Journal of Uncertain Systems
Vol.10, No.1, pp.10-14, 2016
Online at: www.jus.org.uk

Why Modified Exponential Covariance Kernel is Empirically Successful: A Theoretical Explanation

Olga Kosheleva[1],*, Michael Beer[2]

[1] *University of Texas at El Paso, El Paso, Texas 79968, USA*

[2] *Institute for Risk and Uncertainty, School of Engineering, University of Liverpool Liverpool L69 3BX, United Kingdom*

Received 1 June 2015; Revised 25 June 2015

Abstract

It is known that in the first approximation, many real-life stationary stochastic processes are well-described by an exponential covariance kernel $C(u) = \exp(-a \cdot |u|)$. Empirical evidence shows that in many practical situations, a good second approximation is provided by the modified exponential covariance kernel $C(u) = \exp(-a \cdot |u|) \cdot (1 - r \cdot |u|)$. In this paper, we provide a theoretical explanation for this empirical phenomenon.

Keywords: stationary stochastic processes, exponential covariance kernel, modified exponential covariance kernel

1 Formulation of the Problem

How to describe a general stochastic process. Physical quantities change with time. A natural description of this change, a description that takes into account that future values cannot be exactly predicted, is a description by a stochastic process; see, e.g., [5].

In many practical situations, while the quantities themselves change, the rules that describe these changes remain the same. Such situations are described by stationary processes.

A general stationary stochastic process $S(t)$ can be reduced to a zero-mean unit-variance process $X(t)$ by an appropriate linear transformation

$$S(t) = \sigma \cdot X(t) + \mu,$$

where $\mu = E[S(t)]$ is the mean of $S(t)$, and σ is its variance.

Thus, to describe general stationary stochastic processes, it is sufficient to study zero-mean unit-variance processes $X(t)$.

Each such process $X(t)$ can be described by its covariance function

$$C(u) \stackrel{\text{def}}{=} E[X(t) \cdot X(t+u)].$$

In many practical situations, it is sufficient to know the covariance function. In many practical situations, the random physical phenomenon – that we are describing by the stochastic process – is a joint effect of many small independent effects. This is the case, for example, for the measurement errors $S(t)$ at different moments of time.

According to the Central Limit Theorem, under certain reasonable conditions, the distribution of the sum of N small independent random variables tends to Gaussian when N increases. Thus, for situations when the number N of component effects is large, it makes sense to assume that the joint distribution of the variables $S(t_1)$, $S(t_2)$, ..., corresponding to different moments of time is Gaussian (normal). In this case, the re-normalized random variables $X(t_1)$, $X(t_2)$, ... are also normally distributed.

*Corresponding author.
Emails: olgak@utep.edu (O. Kosheleva), mbeer@liverpool.ac.ak (M. Beer).

To describe a normal distribution, it is sufficient to know the means and the covariances. Thus, to describe a zero-mean normally distributed stationary process, it is sufficient to know the covariance $C(u)$.

A good first approximation. For many physical process, the exponential kernel $C(u) = \exp(-a \cdot |u|)$ provides a good approximation for the actual covariance function [5].

A good second approximation. It turns out that in many practical situations, a reasonable second approximation is provided by the *modified exponential covariance kernel* $C(u) = \exp(-a \cdot |u|) \cdot (1 - r \cdot |u|)$ [4].

Formulation of the problem. While the empirical success of the exponential covariance kernel has a good theoretical explanation, there seem to be no good theoretical explanations for the success of the modified exponential covariance kernel.

What we do in this paper. In this paper, we provide a theoretical explanation for the empirical success of the modified exponential covariance kernel.

2 Theoretical Explanation of the Empirical Success of the Exponential Covariance Kernel: Reminder

Let us provide such a reminder. To provide the desired theoretical explanation for the empirical success of the *modified* exponential covariance kernel, let us recall the known theoretical explanation of the empirical success of the original exponential covariance kernel.

For normal distributions, covariance has a clear meaning. We have already mentioned that in many practical cases, we have a normally distributed stochastic process.

For zero-mean unit-variance random variables $X(t_i)$, with mean $\mu_X = 0$ and standard deviation $\sigma_X = 1$, covariance $C(u)$ between the variables $X(t)$ and $X(t_u)$ coincides with correlation $\rho(u)$ between these two variables:

$$C(u) = E[X(t) \cdot X(t+u)] = \frac{E[(X(t) - \mu_X) \cdot (X(t+u) - \mu_X)]}{\sigma_X \cdot \sigma_X} = \rho(u).$$

For normally distributed variables, the fact that their covariance is equal to $\rho(u)$ means that

$$X(t+u) = \rho(u) \cdot X(t) + \eta, \tag{1}$$

where η is independent from $X(t)$.

Resulting derivation of the exponential covariance kernel. For every two positive values $u_1 > 0$ and $u_2 > 0$, formula (1) implies that

$$X(t+u_1) = \rho(u_1) \cdot X(t) + \eta_1 \tag{2}$$

and

$$X((t+u_1)+u_2) = \rho(u_2) \cdot X(t+t_1) + \eta_2. \tag{3}$$

Substituting formula (2) into the expression (3), we conclude that

$$X(t+u_1+u_2) = \rho(u_2) \cdot \rho(u_1) \cdot X(t) + \eta', \tag{4}$$

where $\eta' \stackrel{\text{def}}{=} \eta_2 + \rho(u_2) \cdot \eta_1$ is independent from $X(t)$.

On the other hand, if we directly use the formula (1) with $u = u_1 + u_2$, we conclude that

$$X(t+u_1+u_2) = \rho(u_1+u_2) \cdot X(t) + \eta. \tag{5}$$

By comparing the coefficients at $X(t)$ in the expressions (4) and (5), we conclude that

$$\rho(u_1+u_2) = \rho(u_2) \cdot \rho(u_1). \tag{6}$$

The correlation $\rho(u) = C(u)$ is bounded by 1 from above and by -1 from below. It is well known that the only bounded solution of the functional equation (6) is the solution $\rho(u) = C(u) = \exp(-a \cdot u)$ for some $a \geq 0$; see, e.g., [1].

This explains the empirical success of the exponential covariance kernel.

3 Our Main Result: Theoretical Explanation of the Empirical Success of the Modified Exponential Covariance Kernel

Main idea. To describes the state of a physical system, we usually need to know the values of several physical quantities. To describe the dynamics of these quantities, we need to consider several related random processes $S_1(t)$, $S_2(t)$, ... We can normalize each of them, and consider the resulting zero-mean unit-variance processes $X_1(t)$, $X_2(t)$, ...

Let us show that in this case, an analysis similar to the one from the previous section can help us explain the modified exponential covariance model.

How to describe several related stochastic processes. To describe these processes, in addition to the covariances $C_{ii}(u) = \rho_{ii}(u) = E[X_i(t) \cdot X_i(t_u)]$, we also need to consider mutual covariances $C_{ij}(u) = E[X_j(t) \cdot X_i(t+u)]$.

Let us repeat all the steps of the previous derivation. To get the desired derivation, let us repeat all the steps of the derivation from the previous section, taking into account that now we have several random variables.

For normally distributed variables, we have

$$X_i(t+u) = \sum_j \rho_{ij}(u) \cdot X_j(t) + \eta_i, \tag{7}$$

where η_i are random variables which are independent on all $X_i(t)$. The equation (7) can be described in a matrix form, as

$$X(t+u) = \rho(u) \cdot X(t) + \eta. \tag{8}$$

This description is similar to the above formula (1), except that now $X(t)$ and η are vectors, and $\rho(u)$ is a matrix.

Similarly to the case of a single random variable, we can consider two values $u_1 > 0$ and $u_2 > 0$ and get

$$X(t+u_1) = \rho(u_1) \cdot X(t) + \eta^{(1)}, \tag{9}$$

$$X((t+u_1)+u_2) = \rho(u_2) \cdot X(t+t_1) + \eta^{(2)}, \tag{10}$$

hence

$$X(t+u_1+u_2) = \rho(u_2) \cdot \rho(u_1) \cdot X(t) + \eta', \tag{11}$$

where $\eta' \stackrel{\text{def}}{=} \eta^{(2)} + \rho(u_2) \cdot \eta^{(1)}$ is independent from $X(t)$.

On the other hand, if we directly use the formula (8) with $u = u_1 + u_2$, we conclude that

$$X(t+u_1+u_2) = \rho(u_1+u_2) \cdot X(t) + \eta. \tag{12}$$

By comparing the coefficients at $X(t)$ in the expressions (11) and (12), we conclude that

$$\rho(u_1+u_2) = \rho(u_2) \cdot \rho(u_1). \tag{13}$$

In mathematical terms, this means that the matrices $\rho(u)$ form a *semigroup*. It is known that, if we make an additional reasonable assumption that the dependence $\rho(u)$ is continuous, then the general solution to this equation takes the form $\rho(u) = \exp(-u \cdot A)$ for some matrix A; see, e.g., [3].

This expression can be further simplified if we use the known Jordan normal form representation of a matrix A (see, e.g., [2]), as $A = P^{-1} \cdot J \cdot P$, where P is a linear matrix, and $J = \text{diag}(J_1, J_2, \ldots)$ is a diagonal combination of blocks of the type

$$J_i = \begin{pmatrix} \lambda_i & 1 & 0 & \ldots & 0 & 0 \\ 0 & \lambda_i & 1 & \ldots & 0 & 0 \\ 0 & 0 & \lambda_i & \ldots & 0 & 0 \\ & \ldots & & \ldots & & \ldots \\ 0 & 0 & 0 & \ldots & \lambda_i & 1 \\ 0 & 0 & 0 & \ldots & 0 & \lambda_i \end{pmatrix},$$

where λ_i are eigenvalues of the matrix A (which are, in general, complex-valued).

If all the eigenvalues are different from each other, then each block is simply a 1×1 matrix

$$J_i = \left(\lambda_i\right),$$

and thus, the matrix J is diagonal: $J = \mathrm{diag}(\lambda_1, \lambda_2, \ldots)$:

$$J = \begin{pmatrix} \lambda_1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \lambda_3 & \ldots & 0 & 0 \\ & \ldots & & \ldots & & \ldots \\ 0 & 0 & 0 & \ldots & \lambda_{n-1} & 0 \\ 0 & 0 & 0 & \ldots & 0 & \lambda_n \end{pmatrix},$$

In general, when we have eigenvalues with multiplicity, then the blocks corresponding to these eigenvalues are of larger size.

In general, when a matrix A is represented in the Jordan normal form, we have $\exp(-u \cdot A) = P^{-1} \cdot \exp(-u \cdot J) \cdot P$. Here,

$$\exp(-u \cdot J) = \mathrm{diag}(\exp(-u \cdot J_1), \exp(-u \cdot J_2), \ldots),$$

where each element of the matrix $\exp(-u \cdot J_i)$ is linear combinations of the terms $\exp(-u \cdot \lambda_i)$, $u \cdot \exp(-u \cdot \lambda_i)$, $u^2 \cdot \exp(-u \cdot \lambda_i)$, ...

Each element $\rho_{ij}(u)$ of the matrix $\rho(u) = \exp(-u \cdot A)$ is a linear combination of the elements of the matrix $\exp(-u \cdot J)$; thus, each such element is a linear combination of expressions of the type $u^k \cdot \exp(-u \cdot \lambda_i)$, where $k \geq 0$ is a integer and λ_i is, in general, a complex number.

From the above general case to the second approximation. The first approximation corresponds to the case when we consider a single stochastic process. Thus, a natural second approximation corresponds to the case when we have two processes $X_1(t)$ and $X_2(t)$. In this case, we have three possible situations:

- If the corresponding 2×2 matrix ρ has a single eigenvalue λ of multiplicity, then we conclude that, for each of the variables, the covariance function $C(u)$ is a linear combination of the terms $\exp(-u \cdot \lambda)$ and $u \cdot \exp(-u \cdot \lambda)$. This is exactly the modified exponential kernel that we tried to explain.

- If the matrix ρ has two different real-valued eigenvalues $\lambda_1 < \lambda_2$ then $C(u)$ is a linear combination of two exponential functions $\exp(-u \cdot \lambda_1)$ and $u \cdot \exp(-u \cdot \lambda_2)$; asymptotically, when u is large, the second term can be safely ignored.

- Finally, if the eigenvalues are complex, then they have the form $\lambda_i = a \pm \omega$, so $C(u)$ is a linear combination of the terms $\exp(-a \cdot u) \cdot \cos(\omega \cdot u)$ and $\exp(-a \cdot u) \cdot \sin(\omega \cdot u)$.

4 Conclusions

In the first approximation, many real-life stationary stochastic processes are well described by the exponential covariance kernel $C(u) = \exp(-a \cdot u)$ (for $u > 0$). Empirical evidence show that in many practical situations, a modified exponential covariance kernel $C(u) = \exp(-a \cdot u) \cdot (1 - r \cdot u)$ provides a good second approximation.

In this paper, we have shown that a natural second approximation is either a modified exponential covariance, or a linear combination of two exponential covariance kernels corresponding to different values $a_1 \neq a_2$, or a linear combination of the terms $\exp(-a \cdot u) \cdot \cos(\omega \cdot u)$ and $\exp(-a \cdot u) \cdot \sin(\omega \cdot u)$. The fact that the modified exponential covariance kernel is one of only three possible cases explains why this kernel works well in many practical situations.

Our analysis also helps us find the natural third, fourth, etc. approximations: namely, all such approximations should be linear combinations of the terms $u^k \cdot \exp(-a \cdot u) \cdot \cos(\omega \cdot u)$ and $u^k \cdot \exp(-a \cdot u) \cdot \sin(\omega \cdot u)$, where $k \geq 0$ is a non-negative integer.

References

[1] Aczél, J., and J. Dhombres, *Functional Equations in Several Variables*, Cambridge University Press, 2008.

[2] Golub, G.H., and C.F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 2012.

[3] Hille, E., and R.S. Phillips, *Functional Analysis and Semi-Groups*, American Mathematical Society, Providence, Rhode Island, 1975.

[4] Spanos, P.D., Beer, M., and J. Red-Horse, Karhunen-Loéve expansion of stochastic processes with a modified exponential covarince kernel, *Journal of Engineering Mechanics*, vol.133, no.7, pp.773–779, 2007.

[5] Vanmarcke, E., *Random Fields: Analysis and Synthesis*, World Scientific, Singapore, 2010.