Stat 8501 Lecture Notes
Spatial Gaussian Processes
Charles J. Geyer
April 14, 2014

1 Introduction

We start more or less following Hoel, Port, and Stone (1986, hereinafter HPS), changing what needs to be changed to allow for vector parameters. We are studying second order processes $X(\mathbf{t})$, where now $\mathbf{t}$ denotes a vector. But other than that everything stays the same: for each $\mathbf{t}$ in the domain T, we have a random variable $X(\mathbf{t})$. (I don't really like the vectors are boldface convention, but we will use it in this handout.)

We use the same notation as HPS

$$\begin{aligned} \mu_X(\mathbf{t}) &= E\{X(\mathbf{t})\} \\ r_X(\mathbf{s}, \mathbf{t}) &= \operatorname{cov}\{X(\mathbf{s}), X(\mathbf{t})\} \end{aligned}$$

We say the process is *weakly stationary* or *translation invariant* if

$$\begin{aligned} \mu_X(\mathbf{t} + \mathbf{h}) &= \mu_X(\mathbf{t}) \\ r_X(\mathbf{s} + \mathbf{h}, \mathbf{t} + \mathbf{h}) &= r_X(\mathbf{s}, \mathbf{t}) \end{aligned}$$

for all $\mathbf{h}$, $\mathbf{s}$, and $\mathbf{t}$ for which the expressions make sense. The first implies that μ_X is actually a constant function, and the second implies that $r_X(\mathbf{s}, \mathbf{t})$ is a function of $\mathbf{s} - \mathbf{t}$ only, so (again following HPS) we can define the covariance function as a function of one variable rather than two

$$r_X(\mathbf{t}) = r_X(0, \mathbf{t})$$

As in the one-dimensional case, a covariance function must be a positive definite function, that is for any scalars $a_1, \ldots, a_k$ and vectors $\mathbf{t}_1, \ldots, \mathbf{t}_k$

$$\sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j r_X(\mathbf{t}_i, \mathbf{t}_j) \ge 0$$

when r_X is the covariance function of a (not necessarily translation invariant) second order process, and

$$\sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j r_X(\mathbf{t}_i - \mathbf{t}_j) \ge 0$$

when r_X is the covariance function of a translation invariant second order process.

Bochner's theorem says that r_X is the covariance function of a translation invariant process if and only if $\mathbf{t} \mapsto r_X(\mathbf{t})/r_X(0)$ is the characteristic function of a random vector having a distribution symmetric about zero.

So far, everything is the same as in the one-dimensional carrier case except for some boldface for vectors instead of lightface for scalars. Now comes the different part, but to introduce that we first ask why we would assume translation invariance. The reason is to simplify the model making it easier to estimate when we are doing statistical inference (we won't discuss statistical inference for Gaussian processes in this course, but that is a large part of spatial statistics).

We say a translation invariant process X is *isotropic* or *rotationally invariant* if $r_X(\mathbf{t})$ is a function of $\|\mathbf{t}\|$ only, where $\|\cdot\|$ denotes the Euclidean norm. This is something new that we do not see in the one-dimensional case. There are no rotations in one dimension.

We follow Section 2.3 of Chilés and Delfiner (1999) with some additions from other sources. Write

$$r_X(\mathbf{t}) = c_X(\|\mathbf{t}\|) \tag{1}$$

so we are studying the properties of the function c_X of one nonnegative real variable. We will call the one dimensional covariance functions like c_X *radial covariance functions*. Bochner's theorem still applies: c_X is a valid radial covariance function if

$$\mathbf{t} \mapsto \frac{c_X(\|\mathbf{t}\|)}{c_X(0)}$$

is a valid d-dimensional covariance function. As in the one dimensional case

$$\operatorname{var}\{X(\mathbf{t})\} = c_X(0), \qquad \text{for all } \mathbf{t}.$$

The first issue we deal with is that if $c(r)$ is a valid radial covariance function for processes on $\mathbb{R}^d$, then it is also a valid radial covariance function for processes on $\mathbb{R}^m$ for $m \le d$, but is not necessarily valid for processes on $\mathbb{R}^n$ for $n > d$. For example, the radial covariance function

$$c(r) = \begin{cases} 1 - \frac{r}{a}, & 0 \le r \le a \\ 0, & \text{otherwise} \end{cases}$$

is valid for processes on $\mathbb{R}^1$ but is not valid for higher dimensions.

I could not find simple general conditions for validity of radial covariance functions. There are however, some well known simple examples that are valid in all dimensions. One is the *squared exponential* or *Gaussian* radial covariance function

$$c(r) = A \exp\left(-\frac{r^2}{a^2}\right) \tag{2}$$

for constants $A > 0$ and $a > 0$. Another is the *exponential* radial covariance function

$$c(r) = A \exp\left(-\frac{r}{a}\right) \tag{3}$$

Here are some more complicated models that are also valid for all dimensions

$$c(r) = A \left(1 + \frac{r}{a}\right) \exp\left(-\frac{r}{a}\right)$$
$$c(r) = A \left(1 + \frac{r}{a} + \frac{r^2}{3a^2}\right) \exp\left(-\frac{r}{a}\right)$$

and another with much lighter tails that is also valid for all dimensions, which is called the *generalized Cauchy* radial covariance function,

$$c(r) = A \left(1 + \frac{r^2}{a^2}\right)^{-\alpha}$$

where here, as in all of the above, we have $A > 0$ and $a > 0$ but how also have another parameter $\alpha > 0$.

There is one more quite complicated covariance function that has received some attention in the literature. This is also valid in all dimensions and is called the $K-Bessel$ radial covariance function by Chilés and Delfiner (1999) and the *Matérn* radial covariance function by others (Wikipedia, 2014b)

$$c(r) = \frac{A}{2^{\nu-1}\Gamma(\nu)} \left(\frac{r}{a}\right)^{\nu} K_\nu\left(\frac{r}{a}\right)$$

where here, as in all of the above, we have $A > 0$ and $a > 0$ but how also have another parameter $\nu \geq 0$. Here $\Gamma(\nu)$ denotes the gamma function and $K_\nu(x)$ denotes the modified Bessel function of the second kind of order ν (Wikipedia, 2014a). For those not familiar with Bessel functions, they should just be thought of a another kind of special function, like sines and cosines, exponentials and logarithms, and gamma functions. They are calculated by the `besselK` function in R. Part of the reason for the interest in this class of radial covariance functions is that it includes some of the others. The case $\nu = 1/2$ is the exponential radial covariance function (3), and as

$\nu \to \infty$ it converges to the squared exponential radial covariance function (2).

A Gaussian process with Matérn radial covariance function has sample paths that are $\lfloor \nu - 1 \rfloor$ times differentiable. A Gaussian process with exponential radial covariance function has sample paths that are nowhere differentiable. A Gaussian process with squared exponential radial covariance function has sample paths that are infinitely differentiable.

2 Prediction

Suppose we observe a Gaussian process at locations $\mathbf{t}_1, \ldots, \mathbf{t}_k$ and we want to predict the value at another location t_0 where we have not observed the value of the process.

The best prediction if best means squared error loss is the conditional expectation, and the best prediction if best means absolute error loss is the conditional median, but the conditional distribution is normal for which the mean is equal to the median, so the conditional expectation is best for either case.

If a partitioned random vector

$$\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \tag{4}$$

is multivariate normal with mean vector

$$\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$$

and variance matrix

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

then the conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2$ is multivariate normal with mean vector

$$\boldsymbol{\mu}_1 - \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2) \tag{5}$$

and variance matrix

$$\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \tag{6}$$

(Anderson, 2003, Theorem 2.5.1). This notation assumes that the distribution of $\mathbf{X}_2$ is nondegenerate (so Σ_{22} is invertible). If not, then $\mathbf{X}_2$ is a linear function of some set of its components that do have a nondegenerate

distribution (Anderson, 2003, Theorem 2.4.4 and the rest of Section 2.4), in which case we can rewrite (4) as

$$\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_3 \\ \mathbf{X}_4 \end{pmatrix}$$

where $\mathbf{X}_3$ has a nondegenerate distribution and $\mathbf{X}_4$ is a deterministic function (actually a linear function) of $\mathbf{X}_3$, then the conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2$ is the same as the conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_3$ is the same as and is multivariate normal with mean vector (5) with subscript 2 replaced by 3 and variance matrix (6) with subscript 2 replaced by 3.

So all we need to do prediction is the vector μ and the matrix Σ for the random vector with components $X(\mathbf{t}_i)$, $i = 1, \ldots, k$. The vector μ has i-th component $\mu_X(\mathbf{t}_i)$, and the matrix Σ has i, j-th component $r_X(\mathbf{t}_i, \mathbf{t}_j)$. So we are done with prediction.

3 Integration

Now suppose we want to predict the average of the process over a bounded set A. which we denote

$$X(A) = \frac{1}{m(A)} \int_A X(\mathbf{t}) \, d\mathbf{t}$$

where $m(A)$ is Lebesgue measure of the set A

$$m(A) = \int_A d\mathbf{t}$$

(which is finite because we assumed A is bounded). Again, we use the observations at $\mathbf{t}_i$, $i = 1, \ldots, k$. We already know the means, variances, and covariances for the $X(\mathbf{t}_i)$. All we need to determine is the mean and variance of $X(A)$ and its covariance with each of the $X(\mathbf{t}_i)$.

As in Section 5.2 of HPS, this is done by interchanging the order of integration.

$$\begin{aligned} E\{X(A)\} &= E\left\{ \frac{1}{m(A)} \int_A X(\mathbf{t}) \, d\mathbf{t} \right\} \\ &= \frac{1}{m(A)} \int_A E\{X(\mathbf{t})\} \, d\mathbf{t} \\ &= \frac{1}{m(A)} \int_A \mu_X(\mathbf{t}) \, d\mathbf{t} \end{aligned}$$

Denote that $\mu_X(A)$. Then

$$\begin{aligned}
\operatorname{var}\{X(A)\} &= E\left\{\left[\frac{1}{m(A)}\int_A X(\mathbf{t})\,d\mathbf{t} - \mu_X(A)\right]^2\right\} \\
&= E\left\{\left[\frac{1}{m(A)}\int_A [X(\mathbf{t}) - \mu_X(\mathbf{t})]\,d\mathbf{t}\right]^2\right\} \\
&= E\left\{\frac{1}{m(A)^2}\int_A\int_A [X(\mathbf{s}) - \mu_X(\mathbf{s})][X(\mathbf{t}) - \mu_X(\mathbf{t})]\,d\mathbf{s}\,d\mathbf{t}\right\} \\
&= \frac{1}{m(A)^2}\int_A\int_A E\{[X(\mathbf{s}) - \mu_X(\mathbf{s})][X(\mathbf{t}) - \mu_X(\mathbf{t})]\}\,d\mathbf{s}\,d\mathbf{t} \\
&= \frac{1}{m(A)^2}\int_A\int_A r_X(\mathbf{s},\mathbf{t})\,d\mathbf{s}\,d\mathbf{t}
\end{aligned}$$

and

$$\begin{aligned}
\operatorname{cov}\{X(A), X(\mathbf{t}_i)\} &= \operatorname{cov}\left\{X(\mathbf{t}_i), \frac{1}{m(A)}\int_A X(\mathbf{t})\,d\mathbf{t}\right\} \\
&= E\left\{[X(\mathbf{t}_i) - \mu_X(\mathbf{t}_i)]\frac{1}{m(A)}\int_A [X(\mathbf{t}) - \mu_X(\mathbf{t})]\,d\mathbf{t}\right\} \\
&= E\left\{\frac{1}{m(A)}\int_A [X(\mathbf{t}_i) - \mu_X(\mathbf{t}_i)][X(\mathbf{t}) - \mu_X(\mathbf{t})]\,d\mathbf{t}\right\} \\
&= \frac{1}{m(A)}\int_A E\{[X(\mathbf{t}_i) - \mu_X(\mathbf{t}_i)][X(\mathbf{t}) - \mu_X(\mathbf{t})]\}\,d\mathbf{t} \\
&= \frac{1}{m(A)}\int_A r_X(\mathbf{t}_i,\mathbf{t})\,d\mathbf{t}
\end{aligned}$$

These integrals are not usually easy to do and may need to be done by numerical integration. But they are what has to be done to do this prediction.

4 Differentiation

Now we look at derivatives, when they exist. The first issue is that when $\mathbf{t}$ is multivariate, the first derivative is a vector, the second derivative a matrix, the third derivative a three-dimensional tensor, and so forth. So we only look at first derivatives.

We usually think of the first derivative vector as the vector whose components are the partial derivatives, but this is an oversimplification. The correct definition of multivariate differentiation (Browder, 1996, Definition 8.9)

is that f is differentiable at $\mathbf{x}$ if there exists a linear function l such that

$$\lim_{\mathbf{h}\to 0} \frac{f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) - l(\mathbf{h})}{\|\mathbf{h}\|} = 0 \tag{7}$$

in which case l is uniquely defined and has the form

$$l(\mathbf{h}) = \mathbf{v}^T\mathbf{h}$$

for some vector $\mathbf{v}$, which we call the derivative of f at $\mathbf{x}$, usually written $f'(\mathbf{x})$ or $\nabla f(\mathbf{x})$.

If f is differentiable at $\mathbf{x}$, that is, if there exists l such that (7) holds, then all partial derivatives exist at $\mathbf{x}$, and the $\nabla f(\mathbf{x})$ is the vector of partial derivatives (Browder, 1996, Theorem 8.21). The converse is not true. More is needed than existence of partial derivatives to make the function differentiable. If all partial derivatives exist and are continuous on an open neighborhood of $\mathbf{x}$, then the function is differentiable (Browder, 1996, Theorem 8.23).

We will ignore this distinction between differentiability and existence of partial derivatives in what follows, mostly because we do not have sharp conditions for existence of derivatives. A sharp condition for almost sure existence and continuity of partial derivatives of any order is given by Adler and Taylor (2007, Theorem 1.4.2) but seems complicated.

Staying at the level of HPS, we can consider a partial derivative to be the same as the total derivative of the function restricted to a line, so the results of Section 5.3 in HPS hold. Let us write

$$X_i(\mathbf{t}) = \frac{\partial X(\mathbf{t})}{\partial t_i}$$

(assuming this exists). Each $X_i(\mathbf{t})$ is a Gaussian process. From equation 14 of Chapter 5 of HPS we get

$$\mu_{X_i}(\mathbf{t}) = \frac{\partial}{\partial t_i}\mu_X(\mathbf{t})$$

Then (15) and (16) of Chapter 5 in HPS become

$$r_{YX_i}(\mathbf{s},\mathbf{t}) = \frac{\partial}{\partial t_i} r_{YX}(\mathbf{s},\mathbf{t})$$
$$r_{X_iY}(\mathbf{s},\mathbf{t}) = \frac{\partial}{\partial s_i} r_{XY}(\mathbf{s},\mathbf{t})$$

from which we get (like (20) and (21) of HPS)

$$r_{XX_j}(\mathbf{s},\mathbf{t}) = \frac{\partial}{\partial t_j} r_X(\mathbf{s},\mathbf{t})$$
$$r_{X_iX_j}(\mathbf{s},\mathbf{t}) = \frac{\partial^2}{\partial s_i \partial t_j} r_X(\mathbf{s},\mathbf{t})$$

so that characterizes the first derivative process (or d processes, one for each dimension).

References

Adler, R. J., and Taylor, J. E. (2007). *Random Fields and Geometry*. New York: Springer.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, third ed. Hoboken: John Wiley.

Browder, A. (1996). *Mathematical Analysis: An Introduction*. New York: Springer-Verlag.

Chilés, J.-P., and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. New York: John Wiley.

Hoel, P. G., Port, S. C., and Stone, C. J. (1972). *Introduction to Stochastic Processes*. Boston: Houghton Mifflin. Republished, Waveland Press, Prospect Heights, Illinois, 1986.

Wikipedia (2014a). Bessel function. `http://en.wikipedia.org/w/index.php?title=Bessel_function&oldid=603570290`

Wikipedia (2014b). Matérn covariance function. `http://en.wikipedia.org/w/index.php?title=Mat%C3%A9rn_covariance_function&oldid=590577006`