Derivative info (mean, covariance) of a Stochastic Process with given covariance function

Presentation in video: http://personales.upv.es/asala/YT/V/gradgpEN.html

Objective: we may have observations of the partial derivatives of a stochastic process (measurements of, say, position and speed) so we can improve position measurements, or we might wish to estimate such gradient from position measurements. Covariance between a stochastic process and its partial derivatives is needed to carry out such tasks.

Table of Contents

Preliminaries

Let us consider a stochastic process where $f(x)$ is a random variable, for $x \in \mathbb{R}^n$, and we have a cross-covariance (kernel) function $k(x, x') : \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}$. This makes possible for us to write the conditional mean and variance of the process given observations of f at a given set of points (Kriging, best linear prediction, Kernel regression; these ideas are developed in other materials in this collection).

Under some conditions $f(x)$ may be differentiable... Of course, we need to operate with some sort of caution on what that derivative means in a random process (derivative of a function is a limit, derivative of a stochastic process must be understood in probabilistic terms, if $\frac{1}{h}(f(x+h) - f(x))$ converges with probability one to something). We will NOT discuss such things here and we'll assume existance of these derivatives if the formulae below can be evaluated... you know, we engineers do understand what "velocity" of a "randomly moving thing" is, don't we?

In summary, the partial derivative $\frac{\partial f}{\partial x_k}(x)$ will be another stochastic process, under certain existence conditions, related (not statistically independent) to $f(x)$.

We may have several ways of computing properties of the "derivative".

1. If the process is described by a SDE $dy = Ay \cdot dx + B \cdot dW$, or, well, an stochastic partial differential equation, we might use the solutions of them... (for one-dimensional processes we may denote them as $f(t)$ and name them time-series... position, velocity, acceleration come to mind)
2. In Laplace/Frequency domain, via, say, manipulations of a power spectral density.
3. Directly from the properties of the covariance function $k(x, x')$.

We will concentrate on the third approach, as usual in kernel regression, kriging, etc. However, signal processing and control applications might prefer the first two ones (Kalman filter, transfer functions, ...).

Problem statement: as $f(x)$ will have n input arguments, we wish to compute the expected value of $\frac{\partial f}{\partial x_i}(x')$ as well as the covariance between $f(x)$ and $\frac{\partial f}{\partial x_i}(x')$, $1 \leq i \leq n$, as well as the covariance between $\frac{\partial f}{\partial x_i}(x)$ and $\frac{\partial f}{\partial x_j}(x')$ for $1 \leq i \leq n$, $1 \leq j \leq n$.

Mean (expected value) of the gradient

In the sequel e_i will denote the canonical vector in the i-th direction, $i = 1, \ldots, n$, i.e., the one with all components equal to zero except its i-th component, being that one equal to 1.

Gradient "mean" is just the gradient of the mean function:

$$E\left[\frac{\partial f}{\partial x_i}(x)\right] = E[\lim_{h \to 0} \frac{1}{h}(f(x + he_i) - f(x))] = \lim_{h \to 0} \frac{1}{h}(E[f(x + he_i)] - E[f(x)]) = \frac{\partial}{\partial x_i} E[f(x)]$$

Variance and covariance of the estimated gradient

Let us consider a stochastic process with covariance kernel $k(x, x')$.

In the sequel, we will "abuse the notation", in the sense that, rigorously, $cov(a, b) = E[(a - \overline{a})(b - \overline{b})]$, i.e., covariances are the mean of the products of "increments with respect to the mean". So, when I say $E[ab]$, I should say $E[(a - \overline{a})(b - \overline{b})]$, but, well, if you are familiar with variance computations, everything related with the mean ultimately vanishes so carrying out the variance computations assuming "zero mean" ends up giving the correct results.

With the above abuse of notation, we'll understand $E[f(x)f(x')] = k(x, x')$.

- Let us consider covariance between the stochastic process $f(x)$ and its partial derivatives $\frac{\partial f}{\partial x_i}(x')$

at an arbitrary different point x'.

We'll use finite increments and take limits afterwards:

$$E\left[f(x) \cdot \frac{1}{h}(f(x' + he_i) - f(x'))\right] = \frac{1}{h}E[f(x)f(x' + he_i) - f(x)f(x')] = \frac{1}{h}(k(x, x' + he_i) - k(x, x'))$$

So, letting $h \to 0$ we get

$$cov(f(x), \frac{\partial f}{\partial x_i}(x')) = \frac{\partial k(x, x')}{\partial x_i'}$$

Likewise, with analogous developments:

$$cov(\frac{\partial f}{\partial x_i}(x), f(x')) = \frac{\partial k(x, x')}{\partial x_i}$$

*Due to symmetry of the covariance $k(x, x') = k(x', x)$, the two above expressions ultimately mean exactly the same thing.

- Let us now consider covariance between partial derivatives $E\left[\frac{\partial f}{\partial x_i}(x), \frac{\partial f}{\partial x_j}(x')\right]$

First, let us write covariance between finite increments:

$$E\left[\frac{1}{h_1}(f(x + h_1e_i) - f(x)) \cdot \frac{1}{h_2}(f(x' + h_2e_j) - f(x'))\right] =$$

$$\frac{1}{h_1h_2}E[f(x + h_1e_i)f(x' + h_2e_j) - f(x)f(x' + h_2e_j) - f(x + h_1e_i)f(x') + f(x)f(x')] =$$

$$\frac{1}{h_1h_2}E[f(x + h_1e_i)f(x' + h_2e_j) - f(x)f(x' + h_2e_j) - (f(x + h_1e_i)f(x') - f(x)f(x'))] =$$

$$\frac{1}{h_2}E\left[\frac{1}{h_1}(f(x + h_1e_i)f(x' + h_2e_j) - f(x)f(x' + h_2e_j)) - \frac{1}{h_1}(f(x + h_1e_i)f(x') - f(x)f(x'))\right]$$

taking expectations $E[\cdot]$, we get an expression with the covariances

$$\frac{1}{h_2} \cdot \left(\frac{1}{h_1}(k(x + h_1 e_i, x^{'} + h_2 e_j) - k(x, x^{'} + h_2 e_j)) - \frac{1}{h_1}(k(x + h_1 e_i, x^{'}) - k(x, x^{'})) \right)$$

taking limits when $h_1 \to 0$, we have:

$$\frac{1}{h_2} \cdot \left(\frac{\partial k}{\partial x_i}(x, x^{'} + h_2 e_j) - \frac{\partial k}{\partial x_i}(x, x^{'}) \right)$$

Finally, taking limits when $h_2 \to 0$ we get the main result:

$$E\left[\frac{\partial f}{\partial x_i}(x), \frac{\partial f}{\partial x_j}(x^{'})\right] = \frac{\partial^2 k}{\partial x_i \partial x_j^{'}}(x, x^{'})$$

*The particular case $i = j$ will give the covariance function of the i-th partial derivative considered as an standalone (marginal) stochastic process.