<div align="center">

CALCULATING TOTAL POWER FROM VLBI DATA

Lindy Blackburn[1], Michael D. Johnson[1], Andre Young[1]

</div>

1. TOTAL POWER FROM 2-BIT DATA

Power received by the antenna is proportional to $\sigma_V^2 = \langle V(t)^2 \rangle$. We record 2-bit quantized data $y[n]$ where,

$$y[n] = \begin{cases} 1 & -\infty < V(t) < -V_0 \\ 2 & -V_0 < V(t) < 0 \\ 3 & 0 < V(t) < +V_0 \\ 4 & +V_0 < V(t) < +\infty \end{cases} \quad \text{at } t = n/f_s \tag{1}$$

Here V_0 is a quantization threshold, and for VLBI correlation is often tuned to the rms standard deviation of the signal σ_V. f_s is the Nyquist sampling rate of the band limited signal. We want to estimate changing signal power that leads to small variations in $\sigma_V(t) = \sigma_0 + \delta_\sigma(t)$ by using ensemble statistics of the quantized data $y[n]$ at a fixed 2-bit threshold.

For simplicity, let $x(t)$ measure voltage relative to that measured for the average *SEFD* of the antenna, with $\sigma_x(t) = 1 + \epsilon(t)$. The total power in the signal relative to baseline is $\sigma_x^2 = 1 + 2\epsilon + O(\epsilon^2)$ for small changes about the *SEFD*. The expected fraction of total values that fall within quantization levels 2 and 3 are,

$$p = \int_{-x_0}^{x_0} G(x, \sigma = 1 + \epsilon)\, dx = \int_{-x_0/(1+\epsilon)}^{x_0/(1+\epsilon)} G(x', \sigma = 1)\, dx' \tag{2}$$

with x_0 corresponding to the quantization threshold and $G(x, \sigma)$ representing a zero-mean Gaussian distribution. Expanding p about $\epsilon = 0$ gives,

$$p(\epsilon) = p_0 + \epsilon \frac{dp}{d\epsilon} = p_0 - 2\epsilon x_0\, G(x_0, 1) + O(\epsilon^2) \tag{3}$$

and total power as a function of expected fractional state counts p,

$$\sigma_x^2 = 1 + 2\epsilon = 1 + \frac{p_0 - p}{x_0\, G_0} \tag{4}$$

where $G_0 = G(x_0, 1)$. Given sufficient data, we assume that p_0 can be measured accurately and solve for x_0,

$$x_0 = \sqrt{2}\, \mathrm{erf}^{-1}(p_0) \tag{5}$$

with erf^{-1} representing the inverse error function.

For a series of N independent discrete observations $x[n]$ each with probability p of being counted, the probability of observing a given number of counts k follows a binomial distribution,

$$P(k) = \binom{N}{k} p^k (1-p)^{N-k} \tag{6}$$

The binomial distribution has mean $\langle k \rangle = Np$ and variance $\sigma_k^2 = Np(1-p)$. For a small expected deviation around a large expected $k_0 = Np_0$, we can assume the inferred variance on estimated fraction $\hat{p}$ is constant,

$$\hat{p} = \frac{k}{N} \quad \sigma_{\hat{p}}^2 = \frac{p_0(1-p_0)}{N} + O\left(\frac{\epsilon}{N}\right) \tag{7}$$

This translates into an error on estimated power,

$$\hat{\sigma}_x^2 = 1 + \frac{p_0 - \hat{p}}{x_0\, G_0} \quad \mathrm{var}(\hat{\sigma}_x^2) = \frac{p_0(1-p_0)}{N\,(x_0 G_0)^2} \tag{8}$$

The minimum $\mathrm{var}(\hat{\sigma}_x^2) = 3.07/N$ occurs at $x_0 = 1.42$. At $x_0 = 1$ typical for VLBI, the variance is $3.70/N$.

Equation 8 for total power can be written in terms of constant factors applied to the counts in the inner p and outer $1 - p$ bins,

$$\hat{\sigma}_x^2 = \left[1 - \frac{1 - p_0}{x_0 G_0}\right]\hat{p} + \left[1 + \frac{p_0}{x_0 G_0}\right](1 - \hat{p}) \tag{9}$$

lblackburn@cfa.harvard.edu, mjohnson@cfa.harvard.edu, ayoung@cfa.harvard.edu
[1] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
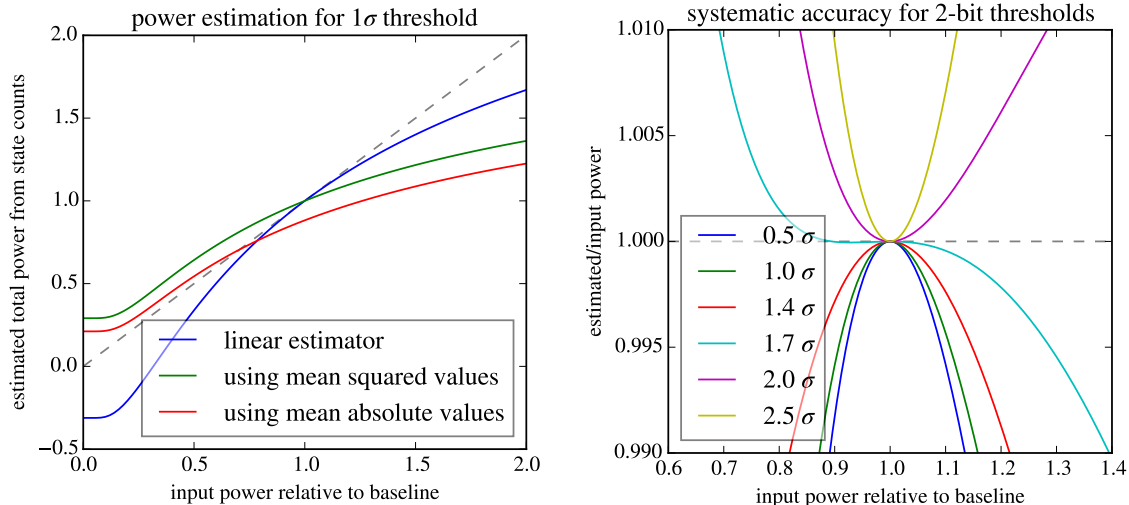
Figure 1. Accuracy of estimated power as input power is varied. The estimated power is a linear function of the 2-bit state counts, with coefficients derived by either linearizing the error function about the threshold value, or by using mean squared values within each quantization level. The use of mean squared values returns the correct baseline power, but underestimates the relative ampltiude of fluctuations, shown by the reduced slope of the green line in the left plot. The choice of quantization threshold affects behavior of the systematic error in the linear approximation. The higher-order terms are minimized at a threshold around 1.7σ. Statistical errors are not shown in these plots. For 32768 samples ($8\mu s$ at 4.096 GHz), equation 8 gives about 1% statistical error at 1σ quantization threshold.

For a quantization threshold of $x_0 = 1$ relative to average baseline power, the expected fractional inner counts is $p_0 = 0.68$, and value of the Gaussian distribution $G_0 = 0.24$. The two factors become -0.31 and 3.82 which are different than the mean squared values of all the samples within each quantization level. They are also different from the square of the mean values,

$$\langle x^2 \rangle = \begin{cases} 0.29 & |x| < 1 \\ 2.53 & |x| > 1 \end{cases} \qquad \langle |x| \rangle^2 = \begin{cases} 0.46^2 = 0.21 & |x| < 1 \\ 1.53^2 = 2.32 & |x| > 1 \end{cases} \tag{10}$$

At baseline power where $p = 0.68$, the mean squared weights give the correct estimate of $\sigma_x^2 = 1$ but understimate the power in small fluctuations about baseline. This is because small fluctuations in power also affect the mean squared values by changing the shape of the distribution, but this is not tracked by the constant factors. The effect is difficult to track because it depends on the very signal energy that is being estimated. Coefficients derived from mean values rather than mean squared values result in a biased estimate of baseline total power. The systematics affecting both direct energy calcualtions get smaller for high-bit data with many narrow bins.

2. TOTAL POWER IN THE CONTINUOUS LIMIT

Consider a random continuous signal with finite bandwidth $\Delta\nu$ that has not been quantized. A signal of duration T is represented by $N/2 = T\Delta\nu$ independent complex Fourier coefficients $\tilde{x}[k]$ or equivalently N real time samples $x[n]$. The total measured signal power $E = \sum x^2[n]$ follows a χ^2 distribution with N degrees of freedom,

$$\langle E \rangle = N\sigma_x^2 \quad \text{var}(E) = 2N\sigma_x^2 \tag{11}$$

As in the quantized case, we assume the baseline noise power is well measured and corresponds to $\sigma_x^2 = 1$. The estimator for total power is,

$$\hat{\sigma}_x^2 = \frac{E}{N} \tag{12}$$

For small deviations from baseline with $\sigma_x^2 = 1 + 2\epsilon$, measurement noise is,

$$\text{var}(\hat{\sigma}_x^2) = \frac{2}{N} + O\left(\frac{\epsilon}{N}\right) \tag{13}$$

The loss in S/N going form the continuous case to the 2-bit quantized case is $\sqrt{2/3.07} = 19.3\%$ for the ideal case, and 26.5% for the typical VLBI 1σ quantization threshold case. The continuous limit can be acheived for high-bit data.