PLN spatial

Raphaelle Momal

May 29, 2019

A variogram in geostatistics is defined for an intrinsically stationary spatial process Z as follows, for sites s and t separated by a distance d:

$$\begin{aligned}\gamma(d) &= \frac{1}{2}\mathbb{E}[(Z(s)-Z(t))^2] \\ &= \frac{1}{2}\left(\mathbb{V}(Z(s))+\mathbb{E}[Z(s))^2+\mathbb{V}(Z(t))+\mathbb{E}[Z(t))^2\right)-\mathbb{E}(Z(s)Z(t))\end{aligned}$$

We wish to write the theoretical variogram formula for the PLN model.

1 Variogram for Y

We recall the PLN model :

$$\begin{aligned}Z_s &\sim \mathcal{N}(0,\Sigma) \\ Y_{sj}|Z_{sj} &\sim \mathcal{P}(e^{x_s\beta_j+Z_{sj}})\end{aligned}$$

In what follows we want to account for a spatial structure of tha data. This means we will need to differentiate between two sites and two species at the same time. Therefore we write $Cov(Z) = [\Gamma]_{stjk}$ where s and t are sites and j and k are species. Two models are under consideration here:

1. $\Gamma_{stjk} = \delta_{st}\sigma_{jk}$
2. $\Gamma_{stjk} = \delta_{st} + \mathbb{1}_{s=t}\sigma_{jk}$

Where δ and σ represent respectively the spatial and inter-species dependence. We assume (a) a stationarity in spatial variance, that is that $\delta_{uu} = cst$: a behavior can be spatially translated. The inter-species dependence only intervenes on a same site, so furthermore there is an assumption (b) that two species in two different sites will behave independently. For clarity sake, we write Γ_{ssjj} as Γ_j, according to the assumption of spatial variance stationarity.

[Est-ce que (b) implique que $\Gamma_{stj} = \Gamma_{st}$? À clarifier dans le modèle 1]

1.1 Y characteristics

We know the expectation, variance and covariance expressions for the PLN model. When only one species is considered, we have:

$$\begin{aligned}\mathbb{E}[Y_{sj}] &= e^{x_s\beta_j-\Gamma_j^2/2} = \mu_{sj} \\ \mathbb{V}(Y_{sj}) &= \mu_{sj}+\mu_{sj}^2(e^{\Gamma_j^2}-1) \\ \mathbb{E}[Y_{sj}Y_{tj}] &= \mu_{sj}\mu_{tj}e^{\Gamma_{stj}}\end{aligned}$$

1.2 Same species in two sites

For sites s and t at a distance d:

$$
\begin{aligned}
\gamma_Y(d) &= \frac{1}{2}\left[\mu_s + \mu_s^2(e^{\Gamma_j^2} - 1) + \mu_s^2 + \mu_t + \mu_t^2(e^{\Gamma_j^2} - 1) + \mu_t\right] - \mu_s\mu_t e^{\Gamma_{stj}} \\
&= \frac{1}{2}\left[(\mu_s^2 + \mu_t^2)e^{\Gamma_j^2} + \mu_s + \mu_t\right] - \mu_s\mu_t e^{\Gamma_{stj}}
\end{aligned}
$$

We then define $\tilde{Y_{sj}} = Y_{sj}/\mu_{sj}$. It comes that

$$
\begin{aligned}
\hat{\gamma}_{\tilde{Y}}(d) &= \frac{1}{2}\mathbb{E}\left[(\tilde{Y_{sj}} - \tilde{Y_{sj}})^2 - (\frac{1}{\mu_s} + \frac{1}{\mu_t})\right] \\
&= e^{\Gamma_j^2} - e^{\Gamma_{st}}
\end{aligned}
$$

$\hat{\gamma}_{\tilde{Y}}$ is a resembles the classical semi-variogram for a stationary spatial process except it is a difference of exponentials rather than a difference of direct variance and covariance. [indice en s ou en sj, il faut choisir]

1.3 Two species in two sites

We recall that

$$\mathbb{E}[Y_{sj}Y_{tk}] = \mu_{sj}\mu_{tk}e^{\Gamma_{stjk}}.$$

Similarly to the above paragraph, we get:

$$\mathbb{E}[(Y_{sj} - Y_{tk})^2] = \left[\mu_{sj} + \mu_{tk} + \mu_{sj}^2 e^{\Gamma_{sj}^2} + \mu_{tk}^2 e^{\Gamma_{tk}^2}\right] - \mu_{sj}\mu_{tk}e^{\Gamma_{stjk}}$$

Then,

$$\hat{\gamma}_{\tilde{Y}_{jk}}(d) = \frac{1}{2}\left(e^{\Gamma_{sj}^2} + e^{\Gamma_{tk}^2}\right) - e^{\Gamma_{stjk}}$$

2 Perspectives

- Culture sur INLA, approximation de Laplace et courbure Gaussienne
- Vraisemblance composite, éventuellement utilisée sur les arbres, pour approcher $p(Z|Y)$.