HERA Memo #099
Received: 27 May 2021

Statistical distribution of a single coherently-averaged delay spectrum bandpower

Phil Bull, May 25, 2021

Abstract

In this note we derive the joint probability density function of the real and imaginary parts of a single delay spectrum bandpower, assuming that the input visibilities have been coherently averaged, and contain only EoR signal and white noise. This distribution has an analytic expression, and differs from what would be expected for a bandpower derived from signal-only visibilities.

Background

One route towards measuring the delay spectrum with HERA involves performing a coherent average of visibilities within each redundant group and then calculating the cross-spectrum of neighbouring ('interleaved') time samples. The idea is that the sky signal will change very little between neighbouring time samples (assuming a $\sim 20-40$ sec integration time per sample), but the noise between the samples will be independent. Coherent averaging has the advantage of beating down the noise more rapidly than incoherent averaging, and also reduces the computational requirements of power spectrum estimation (since only one cross-spectrum is required per redundant group).

We consider a scenario where visibilities have been transformed into delay space, and are uncorrelated between different delay modes (i.e. ignoring the effects of tapering or missing channels due to flagging). With the highest delay modes in mind, we model the delay-space visibility at a given LST as

$$V_a = s + n_a, \tag{1}$$

where s is a mean-zero Gaussian random variate that models the EoR signal, and n_a is also a mean-zero Gaussian r.v. that models the noise. Both are complex valued. If the visibility is formed from a coherent average of visibilities from within the same redundant group, $V_a = (1/n)\sum_p V_p$, the expected distribution of V_a is still a Gaussian r.v. with mean zero, where we have neglected the effects of possible systematics. At a neighbouring time b, we make the approximation

$$V_b \approx s + n_b, \tag{2}$$

where $s = s_a \approx s_b$. We may then form a product of these visibilities as an estimate of the delay spectrum at this delay,

$$P_{ab} = V_a V_b^\dagger. \tag{3}$$

Distribution of a product of correlated complex Gaussian random variates

We wish to derive the probability density function for the real and imaginary parts of the bandpower P_{ab}. Starting from Eq. 3, we can see that this is constructed from the product of two complex Gaussian random variates that are partially correlated (since they both share the same realisation of s, but different noise realisations),

$$P_{ab} = (s + n_a)(s + n_b)^\dagger, \tag{4}$$

where the expectation value of the bandpower is

$$\langle P_{ab} \rangle = \langle V_a V_b^\dagger \rangle = \langle s s^\dagger \rangle. \tag{5}$$

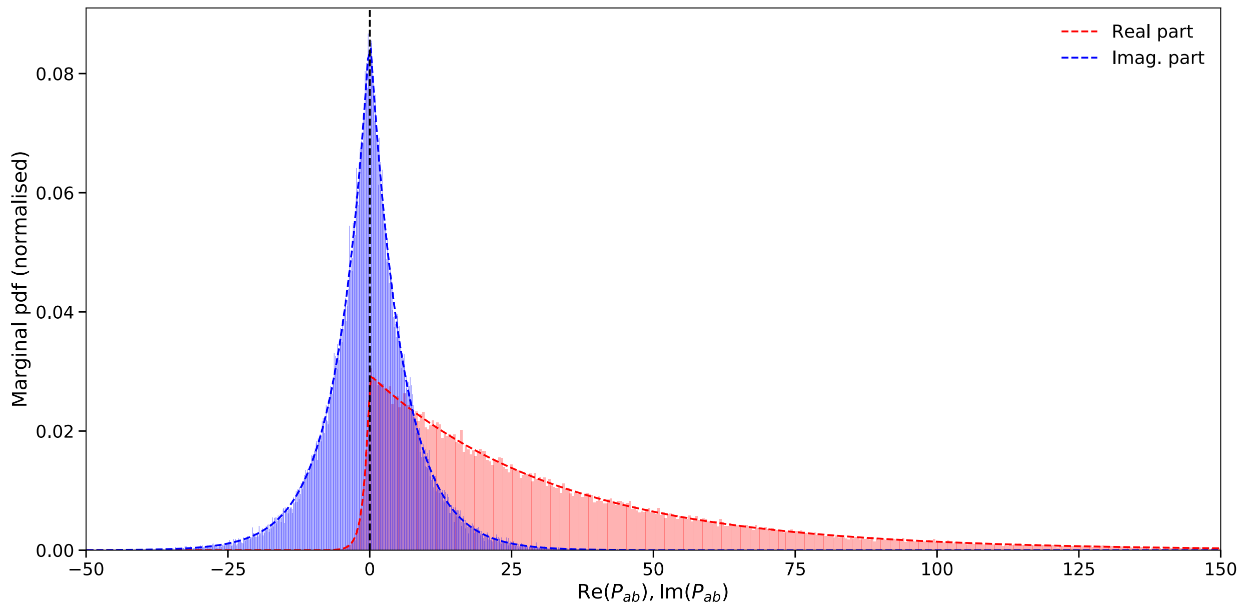

Figure 1: Shaded regions: Histograms of real and imaginary parts of bandpower P_{ab}, calculated from 10^5 random realisations with $C_s = 4^2 + 4^2 = 32$ and $C_{n,a} = C_{n,b} = 1^2 + 1^2 = 2$. Dashed lines: Normalised marginal distributions of the real and imaginary bandpower components calculated according to Eq. 8.

Note that no averaging over time samples has been performed yet. An appropriate probability distribution for a product of correlated complex Gaussian r.v's was derived by [1]; for complex visibilities with (expectation) zero mean, this is

$$p(z_R, z_I) = \frac{2}{\pi C_{aa} C_{bb} \sqrt{1 - |\rho|^2}} \exp\left(\frac{2\,\mathrm{Re}(\rho z^\dagger)}{(1 - |\rho|^2)\sqrt{C_{aa} C_{bb}}}\right) K_0 \left(\frac{2|z|}{(1 - |\rho|^2)\sqrt{C_{aa} C_{bb}}}\right), \tag{6}$$

where $z = z_R + i z_I$ is a particular realisation of P_{ab}. In the expression above, K_0 is a modified Bessel function of the second kind; beware a singularity at $z = 0$. The components of the complex covariance matrices are

$$C_{aa} = C_s + C_{n,a}; \qquad C_{bb} = C_s + C_{n,b}; \qquad C_{ab} = C_s, \tag{7}$$

where the signal and noise covariances are defined as $C_s = \langle s s^\dagger \rangle; C_n = \langle n n^\dagger \rangle$, and the correlation coefficient is $\rho = C_{ab}/\sqrt{C_{aa} C_{bb}}$.

Note that Eq. 6 is the joint pdf of the real and imaginary parts of the bandpower; the marginal distribution for either real or imaginary part only can be calculated as (e.g. for the real part)

$$p(z_R) = \mathcal{N} \int p(z_R, z_I)\, dz_I, \tag{8}$$

where $\mathcal{N}$ is a normalising factor that must be calculated if $p(z_R)$ is to be used as a pdf.

Fig. 1 shows that the expression from Eq. 8 agrees very well with the marginal distributions derived from 100,000 Monte Carlo realisations of the complex bandpowers, in this case with an EoR signal that has several times the variance of the noise components. We have checked that similarly good agreement exists for a range of other values for the signal and noise variances. Note that no fitting or rescaling of the distributions was required; the histograms and marginal distribution calculations were performed completely separately and agree very well.

Inclusion of correlated 'systematics' terms

We will now add a 'systematics' component to the visibilities that is correlated between neighbouring time samples. For simplicity we will model it as having expected mean zero, as otherwise the maths becomes much messier. Our model for visibilities containing a Gaussian systematics r.v. ϵ is $V_a = s + n_a + \epsilon_a$, where

$$\langle \epsilon \rangle = 0; \quad \langle \epsilon_i \epsilon_j^\dagger \rangle = C^\epsilon_{ij}; \quad \langle s\epsilon^\dagger \rangle = \langle n\epsilon^\dagger \rangle = 0, \tag{9}$$

where $i, j \in \{a, b\}$. Since the visibilities remain Gaussian even after the addition of the systematics term, all we need to do is update the covariance terms to obtain a suitable pdf,

$$C_{aa} = C_s + C_{n,a} + C^\epsilon_{aa}; \quad C_{bb} = C_s + C_{n,b} + C^\epsilon_{bb}; \quad C_{ab} = C_s + C^\epsilon_{ab}. \tag{10}$$

References

[1] Y. Li, Q. He R. S. Blum, *On the Product of Two Correlated Complex Gaussian Random Variables*, IEEE Sig. Proc. L 27 (2020).