

# 1 Isotropic Covariance Functions

Let  $\{Z(s)\}$  be a Gaussian process on  $\mathbb{R}^n$ , *i.e.*, a collection of jointly normal random variables  $Z(s)$  associated with  $n$ -dimensional locations  $s \in \mathbb{R}^n$ . The joint distribution of  $\{Z(s)\}$  depends only on the means  $\mu(s) = \mathbb{E}Z(s)$  and the covariances  $C(s, t) = \mathbb{E}(Z(s) - \mu(s))(Z(t) - \mu(t))$ .

The process is called *stationary* or *translation invariant* if the distribution wouldn't change under a rigid translation of the entire collection of locations, *i.e.*, if  $\mu(s) = \mu(s + h)$  and  $C(s + h, t + h) = C(s, t)$  for all  $h$ ; in this case  $\mu(s) \equiv \mu$  is constant and  $C(s, t) = C(s - t, 0)$  can only depend on the difference  $h = (s - t)$  between the two locations, so must be of the form  $C(s, t) = C_0(s - t)$  for some function  $C_0(h) = C(h, 0)$  on  $\mathbb{R}^n$ . Not just any function  $C_0(h)$  can be a covariance function; let's see what the choices are.

It's easy to see that the function  $C_0$  must be *even*, *i.e.*, must satisfy  $C_0(h) = C_0(-h)$ , since  $C(s - t) = \mathbb{E}(Z(s) - \mu(s))(Z(t) - \mu(t)) = C(t - s)$ . But more is true: if  $\{s_j\}$  any collection of locations, then complex linear combinations  $a^\top(Z - \mu) = \sum a_j(Z_j - \mu_j)$  of the centered random variables  $Z_j = Z(s_j)$  (with means  $\mu_j = \mu(s_j)$ ) must have nonnegative squared modulus  $\mathbb{E}|\sum a_j(Z_j - \mu_j)|^2 = \sum a_j C(s_j - s_k) \bar{a}_k \geq 0$  for every set of complex numbers  $\{a_j\} \subset \mathbb{C}$ . A function  $C_0(h)$  is called *positive semi-definite* if it always satisfies the inequality  $\sum_{jk} a_j C(s_j - s_k) \bar{a}_k \geq 0$  for any locations  $s_j$  and complex numbers  $a_j$ ; this is equivalent to asking that  $C(h) = C(-h)$  for every  $h \in \mathbb{R}^n$  and that  $\sum a_j C(s_j - s_k) a_k \geq 0$  for all *real* numbers  $a_j \in \mathbb{R}$ . One way to get a symmetric positive semi-definite function  $C_0(h)$  is by taking the Fourier transform

$$C_0(h) = \int_{\mathbb{R}^n} e^{ih \cdot \omega} G(\omega) d^n \omega$$

of any positive function  $G(\omega)$  on  $\mathbb{R}^n$  or, more generally, of any finite positive measure  $G(d\omega)$ , because then

$$\begin{aligned} \sum_{jk} a_j C(s_j - s_k) \bar{a}_k &= \int_{\mathbb{R}^n} \sum_{jk} (a_j e^{s_j \cdot \omega}) \overline{(a_k e^{s_k \cdot \omega})} G(d\omega) \\ &= \int_{\mathbb{R}^n} \left| \sum_j a_j e^{s_j \cdot \omega} \right|^2 G(d\omega) \geq 0. \end{aligned}$$

It turns out that this is the *only* way to get one— that every positive semi-definite function can be written in this form for some finite positive measure

$G(d\omega)$ , called the *spectral measure* (if  $G(d\omega) = G(\omega) d\omega$  is absolutely continuous,  $G(\omega)$  is called the *spectral density*). Known as “Bochner’s Theorem,” this result is really just the Fourier inversion formula in an unfamiliar setting:

$$G(\omega) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{-ih \cdot \omega} C_0(h) d^n h.$$

Since the process  $\{Z(s)\}$  is real-valued, the spectral density  $G(\omega) = G(-\omega)$  must be an even function and so we can write

$$\begin{aligned} C_0(h) &= \int_{\mathbb{R}^n} \cos(h \cdot \omega) G(\omega) d^n \omega \\ G(\omega) &= (2\pi)^{-n} \int_{\mathbb{R}^n} \cos(h \cdot \omega) C_0(h) d^n h \end{aligned}$$

If the Gaussian process is also *isotropic*, or invariant under rotations, then  $G(\omega) = g(|\omega|)$  must also be invariant under rotations and depend only on the length  $r = |\omega|$  of the vector  $\omega \in \mathbb{R}^n$ . In this case we can simplify these integrals by transforming to polar coordinates.

### 1.1 Polar Coordinates for Probabilists

Polar coordinates are a familiar tool in two-dimensional integrals, where the change of variables from  $x \in \mathbb{R}^2$  to  $r = \sqrt{x_1^2 + x_2^2}$  and  $\theta = \arctan x_2/x_1$  (so  $x_1 = r \cos \theta$ ,  $x_2 = r \sin \theta$ ) and a change from  $d^2x$  to  $r dr d\theta$  lead to simple expressions for the integrals of radial functions. Equivalently, we can let  $\sigma$  have a uniform probability distribution (denoted by  $d\sigma$ ) over the unit circle  $S^1 = \{x : x_1^2 + x_2^2 = 1\}$ , and change variables from  $x = (x_1, x_2)$  to  $(r, \sigma)$ , with  $d^2x = dx_1 dx_2$  replaced by  $2\pi r dr d\sigma$ .

In three dimensions the first polar approach has its analogue in the Euler angles, while the second is simpler with uniform measure for  $\sigma$  on the unit sphere  $S^2 \subset \mathbb{R}^3$ , with  $d^3x = dx_1 dx_2 dx_3$  replaced by  $4\pi r^2 dr d\sigma$ . Notice that  $2\pi r$  and  $4\pi r^2$  are the circumference of the circle and the area of the sphere of radius  $r$ , respectively. In any number  $n$  of dimensions the sphere  $S^{n-1}$  has area  $2\pi^{n/2} r^{n-1} / \Gamma(n/2)$ , and we can again evaluate integrals in polar coordinates with the uniform probability distribution  $d\sigma$  for  $\sigma \in S^{n-1} \subset \mathbb{R}^n$ , and  $d^n x = \frac{2\pi^{n/2}}{\Gamma(n/2)} r^{n-1} dr d\sigma$ . This makes it easy to compute integrals of radial functions; for functions that also depend on one or more of the components  $x_j$ , it is sometimes helpful to note that the squares  $\{\sigma_j^2\}$  have a Dirichlet  $\text{Di}(\frac{1}{2}, \dots, \frac{1}{2})$  joint distribution, so each  $\sigma_j$  is distributed as the square root of a  $\text{Be}(\frac{1}{2}, \frac{n-1}{2})$  random variable.

## 1.2 Evaluating $C_0(h)$

Switching to polar coordinates  $r = |\omega| \geq 0$  and  $\sigma = \omega/|\omega| \in S^{n-1}$  (where  $d\sigma$  denotes the uniform probability measure on the unit sphere  $S^{n-1}$  in  $\mathbb{R}^n$ ), and noting that the component  $\sigma_h = \sigma \cdot h/|h|$  of  $\sigma \in S^{n-1}$  in the direction  $h$  again has the same distribution as the square root of a  $\text{Be}(\frac{1}{2}, \frac{n-1}{2})$  random variable, writing  $\rho$  for  $|h|$ ,

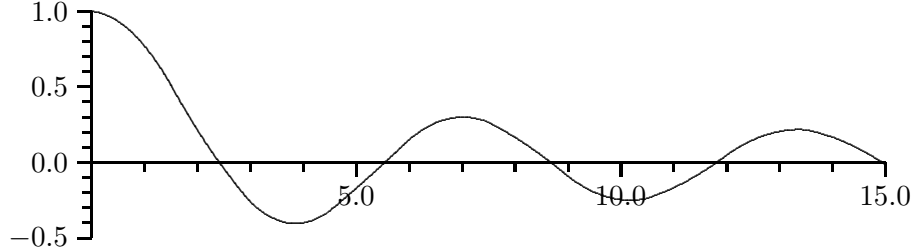
$$\begin{aligned}
 C_0(h) &= \int_{\mathbb{R}^n} \cos(h \cdot \omega) g(|\omega|) d^n \omega \\
 &= \iint_{\mathbb{R}_+ \times S^{n-1}} \cos(r\rho\sigma_h) g(r) \frac{2\pi^{n/2} r^{n-1}}{\Gamma(n/2)} dr d\sigma \\
 &= \int_{\mathbb{R}_+} \int_0^1 \cos(r\rho\sqrt{u}) g(r) \frac{2\pi^{n/2} r^{n-1}}{\Gamma(n/2)} \frac{\Gamma(n/2)}{\Gamma(\frac{1}{2})\Gamma(\frac{n-1}{2})} u^{1/2-1} (1-u)^{(n-1)/2-1} dr du \\
 &= \int_0^\infty \rho (2\pi r/\rho)^{\nu+1} J_\nu(r\rho) g(r) dr, \quad \nu \equiv \frac{n}{2} - 1 \tag{1}
 \end{aligned}$$

$$\begin{aligned}
 &= \int_0^\infty (r\rho/2)^{-\nu} \Gamma(\nu+1) J_\nu(r\rho) \gamma(dr) \tag{2} \\
 &= \begin{cases} \int_0^\infty 2 \cos(r\rho) g(r) dr & \text{if } n = 1 \\ \int_0^\infty 2\pi r J_0(r\rho) g(r) dr & \text{if } n = 2 \\ \int_0^\infty \rho (2\pi r/\rho)^{3/2} J_{1/2}(r\rho) g(r) dr & \text{if } n = 3 \end{cases}
 \end{aligned}$$

where

$$J_\nu(z) = \frac{(z/2)^\nu}{\sqrt{\pi} \Gamma(\nu + 1/2)} \int_0^\pi \cos(z \cos \theta) \sin(\theta)^{2\nu} d\theta$$

is the Bessel function of the first kind of order  $\nu$  (see Watson, 1944). Bessel functions aren't as familiar as sines and cosines, but they're common in engineering and physics and are in the standard C library, the GNU Scientific library (GSL), Maple and Mathematica, MatLab, *etc.*; see Abramowitz and Stegun (1964, Chapter 9) for details. Here's a plot of  $J_0(z)$ :



The plot of  $J_0(z)$  looks a little like a sine or cosine, but falls off like  $1/\sqrt{z}$  as  $z \rightarrow \infty$ .

The most general isotropic covariance is given in (2), with the absolutely continuous measure  $g(r) \frac{2\pi^{n/2}}{\Gamma(n/2)} r^{n-1} dr$  replaced by an arbitrary positive finite measure  $\gamma(dr)$  on  $[0, \infty)$ . Any isotropic covariance function may be approximated by one with a discrete spectral measure  $\gamma(dr) = \sum \gamma_j \delta_{r_j}(dr)$  assigning mass  $\gamma_j$  to finitely many points  $r_j$ :

$$\begin{aligned} C(\rho) &\approx \sum_j (2/r_j \rho)^\nu \Gamma(\nu + 1) J_\nu(r_j \rho) \gamma_j & (3) \\ &= \begin{cases} \sum_j \gamma_j \cos(r_j \rho) & \text{if } n = 1 \\ \sum_j \gamma_j J_0(r_j \rho) & \text{if } n = 2 \\ \sum_j \gamma_j \sqrt{\pi/2r_j \rho} J_{1/2}(r_j \rho) & \text{if } n = 3 \end{cases} \end{aligned}$$

but a more common approach is to choose small parametric families of densities  $g^\theta(r)$  or measures  $g^\theta(dr)$ .

We can recover the spectral density  $g(r) = G(\omega)$  (for  $r = |\omega|$ ) through the Fourier inversion formula, using polar coordinates with  $\rho = |h| \in \mathbb{R}_+$  and  $\sigma = h/|h| \in S^{n-1}$ :

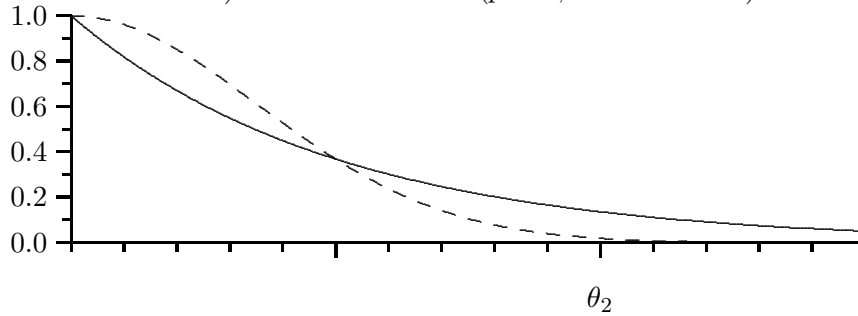
$$\begin{aligned} g(r) = G(\omega) &= \frac{1}{(2\pi)^n} \int \cos(-h \cdot \omega) C_0(h) d^n h \\ &= \frac{1}{(2\pi)^n} \iint_{\mathbb{R}_+ \times S^{n-1}} \cos(-r\rho\sigma_\omega) C(\rho) \frac{2\pi^{n/2} \rho^{n-1}}{\Gamma(n/2)} d\rho d\sigma \\ &= \int_0^\infty r(\rho/2\pi r)^{n/2} J_\nu(r\rho) C(\rho) d\rho, \quad \nu \equiv \frac{n}{2} - 1 & (4) \\ &= \begin{cases} \int_0^\infty \frac{2}{\pi} \cos(r\rho) C(\rho) d\rho & \text{if } n = 1 \\ \int_0^\infty (\rho/2\pi) J_0(r\rho) C(\rho) d\rho & \text{if } n = 2 \\ \int_0^\infty r(\rho/2\pi r)^{3/2} J_{1/2}(r\rho) C(\rho) d\rho & \text{if } n = 3 \end{cases} & (5) \end{aligned}$$

It is hard to imagine what  $C_0(h)$  would look like for different choices of  $g(r)$ ; a simple approach is to take whatever symmetric functions  $G(u)$  whose Fourier transforms we can find, and see what we get. Here are some commonly used covariance families, in  $n = 2$  dimensions; in each case  $\theta_1 = C(0)$  is an overall level parameter and  $\theta_2$  is a distance scale parameter:

- Power family

$$C(\rho|\theta, p) = \theta_1 \exp\{-|\rho/\theta_2|^p\}, \quad 0 < p \leq 2$$

Two notable covariograms in this family are the *exponential* ( $p = 1$ , solid below) and the *Gaussian* ( $p = 2$ , dashed below):

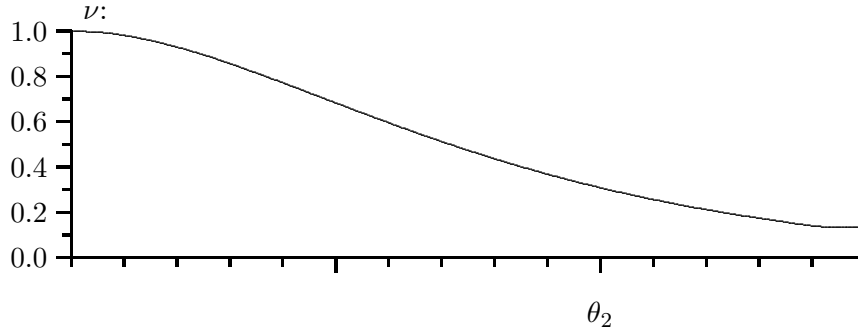


Notice that the exponential has a negative derivative at  $z = 0$ , so it falls off quickly at first, then slowly levels off, while the Gaussian has zero derivative near  $z = 0$  then falls off very quickly. From (5) it follows that the exponential has spectral density function  $g(r) = (\theta_1 \theta_2^2 / 2\pi) / (1 + r^2 \theta_2^2)^{3/2}$ , proportional to a bivariate Cauchy density function, while the Gaussian has spectral density  $g(r) = (\theta_1 \theta_2^2 / 4\pi) \exp(-r^2 \theta_2^2 / 4)$ , proportional to a normal density.

- Matérn

$$C(\rho|\theta) = \frac{2\theta_1}{\Gamma(\theta_3)} \left(\frac{\rho}{2\theta_2}\right)^{\theta_3} K_{\theta_3}(\rho/\theta_2)$$

where  $K_\nu(z)$  is the modified Bessel function of the third kind of order  $\nu$ :



The displayed plot has shape parameter  $\theta_3 = 2$ . The Matérn class is quite flexible and includes the exponential family (with  $\theta_3 = \frac{1}{2}$ ), the Gaussian family (in the limit as  $\theta_3 \rightarrow \infty$ ), and many others. In  $n$  dimensions its spectral density function is

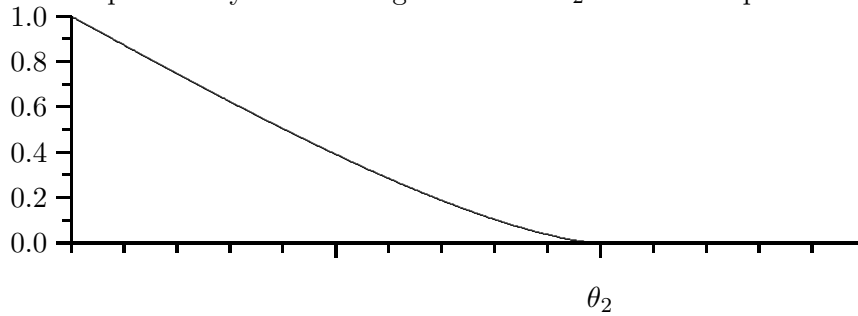
$$g(r) = \frac{\theta_1 \theta_2^n}{\Gamma(\theta_3) \pi^{n/2}} (1 + \theta_2^2 r^2)^{-\theta_3 - n/2},$$

proportional to the familiar  $n$ -variate Student's  $t$  density function with  $2\theta_3$  degrees of freedom and variance scale  $\sigma^2 = 1/2\theta_2^2\theta_3$ . This lends more insight into how the Matérn reduces to the exponential when  $\theta_3 = 1/2$  and to the Gaussian when  $\theta_3 \rightarrow \infty$ .

- Spherical

$$C(\rho|\theta) = \begin{cases} \theta_1 \left[ 1 - \frac{2}{\pi} \left( \frac{\rho}{\theta_2} \sqrt{1 - \left(\frac{\rho}{\theta_2}\right)^2} + \sin^{-1} \frac{\rho}{\theta_2} \right) \right] & \text{for } \rho < \theta_2 \\ 0 & \text{for } \rho \geq \theta_2 \end{cases}$$

The spherical covariance function is proportional to the area of intersection for two discs of diameter  $\theta_2$  with centers separated by distance  $\rho$ . In this model the Gaussian quantities  $Z_j$  and  $Z_k$  at loci  $s_j$  and  $s_k$  separated by a distance greater than  $\theta_2$  will be independent.



This is not quite linear. Like the exponential, it has a negative slope at  $z = 0$  and falls off rapidly at first; like the Gaussian, it falls off rapidly later and in fact reaches zero. The spectral density, while available in closed form, isn't illuminating; it's best to think of the spherical process as a convolution or moving average of Gaussian white noise, integrated at each locus over the surrounding ball of diameter  $\theta_2$ .

A variety of processes may be constructed similarly as kernel integrals of standard Gaussian white noise,

$$Z(h) = \int_{\mathbb{R}^n} k(h - s) \zeta(ds);$$

where "standard" means that  $E[\zeta(ds)] = 0$  and  $E[\zeta(ds)^2] = ds$ . The covariance is

$$C_0(h) = E[Z(0)\overline{Z(h)}] = \int_{\mathbb{R}^n} k(h - s) \overline{k(-s)} ds$$

with spectral density

$$\begin{aligned}
G(\omega) &= (2\pi)^{-n} \int e^{-i\omega \cdot h} C_0(h) dh \\
&= (2\pi)^{-n} \iint e^{-i\omega \cdot h} k(h-s) \overline{k(-s)} ds dh \\
&= (2\pi)^{-n} \left| \int e^{-i\omega \cdot x} k(x) dx \right|^2
\end{aligned}$$

so that an isotropic kernel may be computed from the spectral density as

$$k(x) = (2\pi)^{-n/2} \int e^{i\omega \cdot x} G(\omega)^{1/2} d^n \omega$$

or, in polar coordinates,

$$\begin{aligned}
k(\rho) &= \int_0^\infty r^{\nu+1} \rho^{-\nu} J_\nu(r\rho) g(r)^{1/2} dr \\
&= \begin{cases} \int_0^\infty \sqrt{\frac{2}{\pi}} \cos(r\rho) \sqrt{g(r)} dr & \text{if } n = 1 \\ \int_0^\infty J_0(r\rho) r \sqrt{g(r)} dr & \text{if } n = 2 \\ \int_0^\infty J_{1/2}(r\rho) r^{3/2} \rho^{-1/2} \sqrt{g(r)} d\rho & \text{if } n = 3 \end{cases}
\end{aligned}$$

provided that the square root of the spectral density *is* the Fourier transform of a finite positive function, *i.e.*, is itself positive semidefinite. For the Matérn class, the root spectral density  $\sqrt{g(r)} \propto (1 + \theta_2^2 r^2)^{-(\theta_3 + n/2)/2}$  will be another  $n$ -variate  $t$  density provided  $\theta_3 > n/2$  and in this case, setting  $\epsilon = (2\theta_3 - n)/4 > 0$ , we find

$$k(\rho) = \frac{2\theta_1^{1/2} (2\rho\theta_2)^{-\epsilon - n/2}}{\Gamma(\epsilon + n/2) \sqrt{\Gamma(2\epsilon + n/2)} \pi^{n/4}} K_\epsilon(\rho/\theta_2)$$

leads to a moving-average kernel representation for the Matérn covariance class. In any number  $n \geq 1$  of dimensions the restriction  $\epsilon > 0$  entails  $\theta_3 > n/2 \geq 1/2$ , ruling out the exponential covariance, but the Gaussian covariance (the limiting case as  $\theta_3 \rightarrow \infty$ ) is available in any number of dimensions, with

$$k(\rho) = \theta_1^{1/2} (\pi\theta_2^2/4)^{-n/2} e^{-2\rho^2/\theta_2^2}$$

## References

- Abramowitz, M. and Stegun, I. A., eds. (1964), *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*, volume 55 of *Applied Mathematics Series*, Washington, D.C.: National Bureau of Standards.
- Cressie, N. A. C. (1993), *Statistics for Spatial Data*, New York, NY, USA: John Wiley & Sons.
- Le, N. D. and Zidek, J. V. (1992), “Interpolation with uncertain spatial covariances: A Bayesian alternative to kriging,” *Journal of Multivariate Analysis*, 43, 351–374.
- Mardia, K. and Marshall, R. J. (1984), “Maximum likelihood estimation of models for residual covariance in spatial regression,” *Biometrika*, 71, 135–146.
- Matérn, B. (1986), *Spatial Variation*, volume 36 of *Lecture Notes in Statistics*, New York: Springer-Verlag, second edition, (first edition published 1960 by Statens Skogsforsningsinstitut, Stockholm).
- Ripley, B. D. (1981), *Spatial Statistics*, New York, NY, USA: John Wiley & Sons.
- Yaglom, A. M. (1962), *An Introduction to the Theory of Stationary Random Functions*, New York: Dover, translated and Edited by Richard A. Silverman from 1952 Russian article in *Uspekhi Matematicheskikh Nauk*.