FOURIER TECHNIQUES IN X-RAY TIMING

M. van der Klis[1]

[1] Astronomical Institute "Anton Pannekoek" of the University of Amsterdam
and Center for High Energy Astrophysics,
Kruislaan 403, 1098 SJ Amsterdam, The Netherlands

This article first appeared in the Proceedings of the NATO Advanced Study Institute
Timing Neutron Stars, Çeşme, Turkey, 4-15 April 1988,
H. Ögelman and E.P.J van den Heuvel (eds.), NATO ASI Series C, Vol. 262, p. 27–70.
Second, slightly updated version, May 1994.

FOURIER TECHNIQUES IN X-RAY TIMING

M. van der Klis,
EXOSAT Observatory,
Space Science Department of ESA,
ESTEC, Postbus 299,
2200 AG Noordwijk,
The Netherlands

ABSTRACT. Basic principles of Fourier techniques often used in X-ray time series analysis are reviewed. The relation between the discrete Fourier transform and the continuous Fourier transform is discussed to introduce the concepts of windowing and aliasing. The relation is derived between the power spectrum and the signal variance, including corrections for binning and dead time. The statistical properties of a noise power spectrum are discussed and related to the problems of detection (and setting upper limits) of broad and narrow features in the power spectrum. A "dependent trial" method is discussed to search power spectra consistently for many different types of signal simultaneously. Methods are compared to detect a sinusoidal signal, a case that is relevant in the context of X-ray pulsars.

1. INTRODUCTION

Fourier techniques are widely used in science and engineering, but problems of terminology and differing conventions hamper the flow of information between the various branches. Even within the field of neutron-star timing, radio-, X-ray and high-energy gamma-ray astronomers sometimes have difficulties to compare the techniques they routinely apply.

In the present paper an attempt will be made to explore some of the techniques that are commonly used in timing studies of neutron stars, and that refer to the detection of signals against a background of noise, in the language of the X-ray astronomer.

The regime that I will nearly exclusively be referring to is that of equidistantly binned timing data, the background noise of which is dominated by counting statistics. If there are gaps in the data, they are far apart, and the data are not "sparse"

in the sense that nearly all time bins are empty. This kind of data is eminently suited to analysis with Fast Fourier Transform (FFT) techniques, and the discussed methods will all be based on these techniques. Emphasis will be on the statistics of the detection of weak signals rather than on the characterization of signal shapes, with a special discussion of the basic problem of detecting a strictly periodic signal with a sinusoidal shape. While the discussion is specially geared towards photon counting data such as produced by, for example, an X-ray proportional counter, many of the techniques discussed are also applicable in other regimes.

Section 2 contains a general introduction to the Fourier transform and introduces the power spectrum. Section 3 deals with the problem of detecting a signal in the presence of noise. In Section 4 it is discussed how to quantify the power of the signal and how to estimate its variance using the power spectrum. Section 5 discusses how to search a power spectrum by making use of the basic properties of power spectral statistics only, using "independent trials". At the end of this section there is a summary in "recipe" form of how to simply search a power spectrum for a weak signal. In Section 6, a detailed discussion is given of the specific case of detecting a sinusoidal signal. The subject of Section 7, finally, is a method of searching power spectra for various types of signal simultaneously. The methods discussed in Sections 6 and 7 have in common that because higher demands are made on the tests performed on the power spectrum than in Section 5, the test statistics are no longer simple (in particular, "dependent trials" are considered) and have to be evaluated by simulations.

The present exposition owes much to the paper by Leahy *et al.* (1983). Some of the material discussed is also contained in Chapter 2 of the review about quasi-periodic oscillations by Lewin, van Paradijs and van der Klis (1988), hereafter Paper 1.

2. THE FOURIER TRANSFORM

2.1. Introduction

In this section (2.1) the Fourier transform is introduced in very general terms. We do not yet worry about summation indices and the like; such details are filled in in the following sections.

A Fourier transform gives a decomposition of a signal, say, $x(t)$, into sine waves. At any given frequency ω, one can find a set of values (a, ϕ) or (A, B) such that the sinusoid $a\cos(\omega t - \phi) = A\cos\omega t + B\sin\omega t$ best fits the data $x(t)$[1]. Do this for a sufficient number of different frequencies ω_j, then the signal can be written as

[1] (a, ϕ) and (A, B) are, of course, related by $a = \sqrt{A^2 + B^2}$ and $\tan\phi = B/A$

$$x(t) = \frac{1}{N}\sum_j a_j \cos(\omega_j t - \phi_j) = \frac{1}{N}\sum_j (A_j \cos\omega_j t + B_j \sin\omega_j t)\,. \tag{2.1}$$

The Fourier coefficients A_j and B_j can be straightforwardly computed as

$$\begin{aligned} A_j &= \sum_k x_k \cos\omega_j t_k \\ B_j &= \sum_k x_k \sin\omega_j t_k, \end{aligned} \tag{2.2}$$

where $x_k = x(t_k)$. It can be seen from Eq. 2.2 that A_j and B_j are simply the **correlation** of the signal x_k with a sine or cosine wave of frequency ω_j: if there is a good correlation then the corresponding Fourier coefficient is high and gives a large contribution to the sum in Eq. 2.1 which reconstructs the signal out of sine waves.

For easier handling of the two numbers ((A, B) or (a, ϕ)) which one obtains at each frequency, it is possible to represent the Fourier transform in terms of complex numbers:

$$a_j = \sum_k x_k e^{i\omega_j t_k} \tag{2.3a}$$

$$x_k = \frac{1}{N}\sum_j a_j e^{-i\omega_j t_k}, \tag{2.3b}$$

where $i^2 = -1$. The complex numbers a_j are called the **(complex) Fourier amplitudes**; together they form the Fourier transform of the x_k. Inversely, the x_k form the **inverse Fourier transform** of the a_j. Writing a_j as $|a_j|e^{i\phi_j}$, we see (Eq. 2.3*b*) that the signal x_k is now decomposed into functions of the form $a_j e^{-i\omega_j t_k} = |a_j|e^{-i(\omega_j t_k - \phi_j)} = |a_j|\big(\cos(\omega_j t_k - \phi_j) - i\sin(\omega_j t_k - \phi_j)\big)$, having a non-zero imaginary component. This is nothing to worry about: in this representation both positive and negative frequencies are considered, with $\omega_{-j} = -\omega_j$, and if the x_k are real numbers then one sees from Eq. 2.3*a* that $a_{-j} = a_j^*$ (the asterisk indicating the complex conjugate), so that the imaginary terms at j and $-j$ (*i.e.*, at ω_j and ω_{-j}) cancel out and the end result in the summation (2.3*b*) is $(2/N)|a_j|\cos(\omega_j t_k - \phi_j)$, strictly real.

2.2. The Discrete Fourier Transform

We now get a bit more specific and define our signal as a series of N numbers x_k ($k = 0, \ldots, N-1$); in the applications discussed in this article, x_k will always stand for the number of photons detected in bin k. The discrete Fourier transform a_j ($j = -N/2, \ldots, N/2-1$) decomposes this signal into N sine waves. The following expressions describe the signal-transform pair:

$$a_j = \sum_{k=0}^{N-1} x_k e^{2\pi i j k/N} \qquad j = -\frac{N}{2}, \ldots, \frac{N}{2} - 1 \tag{2.4a}$$

$$x_k = \frac{1}{N} \sum_{j=-N/2}^{N/2-1} a_j e^{-2\pi i j k/N} \qquad k = 0, \ldots, N-1. \tag{2.4b}$$

If, as before, the signal is an equidistant time series of length T, so that x_k refers to a time $t_k \equiv kT/N$, then the transform is an equidistant "frequency series", and a_j refers to a frequency $\omega_j \equiv 2\pi\nu_j = 2\pi j/T$. The time step is $\delta t = T/N$; the frequency step is $\delta\nu = 1/T$, and substituting $2\pi jk/N = \omega_j t_k$, we find back Eq. 2.3.

It is a matter of taste where one puts the factor $1/N$ in Eq. 2.4; definitions where this factor appears in Eq. 2.4*a*, or where both sums are preceded by a factor $1/\sqrt{N}$ are also possible and do, in fact, occur in literature.

Note that the number (N) of input values x_k equals the number of output values a_j; if the x_k are uncorrelated, then the a_j are as well. The discrete Fourier transform gives a complete description of the discrete signal; the highest frequency needed for this complete description is $\nu_{N/2} = \frac{1}{2}N/T$. This frequency, equal to half the "sampling" frequency defined by the spacing of the x_k, is called the **Nyquist frequency**. An oscillation at $\nu_{N/2}$ corresponds to an alternating "up-down" signal in the x_k. Note that $a_{-N/2} = \sum_k x_k e^{-\pi i k} = \sum_k x_k (-1)^k = a_{N/2}$; it does not matter whether one puts the Nyquist frequency at the positive end or the negative end of the Fourier transform. At zero frequency, the result of Eq. 2.4*a* is just the total number of photons detected; $a_0 = \sum_k x_k \equiv N_{\rm ph}$.

2.2.1. The Fast Fourier Transform. The fast Fourier transform (FFT) is a computer algorithm to efficiently compute the discrete Fourier transform. Often, but not always, the data is constrained by these algorithms to have a number of bins N equal to a power of 2. See, *e.g.*, Press *et al.* (1986) for an exposition of the functioning and sample computer codes of FFT algorithms.

2.3. The Continuous Fourier Transform

The continuous Fourier transform decomposes an infinitely extended continuous function $x(t)$ $(-\infty < t < \infty)$ into an infinite number of sine waves:

$$a(\nu) = \int_{-\infty}^{\infty} x(t) e^{2\pi\nu i t}\, dt \qquad -\infty < \nu < \infty \tag{2.5a}$$

$$x(t) = \int_{-\infty}^{\infty} a(\nu) e^{-2\pi\nu i t}\, d\nu \qquad -\infty < t < \infty. \tag{2.5b}$$

When doing analytical calculations, the continuous Fourier transform has a number of pleasing properties (for example, the continuous Fourier transform of a sine wave is a delta function; this is not in general true for the discrete Fourier transform, see Fig. 6.1*a*). Therefore, theoretical predictions of the shape of the Fourier transform of a signal are usually in terms of the continuous Fourier transform.

Unfortunately, in the real world the data are not infinitely extended nor continuous, and one might well ask what is the relation of Eq. 2.5 with the discrete Fourier transform of a discretely sampled **section** of $x(t)$. This question will be adressed in Section 2.5.

2.4. The Power Spectrum

A result known as Parseval's theorem states:

$$\sum_{k=0}^{N-1} |x_k|^2 = \frac{1}{N} \sum_{j=-N/2}^{N/2-1} |a_j|^2. \tag{2.6}$$

This implies that there is a relation between the summed squared modulus of the Fourier amplitudes and the total variance of the data: $\mathrm{Var}(x_k) \equiv \sum_k (x_k - \overline{x})^2 = \sum_k x_k^2 - \frac{1}{N}\left(\sum_k x_k\right)^2 = \frac{1}{N}\sum_j |a_j|^2 - \frac{1}{N}a_0^2$, so that

$$\mathrm{Var}(x_k) = \frac{1}{N} \sum_{\substack{j=-N/2 \\ j\neq 0}}^{N/2-1} |a_j|^2. \tag{2.7}$$

Adopting the normalization used by Leahy *et al.* (1983), we will define the **power spectrum** as

$$P_j \equiv \frac{2}{N_{ph}} |a_j|^2 \qquad j = 0, \ldots, \frac{N}{2}, \tag{2.8}$$

where N_{ph} is again the total number of photons $\sum x_k = a_0$ and a_j is given by Eq. 2.4*a*. Using once more the result that for real data $|a_j| = |a_{-j}|$ and taking

account of the fact that the term at the Nyquist frequency occurs only once in Eq. 2.7, we find for the expression for the total variance in terms of the P_j:

$$\mathrm{Var}(x_k) = \frac{N_{ph}}{N}\left(\sum_{j=1}^{N/2-1} P_j + \tfrac{1}{2}P_{N/2}\right). \tag{2.9}$$

Note the differences in the indexing of a_j and P_j. Computer implementations of the FFT usually employ a storage scheme that is different again (Fig. 2.1).

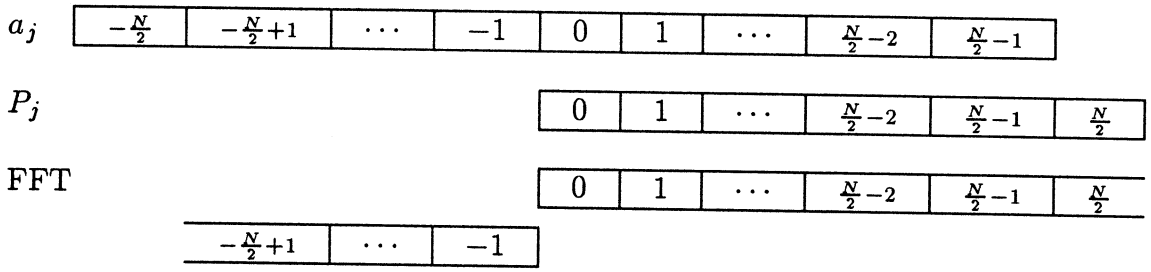

Fig. 2.1. *Storage schemes. The FFT scheme may differ between implementations — in this example bin* $-\frac{N}{2}+1$ *follows bin* $\frac{N}{2}$.

Often the variance is expressed in terms of the fractional root-mean-square (rms) variation in the x_k:

$$r \equiv \frac{\sqrt{\frac{1}{N}\mathrm{Var}(x_k)}}{\overline{x}} = \sqrt{\frac{\sum_{j=1}^{N/2-1} P_j + \frac{1}{2}P_{N/2}}{N_{\mathrm{ph}}}}. \tag{2.10}$$

Sometimes r is expressed in terms of a percentage, and is then also called the "percentage rms variation". A sinusoidal signal at the Fourier frequency ν_j (see Section 4 for the more general case) $x_k = A\sin(2\pi\nu_j t_k)$ will cause a spike at ν_j in the power spectrum with

$$P_{j,\mathrm{sine}} = \frac{1}{2}\frac{N^2}{N_{\mathrm{ph}}}A^2. \tag{2.11}$$

The reason for choosing this apparently rather awkward normalization for the powers lies in the statistical properties of the noise power spectrum, to be described in Section 3.

If the data consist of the sum of a number of independent signals: $x_k \equiv y_k + z_k$, then the so-called superposition theorem ("the transform of the sum is the sum of the transforms") says that if b_j and c_j are the Fourier transforms of y_k and z_k,

respectively, then the Fourier transform of x_k is $a_j = b_j + c_j$. This means, that a similar superposition principle does **not** apply to power spectra:

$$|a_j|^2 = |b_j + c_j|^2 = |b_j|^2 + |c_j|^2 + \text{cross terms.} \tag{2.12}$$

However, if one of the two signals summed consists of random uncorrelated noise, then the cross-terms will tend to average out to zero.

2.5. The Relation between the Discrete and the Continuous Fourier Transform

The answer to the question posed in Section 2.3 about the relation between the discrete and the continuous Fourier transform can be obtained by making use of one of the powerful theorems of Fourier analysis, the convolution theorem. This theorem states, that the Fourier transform of the product of two functions is the **convolution** of the Fourier transforms of these two functions. So, if $a(\nu)$ is the continuous Fourier transform of $x(t)$ and $b(\nu)$ that of $y(t)$ then the continuous Fourier transform of $x(t)y(t)$ is $a(\nu) * b(\nu) \equiv \int_{-\infty}^{\infty} a(\nu')b(\nu - \nu')d\nu'$: "the transform of the product is the convolution of the transforms". The inverse is also true ("the transform of the convolution is the product of the transforms"), and in the case of the discrete Fourier transform analogous theorems apply.

Now suppose that $a(\nu)$ $(-\infty < \nu < \infty)$ is the continuous Fourier transform of the infinitely extended continuous function $x(t)$ $(-\infty < t < \infty)$. Suppose, furthermore, that x_k $(k = 0, \ldots, N-1)$ is a finite discrete time series defined as $x_k = x(t_k)$, where $t_k = kT/N$, *i.e.*, x_k is a discretely sampled section of $x(t)$. Then we see (Fig. 2.2*a*) that the relation of $x(t)$ with x_k is given by a **double multiplication**: $x(t)$ has been multiplied with a "window function"

$$w(t) = \begin{cases} 1, & 0 \le t < T \\ 0, & \text{otherwise,} \end{cases} \tag{2.13}$$

and with a "sampling function"

$$i(t) = \sum_{k=-\infty}^{\infty} \delta\left(t - \frac{kT}{N}\right), \tag{2.14}$$

where $\delta(t)$ is the Dirac delta function.

Consequently, the relation of $a(\nu)$ with a_j is given by a **double convolution**(Fig. 2.2*b*): $a(\nu)$ must be convolved with the Fourier transforms of both the window function and the sampling function.

Because (to be consistent with Eq. 2.4) we have chosen the window function to be asymmetric around $t = 0$, the "window transform" $W(\nu)$ turns out to be

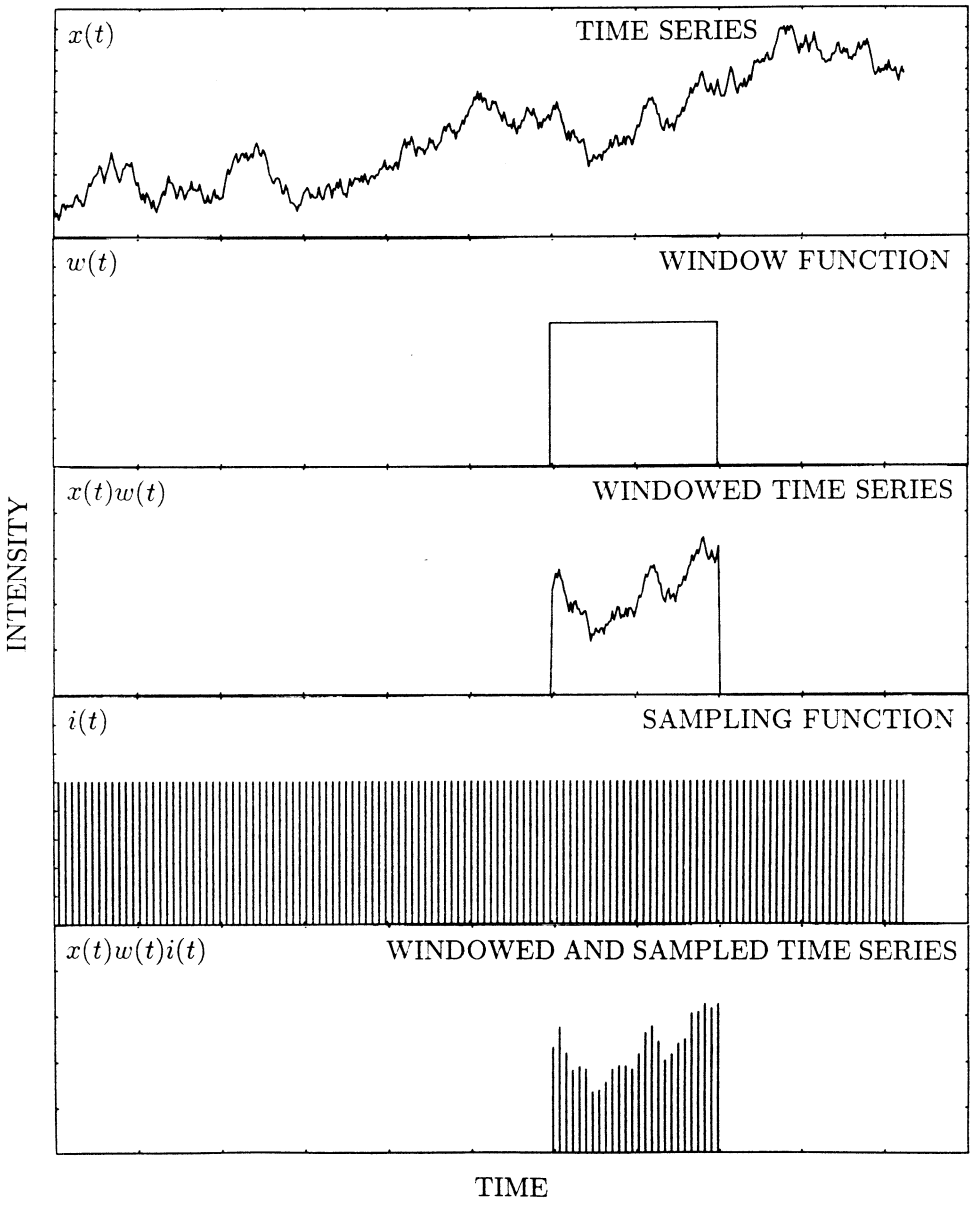

Fig. 2.2. *a) Obtaining the discrete time series x_k as a discretely sampled section of $x(t)$ involves a double multiplication.*

complex. To understand what is going on, it is sufficient to consider the power spectrum of $W(\nu)$:

$$|W(\nu)|^2 \equiv \left| \int_{-\infty}^{\infty} w(t) e^{2\pi\nu i t}\, dt \right|^2 = \left| \frac{\sin \pi\nu T}{\pi\nu} \right|^2 . \tag{2.15}$$

For a symmetric $w(t)$ we would have $W(\nu) = \sin(\pi\nu T)/\pi\nu$. The Fourier transform

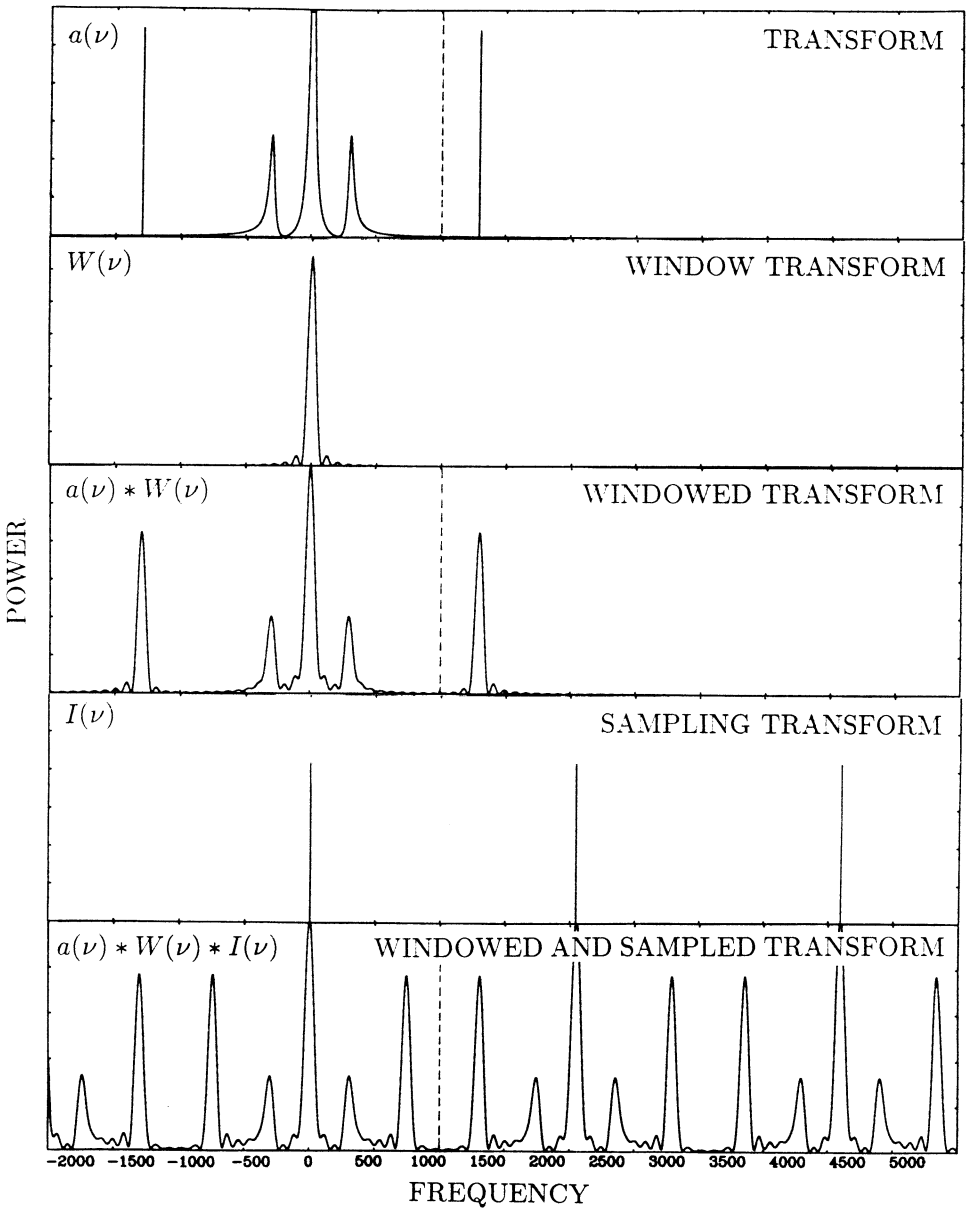

Fig. 2.2. *b) The discrete Fourier transform a_j of x_k is obtained out of the continuous Fourier transform $a(\nu)$ by a double convolution. The figure shows the power spectra corresponding to the various Fourier transforms. Vertical dashed lines indicate the Nyquist frequency.*

of an infinitely extended periodic series of delta functions such as the sampling function $i(t)$ is again an infinite periodic series of delta functions:

$$I(\nu) \equiv \int_{-\infty}^{\infty} i(t) e^{2\pi\nu i t}\, dt = \frac{N}{T} \sum_{\ell=-\infty}^{\infty} \delta\left(\nu - \ell \frac{N}{T}\right). \tag{2.16}$$

The functions $w(t)$ and $i(t)$ and the power spectra corresponding to $W(\nu)$ and $I(\nu)$ are plotted in Fig. 2.2.

The convolution of $a(\nu)$ with $W(\nu)$ causes all features in $a(\nu)$ to become wider; in particular, a delta function can be seen in Fig. 2.2 to change into shifted version of $W(\nu)$; a peak of finite width with side lobes. The convolution of an arbitrary function with a delta function at ν_0 is a shifted version of the original function: $f(\nu) * \delta(\nu - \nu_0) = f(\nu - \nu_0)$. Therefore, the convolution of $a(\nu)$ with $I(\nu)$, which is a series of delta functions with spacing N/T results in a convolved function $a(\nu) * I(\nu)$ that repeats every N/T frequency units.

For a real signal $x(t)$ we have, as before, $a(-\nu) = a(\nu)^*$, so that $|a(\nu)|^2 = |a(-\nu)|^2$: the power spectrum is symmetric around $\nu = 0$. The final result is that the power spectrum of the convolved function $|a(\nu) * I(\nu)|^2$ is reflected around the Nyquist frequency $\nu_{N/2} = \frac{1}{2}N/T$. This causes features with a frequency exceeding the Nyquist frequency by ν_x (so, located at $\nu = \nu_{N/2} + \nu_x$) to **also** appear at a frequency $\nu_{N/2} - \nu_x$, a phenomenon known as **aliasing**; the reflected feature is called the **alias** of the original one.

Using Eqs. 2.13 and 2.14 it is straightforward to show that the discrete Fourier amplitudes a_j are the values at the Fourier frequencies $\nu_j \equiv j/T$ of the windowed and aliased continuous Fourier transform $a_{WI}(\nu)$

$$a_{WI}(\nu) \equiv a(\nu) * W(\nu) * I(\nu) = \int_{-\infty}^{\infty} x(t)w(t)i(t)e^{2\pi i\nu t}\,dt$$
$$= \int_{-\infty}^{\infty} x(t) \sum_{k=0}^{N-1} \delta\left(t - \frac{kT}{N}\right) e^{2\pi i\nu t}\,dt = \sum_{k=0}^{N-1} x\left(\frac{kT}{N}\right) e^{2\pi i\nu kT/N},$$

so that $a_{WI}(j/T) = a_j$. Explicitly performing the convolution of $a(\nu)$ with $I(\nu)$ we finally have:

$$a_j = a_{WI}(j/T) = a_W(j/T) * I(j/T) = \frac{N}{T} \sum_{\ell=-\infty}^{\infty} a_W\left(\nu_j - \ell\frac{N}{T}\right), \qquad (2.17)$$

where we have used Eq. 2.16 and where $\nu_j = j/T$ and $a_W(\nu) \equiv a(\nu) * W(\nu)$.

To summarize, the transition from the continuous Fourier transform to the discrete Fourier transform involves two operations: windowing, a convolution with the function $W(\nu)$, which is essentially a peak with a width $\delta\nu = 1/T$ plus sidelobes, and aliasing, a reflection of features above the Nyquist frequency back into the range $(0, \nu_{N/2})$. Windowing is caused by the finite extent, aliasing by the discrete sampling of the data.

In practice, aliasing is not so much of a problem as one might fear, as the data are not really discretely sampled at intervals $\delta t = T/N$, but rather binned into time bins with a width δt. This is equivalent to **convolving** the data with the "binning window"

$$b(t) = \begin{cases} N/T, & -\frac{T}{2N} < t < \frac{T}{2N} \\ 0, & \text{otherwise} \end{cases} \tag{2.18}$$

before the discrete sampling. Applying the "inverse" convolution theorem, we see that the effect of this on the Fourier transform will be that $a(\nu)$ is **multiplied** with the transform of $b(t)$:

$$B(\nu) = \frac{\sin \pi \nu T/N}{\pi \nu T/N}. \tag{2.19}$$

This function drops from a value of 1 at $\nu = 0$ to 0 at $\nu = N/T$; halfway, at the Nyquist frequency it has a value of $2/\pi$, so that the effect of this multiplication is a considerable repression of the high-frequency features that could be aliased back into the frequency range $(0, \nu_{N/2})$. This is understandable; the effect of the binning is nothing else than averaging the time series over the bin width T/N so that variations with a frequency close to N/T are largely averaged out.

The problems caused by the windowing can be more serious; the "leakage" caused by the finite width of the central peak of $W(\nu)$ and by its side lobes can strongly distort steep power spectra (they become less steep, *e.g.*, Deeter, 1983) and, as we will see later on, it can spread out delta functions over the entire power spectrum.

2.6. Literature

The "handy cookbook" of time series analysis has yet to be written. A good standard reference that covers a large amount of information but is not always easy to follow is Jenkins and Watts (1968). Much easier are the texts by Bloomfield (1976) and Bracewell (1965). A very clear exposition of some basic principles of Fourier analysis stressing intuition rather than mathematics can be found in Press *et al.* (1986).

3. POWER SPECTRAL STATISTICS

3.1. Introduction

The process of detecting something in a power spectrum against a background of noise has several steps. The first thing we need to know is the **probability distribution** of the "noise powers" $P_{j,\text{noise}}$ in a power spectrum of data consisting

only of noise. If one or more of the powers P_j in the observed power spectrum differ significantly from the values expected from noise, then we may conclude that we have detected a "source signal", which is the term I shall use to indicate intrinsic variability in the x_k other than due to background noise.

To quantify the power of the source signal, *i.e.*, to determine what the "signal powers" $P_{j,\mathrm{signal}}$ of the source signal would have been in the absence of noise (or to determine their upper limit), we must consider the **interaction between the noise and the signal powers.**

This quantitative knowledge about the $P_{j,\mathrm{signal}}$ can be directly converted into a statement about the variance (or the rms variation) of the source signal. To say something about other properties of the source signal we need to consider the **expected shape** of the signal power spectrum. The optimal way to detect a given signal will also depend on this expected shape.

In this section, we will consider the first of these steps, signal detection, and consequently we must consider the probability distribution of the noise powers. The problem of quantifying the signal power will be discussed in Section 4. For the interaction between the noise and signal powers, we will follow convention by making the following very simple assumption

$$P_j = P_{j,\mathrm{noise}} + P_{j,\mathrm{signal}}. \tag{3.1}$$

Note that this is an approximation; as we have seen (Section 2.4), if it would be true that $a_j = a_{j,\mathrm{noise}} + a_{j,\mathrm{signal}}$, if the noise is random and uncorrelated, and if many powers are averaged, then Eq. 3.1 is approximately valid. See the paper by Vaughan *et al.* (1994) for convenient recipes of how to deal with the case that the number of powers averaged is **not** large, so that the interaction of noise and signal vectors in the complex plane must be reckoned with (sensitivity and detection levels are not affected, but upper limits and derived signal strengths are).

3.2. The Probability Distribution of the Noise Powers

For a wide range of types of noise, the noise powers $P_{j,\mathrm{noise}}$ follow[2] the χ^2 (chi-squared) distribution with 2 degrees of freedom (dof). The proof of this χ^2 property of the noise powers proceeds approximately as follows (see *e.g.*, Jenkins and Watts, 1968):

The noise power $P_{j,\mathrm{noise}} = A^2_{j,\mathrm{noise}} + B^2_{j,\mathrm{noise}}$, where A_j and B_j are given by Eq. 2.2; A_j and B_j are both linear combinations of the x_k. Therefore, if the x_k follow the normal distribution, then the A_j and B_j do as well, so that P_j, by

[2] With the exception of the power at the Nyquist frequency which follows the χ^2 distribution with 1 dof.

definition, is distributed according to the χ^2 distribution with 2 dof. If the x_k follow some other probability distribution, for example the Poisson distribution, then it follows from the central limit theorem that for "certain" conditions on this other distribution the A_j and B_j will still be approximately normally distributed (for large N), so that the χ^2 property for the P_j still approximately holds.

While this "proof" may seem somewhat unsatisfactory, the conditions for the central limit theorem to be applicable not being spelled out, in practice one finds that noise powers are nearly always χ^2 distributed, not only for Poisson noise, but also for many other types of noise. We shall see examples of this later on in this section.

The normalization of the power spectrum defined by Eq. 2.8 is chosen such, that if the noise in the photon counting data x_k is pure Poissonian counting noise, then the distribution of the $P_{j,\mathrm{noise}}$ is exactly given by the χ^2 distribution with 2 dof, so that the probability to exceed a certain threshold power level $P_{\mathrm{threshold}}$ is given by

$$\mathrm{Prob}(P_{j,\mathrm{noise}} > P_{\mathrm{threshold}}) = Q(P_{\mathrm{threshold}}|2) \qquad (j = 1, N/2 - 1)\,, \tag{3.2}$$

where the integral probability of the χ^2 distribution is defined as

$$Q(\chi^2|\upsilon) \equiv \left[2^{\upsilon/2}\Gamma\left(\frac{\upsilon}{2}\right)\right]^{-1} \int_{\chi^2}^{\infty} t^{\frac{\upsilon}{2}-1} e^{-\frac{t}{2}}\, dt, \tag{3.3}$$

where υ is the number of dof.

Because the $P_{j,\mathrm{noise}}$ follow this distribution, the power spectrum is very noisy; the standard deviation of the noise powers is equal to their mean value: $\sigma_{P_j} = \langle P_j \rangle = 2$. This noisy character of the power spectrum can not be improved by increasing the length T of the data or taking a coarser time step δt; this just changes the number of powers.

Two more or less equivalent methods are often used to decrease the large variance of the $P_{j,\mathrm{noise}}$. One is to rebin the power spectrum, averaging W consecutive frequency bins; the other to divide the data up into M equal segments, transform these segments each individually and then average the resulting M power spectra, each normalized according to Eq. 2.8, where N_{ph} is now the number of photons **in one transform**. Both methods of power spectrum compression, of course, degrade the frequency resolution.

As the time required to calculate the Fourier transform of N data points using an FFT algorithm is proportional to $N \log N$, there is a computational advantage in the second method; the time saving factor is about $1 + \log M / \log N$. In many cases, considerable additional time savings result from the smaller array sizes that need to be handled by the computer. For a variable source, a further advantage of the

second method is that cutting up the data into smaller segments allows one to study variations in the power spectra as a function of, *e.g.*, source intensity by selectively averaging power spectra obtained within certain source intensity intervals and that it allows the construction of two-dimensional images showing the time evolution of the power spectrum. These techniques have proven particularly useful in the detection of transient QPO phenomena. The first method, on the other hand, has the advantage of producing a power spectrum that extends to lower frequencies (the lowest measurable frequency being $1/T$). It is possible to combine both methods; each power in the final spectrum will then be the average of MW original powers.

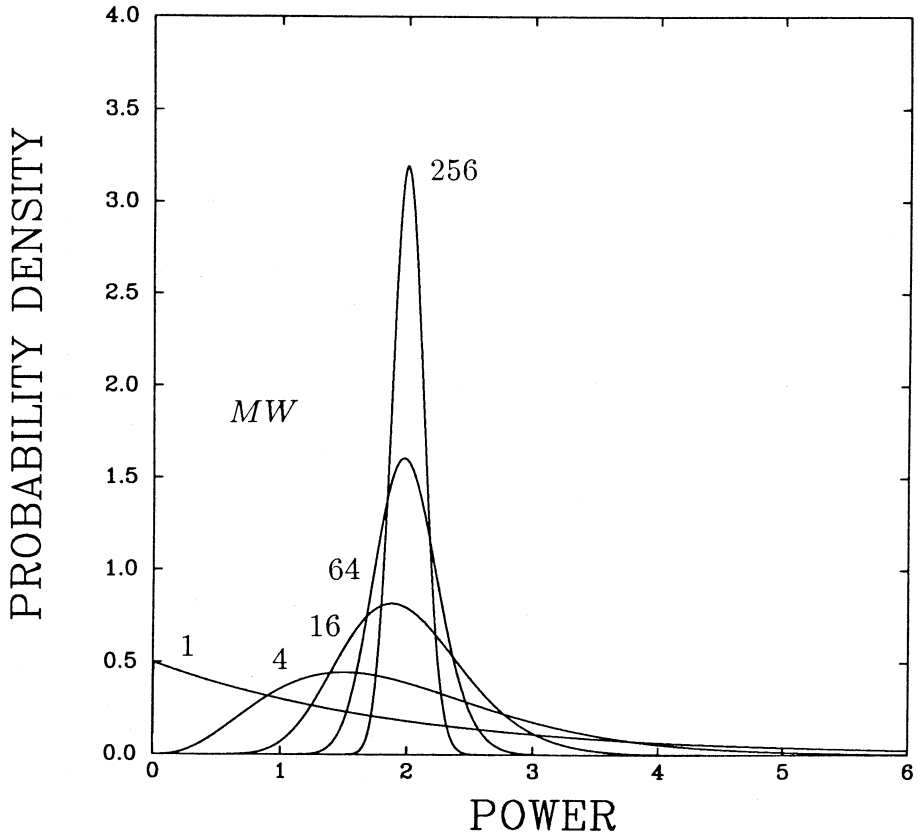

Fig. 3.1. *The probability distribution of average noise powers for different values of the number of powers MW averaged.*

Because of the additive properties of the χ^2 distribution, the sum of MW powers is distributed according to the χ^2 distribution with $2MW$ dof, so that the powers in the averaged spectrum will be distributed according to a χ^2 distribution with $2MW$ dof scaled by a factor $1/MW$. The mean of this distribution is 2, its variance is $4/MW$, and its standard deviation $2/\sqrt{MW}$ so that for large MW the

spectrum becomes much less noisy. The probability for a given power $P_{j,\text{noise}}$ in the averaged spectrum to exceed a $P_{\text{threshold}}$ is given by

$$\text{Prob}(P_{j,\text{noise}} > P_{\text{threshold}}) = Q(MWP_{\text{threshold}}|2MW), \tag{3.4}$$

where $Q(\chi^2|\upsilon)$ is again given by Eq. 3.3. For large MW, this distribution tends asymptotically to a normal distribution (see Fig. 3.1) with a mean of 2 and standard deviation $2/\sqrt{MW}$:

$$\lim_{MW\to\infty} \text{Prob}(P_{j,\text{noise}} > P_{\text{threshold}}) = Q_{\text{Gauss}}\left(\frac{P_{\text{threshold}} - 2}{2/\sqrt{MW}}\right), \tag{3.5}$$

where the integral probability of the normal distribution is

$$Q_{\text{Gauss}}(x) \equiv \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt. \tag{3.6}$$

So, a considerable simplification can be obtained by averaging large numbers of powers, empirically determining mean and standard deviation of the averaged power spectrum to account for non-Poissonian noise in the x_k (see Section 3.4), and then using Gaussian statistics. In the following, we will, unless otherwise stated, assume the more general case described by Eq. 3.4.

3.3. The Detection Level - the Number of Trials

Assuming the χ^2 property for the noise powers (Eq. 3.4), we can now determine how large a power must be to constitute a significant excess above the noise.

Define the $(1-\epsilon)$ confidence detection level P_{detect} as the power level that has only the small probability ϵ to be exceeded by a noise power. So, if there is a power P_j that exceeds P_{detect} then there is a large probability $(1-\epsilon)$ that P_j is not purely due to noise, but also contains signal power (Eq. 3.1).

A crucial consideration, occasionally overlooked, is the number of different P_j values, known as the **number of trials** N_{trial} that one wishes to compare with P_{detect}. N_{trial} can be equal to the total number of powers in the power spectrum, or less than that if only a certain frequency range in the spectrum is considered. The probability to exceed P_{detect} by noise should have the small value ϵ for all powers in the frequency range of interest **together**, so that the chance **per trial** should have the much smaller value of about[3] $\epsilon/N_{\text{trial}}$. So, the detection level P_{detect} is given by

[3] The exact expression can be obtained by setting the joint probability for N_{trial} values of P_j **not** to exceed P_{detect} equal to $1-\epsilon$, which gives a chance to exceed P_{detect} per trial of $1-(1-\epsilon)^{(1/N_{\text{trial}})}$, nearly equal to $\epsilon/N_{\text{trial}}$ for $\epsilon \ll 1$.

$$\frac{\epsilon}{N_{\rm trial}} = Q(MWP_{\rm detect}|2MW) \tag{3.7}$$

In Fig. 3.2, $P_{\rm detect}$ is plotted as a function of $N_{\rm trial}$ for various values of MW and for confidence levels of 90% ($\epsilon = 0.1$) and 99% ($\epsilon = 0.01$). Note that although $P_{\rm detect}$ increases with the number of trials $N_{\rm trial}$, the increase is relatively slow.

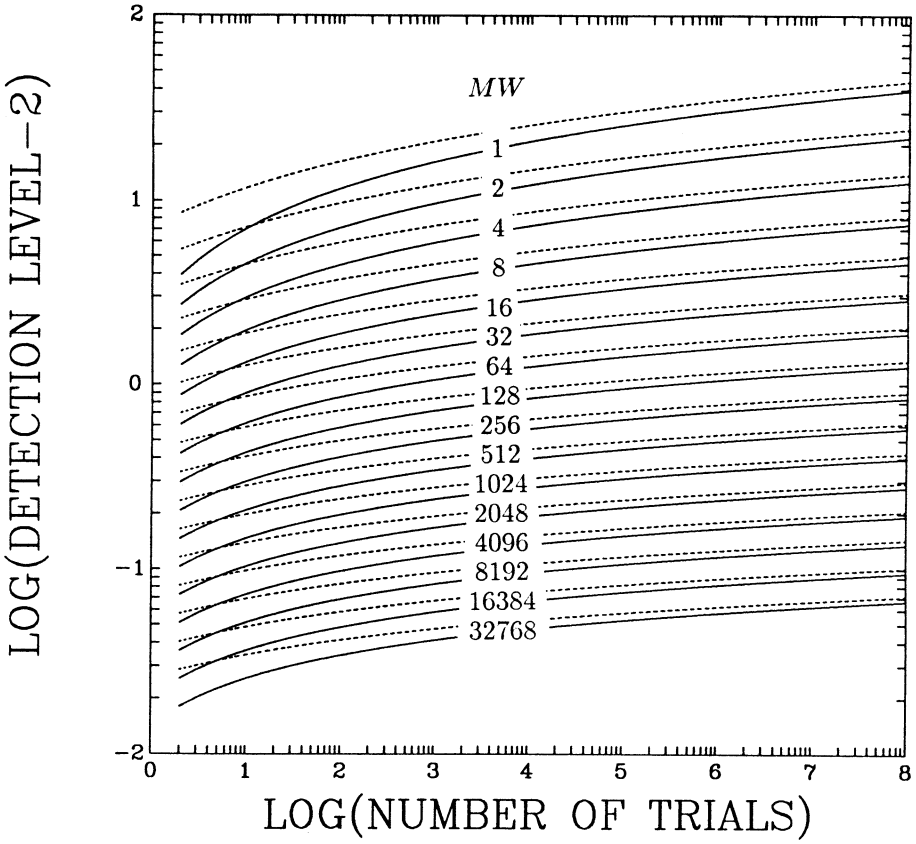

Fig. 3.2. *90% (drawn) and 99% (dashed) confidence detection levels (minus 2) as a function of the number of trials. The number of independent powers, MW, averaged together due to rebinning of the power spectra by a factor W and averaging M different power spectra increases by a factor of 2 in consecutive curves. The trials are assumed to be independent, so no overlaps between the W-bin averages are allowed. As an example, for a power spectrum produced by averaging together 2 "raw" power spectra of 4096 bins each and binning up the resulting spectrum by a factor of 4 to produce a 1024-bin average spectrum the 90% confidence detection level can be read from the curve $MW = 8$ at $N_{\rm trial} = 1024$ to be 5.8.*

3.4. Non-Poissonian Noise

For various reasons the noise in the x_k can differ from purely Poissonian counting noise. The consequence of this is that the $P_{j,\mathrm{noise}}$ are no longer distributed according to Eq. 3.4; in general, correlations introduced by the noise process between the x_k will introduce correlations between the $P_{j,\mathrm{noise}}$, so that the noise power spectrum will not even be flat any more. However, as we have seen in Section 3.2, under "certain" conditions one still expects the noise powers to follow a χ^2 distribution with 2 dof. Even if this is the case, the normalization will be different from that implied by Eq. 3.4 and will in general depend on j.

3.4.1. Dead Time

One reason why noise in the x_k can be non-Poissonian is the occurrence of various instrumental effects summarized under the name "dead time". Various types of dead time exist, all of which considered here have in common that the instrument is unable to detect a photon for a given short interval of time τ_{dead} after a photon has been detected; τ_{dead} may be constant, or depend itself on various instrumental parameters.

The case where τ_{dead} is constant is typical for an X-ray proportional-counter dead time. If the incident count rate is λ and the detected count rate μ, then during an observation of length T the total dead time will be $\mu T \tau_{\mathrm{dead}}$, so that ("incident = detected + missed") $\lambda T = \mu T + \mu T \tau_{\mathrm{dead}} \lambda$, or

$$\mu = \frac{\lambda}{1 + \tau_{\mathrm{dead}} \lambda}. \tag{3.8}$$

This type of dead time introduces a correlation between the x_k: if a photon has been detected in bin k, then there is a certain probability that the dead time interval associated with this photon extends into the next bin $k+1$, and therefore the average chance to detect a photon in bin $k+1$ will be diminuished. This then means that the average chance in bin $k+2$ will be slightly higher, *etc.*, so that the final result is that a quasi-periodic oscillation is introduced into the x_k with a frequency equal to the Nyquist frequency. Consequently, the noise power spectrum will rise towards $\nu_{N/2}$. This constant τ_{dead} process has been simulated by Weisskopf (1985). The result of these simulations is that the expectation value of the noise power spectrum (normalized according to Eq. 2.8, where $N_{\mathrm{ph}} = \mu T$) becomes:

$$\langle P_{j,\mathrm{noise}} \rangle = 2(1 - \mu\tau_{\mathrm{dead}})^2 \left[1 + 2 \left(\frac{\mu\tau_{\mathrm{dead}}}{1 - \mu\tau_{\mathrm{dead}}} \right) \left(\frac{\tau_{\mathrm{dead}}}{T/N} \right) \sin^2 \frac{\pi j}{N} \right]. \tag{3.9}$$

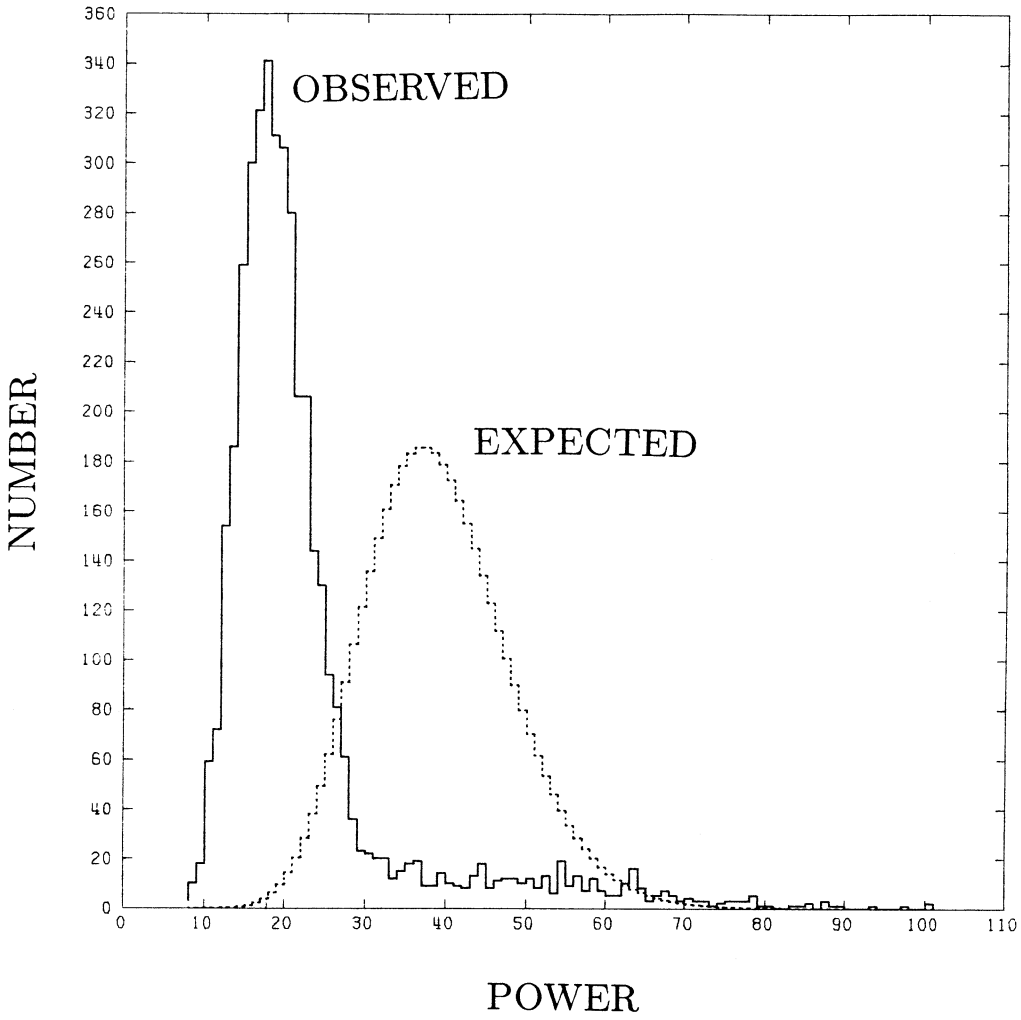

Fig. 3.3. *The complicated dead time process of the EXOSAT ME (see Andrews and Stella 1985, Tennant 1987) strongly modifies the probability distribution of the noise powers (drawn) with respect to that expected for a Poisson noise process. After scaling the distribution with the mean noise power, however, the match with the expected χ^2 distribution (dashed) is good. (Simulated QPO data; sum of 20 power spectra.)*

The amplitude of the frequency-dependent component in the noise power spectrum is seen to be proportional to both the ratio of total dead time to total live time and the ratio of τ_{dead} to the duration of a time bin T/N.

Another simple type of dead-time process is that where the instrument can only detect at most one photon per instrumental "sample" cycle, which has a duration τ_{sample}. So, in this case τ_{dead} is variable and lasts from the time a photon is detected until the end of the sample. If the arrival times of the incident photons are Poisson

distributed, then the chance that **no** photon arrives during $\tau_{\rm sample}$ is $e^{-\tau_{\rm sample}\lambda}$, so that the chance that a photon is detected is $1 - e^{-\tau_{\rm sample}\lambda}$, so

$$\mu_\cdot = \frac{1 - e^{-\tau_{\rm sample}\lambda}}{\tau_{\rm sample}}. \tag{3.10}$$

If each bin x_k contains $N_{\rm sample}$ samples, then the number of counts in x_k is the number of "successes" among $N_{\rm sample}$ trials, where the chance of success is $p = 1 - e^{-\tau_{\rm sample}\lambda}$. This means that the x_k follow (by definition) the binomial rather than the Poisson distribution, with mean $pN_{\rm sample}$ and standard deviation $\sqrt{p(1-p)N_{\rm sample}}$. Note that in this case no correlations are introduced between the x_k. With Eq. 2.9 for the total variance in the x_k and noting that $N_{\rm ph} = \sum x_k = pNN_{\rm sample}$ it can be derived that the average noise power will be

$$\langle P_{j,\rm noise} \rangle = 2e^{-\tau_{\rm sample}\lambda} = 2(1 - \mu\tau_{\rm sample}), \tag{3.11}$$

as compared to 2 in the case of Poisson noise.

In practice, dead-time processes are often much more complex than in the two examples above. In particular, there may be an interaction between dead-time processes in different instrumental channels (see Paper 1). However, usually one finds that the χ^2 property of the noise powers is at least approximately preserved (Fig. 3.3).

3.4.2. Intrinsic Noise

It is very common for the source signal itself to consist (partly) of noise. Such intrinsic noise signals can contain very useful information about the source and are worthwhile to try and detect over the background noise caused by, *e.g.*, counting statistics; examples are red noise and QPO. However, in many cases one wishes to consider such an intrinsic noise component as background against which to detect another source signal component. In such cases it is of particular importance to test empirically the probability distribution of the noise powers.

For example, a theoretical description of red noise as the integral of white noise (*e.g.*, Deeter and Boynton 1982) suggests that the χ^2 property of the noise powers will apply for red noise. However, although any observed power spectrum that rises towards lower frequencies is often called a red noise spectrum, the underlying source variability is not specified by this power spectral property (it is not even necessarily a noise process) and the χ^2 property can not be guaranteed.

Again, in practice one usually finds that also for noise in the source signal the noise power distribution closely matches a χ^2 distribution scaled to the local average power (Figs. 3.4, 3.5).

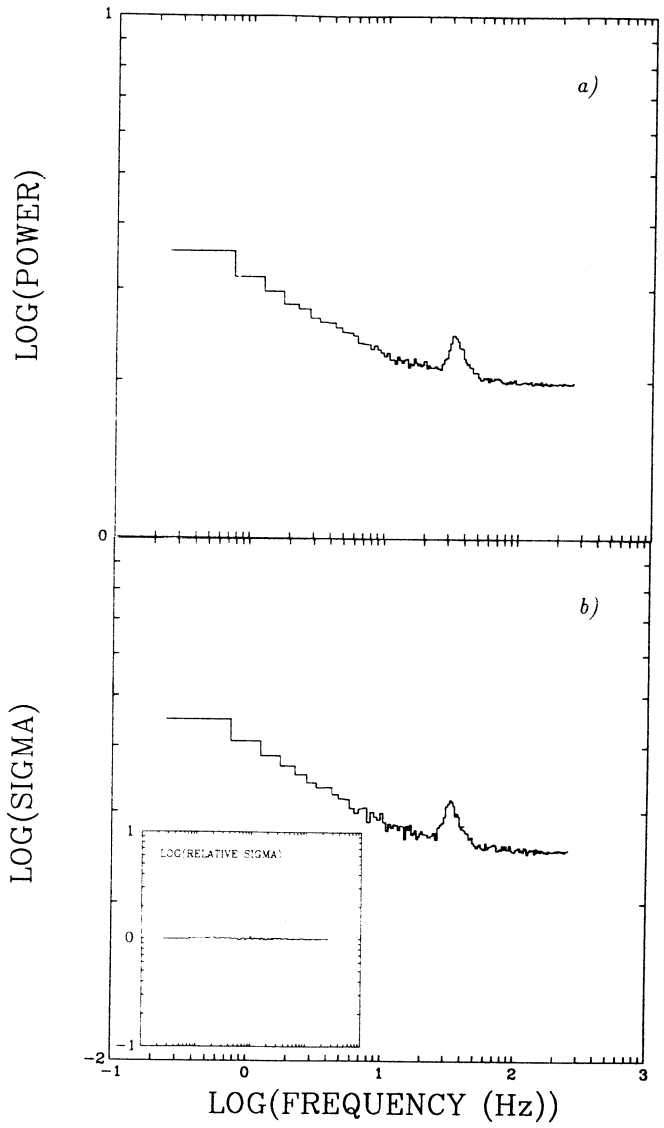

Fig. 3.4. *a) An average of 6166 power spectra of EXOSAT ME data on the source GX 5-1 showing red noise and QPO (quasi-periodic oscillations). b) The standard deviation of the 6166 power values averaged in each frequency bin. Inset: the ratio of b) to a). Standard deviation equals mean power as expected for* χ^2 *distributed powers.*

It is stressed that it is **essential** when looking for weak source signals to take into account the (likely) presence of intrinsic noise. In particular it is completely wrong to use the normalization of the noise power distribution valid for Poisson statistics (Eq. 3.4) when trying to detect something against a background of red noise. Note that red noise can be present even if the quality of the data is not

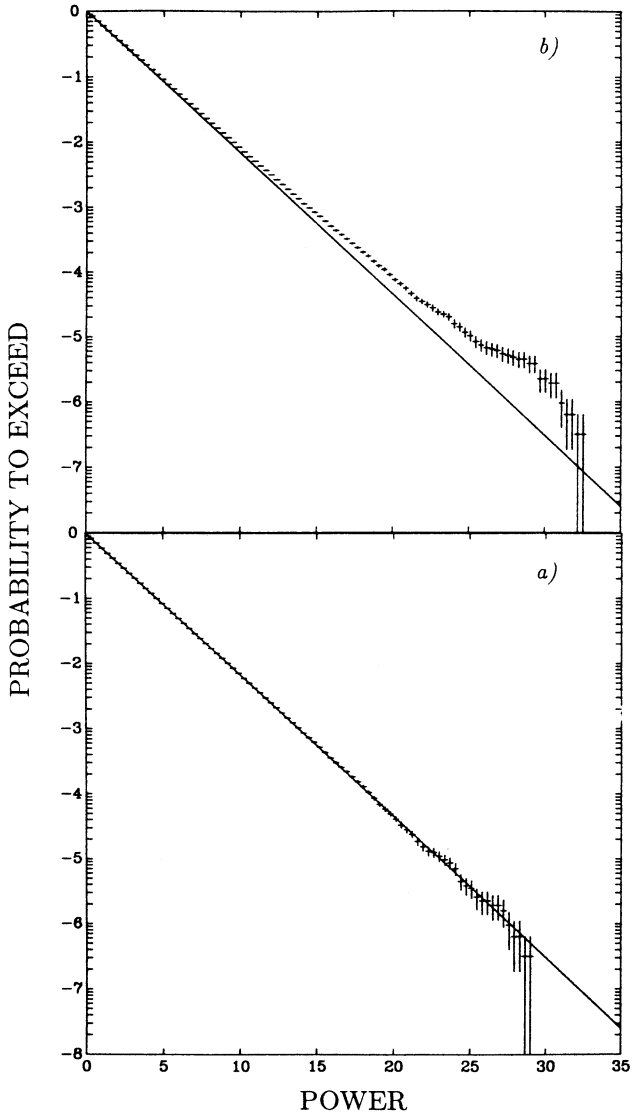

Fig. 3.5. *a) Integral distribution of all individual powers in the 6166 power spectra of Fig. 3.4a (crosses). Strong deviations from a χ^2 distribution (drawn line) due to QPO and red noise are visible. b) Distribution of the same powers as in a), after dividing each individual power spectrum by the average spectrum of Fig. 3.4a. The match to the χ^2 distribution is very close now. Small residual wiggles are due to the intrinsic variations of the shape of the QPO/red noise power spectrum as a function of source intensity.*

sufficient to clearly see the slope in the power spectrum [4].

If the χ^2 property is expected to apply then a correct procedure would for

[4] "Scrambling" techniques where the time order of the x_k is randomized are

example be to **divide** the power spectrum by some average red noise shape (*e.g.*, a best-fit power law) in order to bring all noise powers back to one χ^2 distribution and then evaluate the significance of any excess. Obviously, the uncertainties in the description of the shape of the noise spectrum itself should also be taken in into account in such an analysis.

4. THE SIGNAL POWER

4.1. Introduction

Any quantitative statement one can make about the signal power $P_{j,\text{signal}}$ will be a statement of a probability based on the probability distribution of the noise powers $P_{j,\text{noise}}$, because the only thing one knows for sure is the total power P_j which is (Eq. 3.1) equal to the signal power contaminated with an unknown amount of noise power. In Section 4.2 we will consider this process of quantifying the signal power. In Section 4.3 it will be discussed how to convert a statement about $P_{j,\text{signal}}$ into a statement about the rms variation in the source signal. It is reiterated at this point that to say anything else about the source signal, *e.g.*, about the amplitude of a sine wave, is an entirely different problem for which we need to model the shape of the signal power spectrum. For a sinusoid signal, this problem will be touched upon in Section 6.

4.2. Quantifying the Signal Power

4.2.1. Detected Signal Power. Supposing that we have a detection, *i.e.*, for given j it is true that $P_j > P_{\text{detect}}$, then we ask what is the probable value of the signal power $P_{j,\text{signal}}$ at j.

Determine a "limiting noise power level" $P_{\text{noiselimit}}$ that has only a small probability ϵ' to be exceeded in one trial:

$$\epsilon' = Q(MWP_{\text{noiselimit}}|2MW). \tag{4.1}$$

Then, with confidence $(1-\epsilon')$ we can say that for given j $P_{j,\text{noise}} < P_{\text{noiselimit}}$. Because according to Eq. 3.1 $P_{j,\text{signal}} = P_j - P_{j,\text{noise}}$, this implies that

$$P_{j,\text{signal}} > P_j - P_{\text{noiselimit}} \qquad (1-\epsilon') \text{ confidence.} \tag{4.2}$$

---

sometimes used to evaluate the probability to produce a certain feature in the power spectrum by chance given the distribution of the x_k. These techniques are incorrect when red noise is present, as the scrambling destroys the correlations in the x_k underlying the red noise and artificially converts the red noise into white noise.

Note that a slightly misleading statement about this case occurs in Paper 1 (in the text after Eq. 2.13).

4.2.2. Upper Limit to the Signal Power. If no significant power level has been attained by any of the P_j, then it is useful to determine an upper limit to the signal power. The $(1-\delta)$ confidence upper limit P_{UL} to the signal power is defined as the power level for which with $(1-\delta)$ confidence $P_{j,\mathrm{signal}} < P_{\mathrm{UL}}$ irrespective of where (at which j) in the frequency range of interest this signal power may have occured.

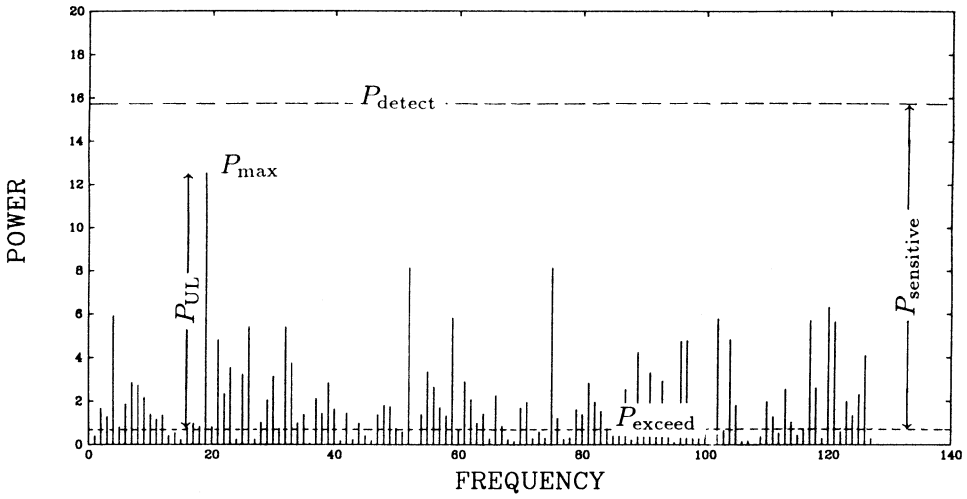

Fig. 4.1. *Relations between the detection level P_{detect}, the "usually exceeded" level P_{exceed}, the maximum observed power P_{max}, the upper limit P_{UL} and the sensitivity level $P_{\mathrm{sensitive}}$ (see text).*

To determine P_{UL} we define a power level P_{exceed} that has the large probability $(1-\delta)$ to be exceeded by a given individual $P_{j,\mathrm{noise}}$:

$$1-\delta = Q(MWP_{\mathrm{exceed}}|2MW). \tag{4.3}$$

So, a fraction of approximately $(1-\delta)$ of all powers considered will exceed P_{exceed} in the absence of a source signal. We now find the largest actually observed power P_{max} in the frequency range of interest, and write

$$P_{\mathrm{UL}} = P_{\mathrm{max}} - P_{\mathrm{exceed}}. \tag{4.4}$$

If for some j there would have been a signal power in the power spectrum with $P_{j,\mathrm{signal}} > P_{\mathrm{UL}}$, then for that j with $(1-\delta)$ confidence P_j would have exceeded $P_{\mathrm{max}} = P_{\mathrm{UL}} + P_{\mathrm{exceed}}$ (Eq. 3.1). As we know that for all j it is true that $P_j \leq P_{\mathrm{max}}$,

we can say with $(1-\delta)$ confidence that indeed for all j $P_{j,\text{signal}} \leq P_{\text{UL}}$. Note that the number of trials N_{trial} does not enter into the calculation of P_{UL}, as one reasons in terms of one hypothetical undetected signal power $P_{j,\text{signal}}$ at **given** j.

4.2.3. Sensitivity to Signal Power. It is sometimes useful to predict the capabilities of a planned experiment in terms of its sensitivity to signal power. The sensitivity level $P_{\text{sensitive}}$ can be calculated on the basis of the (expected) probability distribution of the noise power as

$$P_{\text{sensitive}} = P_{\text{detect}} - P_{\text{exceed}}, \tag{4.5}$$

where P_{detect} and P_{exceed} are defined in Eqs. 3.7 and 4.3, respectively. If there occurs a $P_{j,\text{signal}}$ somewhere in the power spectrum that exceeds $P_{\text{sensitive}}$ then with $(1-\delta)$ confidence it will be detected (at the $(1-\epsilon)$ confidence level associated with P_{detect}). Note that $P_{\text{sensitive}}$ is not the same as P_{UL}; in fact, $P_{\text{sensitive}}$ is in a sense the upper limit to P_{UL} (see Fig. 4.1).

4.3. The rms Variation in the Source Signal

Assuming that the signal power spectrum has been properly separated out from the total power spectrum using the methods described in Section 4.2, we can convert the signal power into the rms variation r of the source signal in the x_k using the expression

$$r = \sqrt{\frac{W \sum_j P_{j,\text{signal}}}{N_{\text{ph}}}}. \tag{4.6}$$

(*c.f.* Eq. 2.10), where P_j is an MW times averaged power and where N_{ph} is the number of photons per transform.

We shall consider two effects that may cause a difference between the value of r obtained in this way and the actual value R applicable to the signal $x(t)$ as emitted by the source. As we have seen in Section 2.5, the binning of the data causes the power spectrum to be suppressed preferably towards the higher frequencies. The correction factor by which r should be multiplied is

$$\left(\frac{R}{r}\right)_{\text{binning}} = \beta = \frac{\pi \overline{\nu} T/N}{\sin \pi \overline{\nu} T/N}, \tag{4.7}$$

(*c.f.* Eq. 2.19), where $\overline{\nu}$ is some appropriate average over the signal feature in the power spectrum.

The second effect is our old friend dead time (see also Section 3.4.1). Provided that the variations in the signal of interest are slow with respect to the dead time

process, the way in which dead time changes the signal is completely described by the instrument-dependent relation between λ and μ, the incident and detected count rates. Examples of such relations are given in Eqs. 3.8 and 3.10; any relation, either theoretical or empirical, can be used.

Expressing the relation as the dead time attenuation factor $f(\lambda) \equiv \mu/\lambda$, one finds that a small change $\delta\lambda$ in the incident rate causes a change $\delta\mu$ in the observed rate given by

$$\frac{\delta\mu}{\mu} = \left(1 + \frac{\lambda}{f}\frac{df}{d\lambda}\right)\frac{\delta\lambda}{\lambda}. \tag{4.8}$$

From this expression we can derive the dead time correction factor

$$\left(\frac{R}{r}\right)_{\text{dead time}} = \alpha = \frac{|\delta\lambda|/\lambda}{|\delta\mu|/\mu} = \left(\left|1 + \frac{\lambda}{f}\frac{df}{d\lambda}\right|\right)^{-1}. \tag{4.9}$$

See Paper 1 for a discussion of the case of dead-time interaction between several instrumental channels.

One minor pitfall remains to be considered, being the practice of describing the signal power spectrum $P_{j,\text{signal}}$ in terms of a function $P_{\text{signal}}(\nu)$ defined such that $P_{\text{signal}}(\nu_j) \equiv P_{j,\text{signal}}$. When integrating this function one should note that $\int P_{\text{signal}}(\nu)\,d\nu = \frac{W}{T}\sum P_{j,\text{signal}}$. Defining the **excess power** in the power spectrum due to the signal as $P_{\text{excess}} \equiv W\sum P_{j,\text{signal}} = T\int P_{\text{signal}}(\nu)\,d\nu$, we finally have for the fractional rms variation in the signal corrected for binning and dead time

$$R = \alpha\beta\sqrt{\frac{P_{\text{excess}}}{N_{\text{ph}}}} = \left|1 + \frac{\lambda}{f}\frac{df}{d\lambda}\right|^{-1}\frac{\pi\overline{\nu}\delta t}{\sin\pi\overline{\nu}\delta t}\sqrt{\frac{\int P_{\text{signal}}(\nu)\,d\nu}{I}}, \tag{4.10}$$

where $I = N_{\text{ph}}/T$ is the count rate, $\overline{\nu}$ the average frequency of the signal feature in the power spectrum, $\delta t = T/N$ the duration of a time bin, and the sum and the integral run over all frequencies where the signal causes a non-negligible power (a power at the the Nyquist frequency should enter with a factor $\frac{1}{2}$, see Eq. 2.10).

5. OPTIMAL DETECTION - INDEPENDENT TRIALS

5.1. Introduction

When deriving the detection level from the noise power distribution in Section 3.3, we implicitely assumed that the trials (*i.e.*, the powers to be tested against the detection level) would all have the same statistical properties, and that they would be statistically independent. In particular, MW was assumed to be equal for all trials and no overlaps were allowed between the averages of W bins. This

approach is not necessarily optimal for all signal shapes, but it has the enormous advantage that the statistical properties of any power excess that is found can easily be evaluated analytically from the known distribution of the noise powers. In this section, we develop the possibilities of this method to its limit by considering the optimal value of the number of bins to average, given the expected signal. In Section 5.4 the method as it has been developed in the previous sections is summarized in a step-by-step style. In Sections 6 and 7, we shall abandon the constraints that the trials should be equal and independent.

5.2. Detecting a Narrow Feature

The detection of a narrow feature in the power spectrum (defined as a feature in which all power is concentrated in one frequency bin) is a fundamental problem because the continuous Fourier transform of a strictly periodic signal consists of one or more delta functions (see, however, Section 4.5 for the case of the discrete Fourier transform). If the signal power of a narrow feature in a full frequency-resolution power spectrum is P_{signal}, then it will drop to $P_{\text{signal}}/(MW)$ after the frequency resolution has been degraded by a factor MW by one of the methods described in Section 3. For the method of averaging W adjacent bins this is immediately obvious; for the method involving division of the data into M equal segments, this can be seen by using Eq. 2.9 (and assuming that the signal remains the same through the observation).

The detection level also drops when the frequency resolution is degraded, both because the probability distribution of the noise powers in the average power spectrum becomes narrower and because the number of trials decreases by a factor MW. In addition, the narrower noise distribution causes P_{exceed} to increase. However, in the final analysis the sensitivity level $P_{\text{sensitive}}$ (Eq. 4.5) always drops more slowly than $1/MW$ (Fig. 5.1), so that the conclusion is that for detecting a narrow feature in the power spectrum the highest sensitivity is reached for the maximum possible frequency resolution, *i.e.*, by choosing $MW = 1$. Note that we have not specified what the source signal should be in order for the signal power spectrum to be narrow; as we shall see, the discrete Fourier transform of even a strictly periodic sinusoidal signal does not approximate a delta function very well.

5.3. Detecting a Broad Feature

Similar reasoning as in Section 5.2 shows that also for a feature of finite width $\Delta\nu$ the signal power **summed** over all frequency bins in the feature will drop proportionally to $1/MW$ when the frequency resolution of the power spectrum is degraded. However, as long as the width of the feature exceeds the frequency resolution: $\Delta\nu > MW/T_{\text{obs}}$, where $T_{\text{obs}} = MT$ is the total length of the observation, the

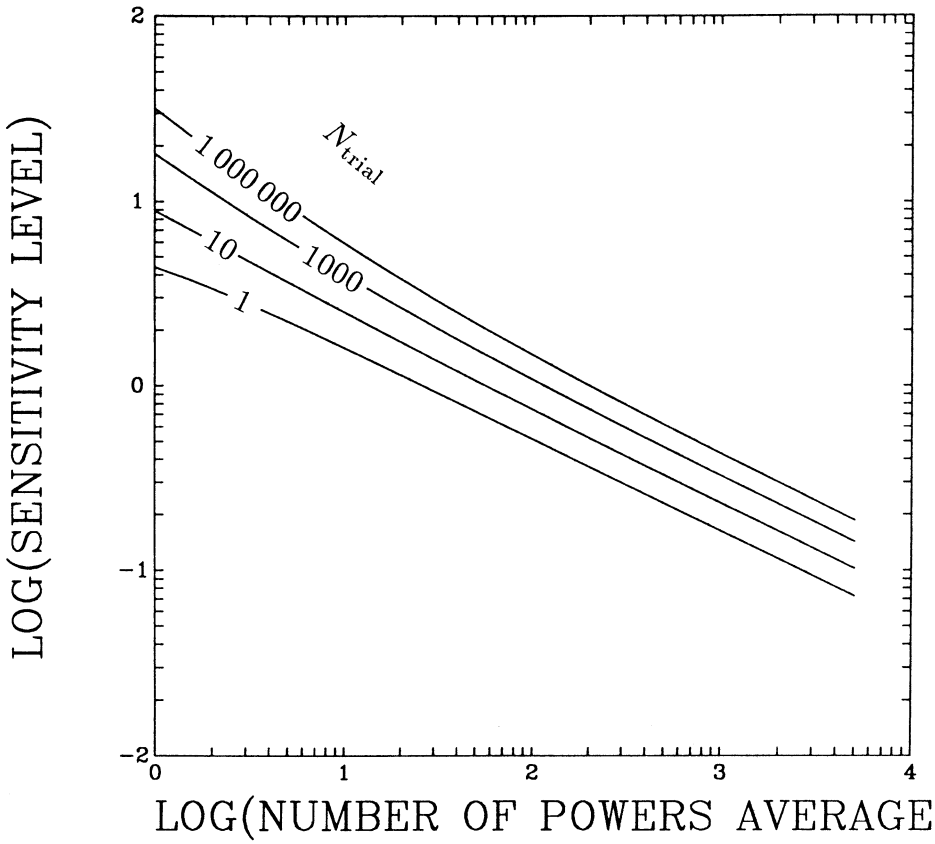

Fig. 5.1. *The sensitivity level as a function of the number of powers averaged (MW) for various numbers of trials. The sensitivity level always decreases more slowly than* $1/MW$.

signal power in one frequency bin within the feature will remain constant (strictly speaking, for a feature with a rectangular profile, and ignoring edge effects), because the number of bins in the feature $\Delta\nu/(MW/T_{\rm obs})$ is also inversely proportional to MW. Because $P_{\rm sensitive}$ drops as a function of MW, this implies that the sensitivity to the feature increases with MW. When $(MW/T_{\rm obs}) > \Delta\nu$, we recover the case of Section 5.2, and the sensitivity begins to drop. So, the optimal value of MW is that which **just** concentrates all power in one bin: $T_{\rm obs}\Delta\nu$.

The above argument ignores the alignment between the feature and the frequency bins; for $MW = T_{\rm obs}\Delta\nu$ one is lucky when all power is in one bin; more likely is a situation where the power is distributed over 2 bins. It is possible to devise tests which take into account various possible alignments and which are optimized for specific shapes of the broad feature; some examples of this will be considered in

Sections 6 and 7.

5.4. Summary: Power Spectral Searches Made Easy

In this section we collect all previous results into a "how to" recipe of testing the power spectrum for a weak signal using equal statistically independent trials.

1. Determine the M and W (Section 3.2). The optimal choice for MW is that which approximately matches the expected width of the power spectral feature one desires to detect, $T_{\rm obs}\Delta\nu$ (Sections 5.2 and 5.3), but gaps in the data or the desire to see the time evolution of the power spectrum may dictate M.
2. Calculate the M power spectra normalized according to Eqs. 2.8 and 2.4*a*. Note that x_k is the number of photons in bin k and $N_{\rm ph}$ the number of photons in **one** power spectrum, $\sum_k x_k$.
3. Average the M power spectra.
4. Think about the noise power distribution (Sections 3.2 and 3.4). Does the noise power spectrum seem to be flat? Is its mean level 2.0? If so, the noise is probably dominated by Poissonian counting statistics – go to step 5. If not, find out why not. Try to determine whether the χ^2 property applies (Section 3.4). If you are satisfied that it does, you can divide the power spectrum by some mean noise power spectral shape and go on step 5. Otherwise, find out what **is** the distribution of the noise powers and determine the detection level accordingly.
5. Determine the detection level (Eq. 3.7, Fig. 3.1).
6. Check the average spectrum for powers exceeding the detection level.
7. Quantify the signal power in terms of a detection (Section 4.2.1) or an upper limit (Section 4.2.2).
8. If necessary, multiply back in the noise power shape you may have divided out in step 4.
9. Convert the signal power into the relative rms variation of the source signal, correcting for the effects of binning and dead time (Section 4.3).
10. To say more about the signal, you need to model its power spectrum. For a sinusoidal signal, see Section 6.

6. DETECTING A SINUSOIDAL SOURCE SIGNAL

6.1. Introduction

When searching for an X-ray pulsar, the first assumption that is made is often that the pulse shape is sinusoidal – and for many X-ray pulsars, of course, this is not a bad assumption. In this section, we first derive the shape of the discrete power spectrum of a sinusoidal signal of arbitrary frequency and phase, and then consider

in some detail how to optimize a power spectral search towards detecting a weak sinusoidal signal. We allow the possibility of dependent trials, which means that detection levels have to be determined by simulations. Throughout this section, we ignore the binning and dead time corrections (Section 4.3).

6.2. The Power Spectrum of a Sinusoidal Signal

The continuous Fourier transform of a sinusoidal signal is a delta function.

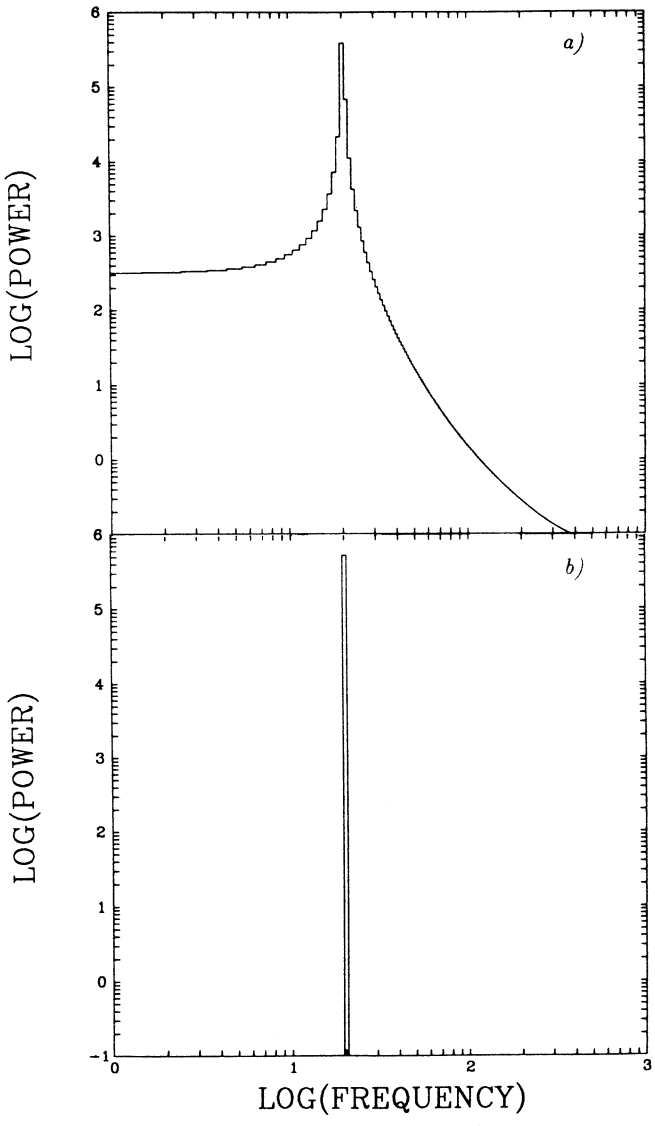

Fig. 6.1. *a) The discrete power spectrum of a sinusoid with an arbitrary frequency. b) Same, with frequency equal to a Fourier frequency.*

However, if one calculates the FFT of some arbitrary sinusoid the result, more likely than not, will be something similar to Fig. 6.1*a*. The reason that this power spectrum does not look very much like a delta function is of course that is was calculated from a discrete Fourier transform. As discussed in Section 2.5, windowing causes a strong distortion of the Fourier transform. (The reason that the spectrum in Fig. 6.1*a* does not look very much like the window transform $W(\nu)$ of Section 2.5 either is that the frequency resolution of a discrete Fourier transform is equal to the width of the side lobes of $W(\nu)$, namely, $1/T$). Only when the frequency ν_{sine} of the sinusoid is equal to one of the Fourier frequencies $\nu_j = j/T$ will all power be concentrated in one bin of the discrete Fourier transform (Fig. 6.1*b*).

The discrete power spectrum of a sinusoidal signal

$$x_k = A\cos(\omega_{\text{sine}} t_k + \phi) \tag{6.1}$$

can be calculated directly from Eq. 2.4*a* by making use of the result

$$\sum_{k=0}^{N-1} e^{i\alpha k} = e^{\frac{1}{2}(N-1)i\alpha} \frac{\sin \frac{1}{2}N\alpha}{\sin \frac{1}{2}\alpha} \tag{6.2}$$

(*e.g.*, Bloomfield 1976). The result of this calculation is

$$|a_j|^2 = \tfrac{1}{4}A^2N^2 \left(\frac{\sin \pi x}{\pi x}\right)^2 \left[\left(\frac{\pi x/N}{\sin \pi x/N}\right)^2 + \left(\frac{\pi x/N}{\sin\left[\pi(2j+x)/N\right]}\right)^2 + \right.$$
$$\left. + 2\left(\frac{\pi x/N}{\sin \pi x/N}\right)\left(\frac{\pi x/N}{\sin\left[\pi(2j+x)/N\right]}\right)\cos\left[(N-1)\left(2\pi(j+x)/N\right) + 2\phi\right]\right], \tag{6.3}$$

where $x = (\nu_{\text{sine}} - \nu_j)T$, the frequency offset of the sinusoid frequency from frequency bin j in units of the Fourier frequency step $1/T$. The dependence of this expression on the phase ϕ of the sinusoid is small except for j close to 0 or $N/2$ (Fig. 6.2). The function describing the signal power at ν_j as a function of x is of course closely related to $|W(\nu)|^2$ (Eq. 2.15). For $x/N \ll 1$ and $0 \ll j/N \ll \frac{1}{2}$, *i.e.*, for frequency bins close to ν_{sine} and not too close to either zero or the Nyquist frequency, expression (6.3) reduces to

$$|a_j|^2 \approx \tfrac{1}{4}A^2N^2 \left(\frac{\sin \pi x}{\pi x}\right)^2. \tag{6.4}$$

This function, normalized to 1 at $x = 0$ is shown as the drawn curve in in Fig. 6.2*b*.

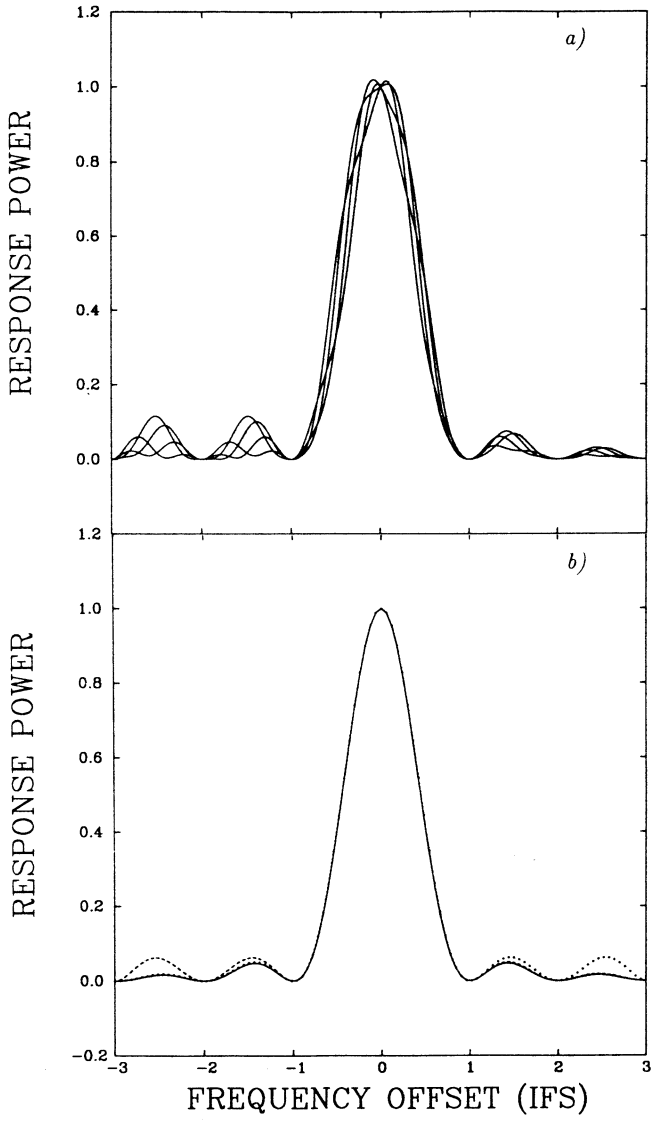

Fig. 6.2. *Response power at ν_j to a sinusoidal signal as a function of $\nu_{\rm sine} - \nu_j$ in units of the independent Fourier frequency step (IFS). Calculations are for $N = 1024$. a) $j = 2$. Curves are for various phases of the sinusoid. b) Phase-averaged curves. Drawn: $j = 200$; dashed: $j = 2$; dotted: $j = 510$.*

6.3. Single-Power Response

The highest power in the signal power spectrum will be obtained at the Fourier frequency ν_j closest to $\nu_{\rm sine}$. Normalized to a power of 1 for $\nu_{\rm sine} \equiv \nu_j \quad (x = 0)$, this power varies between 0.405 and 1, with an average value of 0.773 (Fig. 6.5, drawn curve). In the following we will use the term "input power" for the power that

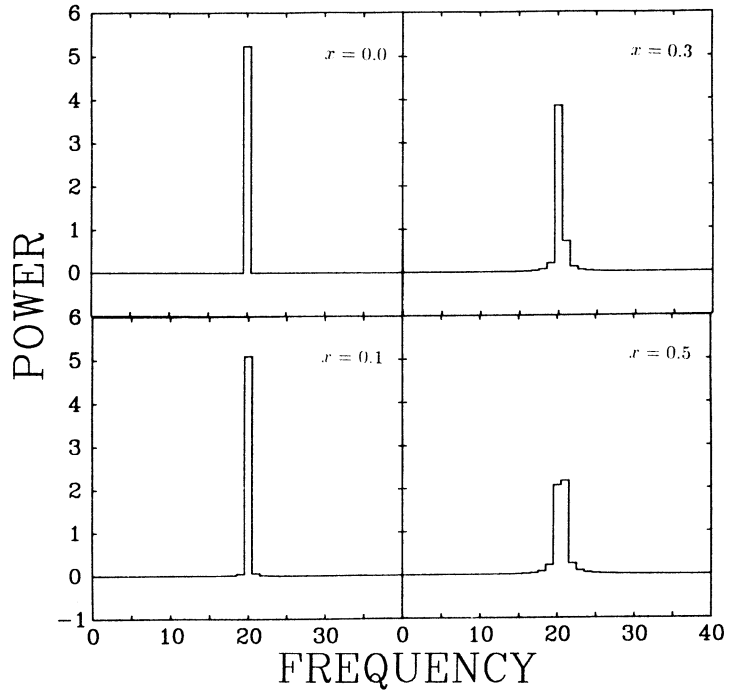

Fig. 6.3. *Discrete power spectra of a sinusoidal signal with frequency offsets from the nearest Fourier frequency as indicated. Nearly all power is in all cases concentrated in the two bins closest to* ν_{sine}.

would have been caused by the sinusoidal signal if $\nu_j = \nu_{\text{sine}}$; the "response power" is the power that is actually produced in the signal power spectrum. The "response function" is the response power divided by the input power as a function of x, and the "response" is the highest response power in the power spectrum divided by the input power. So, the drawn curve in Fig. 6.2*b* is a response function. For this curve, the input power is 1, the response power depends on x and is between 1 and 0 and the response depends on x and is between 1 and 0.405. This "ripple" in the response to a sinusoidal signal implies that we can not exactly predict the highest response power for a sinusoidal signal of given input power and arbitrary frequency: the best we can do is calculate the probability distribution of the response.

This fact should in principle be taken into account when interpreting a power spectrum in terms of the properties of a sinusoidal source signal. For example, to set an upper limit to the amplitude A of the signal when no significant power has been detected, one should take into account that A depends not only on P_j, but also on x, ϕ and, mostly via the binning factor, on j. The canonical method

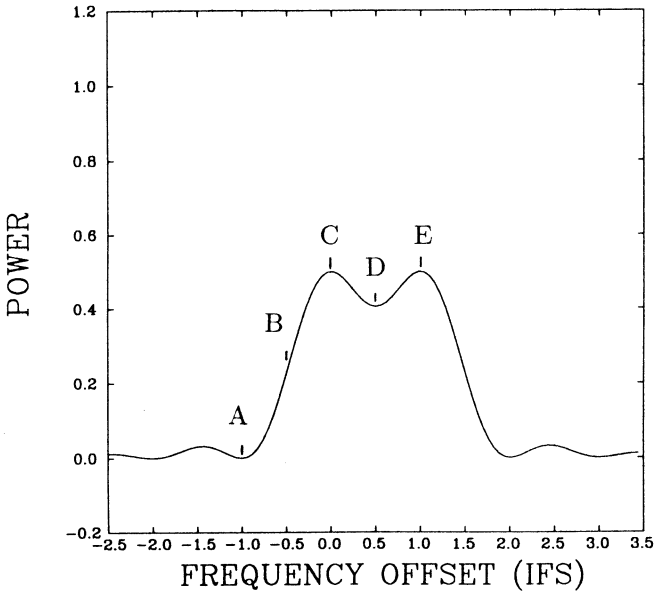

Fig. 6.4. *Averaged response power in two adjacent bins j and $j+1$ as a function of the frequency offset x of the sinusoidal signal from bin j.*

is to first determine an upper limit to $P_{j,\mathrm{signal}}$ using the noise power distribution (Section 4.2.2), and then convert this into an upper limit to A by using the average value (0.773) of the response. However, a signal with an amplitude considerably larger than the upper limit A_{UL} obtained in this way can easily be hidden in the data if its frequency happens to be close to halfway two Fourier frequencies. If there is a $P_{j,\mathrm{signal}}$ somewhere in the spectrum that actually reaches the limiting level P_{UL}, then there is a probability of ~44% that $A > A_{\mathrm{UL}}$. The factor by which A can exceed A_{UL} is at most 1.4, which would in most circumstances probably not be considered a large error. A better way of obtaining A_{UL} would be to consider the bivariate probability function of A in its dependence on $P_{j,\mathrm{noise}}$ and x, (the dependence on ϕ is weak, and that on j is usually kept).

6.4. Optimal Detection of a Sinusoidal Signal

The ripple in the single-power response to a sinusoidal signal will cause one to preferably miss signals with a frequency halfway between two Fourier frequencies.

Inspection of Fig. 6.3 shows that in this case ($x = 0.5$) nearly all power in the signal is divided between the two adjacent bins – power loss to frequency bins further out is relatively small. This suggests taking two-bin averages as a method to diminuish the ripple. Note that degrading the frequency resolution by dividing the data into M segments (Section 3.2) will not diminuish the ripple at all, but will instead lead to a loss of sensitivity similar to that discussed in Section 5.2.

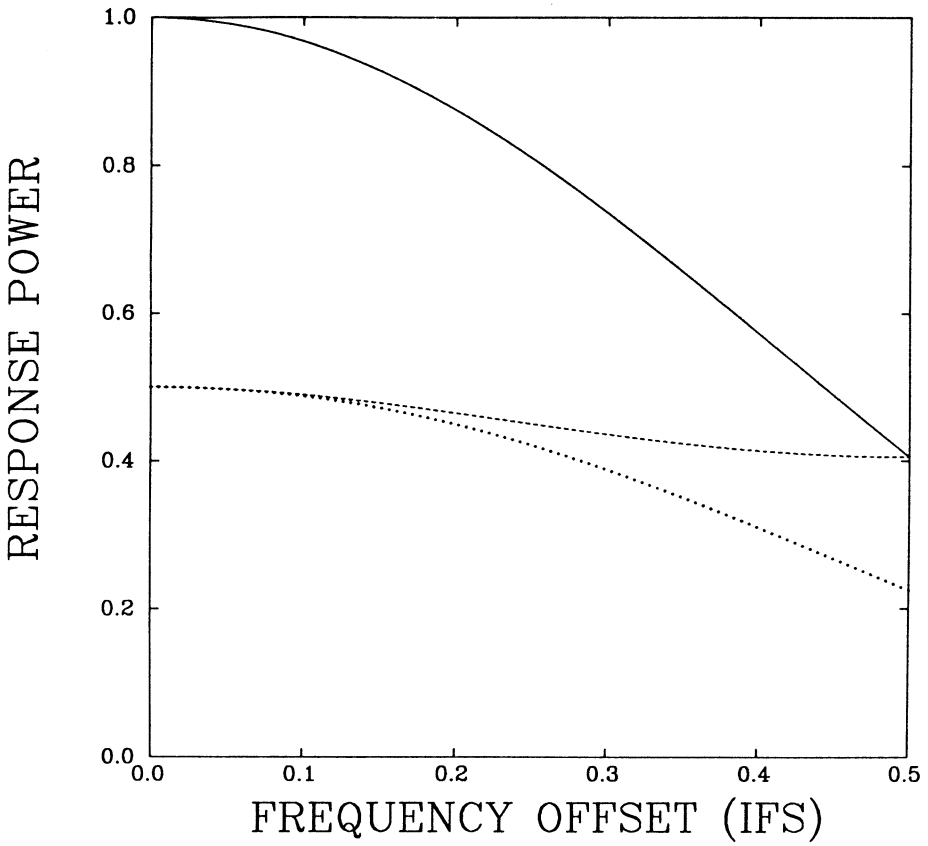

Fig. 6.5. *Comparison of the responses of the single-bin (drawn), two-bin running average (dashed) and two-bin no-overlap (dashed and dotted) schemes.*

Fig. 6.4 shows the average power in two adjacent bins as a function of x. To take full advantage of the flat part (C-E) of the two-bin response, it is necessary to consider **overlapping** averages ($\frac{1}{2}(|a_j|^2 + |a_{j+1}|^2)$; $\frac{1}{2}(|a_{j+1}|^2 + |a_{j+2}|^2)$; ...), otherwise in 50% of the cases the signal would still be in between bins (section B-C of the response). This means that the number of trials will nearly be equal to the single-bin case. In Fig. 6.5 the response of this 2-bin average, 1-bin overlap scheme (to be indicated henceforth as (2,1)) is compared to the single-bin (1,0) approach. The (2,0) response is also indicated. The (2,1) response is a factor 2 lower than the (1,0) one at $x = 0$, but the corresponding detection level is also much lower, the noise having been averaged over two bins. The (2,1) response varies between 0.5 and 0.405 with an average of 0.451.

Because of the overlap between the (2,1) trials, the chance to exceed a certain power level by noise can not be straightforwardly estimated analytically. One can perform simulations of many noise power spectra to determine the chance to exceed a certain power level by noise somewhere in the power spectrum as a function of

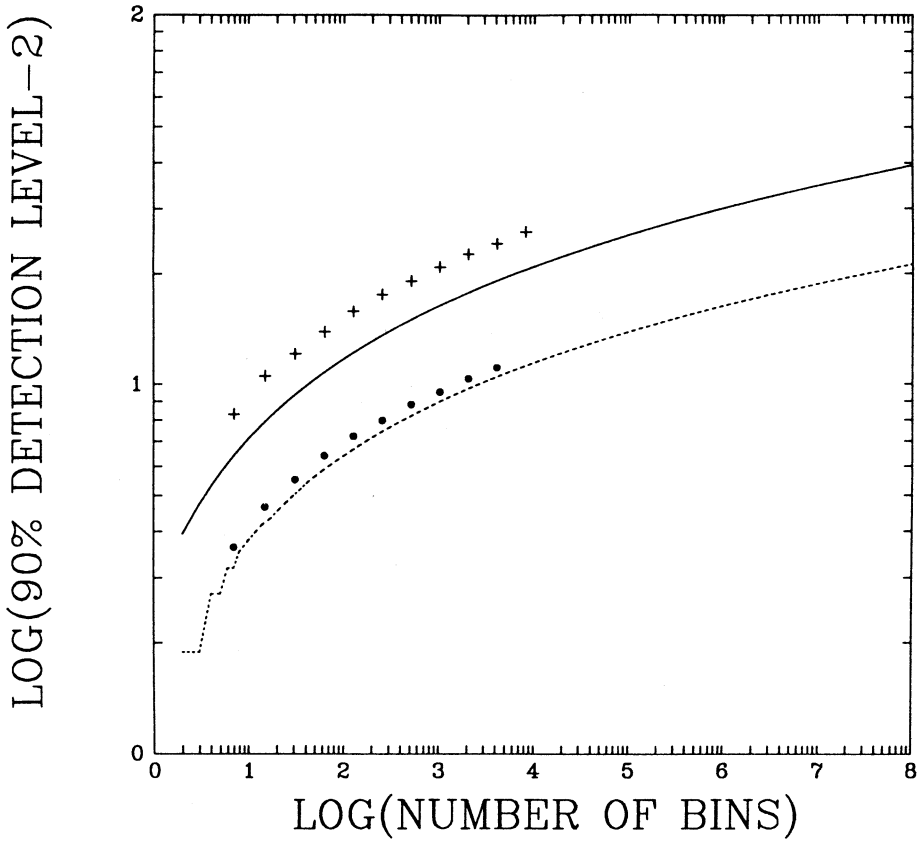

Fig. 6.6. *90% confidence detection levels as a function of the number of frequency bins in the power spectrum for the (1,0) (drawn) and (2,0) (dashed) schemes as determined from Eq. 3.7 (MW = 1), and the (2,1) (filled circles) and B&IB (crosses) schemes as determined from simulations of noise power spectra. Each point corresponds to 10^4 simulated power spectra; error bars are smaller than the points.*

the number of bins in the power spectrum. Results of simulations of this type are given in Fig. 6.6 in terms of the 90% confidence detection level (filled circles). It can be seen that the detection level is scarcely higher than for (2,0) sampling of the power spectrum (dashed curve), despite the twice higher number of trials. The reason for this is the large dependence of the extra trials in the (2,1) scheme with those already in the (2,0) scheme.

Another method of decreasing the ripple, used in pulsar radio astronomy (Backer 1988, priv. comm.) is "interbinning". In this method (Cullers *et al.* 1984), an extra frequency bin is created **in between** each two Fourier frequencies by calculating the difference of the adjacent complex Fourier amplitudes. One calculates the "interbin amplitude"

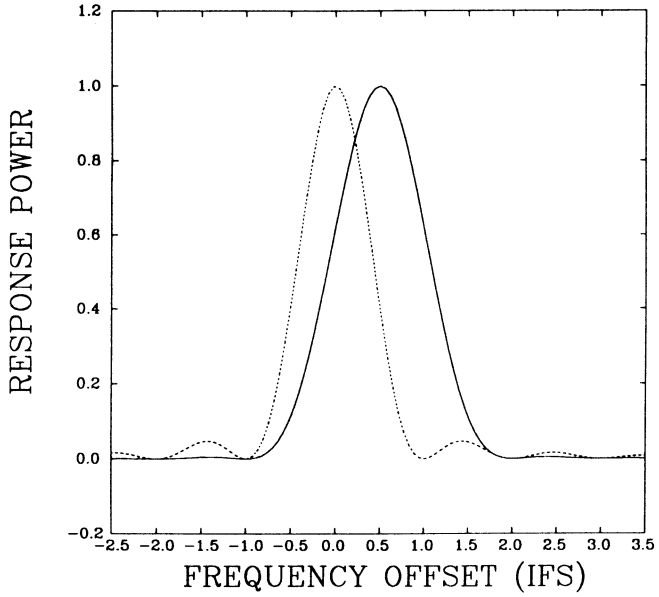

Fig. 6.7. *Interbin response power at $j+\frac{1}{2}$ for an input power of 1 (drawn) compared to response power at j (dashed) as function of the frequency offset x. The interbin reponse has its maximum at $x = \frac{1}{2}$.*

$$a_{j+\frac{1}{2}} = \frac{\pi}{4}(a_j - a_{j+1}). \tag{6.5}$$

The response function of this "filter" is

$$\left(\frac{\pi}{4}\frac{\cos(\pi(x-\frac{1}{2}))}{\pi x(1-x)}\right)^2 \tag{6.6}$$

This function is plotted in Fig. 6.7 together with the response function of one of the adjacent bins. The interbin response function is seen to be wider; its normalization is such that the response is 1 when ν_{sine} is exactly in between Fourier frequencies. The ripple in the response of bins and interbins (henceforth B&IB) together is even less than that in the (2,1) scheme: it varies between 1 and 0.857 with an average of 0.951. The number of trials in the B&IB scheme is twice that in the (1,0) or (2,1) schemes. Again the detection levels of the B&IB scheme must be determined with the help of simulations. In Fig. 6.6 some preliminary results are given (crosses). Note that the choice of one single detection level for bins and interbins together is not necessarily optimal, as the noise distributions of bins and interbins are different. In fact, the reason that the detection levels of the B&IB system are seen to be considerably higher than those corresponding to single-bin trials (1,0), is not so much the twice higher number of trials but the fact that the

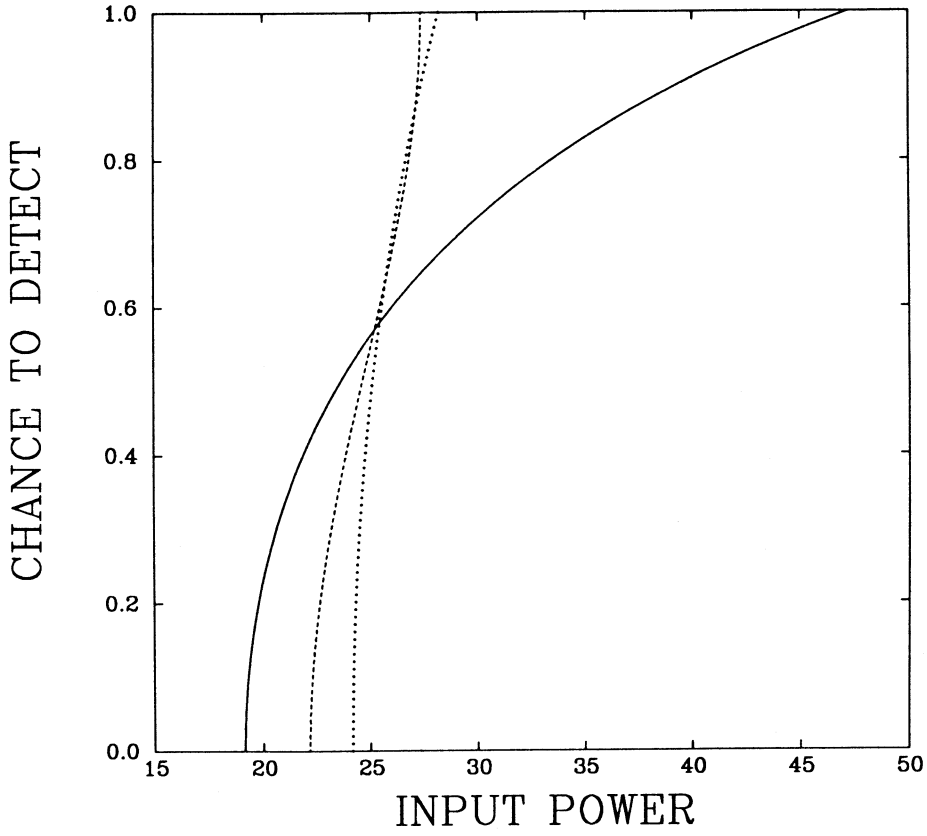

Fig. 6.8. *Comparison of the probability to detect (at the 90% confidence level) a sinusoidal signal of arbitrary frequency as a function of its input power for the single-power (drawn), 2-bin running average (dashed) and bin/interbin (dotted) methods. In this example, $N = 8192$ (4095 bins in the power spectrum). For input powers between 19.1 and 47.2, the average chance to detect the signal is 73, 80 and 77%, respectively.*

average noise level in the interbins is higher than in the bins by a factor ~1.23.

Knowing the detection levels (Fig. 6.6) and the probability distributions of the response to a sinusoidal signal that can be derived from the response functions (Figs. 6.5 and 6.7), one can estimate the probability to detect a sinusoidal signal of given amplitude in the various schemes ((1,0),(2,1),B&IB) as a function of the number of bins in the power spectrum. Fig. 6.8 displays the probability to detect the signal as a function of the input power for a power spectrum containing 4095 bins (the Nyquist frequency was excluded), where the phase-averaged response functions were used, and where the approximation was made that $P_j = P_{j,\mathrm{signal}} + 2$ rather than folding in the noise power distribution; given the large signal powers considered this is probably not a serious approximation. It is seen that the curves of the (2,1)

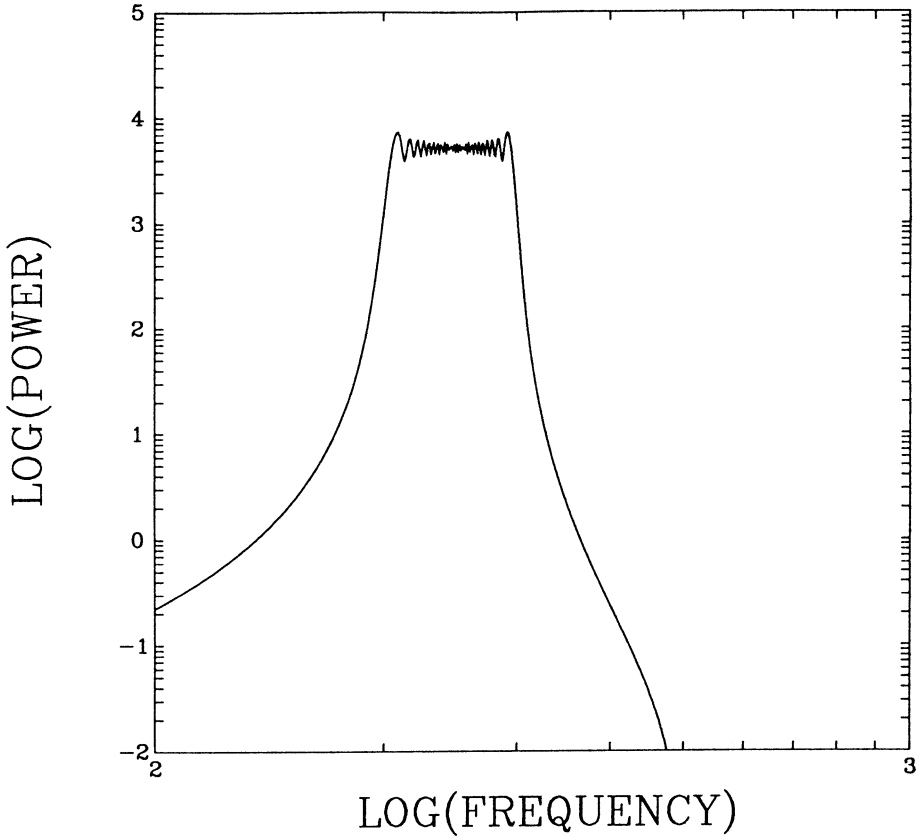

Fig. 6.9. *Power spectrum of a sinusoidal signal with a constant frequency derivative. The sinusoid frequency changes by 100 Fourier frequency steps during the observation.*

(dashed) and B&IB (dotted) schemes are quite similar and much steeper than that of the (1,0) scheme (drawn). The latter method has some probability to detect weak signals which are always missed by the former two, but signals always detected by the (2,1) and B&IB schemes will sometimes be missed by the (1,0) scheme.

When the input power is arbitrary, the methods can be compared by determining the average chance to detect a signal over the entire range of input powers where the three methods give different chances to detect the signal (for very high input powers the chance to detect is 1 for all methods, for very low ones, 0). The results of the preliminary simulations presented here seem to indicate that the differences between the methods are small, that the (2,1) method is best and the (1,0) method worst, with the B&IB method intermediate. However, the differences are sufficiently small for the approximations of the simulations to become relevant.

In any case it seems clear, that even for strictly periodic sinusoidal signals the single-bin approach is not always optimal. If there is any intrinsic broadening of

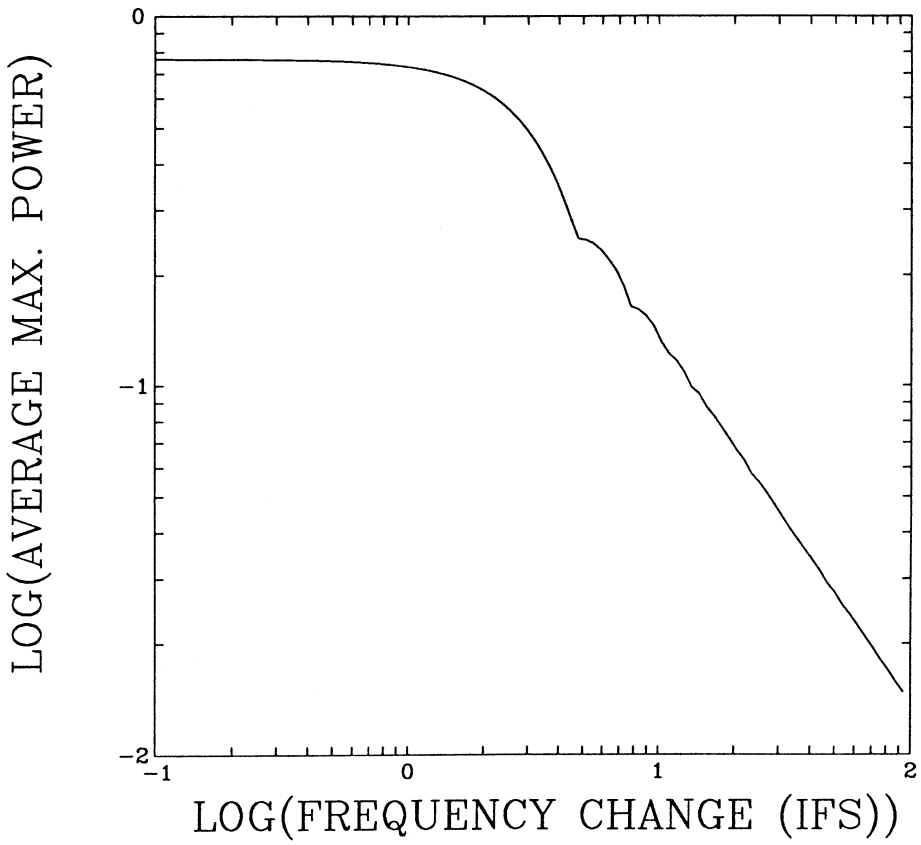

Fig. 6.10. *Highest power in a power spectrum of a sinusoidal signal with a constant $\dot{\nu}$ as a function of $\dot{\nu}T^2$.*

the signal, then the frequency resolution should certainly be degraded to optimize the search.

6.5. Sinusoidal Signal with Varying Frequency

The shape of the power spectrum in the case of a sinusoidal signal with a constant frequency derivative $\dot{\nu}$ is mainly determined by the number of bins in the power spectrum that the frequency changes during the observation ($\dot{\nu}T^2$). In general, the power spectral peak will have a width equal to this; the power will be more or less equally spread out over the peak (Fig. 6.9).

Averaging over all possible phases of the sinusoid and the over the initial frequency offset x one finds that the highest power in the power spectrum drops rapidly for $\dot{\nu}T^2 > 2$ (Fig. 6.10). As we have seen in Section 5.3, the frequency resolution of the power spectrum should be degraded to match the width of the peak for optimal sensitivity.

Another type of frequency variations that is encountered in practice is that caused by orbital motion. In this case, the frequency will change periodically with an amplitude $\nu_0 K/c$ (where ν_0 is the rest frequency, K the velocity amplitude of the orbital motion and c light speed) and a period equal to the orbital period $P_{\rm orb}$. If the observation is much shorter than the orbital period: $T \ll P_{\rm orb}$, then we approximately recover the previous case with $\dot{\nu} \approx \nu_0 (K/c)(2\pi/P_{\rm orb}) \cos \Phi_{\rm orb}$, where $\Phi_{\rm orb}$ is the orbital phase.

If on the other hand $T \gtrsim P_{\rm orb}$ then the power spectrum gets quite complicated, a central peak at ν_0 being accompanied by many side lobes with a separation $1/P_{\rm orb}$. Often, one tries to deal with the complexities of this situation by cutting up the observation into M segments and demanding that the change of frequency caused by the orbital motion is less than one frequency bin in any individual power spectrum. This implies

$$T < \sqrt{\frac{cP_{\rm orb}}{2\pi K \nu_0}}. \tag{6.7}$$

However, this condition does not prevent the power spectral peak from moving from one bin to the next in successive power spectra, so that a broad peak in the average spectrum in any case results.

An in principle very powerful method to deal with variable-frequency signals is "precorrecting": correcting the arrival times of the data for different assumed binary orbital parameters or frequency derivatives in the hope to recover a constant-frequency signal. A large gain in sensitivity can in principle be attained with this method. However, often the number of different sets of parameters one needs to try is prohibitive.

7. COMPREHENSIVE POWER SPECTRAL SEARCHES

In the previous sections we have considered how to optimize power spectral searches to finding specific features of known width. Often, the signal(s) one is looking for are not sufficiently closely constrained to predict the width of the feature in the power spectrum, or one wishes to look for source variability in general. In that case, one needs to perform a search of the power spectrum that is sensitive to narrow as well as broad features.

A reasonable approach seems to search the power spectrum several times, using different values of W. The trials in these searches will obviously be strongly interdependent, so that the detection levels will again have to be determined using simulations, and should take into account the total number of trials in the entire search. The degree of interdependence of the trials, and therefore the detection levels, will depend on the particular **search scheme** that is employed, and in particular, on the values of W and on the overlap between the trials. As different

values of W will occur in a search, the distribution of the noise powers will differ between the trials. It seems reasonable, therefore, to choose the detection level different for each value of W, in such a way that the *a priori* chance ϵ_{scheme} **in one trial** to exceed the detection level is equal for all trials in the entire search. Instead of one detection level P_{detect}, there will now be several different ones $P_{\text{detect},W}$ corresponding to the different values of W that occur. These detection levels are given by

$$\epsilon_{\text{scheme}} = Q(MWP_{\text{detect},W}|2MW). \tag{7.1}$$

Note that the number of times the detection levels are actually exceeded in the search of a pure noise spectrum is **not** equal to $\epsilon_{\text{scheme}}N_{\text{trial}}$ because of the strong interdependence of the trials. Therefore, ϵ_{scheme} can not be simply related to the required confidence level of detection, but has to be determined from simulations of the search scheme. ϵ_{scheme} is just a compact way of summarizing the different detection levels $P_{\text{detect},W}$ applicable to the particular search scheme employed.

In Fig. 7.1 ϵ_{scheme} corresponding to a 90% confidence detection level is plotted for two different search schemes *vs.* the number of independent powers in the power spectrum. In both schemes the values chosen for W were (1,2,4,8,...,$N/2$). In one scheme (indicated by filled circles) a 50% overlap between trials of the same width was used; in this scheme $N_{\text{trial}} \approx 3N/2$, (the number of powers in the spectrum, or the number of trials in a (1,0) scheme is $N/2$). In the other scheme (crosses) no overlap was allowed; in this scheme $N_{\text{trial}} \approx N$.

The values of ϵ_{scheme} in Fig. 7.1 were determined from simulations of many noise power spectra by assuming, for each power spectrum, a series of different values of ϵ_{scheme}, calculating for each value of ϵ_{scheme} the detection levels $P_{\text{detect},W}$ corresponding to each value of W using Eq. 7.1, and then testing the power spectrum for averaged powers exceeding these levels. The value of ϵ_{scheme} corresponding to a 90% confidence detection was then determined as that value for which only 10% of the spectra showed an excess over a $P_{\text{detect},W}$.

The way in which one would employ in practice the results of such simulations would be to use the ϵ_{scheme} obtained by simulations of the appropriate search scheme to calculate the various $P_{\text{detect},W}$ values with Eq. 7.1, and then test the spectrum, following the search scheme, against these detection levels. An excess would have well-defined statistical properties (it would constitute a detection at the required confidence level), contrary to the method where likely-looking excesses are selected by eye from plots of the power spectrum and tested *a posteriori* for significance using rules-of-thumb. A method such as described here has the further advantage that it is sensitive to "everything". It is of course somewhat less sensitive to a specific signal than a specialized test.

Once the simulation process has been set up, search schemes of this kind can

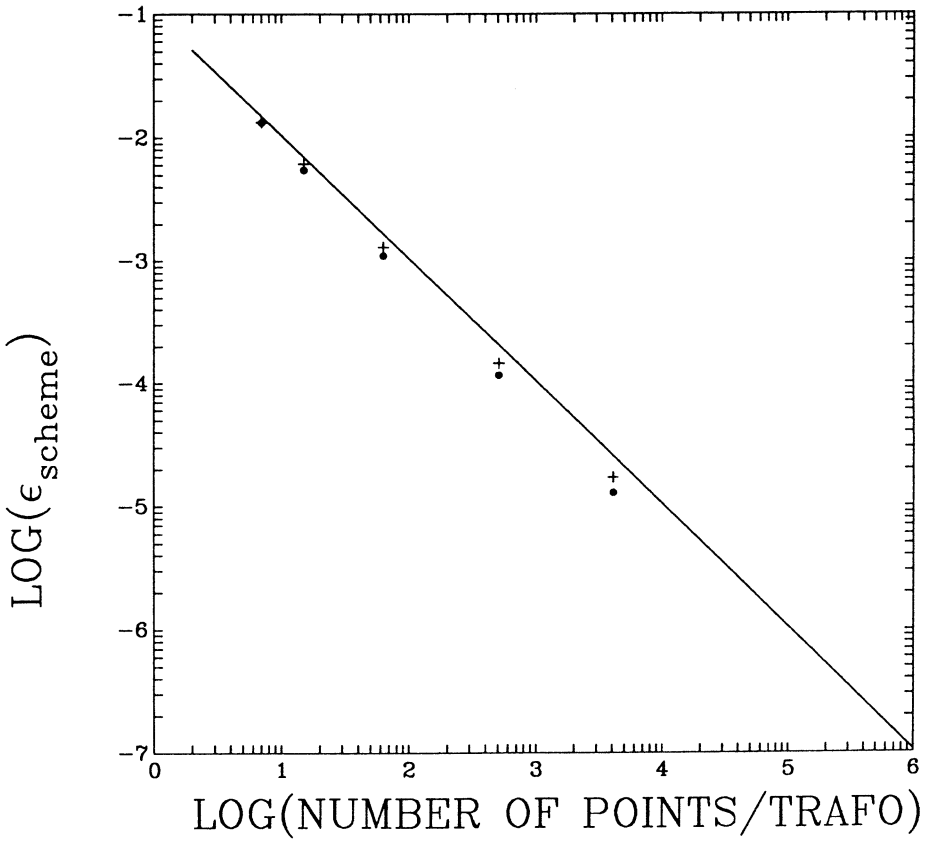

Fig. 7.1. *The a priori chance to exceed the detection level in one trial, from which one must calculate the detection levels as a function of W to obtain an effective 90% confidence detection level. Two "dependent sample" schemes (symbols) and the (1,0) scheme (curve) are illustrated. See text for a description of the dependent sample schemes. Each point corresponds to 10^4 simulated power spectra. Error bars are smaller than the points.*

easily be extended to include searches where W itself depends on j (for example, because the features searched for have a known $\Delta\nu/\nu$) or where searches are performed of two-dimensional arrays of power spectra (for example when searching dynamic power spectra for transient phenomena).

ACKNOWLEDGEMENTS

This paper benefitted from stimulating discussions with many of the participants of this ASI, notably Don Backer and Ocke de Jager, and with Luigi Stella.

REFERENCES

Andrews, D., Stella, L., 1985, *EXOSAT Express* **10**, 35.

Bloomfield, P., 1976, *Fourier Analysis of Time Series: an Introduction*, (John Wiley & Sons – New York).

Bracewell, R., 1965, *The Fourier Transform and its Applications*, (McGraw-Hill).

Deeter, J.E., 1983, *Astrophys. J.* **281**, 482.

Cullers, D.K., Oliver, B.M., Day, J.R. and Olsen, E.T., 1984, *NASA Tech. Paper* **2244**, 49.

Deeter, J.E., and Boynton, P.E., 1982, *Astrophys. J.* **261**, 337.

Jenkins, G.M. and Watts, D.G., 1968, *Spectral Analysis and its Applications*, (Holden-Day – Oakland).

Leahy, D.A., Darbro, W., Elsner, R.F., Weisskopf, M.C., Sutherland, P.G., Kahn, S. and Grindlay, J.E., 1983, *Astrophys. J.* **266**, 160.

Lewin, W.H.G., van Paradijs, J. and van der Klis, M., 1988, *Space Science Reviews*, in press (Paper 1).

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T., 1986, *Numerical Recipes*, (Cambridge University Press).

Tennant, A.F., 1987, *MNRAS* **226**, 963.

Vaughan, B.A., van der Klis, M., Wood, K.S., Norris, J.P., Hertz, P., Michelson, P.F., van Paradijs, J., Lewin, W.H.G., Mitsuda, K., Penninx, W., 1994, *Astrophys. J.*, in press.

Weisskopf, M.C., 1985, talk presented at Workshop *Time Variability in X-Ray and Gamma-Ray Sources*, Taos NM, USA. ex ex