to predictability, is discussed next.
1.4 Analysis - Estimation of $\alpha$

Visual inspection of the data in Figure 1.8 does not reveal any obvious pattern in the epileptic inter-event times. The data looks random - although the distributions are not clear and not necessarily the same for all datasets. The data also looks non-stationary - obvious changes in mean can be seen in Datasets 1, 3 and 5, whilst obvious changes in variance can be identified in datasets 2, 4 and 5. These changes may be due to mechanisms of the generators of epileptic activity in each case, external factors such as drug dosage changes, or in the case of dataset 5 may simply be the difference between day and night time. In any case the presence of these changes promotes the use of wavelet analysis tools.
Let us first see if long memory may exist in the data by drawing inter-event probability histograms (IPH) for each dataset. These can be seen in Figure 1.9, in log-log plots, along with comparative exponential (fast) decay. In all cases it is clear that the probability of large events decays slower than exponential, and in all cases it is also possible to identify a straight line that can be fitted to the IPH. Thus a power-law exists, and this heavy tail indicates that long memory may be present. Further analysis is validated for all datasets. Note that in theory it is possible to use this power-law directly to estimate the scaling exponent $\alpha$, but like other traditional estimation methods the results can vary greatly with choice of histogram intervals, and much more care needs to be taken in the interpretation of results.

Before estimating $\alpha$, the analysis is quickly performed for different order $O$ of the Daubechies wavelets. This is shown in Figure 1.10 for $O = 3, 4$ and $5$. In all cases the results are stable under the choice of $O$, thus the smallest order ($O = 3$) is chosen and the DB-3 wavelet is used throughout the remainder of analysis.

Figure 1.11 shows the estimated $\alpha$ for all datasets. Each graph shows $y_k$ versus $k$ and the line of regression used to estimate $\alpha$, as well as the range of $k$ for which it was calculated. Also shown are the resolutions used, the estimated $\alpha$ including error bounds, and the goodness of fit parameter $Q$ (greater than 0.05 in all cases, indicating a reasonable fit). Recall that a value of $0 < \alpha < 1$ at large scales indicates LRD, whilst a value of $\alpha = 0$ indicates no memory. The problem is sometimes in selecting the correct region of alignment for which $\alpha$ should be estimated. Consider Datasets 4 - two regions of alignment are shown, one in which $\alpha \approx 0$ and one in which $\alpha = 0.65$. Should the larger range be chosen, because the result is more likely to be correct? What if the LRD only presents itself at larger scales, and it is only the un-availability of data which means that larger scales are not there? In both cases the previously mentioned minimum of 4 scales are used, thus both are valid observations. However when one starts examining the error bars it becomes clear that the $\alpha = 0.65$, although suggestive of long memory, has error bars ($\pm 0.62$) that span the entire range of $0 < \alpha < 1$. This not only means that if there is long memory its effect cannot be estimated, but also that randomness cannot be rejected because a result very close to $\alpha = 0$ is possible. The error bars are important in determining the confidence to which results can be interpreted.

Similar observations can be made about Dataset 2, in which several regions of alignment are possible. Initially it is tempting to say that there is evidence for scale-invariance between scales $4 \leq k \leq 7$. However this dataset is known to have seizures that occur in clusters, and the time scales at which these clusters occur correspond to these scales. It could be that within the clusters memory exists, and this is indicated by this region of alignment - once a seizure begins this data could be used to determine how likely it is that more seizures will occur shortly.
Figure 1.9: Inter-event probability histograms (IPH) for each of the 6 datasets in a log-log plot. The probability of large inter-event times in all cases is non-negligible. Moreover at least visually it seems that all these distributions obey a power-law because the data follows a linear trend. Because this power-law exists, further analysis in search for the existence of memory is justified. Exponential decay is plotted in each graph to emphasise this linear trend.

However this cannot be used to infer long-range memory beyond these clusters because an $\alpha \approx 0$ is estimated for larger scales. It is important to use knowledge of the nature of the data or the data collection techniques when interpreting results. Although it seems that a region of alignment with non-zero gradient may develop at larger scales, this is only observed for 3 values of $k$ - an insufficient number to be taken seriously. More data would be necessary to infer memory at longer scales.

The remainder of the datasets seem a little more straight forward. An $\alpha = 0$ seems a logical conclusion for Datasets 2, 3 and 4, although in all cases neither
1.4. ANALYSIS - ESTIMATION OF $\alpha$

![Figure 1.10: Calculations of $y_k$ using Daubechies wavelets DB-3, DB-4, and DB-5 on Datasets 1-6. The above graphs validate the use of the DB-3 wavelet for further analysis because the results are stable under changes in the wavelet order. Using wavelet orders higher than necessary would reduce number of scales available for analysis - this is not desirable when the datasets are already so short.](image)

the existence or lack of memory can be rejected seeing as the error bars resulting from insufficiently long data series span both the $\alpha = 0$ and significant proportions of the $0 < \alpha < 1$ range. It does tell us that further analysis is fruitless, and these data should not be used.

Datasets 1, 5 and 6 are a different story because the estimated $\alpha$ in all of them indicates LRD with errors that exclude the possibility of $\alpha = 0$. It would be tempting to suggest that long memory exists in all of these, but first there is need to evaluate how robust these results are. A stationarity test is performed for each of these 3 datasets in Figure 1.12 by breaking down the time series into 3 consecutive segments and comparing the results of each segment to the originally
CHAPTER 1. ON THE PREDICTABILITY OF SEIZURES

Figure 1.11: Calculations of $\alpha$ for Datasets 1-6. A clear trend is identified for Datasets 1, 3, 5 and 6, but the selection of an appropriate range to calculate $\alpha$ from is less clear in Datasets 2 and 4, where different ranges yield different gradients. In Datasets 2-4 the more likely conclusion is that $\alpha = 0$ - that is, no memory exists. However the small number of data points used in this analysis means that error bars are too large to conclusively reject the existence of LRD - further analysis using this data is futile. Datasets 1, 5 and 6 on the other hand seem to indicate that memory exists because within the confidence intervals $0 < \alpha < 1$. Further validation tests are necessary before conclusions can be drawn.
computed $\alpha$. This is particularly important for Dataset 1 which is known to be of a 25 year duration - again by using knowledge of where the data comes from we can see that a stationarity test is important because the data is likely to be affected by countless factors occurring over such a very long time. Figure 1.12(a) reveals that the scepticism is justified - there are large fluctuations occurring between each of the 3 tests. Even though the error bars are quite large and in most cases they overlap each other, this overlap is very small. Thus Dataset 1 is not very robust in informing us whether LRD is present - pursuing analysis on this data is not likely to validate or in-validate the existence of LRD.

This is not so for Datasets 5 and 6. The stationarity test shows that similar results are observable in all segments, with error bars spanning largely the same space in all cases. This suggests that the assumption of stationarity is valid and results can be taken seriously. Further robustness tests are seen in Figure 1.13, where events were removed at random at different removal rates. No change is observed in the estimation of $\alpha$. This test is performed because even though both these datasets are less likely to contain missing events than Datasets 1-4, EEG records were marked manually and errors are possible. This gives further confidence in the existence of long memory. For completeness, a comparison is made in this same figure to the case in which only long events are removed. Visible changes in the characteristics of the calculated $y_k$, as well as the corresponding gradient, are observable in both cases - the long range memory is destroyed by the removal of these events.

It is important to note that the presence of LRD cannot be rejected for Datasets 1-4 - but neither can a pure noise model. The errors caused by insufficient data (because events are rare) with possible abrupt non-stationarities causes large errors in the estimation that make the use of this type of data unsuitable. The analysis in this section can validate an $\alpha$ indicative of LRD for Datasets 5 and 6 only. It is interesting to see that in both these datasets, which were extracted from EEG, the length of the events are known and can also be seen to follow a power-law that is indicative of long memory. This can be seen in Figure 1.14. Although this may be useful for modelling purposes in other studies, it does not add any value to the determination of how predictable seizures are and is only included here for interest’s sake.

An $\alpha$ value consistent with long memory has been found in 2 out of the 6 datasets. Memory may exist in the processes described by the other 4 datasets, but the data is unable to confirm or deny this - either because of the nature of recordings are inadequate, because clinical events are not frequent enough to provide sufficient amounts of data, or because memory simply does not exist.
Figure 1.12: Stationarity analysis for Datasets 1, 5 and 6. Datasets were partitioned into 3 consecutive blocks and compared to the initial estimation of $\alpha$ as shown in Figure 1.11. The above shows that Datasets 5 and 6 are likely stationary because each partition give estimates of $\alpha$ similar to the original calculation. Dataset 1, on the other hand, is more volatile and stationarity is unlikely. This is expected because Dataset 5 and 6 span only hours and days, whereas Dataset 1 spans 25 years over which many changes in the characteristics of the epilepsy likely occur. Because of this lack of stationarity further analysis is unlikely to yield robust results, and thus Dataset 1 must be rejected as a candidate for LRD.

The implications of the discoveries listed here are discussed next.
1.5 Memory and Predictability of Seizures

What do these findings tell us about how predictable seizures are? In actuality, the fact that $\alpha$ was indicative of long memory does not make it true. The results
simply tell us that the data is consistent with a model in which LRD is present. It is always possible that the structure discussed in Section 1.4 is due to something else entirely. However, since results have proved stationary over time and because of the nature of the data they do tell us that a purely random model of the generation of epileptic activity is not suitable. A richness of structure exists that may suitably be modelled by stochastic processes with long memory, but the existence of another model equally capable of describing this phenomena is not ruled out [13].

Furthermore, even if the generation of events is due to long memory, the results do not in any way tell us what physical mechanisms are responsible for the existence of this structure. Is it a phenomena brought about from the neural network architecture in the brain? Perhaps it is due to sensory input? More likely it is to do with both of these, and more. In any case the existence of the scale-invariance has implications for the types of models chosen to model brain behaviour - they need to be capable of reproducing this type of behaviour. It should be kept in mind that the models presented in Chapter ?? do not contain any mechanisms in which memory longer than a few milliseconds is possible - if these local models are accepted as suitable for small scales then the development of global models need to somehow incorporate it.

If this work were to be used to develop these models then it is only the short-term studies acquired from clinical records that are usable. In long term qualitative studies the number of events is small and records are unreliable under stationarity tests. This is both good news and bad: short-term records are more easily maintained, but most epileptic patients do not experience a suitably large number of epileptic events in these time frames. Thus if prediction models are derived there is a limit in the type of patients that may be helped, once again emphasising the patient-specific nature of a likely predictor. However although this rare type of data, or even data derived from rat EEG, may not be directly applicable as a practical prediction mechanism, they may be used to provide insight into the development of more appropriate physical models usable for this task - an improvement on the black box predictors available today. The relationships in the length of events briefly introduced in the previous section may also be used to this end.

In any case, the aim of this work is not to develop these models - it is simply to identify the existence of scaling to try and understand if seizures are predictable. The results show that there appears to be evidence of long range memory in the system. This is apparent for studies that span hours and days, but not months or years. Perhaps when one thinks of the capabilities of the brain it is not surprising that such memory is possible in the brain - what is surprising is that this memory exists between epileptic events, and that it is not short.
For prediction - is this type of memory a good thing or a bad thing? The fact that memory exists at all in epilepsy, beyond very short times, is a positive discovery - at least seizures could be predictable! If no memory was observed then no clear evidence exists that seizures are at all predictable. Because event times are not purely random information about the timing of a current seizure could give information about future seizures. The more information of this type that is available, the more predictable seizures may be. This is a simplistic argument that holds only for cases in which many epileptiform bursts are observed in short enough time period so that data is roughly stationary. Perhaps the type of data to be used is made clearer by these findings - global events measurable by scalp or intra-cranial EEG are rare, but does this memory also exist in local events such as the discharges from focal lesions? These events occur much more frequently, more data is available, and records could be obtained using intra-cranial microelectrodes. Predictors may be able to use this type of information that to a large extent has been ignored in the past.

That this memory appears to be so long casts further doubt on how current predictors of epilepsy work because they typically depend on at most a few hours of activity, often restricting themselves to inter-seizure EEG. Of course the type of data they are looking at is different than the one here and perhaps it is sufficient to achieve success, but given their limited performance it still seems unlikely that this information can be ignored. Why don't current techniques look at a few days rather than a few hours? In theory better baselines can be established in this way, although in practice such large volumes of data cannot be processed easily and the types of features that are extracted are already too computationally demanding. Perhaps the only moral to come out of this study is that the type of data used for prediction should shift the focus from the macro to the micro scale, as is already being done in some detection-based predictors such as [3].

The identification of scaling regions in data spanning time scales of up to days implies that the generation of the next epileptic event relies in part on events up to days old. This means that to predict seizures it may be necessary to include data representative of these time scales. Current predictors of epilepsy do not utilise such long memory.

However, the fact that memory is likely gives hope that seizures are predictable at all. Future efforts should move toward the development of physical rather than black box models using information such as the one discovered here.
1.6 Conclusions

If prediction algorithms are likely to ever evolve beyond black box methods then information must be sought beyond the traditional EEG records. The work presented here has done this by inspecting the point-processes of epileptic events at different time scales.

Epileptic event times, although seemingly random, were shown here to contain some structure with memory found between events up to a scale of days. The type of structure found indicates that long range memory exists in the generators of epileptic seizures.

These findings suggest that current prediction algorithms may be using insufficient amounts of data that do not account for the presence of such long memory. Future efforts should consider this, although the results presented here are not practical for the implementation of a seizure prediction algorithm because the number of data points required is often prohibitive.

For this work to be usable directly in prediction algorithms research may need to move to a different of data in which epileptic events are very frequent. This could simply mean a reduction in the scale at which the EEG is recorded, perhaps single microelectrodes as opposed to more global electrode arrays could be used.

Although it is unlikely that this data by itself will ever be used as a predictor it is feasible that it may be used instead as a support mechanism to make current predictors more robust. This support could be as simple as the introduction of expected prediction errors.

The identification of scaling in epilepsy could lead to an alternate route of research in which this information is used for the validation or the development of models capable of replicating this behaviour. These models could subsequently be used in the prediction process, and although it is not the point of this text to suggest how this may be done, a predictor and even a detector based on the properties of network architecture, neuron properties and other such factors has a clear advantage over traditional black box time-series analysis. Until the generators of epileptic activity are better understood it is the belief of the authors that it is unlikely that implantable devices will ever move beyond the detection regime. This is of course still a workable solution - preventing seizures from occurring is the goal after all - but is unlikely to serve the wide variety of epilepsies that exist.


