INTRODUCTION TO STATISTICAL NEURODYNAMICS:
EFFECTS OF ADDITIVE AND PARAMETRIC NOISE

John Milton
Department of Neurology
Committees on Neurobiology &
Computational Neuroscience
The University of Chicago

In biology, an experimentalist typically repeats the same experiment several times. The fact that the results obtained for each trial are never precisely the same leads to those published figures with “error bars” surrounding each data point that are familiar to all who read the scientific literature. What is the source of this variability? Of course some of the variability might be attributable to measurement error (Rabinovitch, 1995). However, with modern technology the uncertainty in making a measurement is typically much smaller than the observed trial-to-trial variability. Here we consider the possibility that this variability reflects the effects of uncontrolled perturbations on the dynamical system under study. These uncontrolled perturbations are referred to herein as “noise”.

The nervous system is continuously subjected to the effects of noise [Areili, et al, 1996; Borsellino, et al, 1972; Cabrera and Milton, 2002; Hunter, et al, 2000; Stark, et al, 1958; Verveen and De Felice, 1974]. Historically, analysis of fluctuations of membrane current first attracted the attention of neurobiologists to the presence of ion channels, or pores, in neuronal membranes [Verveen and DeFelice, 1974]. Fluctuations in inter-spike intervals led to explorations of the nature of spike generating mechanisms [Tuckwell, 1989]. Just 50 years ago the concern was that noise might be detrimental to neural function [Calvin and Stevens, 1968; Fatt and Katz, 1950]. This worry was based on the fact that the root-mean-square voltage of the spontaneous fluctuations in membrane potential at a nerve ending with diameter 0.1 \( \mu m \) is of the order of 1 mV; a magnitude sufficient to affect the timing of neural spikes. However, since this time there has been the growing realization that noise plays essential roles in the sensory, motor and cognitive activities of the nervous system [Borsellino, et al, 1972; Cabrera and Milton, 2002; Cabrera, et al, 2004; Chance, et al, 2002; Douglas, et al, 1993; Eurich and Milton, 1996; Harris and Wolpert, 1998; Lindner, et al, 2004; Longtin, et al, 1990, 1991; Milton, et al, 2004; Milton and Mackey, 2000; Tuckwell, 1989]. Indeed, the addition of noise generated by an external noise generator into the nervous system can be used to therapeutic advantage [Moss and Milton, 2003; Priplata, et al, 2003]. Yet despite these observations few neurobiologists understand what noise is, how it is described, and the nature of its beneficial effects on neural function.
From a mathematical point of view there are only two ways that noise can affect the behavior of a dynamical system. First we can add noise to the solution of the equations that describe the time evolution of the stochastic dynamical system. This effect of noise is typically discussed under the heading “measurement noise” [Bendat and Piersol, 1986; Rabinovitch, 1995] and will only briefly be discussed here. Our attention is focused on the second possibility, namely that noise is part of the governing equations for the dynamical system. In this case noise plays an essential role in determining the solution of the stochastic differential equations. The noisy perturbations can either be added as a separate term to the governing equations (additive noise) or to the parameters (parametric noise). Our goal in these lectures is to discuss some effects of additive and parametric noise which have significant effects on neurodynamics. It is important to note that simulations using electronic analog circuits have played a major role in understanding the effects of noise on dynamical systems particularly when the governing equations are nonlinear [see, for example, Fauve and Heslot, 1983; Foss, et al, 1997; Gammaitoni, et al, 1998; Horowitz and Hill, 1993; Losson, et al, 1993; Sato, et al, 2000; Sinha and Moss, 1989; Yu and Lewis, 1989].

The outline of these lectures is as follows: First, we introduce the notion of a probability distribution and its associated probability density. Recent observations have emphasized that in many neurophysiological contexts the observed probability densities have much “fatter tails” that can be accounted for by a Gaussian probability density. Second, we introduce the study of the effects of additive noise by considering its effects on dynamical systems that have thresholds. The obvious application to neurodynamics arises in the description of the input-output relationships of spiking neurons. Finally, we examine the effects of parametric noise. Parametric noise is increasingly being recognized as having important roles, for example, in motor control. An interesting property of the effects of parametric noise in the vicinity of a threshold is the generation of probability densities with fat tails even in the case that the underlying dynamics are linear and deterministic!

I. Random variables and their properties

a). Introductory concepts

Useful discussions of noise and random processes1 are Bailey [1964], Davenport and Root [1987], Evans, et al [1993], Gardiner [1990], Horsthemke and Lefever [1984], Jenkins and Watts [1968], Lasota and Mackey [1994], and MacDonald [1962].

Identify the outcome of each experiment as a point (possibly n-dimensional), s, in a sample space S. This sample space represents the

---

1 We restrict our attention to stationary stochastic processes, i.e. processes whose statistical properties, e.g. mean, standard deviation, do not change with time.
aggregate of all possible outcomes of the experiment. A **random variable**, \( x(s) \), defined on a sample space of points \( s \) will be called a random variable if for every real number \( A \), a set of points \( s \) for which \( x(s) \leq A \) is one of the class of admissible sets for which the probability is defined. For example the sample space, \( S \), for two coin flips is \( S = \{ HH, HT, TH, TT \} \), where \( H \) indicates a head and \( T \) a tail. The random variable “number of heads” associates the number 0 with the set \( s_0 = \{ TT \} \), the number 1 with the set \( s_1 = \{ HT, TH \} \), and the number 2 with the set \( s_2 = \{ HH \} \). Of course we could choose other random variables, e.g. “number of tails”, “two heads”, “two tails”.

The **probability** \( P(A) \) that the outcome of a given experiment will be the event \( (A) \) can be expressed as the probability \( P(S_A) \) that the sample point \( s \), corresponding to the outcome of the experiment, falls in the subset of sample points \( S_A \) corresponding to the event \( A \):

\[
P(A) = P(s \in S_A) = P(S_A)
\]

where the notation \( s \in S_A \) means that a point \( s \) is an element of the point set \( S_A \).

A random variable may be discrete or continuous. A **discrete random variable** \( x \) takes on only a finite number of values on a finite interval (Figure 1). Figure 1b shows an example of a discrete random variable that takes on only five possible values, \( \{ x_1, x_2, x_3, x_4, x_5 \} \) on the interval \([ -\infty, +\infty ]\). The probability associated with each \( x_j \) is \( P(x_j) \). The function of \( X \) whose value is the probability \( P(x \leq X) \) that random variable \( x \) is less than or equal to \( X \) is called the **probability distribution function**

\[
P(x \leq X) = \sum_{x_j \leq X} P(x_j)
\]  

(1)

From Figure 1a we see that \( P(x \leq X) \) is a staircase function that is non-decreasing and is bounded between 0 and 1. It is perhaps easier to think of \( P(x \leq X) \) as the cumulative probability function as \( X \) goes from left to right; however, this terminology is currently regarded as being redundant.
In general any random variable may be thought of as having both a discrete and a continuous part. This is not surprising since even a continuous function can be represented as a staircase function with very closely spaced steps. However, in experimental situations we most often encounter the situation that the probability distribution is not only continuous but also differentiable with a continuous derivative, except, perhaps, at a discrete set of points. In this case more direct methods of analysis are possible since we can define a probability density function, $p(x)$, as

$$p(X) = \frac{dP(x \leq X)}{dX}$$

where $p(x)$ satisfies

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

The above condition is the continuous random variable expression of the fact that the probability of a certain event is unity.
To illustrate the difference between $P(x \leq X)$ and $p(x)$ consider the example of choosing a point at random on the interval $[a,b]$ (Figure 2). Then $P(x \leq X)$ is (Figure 2a)

$$P(x \leq X) = \begin{cases} 
0 & x < a \\
\frac{x-a}{b-a} & a \leq x \leq b \\
1 & x > b 
\end{cases} \quad (4)$$

and $p(x)$ (Figure 2b) is

$$p(x) = \begin{cases} 
(b-a)^{-1} & a \leq x \leq b \\
0 & otherwise 
\end{cases} \quad (5)$$

Together these functions describe the uniform probability distribution.
The probability that a continuous random variable has a value falling in the interval \((a < x \leq b)\) is given by the integral of the probability density function over that interval\(^2\)

\[
P(a < x \leq b) = \int_{a}^{b} p(x)dx
\]  

(6)

From this relation we see that the probability that a continuous random variable takes on any specific value is zero. In other words, in a random dynamical system we cannot predict specific outcomes but rather only “an average” outcome following the performance of a very large number of experiments. Herein is the Achilles heel for the study of stochastic dynamical systems. We must be prepared to either perform a large number of experiments or a large number of numerical simulations in order to identify the behavior “on average”. The importance of the above defined probability functions, in particular \(p(x)\), is that they describe the long-run behavior of random variables.

b). Moments

The various averages of random variables (e.g. the mean, variance, etc) measured experimentally can be determined from the probability density function, \(p(x)\). One group of averaged properties that are of considerable significance are the moments of \(p(x)\). The \(n\)-th moment of \(p(x)\) is the statistical average of the \(n\)-th power of \(x\), i.e.

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

(7)

For a stationary random process the first moment gives the mean or DC-component of the process. The second central moment gives the intensity of the varying component of the random process, i.e. the AC component. The positive square root of the variance is the standard deviation which is typically referred to as the root-mean-square (rms) value of the AC component of the random process.

For example the mean of the uniform distribution is

\[
\mu = \int_{-\infty}^{\infty} xp(x)dx = \frac{1}{b-a} \int_{a}^{b} x dx = \frac{1}{b-a} \cdot \frac{(b^2-a^2)}{2} = \frac{b+a}{2}
\]

(8)

and the variance is

\[\text{Note that} \ a < x \leq b \text{ rather than } a \leq x \leq b \text{ is because we must exclude the case } a = b \text{ (see below).}\]
Another averaged quantity that is of considerable importance is the characteristic, or moment generating, function, which is the statistical average of exp(\(ivx\))

\[
M_{x}(iv) = \int_{-\infty}^{+\infty} \exp(ivx)\rho(x)dx
\]  

(10)

where \(v\) is real and \(i = \sqrt{-1}\). Since \(\rho(x)\) is nonnegative and \(\exp(ivx)\) has unit magnitude,

\[
\left|\int_{-\infty}^{+\infty} \exp(ivx)\rho(x)dx\right| \leq \int_{-\infty}^{+\infty} \rho(x)dx = 1
\]  

(11)

and hence the characteristic function always exists.

It follows from the definition of the Fourier integral\(^3\) that the characteristic function of the probability distribution of a random variable \(x\) is the Fourier transform of the probability density function of that random variable [Bracewell, 1986; Brigham, 1988; Davenport and Root, 1987]. Therefore we may use the inverse Fourier transformation

\[
p(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} M_{x}(iv)\exp(-ivx)dv
\]  

(12)

to obtain the probability density function of a random variable when we know its characteristic function. It is often easier to determine the characteristic function first and then transform it to obtain the probability density than it is to determine the probability density directly (see below).

c). Characterization of noise

The specification of a set of experiments with the corresponding probability functions and random variables defines a random, or stochastic, process. Experimentally we measure three attributes to characterize the stochastic

\(^3\) The Fourier integral is defined as

\[
H(f) = \int_{-\infty}^{+\infty} h(t)e^{i2\pi ft}dt
\]

If the integral exists for every value of the parameter \(f\) then \(H(f)\) is the Fourier transform of \(h(t)\).
process that generates the noisy inputs: 1) intensity; 2) probability density distribution; and 3) correlation time.

i). Intensity: The intensity of noise is given by its variance. However, engineers and scientists typically refer to the intensity of noise in terms of its root-mean-square value which is equivalent to its standard deviation. It is important to note that whereas the variance of two independent noise signals is the sum of their variances, the rms value is not equal to the sum of the rms values, but is equal to

\[ V_{\text{rms}}(a+b) = \sqrt{V_{\text{rms}}^2(a) + V_{\text{rms}}^2(b)} . \quad (13) \]

ii). Correlation time: The correlation time \( t_{\text{cor}} \) of a stationary stochastic process with density \( p(x) \) is defined by [Bendat and Piersol, 1986]

\[ t_{\text{cor}} = \frac{1}{C(0)} \int_0^\infty C(\tau) d\tau \quad (14) \]

where \( C(\tau) \) is the autocorrelation function of the noise process\(^4\) and \( C(0) \) is the variance. The rate at which \( C(\tau) \) goes to zero as \( \tau \to \infty \) is a measure of the memory time of the stochastic process. A process has short memory if \( C(\tau) \) decays rapidly in relation to the characteristic relaxation time\(^5\) of the system, \( t_{\text{sys}} \). The relaxation time of a system is estimated from its response to a brief perturbation. In the case of the dynamical system \( \dot{x}(t) + kx(t) = 0 \) we have \( t_{\text{sys}} = k^{-1} \).

All real noise is colored, i.e. \( t_{\text{cor}} \) is finite. In this lecture we restrict our attention to “white noise”, i.e. the special case for which \( t_{\text{cor}} = 0 \). For white noise \( C(\tau) \) is a Dirac delta function, hence white noise is often referred to as \( \delta \) -correlated noise. The fact that the power spectrum of a white noise process is constant implies infinite power. Thus white noise is not physically realistic. However, it is found that white noise describes real noise surprisingly well when \( t_{\text{cor}} \ll t_{\text{sys}} \). The use of the white noise approximation greatly facilitates the mathematical analyses of noisy dynamical systems.

\(^4\) The autocorrelation function, \( R(\tau) \), for a random process is

\[ R(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_0^T y(t)y(t+\tau)dt \]

\(^5\) The relaxation time of a system is estimated from its response to a brief perturbation. In the case of the dynamical system \( \dot{x}(t) + kx(t) = 0 \) we have \( t_{\text{sys}} = k^{-1} \) and is thus the time for the perturbation to decay to \( e^{-1} \) of its initial value.

![Figure 3: Examples of the probability density functions for Lévy-stable distributions as a function of the Lévy index, \( \alpha \).](image)

At first glance these observations are quite disturbing. The Central Limit Theorem states that the limit of normalized sums of independent identically distributed terms with finite variance must be Gaussian. The solution to this paradox is provided by relaxing the requirement that the variance is finite. The Generalized Central Limit Theorem states that the only possible nontrivial limits of normalized sums of independent identically distributed terms are Lévy-stable. All Lévy-stable, or \( \alpha \)–stable, distributions have the property that if \( x \) and \( y \) are random variables from this family, then the random variable formed by \( x + y \) also has a distribution from this family. The Gaussian distribution is the only Lévy-stable distribution that has finite variance. As better measuring devices and larger data sets have become available it is being recognized that non-Gaussian Lévy distributions are prevalent in nature [Tsallis, et al, 1995].
Lévy-stable distributions represent an example of a situation in which the characteristic function is particularly useful. The characteristic function, $M(z)$, for the symmetric Lévy distributions is (Woyczynski, 2001)

$$M(z) = e^{-t|z|} = \exp \left\{ \gamma \int_{-\infty}^{\infty} \left( e^{ixz} - 1 \right) \frac{dx}{|x|^{\alpha+1}} \right\}$$

(15)

The parameter $0 < \alpha \leq 2$ is an index of the peakedness of a distribution, and the parameter $\gamma \geq 0$ is a scale factor that must be nonnegative. Note that the mean of the distribution, $\mu$, exists only if $1 < \alpha \leq 2$.

A characteristic feature of Lévy-stable distributions is that for asymptotically large displacements there is a power law decay of its density

$$p(x) \approx x^{-\beta}$$

(16)

where $\beta = \alpha + 1$.

A practical difficulty in working with Lévy-stable distributions is that closed form expressions for $p(x)$ exist only for three special certain choices of the parameters (Figure 3). For example the Gaussian distribution corresponds to the Lévy distribution with $\alpha = 2$, $\gamma = \sigma^2 / 2$ and $\mu =$ mean; the Cauchy distribution corresponds to the case $\alpha = 1$; and the Lévy distribution (a confusing choice of terms) to the case $\alpha = 0.5$. However, since all of the Lévy-stable distributions are real-valued and symmetric about the mean (i.e. they are even functions), it follows that the Fourier transform and its inverse are real-valued functions. Thus the development of the Fast Fourier Transform and fast computers has enabled investigators to readily study models based on Lévy-stable models.

There are also a number of experimental challenges that arise in the measurement of fluctuations described by the non-Gaussian variants of the Lévy-stable distribution. From a practical point of view probability densities with tails fatter than Gaussian means that events that are many times the standard deviation occur more frequently than for a Gaussian distributed variable. However, despite the increased frequency of these events they are still rare! Consequently very large data sets must be collected in order to ensure that these rare events are detected, e.g. $\geq 10^6$ data points [Weron, 2001]. In collecting time series these rare events can be missed unless proper attention is given to the choice of the frequency cut-offs of the low-pass and high-pass filters. Moreover it is crucially important that the sampling frequency is sufficiently high. It can be shown that as the sampling frequency decreases the distribution of the fluctuations necessarily must approach Gaussian [Mantegna and Stanley, 1995].
Recent excitement concerning Lévy phenomena in neurobiology stems from the relationship between a Lévy flight and a Lévy-stable distribution. A Lévy flight is a type of random walk that historically arose in the description of anomalous diffusion. It can be shown that after many steps the probability density of the displacement of a Lévy flight converges to a Lévy stable distribution. Surprisingly, measurements of the foraging movements of animals and birds [Viswanathan, et al, 1996], eye movements during reading (Brockman and Geisel, 2000) and finger movements during stick balancing at the fingertip [Cabrera and Milton, 2004b] all are characterized by Lévy distributions with $\alpha \leq 1$. Lévy flights with this index are typical for super-diffusion [Brockman and Geisel, 2003; Rangarajan and Ding, 2000]. In particular, it has been shown that Lévy flights with $\alpha \approx 1$ are optimal for random search patterns [Brockman and Geisel, 2003; Buldyrev, et al, 2001, Viswanathan, et al, 1999].

II. Additive noise

The paradigm for the effects of additive noise on the fluctuations of a fixed-point is the Langevin equation

$$\frac{dx(t)}{dt} + bx(t) = \sigma \xi(t), \quad x(0) = x_0$$

(17)

where $b$, $\sigma$ are constants and $\xi(t)$ is Gaussian-distributed white noise [for readable discussions of the analyses of the Langevin equation see Gardiner, 1990; Lasota and Mackey, 1994]. This equation will be studied in more detail in Professor Swain’s lecture tomorrow. Here we point out that physicists and mathematicians tend to forget that this first-order stochastic differential equation can be thought of as describing the effects of a low pass filter (left hand side of (17)) on a noisy input (right hand side of (17)). This is an important insight since it implies that integral methods, such as those involving the Fourier integral, can also be used to understand the behavior of (17) [Schwarzenbach and Gill, 1992]. It is known that the result of low pass filtering, Gaussian distributed noise with a linear filter results in noise with a Gaussian distribution is to produce noise with a Gaussian probability distribution. Of course the variance of the Gaussian distribution changes as a result of the filtering to $\sigma^2 / b$.

a). Linearization with noise

---

6 For a thorough discussion of the numerical integration of stochastic differential equations under the influence of either additive or parameter noise the reader is referred to Kloeden and Platen [1992].

7 Indeed analog noise is typically low-pass filtered using a simple RC circuit to produce band-limited white noise with a Gaussian distribution [Horowitz and Hill, 1993].
One way to characterize neurons is by the relationship between their input and their output. For a large class of spiking neurons the output consists of spike trains in which the inter-spike interval is a random variable. It is observed that variations in the instantaneous rate of the spike input, i.e. \((\text{inter-spike interval})^{-1}\), are translated into a continuous modulation of the instantaneous output spiking rate. Living neurons exhibit a wide dynamic range: input mean spike rates can be modulated over a nearly 1000-fold range, i.e. from nearly 0 spikes/s to almost 1000 spikes/s. The relationship between the input to a neuron to its output is called the input-output relationship. All neural network modelers are familiar with sigmoidal approximations of the input-output relationship of neurons, for example, as \(\tanh(x)\) (Figure 4a). However, few realize that the validity of this approximation represents an effect of noise!

The large dynamic range for neurons presented a problem for modelers. Figure 4b (solid line) shows the calculated input-output relationship for a 'leaky' integrate-and-fire (IF) neuron [Feldman and Cowan, 1975; Milton, 1996]. An IF neuron does not, by definition, spike unless the integrated membrane potential exceeds a spiking threshold. Thus an IF neuron cannot modulate input spiking rates that are lower than a certain critical rate. How is it possible to explain the observation that the lower end of the dynamic range for a living neuron approaches 0 spikes/s? Surprisingly the answer to this dilemma involves the linearizing effects of noise (additive) on a threshold.
A simple way to understand the effects of noise at a threshold is to consider the circuit shown in Figure 5a. \( U(t) \) is the input to the system and \( Y(t) \) is the output. \( N(t) \) represents noise and is assumed to be uncorrelated with \( U(t) \). \( X(t) \) is the sum of \( U(t) \) and \( N(t) \), and \( x \) is the instantaneous value of \( X(t) \). \( F(x) \) represents a static threshold function

\[
F(x) = \begin{cases} 
k & x \geq C, \ k = \text{constant} \\
0 & x < C
\end{cases}
\]

(18)

The instantaneous value \( z \) of \( Z(t) \) is a random variable and \( E[z] \) is its expectation.

Suppose that \( N(t) = 0 \). Then

\[
Y(t) = E[F(U(t))] = \begin{cases} 
k & U(t) \geq C \\
0 & U(t) < C
\end{cases}
\]

(19)

In this case we see that the output, \( Y(t) \), is a step function of the amplitude of \( U(t) \).
Figure 6: Left: Averaged system output \( \langle y \rangle \) versus \( x \) for a threshold device as a function of uniform noise of different intensities. Right: the same thing but for Gaussian distributed noise of different intensities. Figures from Gammaitoni [1995].

In these calculations the dynamical system was

\[
y(t) = \begin{cases} 
0 & \text{for } x(t) + \xi(t) < 0.5 \\
1 & \text{for } x(t) + \xi(t) \geq 0.5
\end{cases}
\]

where \( 0 \leq x \leq 1 \).

Now let us assume that \( N(t) \) is uniformly distributed in the interval \([-a,a]\). The relationship between \( Y(t) \) and \( U(t) - C \) is shown in Figure 5b and we have

\[
Y(t) = \begin{cases} 
0 & U(t) < C - a \\
\frac{k}{2} + \left( \frac{k}{2a} \right)(U(t) - C) & C - a \leq U(t) \leq C + a \\
k & U(t) \geq C + a
\end{cases}
\]

Clearly \( Y(t) \) and \( U(t) \) are linearly related over the range \( U(t) \in [C - a, C + a] \)! The range of this linearization increases with increasing noise amplitude (Figure 6-left). For example, since \( N(t) \) is the uniform distribution, the variance, \( \sigma_y^2 \), of \( N(t) \) is

\[
\sigma_y^2 = \frac{(2a)^2}{12} = \frac{a^2}{3}
\]

and thus we have
\[ a = \sqrt{3\sigma_v} \]
\[ [C-a, C+a] = [C-\sqrt{3\sigma_v}, C+\sqrt{3\sigma_v}] \] (22)

Since \( \sigma_v \) is the rms deviation of the instantaneous amplitude of \( N(t) \) from its mean (equal zero in this case) it follows that \( \sigma_v \) is the rms amplitude of the noise.

The same result holds if \( N(t) \) is a stationary Gaussian random process with zero mean and standard deviation \( \sigma \) (Figure 6-right). In this case we have

\[ Y(t) = \frac{k}{2} + k \int_0^{U(t)-C} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( \frac{x^2}{2\sigma^2} \right) dx \] (23)

Now if \( \sigma \) is sufficiently large we have

\[ \exp \left( \frac{x^2}{2\sigma^2} \right) \approx 1 - \frac{x^2}{2\sigma^2} + \ldots \]
\[ \approx 1 \] (24)

and we obtain

\[ Y(t) = \frac{k}{2} + \sqrt{\frac{1}{2\pi\sigma^2}} (U(t) - C) \] (25)

Figure 4b shows the linearizing effects of additive noise on the input-output relationship of a neuron. In this numerical simulation the input to the neuron is the sum of a DC current plus uniformly distributed white noise. Clearly noise extends the bottom limit of the input-output relationship modulation range to zero yielding the familiar sigmoidal neuronal input-output relationship (Figure 4a).

b). Dithering

A practical application of the linearizing effects of noise on a threshold arise in the situation that data is collected directly into a computer via an A/D board is now common place [Gammaitoni, 1995]. This process involves two steps: 1) time discretization, and 2) amplitude quantization. Time discretization is the error free step. Two errors enter during amplitude quantization: 1) signal quantization leads to unavoidable distortion, i.e. the presence of spurious signals in a frequency band different from the original one, and 2) loss of signal detail that is small compared to the quantization step (the dynamic range of the digital signal is finite).
Define the quantization error, $\eta$, as

$$\eta = y - x$$  \hspace{1cm} (26)$$

where $x$ is the analog input to the A/D board and $y$ is the digital representation of the analog signal inputted from the A/D board to the computer. If we had a linear response of the A/D board, i.e. $y=x$, then $\eta = 0$. It has been shown that the minimum loss of statistical data from the input $x$ occurs when the quantization error can be made independent of $x$. A simple way to accomplish this is to add noise to the signal before quantization. This technique is referred to as dithering. A great deal of work has been devoted to determining the best choices of the dithering signal. The main conclusions of these studies are 1) the addition of a proper dither signal can cause the independence and whitening of the quantization error resulting in both a reduction of signal distortion and an improvement of the system dynamic range, and 2) the best choice for the dither signal is a random dither that is uniformly distributed within an interval of amplitude equal to the quantization step.

In order to appreciate the difference between uniformly and Gaussian distributed noise define (see also legend to Figure 6)

$$D = \sqrt{\int_0^1 \left(\langle y \rangle - x \right)^2 dx}$$  \hspace{1cm} (27)$$

Figure 7 compares $D$ versus the rms value of the noise intensity, $\sigma$, for uniform and Gaussian distributed noise. In both cases $D$ approaches a minimum but only in the case of uniform noise does the minimum equal zero. Thus if the noise is properly matched, the averaged output is indistinguishable from the output. It is not difficult to appreciate why uniform distributed noise is optimal. Suppose we had a constant input signal which has amplitude 0.2 and the threshold is 0.5. To achieve a correct representation of this signal at the output, we require that the system have an averaged output of 0.2. This result can be obtained if the signal crosses the threshold two times out of ten, i.e. 2 ‘1’s and 8 ‘0’s averaged equals 0.2. To reach linearity such a probability should be kept equal for all the input range, i.e. we need a dither signal that is uniform over the whole quantization step with an amplitude equal to the quantization step. The problem with Gaussian noise is that the PDF is not uniform. With such a noise source, the better performance is reached in a range where the probability density function is more uniform, i.e. around the mean or when $\sigma \rightarrow \infty$. However, in these cases the output range is dramatically reduced.

The essential requirement for dithering is that the time series be oversampled in the presence of noise. If the frequency content of the noise is
much higher than that for the primary signal, then over short intervals the average population of the two quantized levels is a measure of how close the true signal is to the threshold that separates two adjacent bins. Thus it follows that the true signal can be recovered by low-pass filtering (equivalent to averaging) the quantized data. This procedure is illustrated for a noisy pupillometer in Figure 8 [Hunter, et al, 2000]. An analog input signal proportional to pupil diameter (Figure 8a) is discretized in the presence of noise at 25 Hz (Figure 8b). It should be noted that the major frequency components of the open-loop irregular fluctuations in pupil size are \( \leq 0.3 \) Hz [Longtin, et al, 1990; Stark, et al, 1958. The quantized signal is then low-passed filtered to yield a time series that closely resembles the original time series (compare Figure 8a to 8c). The main difference is that the low-pass filtering removes the high-frequency components of original signal.

![Figure 8](image)

**Figure 8:** Application of dithering for measuring pupil size in a noisy pupillometer with low resolution. See text for discussion. Data from Hunter, et al [2000].

**c). Stochastic resonance**
Although the linearizing effects of noise on a threshold are useful in some contexts, this phenomenon can be a nuisance in situations where a threshold crossing is used to trigger another event. This is because in the presence of noise there will be multiple threshold crossings when the signal is close to the threshold (Figure 9a). Electrical engineers solved this problem by developing a special type of voltage comparator called a Schmitt trigger (Figure 10a). A Schmitt trigger is a bistable device that has two thresholds which in turn depend on the output state, i.e. there is a positive feedback loop in the circuit (Figure 10b). As is shown in Figure 9b this circuit is less likely to produce multiple triggerings. Fauve and Heslot [1983] studied the behavior of a Schmitt trigger that simultaneously received a weak periodic voltage and noise. Surprisingly, they observed the information in the input could be amplified and optimized in the output by adjusting the intensity of the noise. This phenomenon is called stochastic resonance. It can be readily appreciated that dithering is the special case of stochastic resonance in which the two voltage threshold levels in the Schmitt trigger coincide [Gammaitoni, et al, 1998].

Figure 10: a) Circuit diagram for a Schmitt trigger. The essential difference between a Schmitt trigger and a simple voltage comparator is the presence of the positive feedback loop. b) The output of the Schmitt trigger depends on the input voltage and on its recent history (hysteresis). Figures adapted from Horowitz and Hill [1993].

Priplata, et al, 2003. Consider the properties of a general bistable system driven by a sinusoidal function with added noise. A simple example of such a dynamical system is

\[
\frac{dx(t)}{dt} = -\frac{dU(x)}{dx} + \xi(t) + m\sin\omega t
\]  

(28)

where \( U(x) \) is the double well potential shown in Figure 11 and \( \xi(t) \) describes Gaussian distributed white noise. The effect of the sinusoidal modulation is to rock the potential surface to and fro (Figure 11). In the absence of the sinusoidal forcing switches between the two basins of attraction are possible through the effects of noise. For small intensity noise and a sufficiently high barrier between the two basins of attractions such switches occur rarely. However, for \( m > 0 \), the switching times between the two potential wells become correlated with the periodic forcing, and hence the transition probabilities become periodically modulated. It is easy to see that switches become more probable once the intensity of the noise exceeds to height of the potential barrier that separates the two wells.

This stochastic resonance mechanism predicts two phenomena which have been observed experimentally. First, it explains two curious properties of multimodal inter-spike histograms that have been recorded for a variety of
neurons including retinal ganglion neurons in the cat [Sanderson, et al, 1973] and in auditory fibers of monkeys [Rose, et al, 1967] (Figure 12a): 1) the modes occur at an integer multiple of a fundamental period; and 2) the mode amplitude decreases approximately exponentially.

Figure 11: A suitable dose of noise (i.e. when the period of the driving approximately equals twice the noise-induced escape time) will make the “sad” face happy by allowing synchronized hopping to the attractor with the lower energy. Strictly speaking, this holds true only for the statistical average. Figure taken from Gammaitoni, et al [1998]. Why? It’s really a fun figure!

Second, the signal to noise ratio (SNR) for a stochastically resonant dynamical system exhibits a non-monotone dependence on noise intensity. As the intensity of the noise increases the output of the sensory neuron increases (Figure 12b). This is because switches between the basins of attraction are possible only if noise is present. The SNR reaches a maximum at roughly the point where the noise intensity is on average sufficient to cause a transition at the point in time when the potential barrier between the basins of attraction is most shallow. Thereafter, further increases in noise intensity deteriorate the SNR (as is always the case in dynamical systems that do not exhibit stochastic resonance).
d). Spike time reliability

Typically the input of a neuron does not contain a strong periodic component but is a rather complex aperiodic signal. Such inputs can be represented as having both a DC, or constant, component and an AC, or time-varying, component. The
neuron's firing rate ($f_{DC}$) in response to the DC component places constraints on the types of signals that may be encoded reliably in the presence of noise. This effect is currently studied under the heading "spike timing reliability".

Spike timing reliability refers to the precision of spike timing when a spiking neuron is repeatedly given the same time-varying input. In 1995 Mainen and Sejnowski made a surprising observation concerning the spike-timing reliability of neocortical neurons. Namely, precisely timed spike trains are produced by neurons in response to aperiodic input signals in which current fluctuations resemble synaptic activity, but not when they receive a constant-current stimulus. Again the interpretation of this phenomenon involves considerations of the effects of noise on the input-output relationship of a neuron.

![Figure 13: Spike timing reliability (A, C) as a function of the magnitude of current fluctuations for a "leaky" integrate-and-fire model neuron (e.g. (29)) receiving B a broadband input and D a band-stop current input lacking frequencies around $f_{DC}$. Modulation intensity of these currents was periodically increase to mimic situations in which neurons receive bursts of large amplitude inputs, as in regularly spiking cortical neurons during sleep. Figure from Hunter, et al [1998].](image)

An important consideration is the magnitude of the fluctuations (AC component) in the input relative to the current (DC component) necessary to cause neural spiking (Figure 13). It is not difficult to appreciate that spike timing
will be reliable when the relative magnitude of the input fluctuations is high. In the presence of noise, the width of inter-spike interval distributions is inversely proportional to the slope of the membrane potential at threshold [Goldberg, et al, 1984; Stein, 1967]. A current with a large-amplitude fluctuating component will cause threshold crossings with a steeper slope than will a constant current and thus generate more reliably timed spikes in the presence of noise. An analogous slope condition exists for the synchronization of a network of neurons to a coherent input [Gerstner, et al, 1996] and also explains the increased variability of the period of pupil cycling when the threshold is set close to the maximum pupil area [Longtin, et al, 1990; Milton and Longtin, 1990].

![Figure 14: 'Leaky' integrate-and-fire model neuron (e.g. (29)) with constant and periodic inputs. A: membrane current in response to a DC input current, B: membrane potential with current modulated at $f/f_{DC} = 0.65$, and C: trajectory modulated by $f/f_{DC} = 1$. The dashed horizontal lines indicate the threshold for spiking, D-F the distribution of spike times obtained for the currents in G-I in the presence of noise. For each input, initial conditions were chosen to trigger a spike at time 0 in the absence of noise resulting in a peaked spike time distribution around 0 when noise is added. Note that the first peak in the spike distribution is sharper in response to the sinusoidal currents (E and F) than in response to a DC input (D) because the slope of the membrane potential at threshold crossing is steeper since the initial conditions were chosen so that the first spike in the deterministic case would occur at the same phase of the sinusoids. At the second threshold crossing, the slope of the membrane potential in C is steeper than in A or B. This results in a narrower spike time distribution in the presence of noise. Figure from Hunter, et al [1998].](image)

However, when the fluctuating component of the input is relatively small it is also possible for spike-timing reliability to arise from a resonance phenomenon.
Suppose at time 0 a spike occurs on the rising phase of an input signal and the membrane potential is reset to rest. If the time-varying component of the input is small compared with the DC component, then the next spike will not occur until some time approximately equal to $T_{dc}$, the inverse of $f_{dc}$. If the time-varying component of the input signal is increasing around time $T_{dc}$, the spike time distribution will be narrower than if the time-varying component is decreasing at this time because of the above slope considerations.

It follows that spike-timing reliability will be highest when the aperiodic input contains frequencies of the order of $f_{dc}$ [Bryant and Segundo, 1976; Collins, et al, 1995; Hunter, et al, 1998; Knight, 1972; Jensen, 1998]. Figure 15 shows that this is true. This figure shows spike timing reliability of an Aplysia motoneuron to three different aperiodic signals: 1) broadband, low-pass filtered noise (Figure 15a); 2) broadband, low-pass filtered noise in which the frequencies around and including $f_{dc}$ were removed (Figure 15b); and 3) a broadband, low-pass filtered noise in which a frequency band below $f_{dc}$ was removed (Figure 15c). Spike timing reliability is highest provided that the input contains frequencies around $f_{dc}$.

Figure 15: Spike time reliability (right) for an Aplysia motoneuron to different aperiodic inputs (left). See text for discussion. Figure from Hunter and Milton [2002].

---

8 A convenient measure of spike time reliability is the reliability statistic, $R$, calculated as the time-series variance normalized so that it ranges from zero to one [Hunter, et al, 1998; Hunter and Milton, 2003].
The same approach that was used in Figure 15 can be used to demonstrate the amplitude dependence of spike timing reliability [Hunter and Milton, 2003]. In this experiment the Aplysia motoneuron was stimulated with two types of periodic signals: broadband noise constructed by low-pass filtering Gaussian-distributed white noise and the same signal with the frequency $f_{DC}$ removed (Figure 16). The signals were normalized so that the total power was the same. For very small relative amplitude inputs (coefficient of variation (CV) = 0.0025), neither signal generated reliable spiking (Figure 16a). When the relative amplitude of the aperiodic input was moderate (CV = 0.15), the broadband signal containing $f_{DC}$ generates higher reliability than the input lacking that frequency band (Figure 16b). Finally, for a high relative amplitude input (CV = 0.35), the reliability is higher still but does not depend on the presence or absence of $f_{DC}$ (Figure 16c).

Figure 16: Spike time reliability as a function of the relative amplitude (measured as the coefficient of variation (CV) = standard deviation/ mean) of the input to an Aplysia motoneuron. For each CV the spike time reliability (10 superimposed trials) we compared a broad band input to one lacking $f_{DC}$. Figure from Hunter and Milton [2003].
The observations in Figure 16 can be reproduced by studying the response of a leaky integrate-and-fire neuron model to inputs, $I(t)$, constructed from modulated Poisson processes [Hunter and Milton, 2003]

$$ RC \frac{dV}{dt} = -V(t) + RI(t) + \sigma \xi(t) \quad (29) $$

where $V(t)$ is the membrane potential, $R$ is the membrane resistance, $C$ is the membrane capacitance, $\xi(t)$ is Gaussian distributed noise with unit standard deviation scaled by $\sigma$. In integrating (29) it is assumed that when $V \geq \theta$ the neuron produces a spike and the membrane potential is reset to its resting potential. Modulated Poisson processes arise in the description of neural spike trains in which the probability of neural firing varies with time [Burkitt and Clark, 2000; Perkel and Bullock, 1968; Tuckwell, 1989]. The advantage of this signal is that the power spectra and SNR at the modulation frequency are known in closed form [Bartlett, 1963; Bayly, 1968; Knox, 1970]. The modulation rate is given by

$$ rate(t) = \lambda_0 \left(1 + m \sin(2\pi f_m t)\right) \quad (30) $$

where $\lambda_0$ is the carrier rate, $0 \leq m \leq 1$ is the modulation amplitude, and $f_m$ is the modulation frequency. It is assumed that the carrier rate is sufficiently low that the probability of spiking twice in a single time interval is small. In order to introduce a periodically modulated spike train input into (29), we took

$$ I(t) = DC + i \left( \sum_k \delta(t - t_k) - \lambda_0 \right) \quad (31) $$

where $\delta(t)$ is the Dirac-delta function, $t_k$ are the spike times from the Poisson process with modulated given by (30), and $I$ is the event amplitude. It should be noted that changes in the carrier rate affect the average spiking rate of the neuron and hence the location of the resonance peak. The offset $\lambda_0$ removes this confounding factor.

A realization of a neural spike train of a neural spike train generated by a periodically modulated Poisson process is show in Figure 17A, and its power spectral density in Figure 17B. The power spectral density, $S(f)$, is

$$ S(f) = \lambda_0 + \lambda^2_0 \delta(f) + \frac{1}{4} \lambda^2_0 m^2 \delta \left(|f| - f_m\right) \quad (32) $$

where $\delta$ is the Dirac delta function. The broadband, background power is the same at all frequencies and equals $\lambda_0$. The power at the modulation frequency is $\lambda_0 m^2 / 4$ and hence the SNR is proportional to $\lambda^2_0 m^2$. The spike time reliability for
the IF neurons receiving the modulated Poisson input plus additive noise is shown in Figure 17C. As can be seen there is a monotonic relationship between \( R \) and \( \lambda \). An analogous effect is seen if the modulation depth is held constant and the modulation frequency is varied. In this case there is a resonance peak seen only for moderate amplitudes (Figure 17D).

Figure 17: Modulated Poisson input current generates three relative amplitude regimes for spike-timing reliability. See text for discussion. Figure from Hunter and Milton [2003]

The important point about spike timing reliability is that it emphasizes that the encoding properties of a neuron depends on both the intrinsic properties of the neuron and the nature of the input. Thus the same neuron can be either a rate or a spike time encoder. Moreover there exists a range of input modulation amplitudes for which small modifications in either the frequency content of the input or the firing rate of the neuron can dramatically alter spike timing reliability. This leads to a rate-control mechanism for neural synchrony [Hunter and Milton, 2002, 2003].

III. Parametric noise

In the nervous system the effects of noise typically are state-dependent. For example, membrane noise reflects fluctuations in conductance and hence its effect on current (proportional to the product of conductance and the driving potential) is state-dependent [Chance, et al, 2002; Verveen and De Felice, 1974]. State-dependent noise underlies the spontaneous fluctuations in pupil size [Stark, et al, 1958] and plays important roles in motor [Harris and Wolpert, 1998; Jirsa, et al, 2000] and balance [Cabrera and Milton, 2002; Cabrera, et al, 2004] control. Indeed it is quite difficult to think of an effect of noise on the nervous system that is not state dependent!
One way to account for these state-dependent effects of noise is to assume that the noise enters through parameters. It must be remembered that in modeling those variables that change much more slowly than others are typically held constant (hence are parameters). However, experimentally it is typically found that those variables designated as parameters in a model themselves vary (see, for example, Milton, et al, 1989) thus blurring the traditional distinction between variable and parameter. It is obvious that parametric noise will have much greater effects on a dynamical system than additive noise. This is because the parameters determine the shape of the "potential surface" on which the dynamics evolve. If parameters change as an effect of the noise, then this means that this potential surface is itself continually changing. Here lies the problem.

Many of the same effects observed for dynamical systems subjected to additive noise have their analogues in dynamical systems subjected to parametric noise [Milton, et al, 2004]. For example, although we have discussed stochastic resonance in the context of additive noise, the same phenomena occurs in dynamical systems subjected to the effects of parametric noise (for a review see [Gammaitoni, et al 1998]). An exception arises in the case that a parameter is stochastically forced back and forth across a stability boundary [Cabrera and Milton, 2002; Heagy, et al, 1994; Platt, et al, 1993]. In this case phenomena arise which do not have an analog in dynamical systems with additive noise. These phenomena are discussed here under the heading “on-off intermittency”.

To illustrate the added complexities that arise consider parametric Langevin equation

$$\frac{dx(t)}{dt} = \eta(t)x(t) + \xi(t)$$  \hspace{1cm} (33)

where x(t) is a dynamic variable, $\eta(t)$ is parametric white noise with mean $\overline{\eta}$ and intensity $D_\eta$, and $\xi(t)$ is additive white noise with mean zero and intensity $D_\xi$. It has been shown [Deutsch, 1994; Venkataramani, et al, 1996; Nakao, 1998; Sato, et al, 2000a; Takayasu, et al, 1997] that the stationary probability density $p(x)$ is

$$p(x) \approx \left(D_\xi + D_\eta x^2\right)^{-\frac{1}{2}}$$  \hspace{1cm} (34)

and that the tails for large x exhibit power laws of the form

$$p(x) \approx |x|^{-\beta-1}$$  \hspace{1cm} (35)
where \( \beta = -\eta / D_\eta \). There are two important observations: 1) the power law exponent depends only on the statistical properties of the parametric noise, in particular its mean and intensity; and 2) the power laws can be observed for a completely linear dynamical system.

\[
\eta \propto -\eta / D_\eta
\]

![Analog electrical circuit](image)

Figure 18: a) Analog electrical circuit for (37). b) Example of output of this circuit. Note that fluctuations whose amplitude is much larger than “on average” are seen infrequently. Figure adapted from Sato, et al [2000b].

The predictions given by (34)-(35) have been verified in a simple analog electronic circuit that contains both negative and positive feedback (Figure 18a) and which is described by the differential equation [Sato, et al, 2000b]

\[
\frac{dv(t)}{dt} = \left( \frac{1}{R_1 C} + \frac{k}{R_2 C} \right) \eta(t) v(t) + \xi(t)
\]  

(36)

Figure 18b demonstrates that the time series exhibit intermittency, i.e. relatively long intervals of lower amplitude fluctuations are interrupted intermittently by brief high amplitude fluctuations. The brief high amplitude fluctuations are referred to as “bursts” and the intervals of low amplitude fluctuations as the “laminar phases”. These observations are particularly intriguing given the widespread occurrence of recurrent inhibitory loops (mixed feedback) in the nervous system [Mackey and an der Heiden, 1984; Foss and Milton, 2000].

Current interest has focused on the kind of intermittency shown in Figure 18b which is referred to as on-off intermittency\(^9\). On-off intermittency arises

\[^9\] Do not confuse this exponent with the -3/2 exponent for the laminar phases of on-off with the exponent in (35).
because of the stochastic or chaotic forcing of a parameter back and forth across a stability boundary. This type of intermittent behavior is distinct from the intermittency observed in deterministic dynamical systems such as Pomeau-Manneville type-I, -II, and –III intermittency and crisis-induced intermittency. These types of intermittency occur for fixed values of the parameters while in on-off intermittency the bifurcation parameter is itself a dynamical variable. The implication of the detection of on-off intermittency experimentally is that it implies that the underlying dynamical system must be tuned in parameter space to be close to a stability boundary. How close? This depends on the intensity of the parametric noise.

A particularly insightful discussion into the effects of parametric noise on discrete time dynamical systems was developed by Heagy, Platt and Hammel [1994]. Here we briefly outline their presentation. The interested student who wishes more details should consult the original source. We then discuss on-off intermittency in an experimental system, namely, stick balancing at the fingertip, and then briefly discuss the rather surprising implications for the neural control of movement.

a). On-off intermittency in the quadratic map

Consider a quadratic map with parametric noise, i.e.

\[ x_{t+1} = \lambda(\xi_t) x_t (1 - x_t) \]  \hspace{1cm} (37)

where

\[ \lambda(\xi_t) = ay(\xi_t) \]  \hspace{1cm} (38)

where a is a constant and \( y(\xi_t) \) is a random variable drawn from a uniform density on the interval [0,1]. This map produces intermittency for an appropriate value of a (Figure 19). Numerical experiments indicate that the longer the laminar phases, the smaller the magnitude of the fluctuations. This observation was interpreted as implying that the dynamics are almost completely determined by the linear part of the map. The nonlinear terms serve only to bound or re-inject the dynamics back toward small values of \( x \). In other words the nonlinearities are essential for sustaining the intermittent behavior, but not for initiating the bursts.

If we consider only a linear map,

\[ x_{t+1} = \lambda(\xi_t) x_t, \]  \hspace{1cm} (39)
then we know that its long time behavior is given by

\[ x_n = \prod_{j=0}^{n-1} \lambda_j x_0 \]  

(40)

where \( x_0 \) is the value of \( x \) at \( t = 0 \). Consequently the long term behavior is determined by the asymptotic behavior of the random product

\[ \Psi_n = \prod_{j=0}^{n-1} \lambda_j \]  

(41)

Taking logarithms and making use of the law of large numbers\(^{10}\) we obtain

\[ \ln \Psi_n = \sum_{j=0}^{n-1} \ln \lambda_j \approx n \langle \ln \lambda \rangle \]  

(42)

Since we have assumed that the density of the distribution for the parametric noise is uniform we can readily calculated the statistical average as

\[ \langle \ln \lambda \rangle = \int_0^a p_\nu (\lambda) \ln \lambda d\lambda = \int_0^a \ln \lambda d\lambda = \ln a - 1 \]

(43)

Thus we see that the asymptotic solution of (39) is

\(^{10}\) The law of large numbers states that the arithmetic average of a very large sum of independent observations of a random variable \( x(s) \) is equal to the mean of \( x(s) \). It is a corollary of the Central Limit Theorem (Chung, 1975).
\[ x_n \approx e^{n \langle \ln \lambda \rangle} x_0 = \left( \frac{a}{e} \right)^n x_0 \]  \hspace{1cm} (44)

It follows that the condition for intermittent behavior is \( \langle \ln \lambda \rangle = 0 \). When \( \langle \ln \lambda \rangle > 0 \), the fixed-point \( x = 0 \) is, on average, exponential unstable. This instability is the source of the intermittent bursts. However, we must remember that this is just a local result. Thus this instability does not preclude the occurrence of long orbit segments in the neighborhood of \( x = 0 \), i.e. the laminar phases. The critical value of \( a \) for the onset of intermittency is

\[ a \approx 2.71828 \ldots \]  \hspace{1cm} (45)

b). \( \frac{3}{2} \) – power laws

Taking the logarithm of (39) we obtain

\[ y_{n+1} = \hat{\lambda}_n + y_n \]  \hspace{1cm} (46)

where \( y = \ln(x) \) and \( \hat{\lambda} = \ln(\lambda) \). A laminar phase of length \( n \) is defined by

\[ [y_1 \leq y_0, y_2 \leq y_0, \ldots, y_n \leq y_0, y_{n+1} > y_0] \]  \hspace{1cm} (47)

Since (46) is translationally invariant we can define without loss of generality the threshold as \( y_0 = 0 \). The probability that a laminar phase has exactly length \( n \) is given by the conditional probability

\[ \Lambda_n = \text{Prob} \left[ \bigcap_{j=1}^{n} y_j \leq 0 \bigcap y_{n+1} > 0 \bigcap y_1 \leq 0 \right] \]  \hspace{1cm} (48)

The laminar phase of length \( n \) corresponds to an event for which the product in (41) remains less than or equal to zero for exactly \( n \) iterations and becomes greater than one on the \( n+1 \)st iteration.

Using the decomposition rule

\[ \text{Prob}(A | B) = \frac{\text{Prob}(A \cap B)}{\text{Prob}(B)} \]

we can rewrite (48) as
If we define

$$g_n = \text{Prob}\left(\bigcap_{j=1}^{n} s_j \leq 0\right)$$

Then we can rewrite (49) as

$$\Lambda_n = \frac{g_n - g_{n+1}}{\text{Prob}(y_1 \leq 0)}$$

(50)

Since the computation of $g_n$ is an example of a recurring event, we can use a generating function\textsuperscript{11}, $G(t)$, to estimate the probability $g_n$

$$G(t) = \sum_{n=0}^{\infty} g_n t^n$$

(51)

It is shown in Feller [1971] that

$$\ln[G(t)] = \sum_{n=1}^{\infty} \frac{t^n}{n} \text{Prob}(y_n \leq 0)$$

(52)

Since $\lambda$ is chosen from a uniform distribution we know that the probability density is symmetric about the origin (zero in this case). Thus we have

$$\text{Prob}(y_n \leq 0) = \frac{1}{2}$$

(53)

\textsuperscript{11} Generating functions frequently arise in the description of recurrent events, i.e. stochastic processes that are generated by a repetitive series of experimental trials such as coin flipping [Bailey, 1963]. If we have a sequence of real numbers $a_0, a_1, a_2, \ldots$, then we can write

$$A(x) = a_0 + a_1 x + a_2 x^2 + \cdots = \sum_{j=0}^{\infty} a_j x^j$$

where $x$ is a dummy variable. If this series converges in some interval $[-x_0, x_0]$, then the function $A(x)$ is called the generating function of the sequence $\{x_j\}$. The generating function can be regarded as a transformation carrying the sequence $\{x_j\}$ into the function $A(x)$. 

33
and (52) becomes

\[
\ln[G(t)] = \frac{1}{2} \sum_{n=1}^{\infty} \frac{t^n}{n} = \ln \left[ \frac{1}{\sqrt{1-t}} \right]
\]

or

\[
G(t) = \frac{1}{\sqrt{1-t}} \tag{54}
\]

The coefficients in (52) can be evaluated from

\[
g_n = \frac{1}{n!} \frac{\partial^n G(t)}{\partial t^n} \bigg|_{t=0} = \frac{(2n)!}{2^{2n} (n!)^2}
\]

The numerator of (50) becomes

\[
g_n - g_{n+1} = \frac{g_n}{2(n+1)} \sim \frac{1}{2\sqrt{\pi}} n^{-3/2} \tag{55}
\]

where the last relation follows from Stirling’s approximation. Thus we see that at onset the laminar phases, \( \Lambda_n \), decay as a power law with exponent -3/2. This result has been extended to include all probability densities for the distribution \( \hat{\lambda} \) that have zero mean and finite variance [Heagy, et al, 1994].

c). Stick balancing at the fingertip

The study of stick balancing at the fingertip is important for understanding the neural control of balance [Cabrera and Milton, 2002; Foo, et al, 2000; Mehta and Schaal, 2002; Treffner and Kelso, 1999] and the development of expertise [Cabrera and Milton, 2004b; Cabrera, et al, 2004; Mah and Mussa-Ivaldi, 2003a,b; Milton, et al, 2004]. This task is particularly relevant to our discussion of the effects of parametric noise since it is well known that when the pivot point of an inverted pendulum moves, either periodically [Bogdanoff, 1962; Bogdanoff and Citron, 1964] or noisily [Landa, et al, 1998], the perturbations enter the equations through their effects on a parameter. There are also a number of

\[12\] This result is obtained by differentiating the sum on the right hand side of (53), then summing the resultant geometric series, and integrating.

\[13\] Stirling’s approximation is \( n! \sim \sqrt{2\pi n} e^{-n} n^n \).

\[14\] For example, if the pivot point of an inverted pendulum is made to vibrate up and down then the effect of gravity is lessened by an amount \( g - \frac{d^2 h}{dt^2} \), i.e.
advantages offered by the study of stick balancing at the fingertip as an experimental paradigm: 1) the movements can be measured with high precision using 3-D motion analysis techniques [Cabrera and Milton, 2002, 2004b]; 2) mathematical models for the control of inverted pendulums with time-delayed feedback have been extensively studied [Landray, et al, 2004; Stépán, 1989; Stépán and Kollár, 2000]; 3) virtual stick balancing tasks that involve a human interacting with a computer have been developed which enable critical parameters to be precisely manipulated [Bormann, et al, 2004; Cabrera, et al, 2004; Mah and Mussa-Ivaldi, 2003b; Mehta and Schaal, 2002]; 4) skill level can be objectively determined by, for example, measuring the fraction of sticks that can be balanced longer than a certain time [Cabrera and Milton, 2004a,b; Cabrera, et al, 2004]; and 5) skill level can be significantly increased with just a few days of intensive practice [Cabrera and Milton, 2004b; Cabrera and Milton, 2003, 2004].

The importance of the time delay for stick balancing is demonstrated by the observation that longer sticks are easier to balance than shorter ones: once the stick becomes sufficiently long its rate of movement becomes slow relative to the time required by the nervous system to make corrective movements. Indeed mathematical models for the stabilization of an inverted pendulum with time delayed feedback suggest that in order to balance a stick of length, \( \ell \), the time delays must be less than a critical value proportional to \( \sqrt{\ell} \) [Stépán, 1989; Stépán and Kollár, 2000].

The fluctuations in the controlled variable, \( \theta \), exhibit on-off intermittency [Cabrera and Milton, 2002]. This observation is supported by the presence of a \(-1/2\)-power law [Venkataramani, et al, 1996] in the power spectrum of the fluctuations in \( \theta \) (Fig. 19a) and a \(-3/2\) power law for the laminar phases (Fig. 19b). The observation of on-off intermittency suggests that the neural control mechanism for stick balancing is tuned at the edge of stability: this phenomenon arises because of the stochastic or chaotic forcing of an important parameter across a stability boundary [Cabrera and Milton, 2002; Heagy, et al, 1994; Platt, et al, 1993]; the proximity to the stability boundary depending on the intensity of the noise.

The observations in Fig. 19 can be reproduced by the equation of motion for an inverted pendulum stabilized by time-delayed feedback with a noisy gain [Cabrera and Milton, 2002], i.e.

\[
\frac{d^2\theta}{dt^2} + \left( \frac{g}{\ell} \frac{d^2h}{dt^2} \right) \sin \theta = 0
\]

where \( d^2h/dt^2 \) denotes the downward acceleration of the pivot and \( g \) the acceleration due to gravity [Acheson, 1997]. This is why we feel we feel less heavy than usual if we are in an elevator which is accelerating downwards!
\[
\frac{d\theta^2(t)}{dt^2} = -e\frac{d\theta(t)}{dt} + q\sin\theta - (r_0 + \xi(t))\theta(t - \tau)
\]  
(56)

where \( e, q, r_0 \) are constants and \( \xi(t) \) is Gaussian-distributed, white noise.

Although a feedback term that does not contain a velocity term might initially concern investigators who study motor control, it should be noted that a delayed controller shares features with derivative feedback in modifying the behavior of the system\(^{15}\). In contrast to an inverted pendulum, an inverted pendulum with delayed feedback has a stable upright position [Cabrera and Milton, 2002; Landray, et al, 2004; Stépán, 1989; Stépán and Kollár, 2000]. The essential condition for (56) to reproduce the observations in Fig. 19 is that the parameters \( r_0 \) and \( q \) be chosen sufficiently close to the appropriate stability boundary [Cabrera and Milton, 2002].

The observation that on-off intermittency arises in stick balancing at the fingertip has several implications concerning how the nervous system might

\(^{15}\) From the definition of a derivative we have \( \theta(t - \tau) = \theta(t) - \tau \frac{d\theta(t)}{dt} \).
control this task. First, on-off intermittency provides a mechanism that enables corrective movements to be made on time intervals shorter than the neural delay (> 98% of the $\delta t$ in Figure 20 are shorter than the delay - see figure legend and [Cabrera and Milton, 2002]). Second, on-off intermittency provides a mechanism of non-predictive control for stick balancing, i.e. the fluctuations in $\theta$ resemble a random walk for which the mean value of $\theta$ is approximately zero, i.e. the upright position has been statistically stabilized [Cabrera and Milton, 2002, 2003, 2004a,b]. The fact that the movements of the fingertip during stick balancing can be described as a Lévy flight with $\alpha \approx 0.9$ emphasizes the non-predictive aspects of the control strategy [Cabrera and Milton, 2004b]. Finally, the existence of a non-predictive control mechanism for stick balancing reduces the demands on intentionally directed control mechanisms and thus “frees up” the conscious and attentional mechanisms of the brain for other tasks [Cabrera and Milton, 2004a,b; Cabrera, et al, 2004; Milton, et al, 2004].

IV. Concluding Remarks:

It must be remembered that comparing the prediction of a model to experimental observation is an essential component of the scientific method. Such comparisons are not particularly useful unless the effects of noise on the predictions of the model are carefully investigated. Moreover, the fact that measurements of the statistical properties of the fluctuations can provide important non-invasive insights into the structure of the underlying dynamical system is likely to have increasing important as regulatory burden placed on human subject research increase, particularly in the study of normal physiology.

Acknowledgements

The author thanks John D. Hunter for proof reading this manuscript and helpful comments. Research was supported by grants from the National Institutes of Mental Health and the Brain Research Foundation.

References:


Press, Toronto.
Stanley HE, Amaral LAN, Andrade JS, Buldyrev SV, Havlin S, Makse HA, Peng C-K, Suki B and
Stanley HE, Amaral LAN, Buldyrev SV, Havlin S, Leschhorn H, Maass P, Sallinger MA and
804-806.
Stark L, Campbell FW and Atwood J (1958). Pupillary unrest: An example of noise in a biological
stabilization, Ecol. Psychol. 11: 103-137.
Venkataramani SC, Antonsen, Jr TM, Ott E and Sommerer JC (1996). On-off intermittency:
Viswanathan GM, Afanazyev V, Buldyrev SV, Murphy EJ and Stanley HE (1996). Lévy flight
and Applications (Barndorff-Nielsen OE, Mikosch T and Resnick I, eds). Birkhäuser,
Boston, pp. 241-265.
Weron R (2001). Levy-stable distributions revisited: Tail index > 2 does not exclude the Levy-