Draft version September 25, 2019
Typeset using LaTeX **modern** style in AASTeX63

Optimal Detection of Stochastic Signals

Will M. Farr[1,2]

[1] *Center for Computational Astronomy, Flatiron Institute, New York NY 10010, United States*
[2] *Department of Physics and Astronomy, Stony Brook University, Stony Brook NY 11794, United States*

ABSTRACT

I derive an optimal statistic for detection of a common stochastic signal in independent data streams. I compute the SNR in terms of the power spectral density (PSD) of the common signal and each stream's noise, and give several useful limits.

1. FORMALITIES

Imagine we have a stochastic stationary zero-mean Gaussian signal, $h(t)$, whose PSD is $P_h(f)$:

$$P_h(f) = \lim_{T \to \infty} \frac{1}{T} \left| \int_{-T}^{T} \mathrm{d}t \, e^{-2\pi f t} h(t) \right|^2 . \tag{1}$$

The signal is linearly projected into two data streams[1] with independent stochastic stationary zero-mean Gaussian noise:

$$s_1(t) = A_1 h(t) + n_1(t) \tag{2}$$

and

$$s_2(t) = A_2 h(t) + n_2(t), \tag{3}$$

where the PSD of each stream's noise is

$$P_i(f) = \lim_{T \to \infty} \frac{1}{T} \left| \int_{-T}^{T} \mathrm{d}t \, e^{-2\pi f t} n_i(t) \right|^2 , \tag{4}$$

and the signal and both noise streams are independent:

$$p\left(h, n_1, n_2\right) = p\left(h\right) p\left(n_1\right) p\left(n_2\right). \tag{5}$$

Stationarity implies that the Fourier components are independent; Gaussianity and the PSDs imply that

$$\tilde{h}(f) \sim N\left[0, \sqrt{T P_h(f)}\right] \tag{6}$$

wfarr@flatironinstitute.org

[1] The formalism here generalizes easily to an arbitrary number of data streams.

and

$$\tilde{s}_i(f) \sim N\left[A_i\tilde{h}(f), \sqrt{TP_i(f)}\right], \tag{7}$$

where $N[\mu, \sigma]$ is a Gaussian distribution with mean μ and standard deviation σ. Integrating out the (unknown) signal h[2], we have the marginal likelihood for the data streams

$$\mathcal{L}\left(s_1(f), s_2(f) \mid P_h(f)\right) = \int \mathrm{d}h(f)\, N\left[A_1\tilde{h}(f), \sqrt{TP_1(f)}\right](s_1(f)) \\ \times N\left[A_2\tilde{h}(f), \sqrt{TP_2(f)}\right](s_2(f))\, N\left[0, \sqrt{TP_h(f)}\right](h(f)). \tag{8}$$

A bit of algebra reveals

$$\mathcal{L}\left(s_1, s_2 \mid P_h\right) = \\ \frac{1}{2\pi\sqrt{S_h\left(A_2^2S_1 + A_1^2S_2\right) + S_1S_2}} \exp\left[-\frac{\left(A_2s_1 - A_1s_2\right)^2 S_h + s_2^2S_1 + s_1^2S_2}{2\left(\left(A_2^2S_1 + A_1^2S_2\right)S_h + S_1S_2\right)}\right], \tag{9}$$

where we have suppressed the dependence on f for the moment, and absorbed the T-dependence into S via

$$S_x \equiv TP_x \tag{10}$$

(i.e. S_x is the variance of $\tilde{x}$).

The maximum-likelihood estimator of P_h is

$$\hat{P}_h \equiv \frac{1}{T}\frac{\left(A_2S_1s_2 + A_1S_2s_1\right)^2 - S_1S_2\left(A_2^2S_1 + A_1^2S_2\right)}{\left(A_2^2S_1 + A_1^2S_2\right)^2}. \tag{11}$$

The expected value of $\hat{P}_h$ is

$$\left\langle \hat{P}_h \right\rangle = P_h \tag{12}$$

(i.e. $\hat{P}$ is an un-biased estimator of P_h—whatever that is worth) and the variance is

$$\operatorname{var}\hat{P}_h = \frac{2\left(S_h\left(A_2^2S_1 + A_1^2S_2\right) + S_1S_2\right)^2}{T^2\left(A_2^2S_1 + A_1^2S_2\right)^2} = \frac{2\left(P_h\left(A_2^2P_1 + A_1^2P_2\right) + P_1P_2\right)^2}{\left(A_2^2P_1 + A_1^2P_2\right)^2} \tag{13}$$

Note that the variance in Eq. (13) is independent of T, so that the signal to noise ratio (SNR) *in a single frequency bin* is independent of time.

Typically we will be estimating the power in some range of frequencies, not just in a single bin. Because the different frequencies are independent (stationarity), the optimal estimator for the integral of $P_h(f)$ over some range is given by

$$\int_{f_0}^{f_1} \mathrm{d}f\, \hat{P}_h(f) \simeq \sum_{f=f_0}^{f_1} \Delta f \hat{P}_h(f), \tag{14}$$

[2] See Cornish & Romano (2013).

where $\Delta f = 1/T$ is the frequency resolution implied by an observation over a time T. The mean of this estimator is the integral of P_h over the corresponding interval; the variance of this estimator is given by

$$\sum_{f=f_0}^{f_1} \Delta f^2 \operatorname{var} \hat{P}_h(f), \tag{15}$$

and therefore the SNR of such a measurement is given by

$$\rho = \frac{\sqrt{T} \int_{f_0}^{f_1} \mathrm{d}f \, P_h(f)}{\sqrt{\int_{f_0}^{f_1} \mathrm{d}f \, \operatorname{var} \hat{P}_h(f)}} \tag{16}$$

with $\operatorname{var} \hat{P}_h$ given by Eq. (13). We see that the SNR for a measurement over some finite bandwidth grows with $\sqrt{T}$, as it should.

2. USEFUL LIMITS

2.1. *Signal-Dominated Limit*

If $A_{1,2}^2 P_h \gg P_{1,2}$ over the relevant range of frequencies, we have

$$\operatorname{var} \hat{P}_h \simeq 2P_h^2, \tag{17}$$

and the per-bin SNR asymptotes to 1/2. The finite-bandwidth SNR, Eq. (16), becomes

$$\rho \simeq \sqrt{\frac{T\,(f_1 - f_0)}{2}} \frac{\langle P_h \rangle}{\sqrt{\langle P_h^2 \rangle}} \tag{18}$$

where angle brackets indicate an average over frequencies $f_0 \leq f \leq f_1$. Note that in this limit the SNR is independent of the amplitude of P_h.

2.2. *Noise-Dominated Limit*

In the opposite limit, $A_{1,2} P_h \ll P_{1,2}$, we have

$$\operatorname{var} \hat{P}_h \simeq \frac{2P_1^2 P_2^2}{\left(A_2^2 P_1 + A_1^2 P_2\right)^2} = \frac{2}{\left(A_1^2/P_1 + A_2^2/P_2\right)^2}, \tag{19}$$

and the finite-bandwidth SNR becomes

$$\rho \simeq \sqrt{\frac{T\,(f_1 - f_0)}{2}} \frac{\langle P_h \rangle}{\sqrt{\left\langle \frac{1}{\left(A_1^2/P_1 + A_2^2/P_2\right)^2} \right\rangle}} \tag{20}$$

and the SNR is reduced compared to the signal-dominated case by a factor

$$\alpha \sim \frac{A_1^2 P_h}{P_1} + \frac{A_2^2 P_h}{P_2}. \tag{21}$$

REFERENCES

Cornish, N. J., & Romano, J. D. 2013, PhRvD, 87, 122003, doi: 10.1103/PhysRevD.87.122003