Gaussian processes in spatial statistics

Emiko Dupont

15 November 2017

What is a Gaussian process/Gaussian random field (GRF)?

Stochastic process $\{Z(s) | s \in D\}$, $D \subset \mathbb{R}^d$

- Mean: $\mu(s) = \mathrm{E}(Z(s))$
- Variance: $\mathrm{Var}(Z(s)) < \infty$

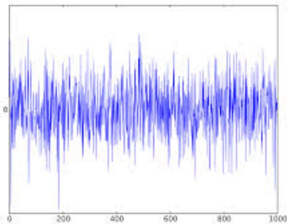
Any finite collection $\{Z(s_1), \ldots, Z(s_k)\}$ is multivariate normal:

$$\begin{bmatrix} Z(s_1) \\ \vdots \\ Z(s_k) \end{bmatrix} \sim N\left(\begin{bmatrix} \mu(s_1) \\ \vdots \\ \mu(s_k) \end{bmatrix}, \begin{bmatrix} & \mathrm{Cov}(Z(s_i), Z(s_j)) & \end{bmatrix} \right)$$

What is a Gaussian process/Gaussian random field?

Time series: $\{Z(t) | t = 0, 1, \ldots\}$

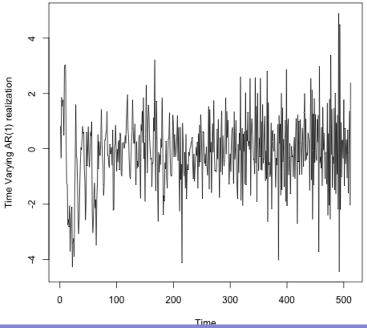
- White noise
  - $Z(t) \sim_{\text{iid}} N(0, \sigma^2)$
  - Any finite collection $\{Z(t_1), \ldots, Z(t_k)\} \sim N(\mathbf{0}, \sigma^2 I)$



What is a Gaussian process/Gaussian random field?

Time series: $\{Z(t) | t = 0, 1, \ldots\}$

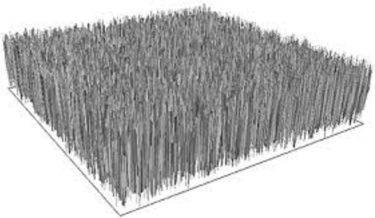
- AR(1) e.g. $Z(t)$ = closing stock price on day t
  - $Z(0) \sim N(0, \frac{\sigma^2}{1-\phi^2})$
  - $Z(t) = \phi Z(t-1) + \epsilon(t), \quad \epsilon(t) \sim_{\text{iid}} N(0, \sigma^2), \quad$ for $t = 1, 2, \ldots$



What is a Gaussian process/Gaussian random field?

Spatial field: $\{Z(s) | s \in D\}$, $D \subset \mathbb{R}^2$

- White noise
  - $Z(s) \sim_{\text{iid}} N(0, \sigma^2)$
  - Any finite collection $\{Z(s_1), \dots, Z(s_k)\} \sim N(\mathbf{0}, \sigma^2 I)$

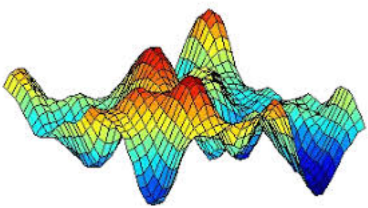


What is a Gaussian process/Gaussian random field?

Spatial field: $\{Z(s) | s \in D\}$, $D \subset \mathbb{R}^2$

- $Z(s) =$ concentration of mineral at location s
  - $\mu(s) = \mu$
  - $\mathrm{Cov}(Z(s_1), Z(s_2)) = C(|s_2 - s_1|)$ where

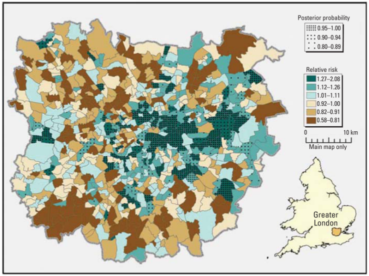
$$C(r) = exp(-r^2/R^2), \quad \text{for some range parameter } R$$



What is a Gaussian process/Gaussian random field?

Spatial field: $\{Z(i) | i = 1, \ldots, N\}$, N regions

- $Z(i) =$ relative risk of lung cancer in region i
- Covariance: Neighbouring regions more similar than those far apart



What are Gaussian processes used for?

Improve inference

- Identify spatial correlation structure/clustering
- More powerful inference by pooling data (e.g. identify time trend in river flow data)

Prediction: Given observations $Z(s_1), \ldots, Z(s_n)$

- Reconstruct entire field $Z(s)$ (e.g. global sea surface temperature)
- Estimate $\int_A Z(s) ds$ (e.g. total quantity of ore across region A from observed densities)
- Assess uncertainty of these estimates

What are Gaussian processes used for?

Applications in

- environmental sciences (e.g. assessing time trends/spatial trends in flood risk/sea ice concentration, sea temperature..., forecasting)
- geology (e.g. estimating mineral concentration for mining)
- ecology (e.g. assess fish stock to avoid overexploitation)
- epidemiology (e.g. understanding spatial distribution of diseases)
- econometrics (e.g. financial time series modelling)
- ...

Gaussian process models

A Gaussian process is completely determined by its mean and covariance

Additional model assumptions could be:

- Stationarity: Process depends only on $s_1 - s_2$
  $\mu(s) = \mu$ and $\mathrm{Cov}(Z(s_1), Z(s_2)) = C(s_1 - s_2)$
- Isotropy: Covariance depends only on $|s_1 - s_2|$
  $\mathrm{Cov}(Z(s_1), Z(s_2)) = C(|s_1 - s_2|)$

Examples of isotropic covariance functions:

- Exponential (range parameter R)
- Spherical (range parameter R) (resulting field quite spiky)
- Matérn (range parameter and smoothness parameter) (very flexible!)

(all reflect the idea that nearby observations are most similar)

Parameter estimation - method 1 (MLE)

Given $z = (z_1, \ldots, z_n)$ observations of $Z(s)$ at locations $s_1, \ldots, s_n$.

Model

$$z \sim N(X\beta, \Sigma) \quad \text{where } \Sigma = \alpha V(\theta)$$

$X =$ observed covariates at locations $s_1, \ldots, s_n$
$\beta =$ unknown coefficients of covariates
$\alpha =$ unknown overall degree of smoothing
$\theta =$ unknown parameters of the chosen covariance function

Maximum likelihood estimate of parameters

$$(\hat{\beta}, \hat{\alpha}, \hat{\theta}) = \text{argmax } l(\beta, \alpha, \theta)$$

where

$$l(\beta, \alpha, \theta) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\alpha - \frac{1}{2}\log|V(\theta)| - \frac{1}{2\alpha}(z - X\beta)^T V(\theta)^{-1}(z - X\beta)$$

Confidence intervals from asymptotic properties of MLE

Parameter estimation - method 2 (Bayesian)

Given $z = (z_1, \ldots, z_n)$ observations of $Z(s)$ at locations $s_1, \ldots, s_n$.

Model

$$\begin{aligned} z|\beta, \alpha, \theta &\sim N(X\beta, \Sigma) \quad \text{where } \Sigma = \alpha V(\theta) \\ (\beta, \alpha, \theta) &\sim \text{some prior distribution} \end{aligned}$$

$X =$ observed covariates at locations $s_1, \ldots, s_n$
$\beta =$ unknown coefficients of covariates
$\alpha =$ unknown overall degree of smoothing
$\theta =$ unknown parameters of the chosen covariance function

Posterior distribution

$$f(\beta, \alpha, \theta|z) \propto f(z|\beta, \alpha, \theta) f(\beta, \alpha, \theta)$$

For certain priors, posterior modes correspond to frequentist REML estimates

Prediction

Goal

Given: $z = (z_1, \ldots, z_n)$ observations of $Z(s)$ at locations $s_1, \ldots, s_n$
Estimate: $z_0 = Z(s_0)$ where $s_0 \notin \{s_1, \ldots, s_n\} \implies \hat{z}_0$, uncertainty of $\hat{z}_0$
(generalises to estimation of the entire field $Z(s)$ or $\int_A Z(s)ds$)

Methods

- Kriging (known covariance structure)
  - $\hat{z}_0 = \lambda^T z$ (weighted average of observations)
- Bayesian method (covariance structure with unknown parameters)
  - Posterior distribution of parameters
  - Posterior distribution $z_0|z$ (estimate is mean/mode/median)

Kriging

Given: $z = (z_1, \dots, z_n)$ observations of $Z(s)$ at locations $s_1, \dots, s_n$ and known covariance structure

Idea:

- Predict $z_0 = Z(s_0)$ as a weighted average $\hat{z}_0 = \lambda^T z$
- Weights λ are determined by spatial correlation structure

Kriging

Given: $z = (z_1, \ldots, z_n)$ observations of $Z(s)$ at locations $s_1, \ldots, s_n$ and known covariance structure

Model

$$z \sim N(X\beta, \Sigma), \quad z_0 \sim N(x_0^T\beta, \sigma_0^2), \quad \mathrm{Cov}(z, z_0) = \tau$$

$x_0, X =$ observed covariates at locations $s_0, s_1, \ldots, s_n$
$\beta =$ unknown coefficients of covariates
$\sigma_0^2, \tau, \Sigma =$ known covariances

Prediction: Choose $\hat{z}_0 = \lambda^T z$ so that

- $\hat{z}_0$ is unbiased ($\mathrm{E}(\hat{z}_0) = z_0$)
- Mean squared prediction error $\mathrm{E}((z_0 - \hat{z}_0)^2) = \mathrm{Cov}(\hat{z}_0)$ is minimised

Result

$$\hat{z}_0 = x_0^T\hat{\beta} + \tau^T\Sigma^{-1}(z - X\hat{\beta}) \quad \text{where } \hat{\beta} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}z$$

$$\mathrm{Cov}(\hat{z}_0) = \sigma_0^2 - \tau^T\Sigma^{-1}\tau + (x_0 - X^T\Sigma^{-1}\tau)^T(X^T\Sigma^{-1}X)^{-1}(x_0 - X^T\Sigma^{-1}\tau)$$

Bayesian method for prediction

Given: $z = (z_1, \ldots, z_n)$ observations of $Z(s)$ at locations $s_1, \ldots, s_n$

Model

$$\begin{aligned}
\begin{bmatrix} z|\alpha,\beta,\theta \\ z_0|\alpha,\beta,\theta \end{bmatrix} &\sim N\left(\begin{bmatrix} X\beta \\ x_0^T\beta \end{bmatrix}, \begin{bmatrix} \Sigma & \tau \\ \tau^T & \sigma_0^2 \end{bmatrix}\right), \\
& \Sigma = \alpha V(\theta), \tau = \alpha w(\theta), \sigma_0^2 = \alpha v_0(\theta) \\
& x_0, X = \text{ observed covariates at locations } s_0, s_1, \ldots, s_n \\
\alpha, \beta, \theta &\sim \text{some prior distribution}
\end{aligned}$$

Prediction: Posterior distribution

$$f(z_0|Z) = \int f(z_0|z,\alpha,\theta,\beta) f(\beta|z,\alpha,\theta) f(\alpha|z,\theta) f(\theta|z) d\beta d\alpha d\theta$$

$$\hat{z}_0 = \text{mean/median/mode}$$

Tools for estimation and prediction of Gaussian processes

Frequentist methods

- Directly using `optim` to optimise likelihood/REML/prediction error
- `nlme` (linear mixed model formulation of Gaussian process) (uses ML or REML)
- `mgcv` (GAM formulation) (uses penalised likelihood method)

Bayesian methods

- Markov Chain Monte Carlo
- `INLA` for Gaussian Markov random fields (GMRFs) (uses integrated nested Laplace approximation)

Tools for estimation and prediction of Gaussian processes

For large datasets z, calculation of Σ^{-1} and $|\Sigma|$ is difficult and requires numerical methods, e.g.

- tapering (sparse Σ by setting small values equal to 0)
- other likelihood approximations e.g. in the spatial domain (condition on subvectors of z rather than full z), or in the spectral domain (truncate spectral density of z)
- fixed rank kriging (particular covariance structures - invert Σ by inverting smaller fixed rank matrices)
- GMRF representation of e.g. Matérn fields on triangulated lattice (sparse precision matrix, cholesky decomposition)

References

C. Paciorek, *Technical vignette 3: Kriging, interpolation, and uncertainty: Department of biostatistics*, Harvard School of Public Health, Version, 1 (2008).

H. Rue, A. Riebler, S. H. Sørbye, J. B. Illian, D. P. Simpson, and F. K. Lindgren, *Bayesian computing with inla: A review*, arXiv preprint arXiv:1604.00860, (2016).

R. L. Smith, *Environmental statistics*, Facultad de Ciencias Económicas, Universidad Nacional del Cuyo, 1999.

Y. Sun, B. Li, and M. G. Genton, *Geostatistics for large datasets*, in Advances and challenges in space-time modelling of natural events, Springer, 2012, pp. 55–77.

S. N. Wood, *Generalized additive models: an introduction with R*, CRC press, 2017.