Correlation function of a clipped Gaussian field

S. Plaszczynski

December 21, 2017

Let X follow a central normal distribution. We wish to avoid the unphysical region $X < -1$ by applying to the random variable some local transform

$$Y = L(X) \tag{1}$$

Since $X(\vec{r})$ actually represents a Gaussian field with a given covariance matrix ($\mathbf{\Sigma}$,) a pair of values follows the (central) bivariate Gaussian distribution

$$(X_1, X_2) \sim \mathcal{N}_2(0, \mathbf{\Sigma}), \tag{2}$$

where

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_G^2 & \xi_G \\ \xi_G & \sigma_G^2 \end{pmatrix}. \tag{3}$$

The covariance of the transformed Y variable

$$\xi_Y = \mathbb{E}\left[y_1 y_2\right] - \mathbb{E}^2[y] \tag{4}$$

can be computed using probability conservation

$$\begin{aligned} \mathbb{E}\left[y_1 y_2\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1 y_2 f_Y(y_1, y_2) dy_1 dy_2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} L(x_1) L(x_2) \mathcal{N}_2(x_1, x_2) dx_1 dx_2. \end{aligned} \tag{5}$$

Eq.(5) is rarely analytical (but for the log-normal case why it is often used).

Bel et al. (2016) introduces a very interesting method to compute this integral numerically. Let us first assume that $\sigma = 1$. Then one can use a formula from an obscure German mathematician (Mehler), which quoted in a probabilistic language (Kibble 1945) with our notations reads

$$\mathcal{N}_2(x_1, x_2) = \mathcal{N}(x_1)\mathcal{N}(x_2) \sum_{n=0}^{\infty} H_n(x_1) H_n(x_2) \frac{\xi_G^n}{n!}, \tag{6}$$

$\mathcal{N}$ representing here the Normal distribution, and H_n the (probabilistic) Hermite polynomials which are orthogonal wrt to the Gaussian measure:

$$\int_{-\infty}^{+\infty} H_n(x) H_m(x) \mathcal{N}(x) dx = n! \delta_{nm} \tag{7}$$

Then if we decompose the local transform onto them

$$L(x) = \sum_{n=0}^{\infty} h_n H_n(x), \tag{8}$$

we obtain the series expansion

$$\xi_Y = \sum_{n=0}^{\infty} \alpha_n \xi_G^n, \tag{9}$$

where (here) $\alpha_n ! h_n^2$.

The h_n's are obtained from the orthogonality relation through:

$$h_n = \frac{1}{n!} \int_{-\infty}^{+\infty} L(x) H_n(x) \mathcal{N}(x) dx \tag{10}$$

Now if $\sigma^2 \neq 1$ one rescales the variables $x'_1 = x_1/\sigma, x'_2 = x_1/\sigma$ so that Eq.(5) can be written as

$$\mathbb{E}\left[y_1 y_2\right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} L(\sigma x'_1) L(\sigma x'_2) \mathcal{N}_2(x'_1, x'_2) dx'_1 dx'_2 \tag{11}$$

but now

$$(x'_1, x'_2) \sim \mathcal{N}_2 \left[0, \begin{pmatrix} 1 & \frac{\xi_G}{\sigma^2} \\ \frac{\xi_G}{\sigma^2} & 1 \end{pmatrix}\right]. \tag{12}$$

We can now safely apply Mehler's formula and the generalized Eq.(9) coefficients read:

$$\alpha_n = \frac{n! h_n^2}{\sigma^{2n}} \tag{13}$$

$$h_n = \frac{1}{n!} \int_{-\infty}^{+\infty} L(\sigma x) H_n(x) \mathcal{N}(x) dx. \tag{14}$$

The $n = 0$ coefficient is

$$h_0 = \int_{-\infty}^{+\infty} L(\sigma x) H_0(x) \mathcal{N}(x) dx = \frac{1}{\sigma} \int_{-\infty}^{+\infty} L(x) \mathcal{N}(x/\sigma) dx \tag{15}$$

since $H_0 = 1$. The 1D conservation of probability being

$$p(y) dy = \mathcal{N}(x) dx = \sigma \mathcal{N}(x/\sigma) dx, \tag{16}$$

and since $L(x) = y$, we find that

$$h_0 = \mathbb{E}\left[y\right]. \tag{17}$$

Recalling Eq.(4) we see that the squared mean value cancels with h_0^2 so that Eq.(9) with $\alpha_0 = 0$ is actually valid for *any local transform*. This makes sense since we introduced no stochasticity that could add some extra DC power.

Log-normal case

$$L(x) = e^{x - \frac{\sigma^2}{2}} - 1 \tag{18}$$

Eq.(5) can be computed directly (using for instance the multi-normal moments generating function) which gives

$$\xi_Y = e^{\xi_G} - 1 \tag{19}$$

Let's see how the previous approach works:

$$\begin{aligned} h_n &= \frac{1}{n!}\int_{-\infty}^{+\infty}(e^{\sigma x-\frac{\sigma^2}{2}}-1)\mathcal{N}(x)H_n(x)dx \\ &= \frac{1}{n!}\left[\int_{-\infty}^{+\infty}\mathcal{N}(x-\sigma)H_n(x)dx-\int_{-\infty}^{+\infty}\mathcal{N}(x)H_n(x)dx\right] \\ &= \frac{1}{n!}[\sigma^n-\delta_{n0}]. \end{aligned} \tag{20}$$

where we used the fact that $H_0(x)=1$ and $H_n(x+\sigma)=\sum_{k=0}^{n}\binom{n}{k}\sigma^k H_{n-k}(x)$.

The series expansion then reads (Eq.(13))

$$\alpha_n = \left[\frac{\sigma^n}{n!}\right]^2\frac{n!}{\sigma^{2n}}-\delta_{n0} \tag{21}$$

$$\rightarrow \xi_Y = \sum_{n=1}^{\infty}\frac{\xi_G}{n!} = e^{\xi_G}-1 \tag{22}$$

Clipping

$$L(x) = \begin{cases} x & x\geq -1 \\ -1 & x<-1 \end{cases} \tag{23}$$

Eq.(5) is not analytical.

The Hermite coefficients are

$$h_n = \frac{1}{n!}\left[\int_{-1/\sigma}^{+\infty}\sigma x\mathcal{N}(x)H_n(x)dx-\int_{1/\sigma}^{+\infty}\mathcal{N}(x)H_n(-x)dx\right] \tag{24}$$

which has no general analytical form. However only the very first coefficients (but 0) contribute significantly as shown on Fig 1 for $\sigma=1$.

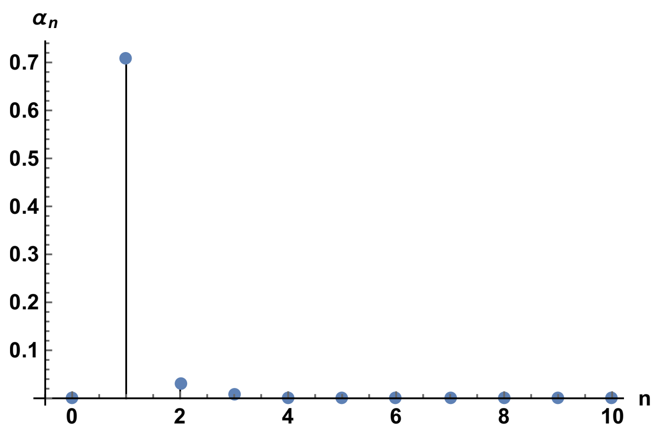

Figure 1: Coefficients entering the series expansion Eq.(9) for the -1 clipping transform ($\sigma=1$)

- $n=1$ is the linear component that is majority,
- the non-linear coefficients $n\geq 2$ drop rapidly.

This is observed for $\sigma = 1$. When σ is lower (which is in practice always the case in LSS) the transformed variable should be even more linear (less clipping) so that keeping only $n = 1$ and 2 terms is largely sufficient.

With the explicit expression of Hermite polynomials, one obtains

$$\alpha_1 = \frac{1}{4}\left(\mathrm{erf}(\frac{1}{\sqrt{2}\sigma})\right)^2 \tag{25}$$

$$\alpha_2 = \frac{e^{-\frac{1}{\sigma^2}}}{4\pi\sigma^2}, \tag{26}$$

and the correlation function of the -1 clipped field ($\sigma \leq 1$) is simply

$$\xi_y \simeq \alpha_1 \xi_G + \alpha_2 \xi_G^2 \tag{27}$$

References

Bel, J., Branchini, E., Di Porto, C., et al. 2016, A&A, 588, A51

Kibble, W. F. 1945, Mathematical Proceedings of the Cambridge Philosophical Society, 41