



WIKIPEDIA
The Free Encyclopedia

WIKIPEDIA

Gaussian process

In probability theory and statistics, a **Gaussian process** is a stochastic process (a collection of random variables indexed by time or space), such that every finite collection of those random variables has a multivariate normal distribution. The distribution of a Gaussian process is the joint distribution of all those (infinitely many) random variables, and as such, it is a distribution over functions with a continuous domain, e.g. time or space.

The concept of Gaussian processes is named after Carl Friedrich Gauss because it is based on the notion of the Gaussian distribution (normal distribution). Gaussian processes can be seen as an infinite-dimensional generalization of multivariate normal distributions.

Gaussian processes are useful in statistical modelling, benefiting from properties inherited from the normal distribution. For example, if a random process is modelled as a Gaussian process, the distributions of various derived quantities can be obtained explicitly. Such quantities include the average value of the process over a range of times and the error in estimating the average using sample values at a small set of times. While exact models often scale poorly as the amount of data increases, multiple approximation methods have been developed which often retain good accuracy while drastically reducing computation time.

Definition

A time continuous stochastic process $\{X_t; t \in T\}$ is Gaussian if and only if for every finite set of indices t_1, \dots, t_k in the index set T

$$\mathbf{X}_{t_1, \dots, t_k} = (X_{t_1}, \dots, X_{t_k})$$

is a multivariate Gaussian random variable.^[1] That is the same as saying every linear combination of $(X_{t_1}, \dots, X_{t_k})$ has a univariate normal (or Gaussian) distribution.

Using characteristic functions of random variables with i denoting the imaginary unit such that $i^2 = -1$, the Gaussian property can be formulated as follows: $\{X_t; t \in T\}$ is Gaussian if and only if, for every finite set of indices t_1, \dots, t_k , there are real-valued σ_{lj} , μ_l with $\sigma_{jj} > 0$ such that the following equality holds for all $s_1, s_2, \dots, s_k \in \mathbb{R}$,

$$\mathbb{E} \left[\exp \left(i \sum_{\ell=1}^k s_{\ell} X_{t_{\ell}} \right) \right] = \exp \left(-\frac{1}{2} \sum_{\ell, j} \sigma_{\ell j} s_{\ell} s_j + i \sum_{\ell} \mu_{\ell} s_{\ell} \right),$$

or $\mathbb{E} \left[e^{i s (\mathbf{X}_t - \mu)} \right] = e^{-s \sigma s / 2}$. The numbers $\sigma_{\ell j}$ and μ_{ℓ} can be shown to be the covariances and means

of the variables in the process.^[2]

Variance

The variance of a Gaussian process is finite at any time t , formally^{[3]:p. 515}

$$\text{var}[X(t)] = \mathbb{E} \left[|X(t) - \mathbb{E}[X(t)]|^2 \right] < \infty \quad \text{for all } t \in T.$$

Stationarity

For general stochastic processes strict-sense stationarity implies wide-sense stationarity but not every wide-sense stationary stochastic process is strict-sense stationary. However, for a Gaussian stochastic process the two concepts are equivalent.^{[3]:p. 518}

A Gaussian stochastic process is strict-sense stationary if and only if it is wide-sense stationary.

Example

There is an explicit representation for stationary Gaussian processes.^[4] A simple example of this representation is

$$X_t = \cos(at) \xi_1 + \sin(at) \xi_2$$

where ξ_1 and ξ_2 are independent random variables with the standard normal distribution.

Covariance functions

A key fact of Gaussian processes is that they can be completely defined by their second-order statistics.^[5] Thus, if a Gaussian process is assumed to have mean zero, defining the covariance function completely defines the process' behaviour. Importantly the non-negative definiteness of this function enables its spectral decomposition using the Karhunen–Loève expansion. Basic aspects that can be defined through the covariance function are the process' stationarity, isotropy, smoothness and periodicity.^{[6][7]}

Stationarity refers to the process' behaviour regarding the separation of any two points \mathbf{x} and \mathbf{x}' . If the process is stationary, the covariance function depends only on $\mathbf{x} - \mathbf{x}'$. For example, the Ornstein–Uhlenbeck process is stationary.

If the process depends only on $|\mathbf{x} - \mathbf{x}'|$, the Euclidean distance (not the direction) between \mathbf{x} and \mathbf{x}' , then the process is considered isotropic. A process that is concurrently stationary and isotropic is considered to be homogeneous;^[8] in practice these properties reflect the differences (or rather the lack of them) in the behaviour of the process given the location of the observer.

Ultimately Gaussian processes translate as taking priors on functions and the smoothness of these priors can be induced by the covariance function.^[6] If we expect that for "near-by" input points \mathbf{x} and \mathbf{x}' their

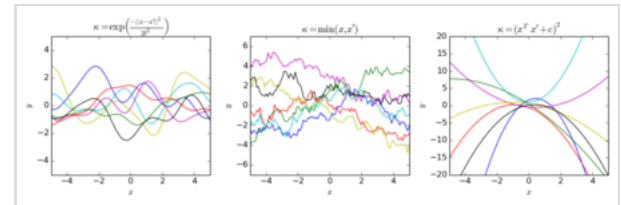
corresponding output points \mathbf{y} and \mathbf{y}' to be "near-by" also, then the assumption of continuity is present. If we wish to allow for significant displacement then we might choose a rougher covariance function. Extreme examples of the behaviour is the Ornstein–Uhlenbeck covariance function and the squared exponential where the former is never differentiable and the latter infinitely differentiable.

Periodicity refers to inducing periodic patterns within the behaviour of the process. Formally, this is achieved by mapping the input \mathbf{x} to a two dimensional vector $\mathbf{u}(\mathbf{x}) = (\cos(\mathbf{x}), \sin(\mathbf{x}))$.

Usual covariance functions

There are a number of common covariance functions:^[7]

- Constant : $K_C(\mathbf{x}, \mathbf{x}') = C$
- Linear: $K_L(\mathbf{x}, \mathbf{x}') = \mathbf{x}^\top \mathbf{x}'$
- white Gaussian noise: $K_{GN}(\mathbf{x}, \mathbf{x}') = \sigma^2 \delta_{\mathbf{x}, \mathbf{x}'}$
- Squared exponential: $K_{SE}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{d^2}{2\ell^2}\right)$
- Ornstein–Uhlenbeck: $K_{OU}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{d}{\ell}\right)$
- Matérn: $K_{\text{Matern}}(\mathbf{x}, \mathbf{x}') = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}d}{\ell}\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}d}{\ell}\right)$
- Periodic: $K_P(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{2}{\ell^2} \sin^2(d/2)\right)$
- Rational quadratic: $K_{RQ}(\mathbf{x}, \mathbf{x}') = (1 + d^2)^{-\alpha}, \quad \alpha \geq 0$



The effect of choosing different kernels on the prior function distribution of the Gaussian process. Left is a squared exponential kernel. Middle is Brownian. Right is quadratic.

Here $d = |\mathbf{x} - \mathbf{x}'|$. The parameter ℓ is the characteristic length-scale of the process (practically, "how close" two points \mathbf{x} and \mathbf{x}' have to be to influence each other significantly), δ is the Kronecker delta and σ the standard deviation of the noise fluctuations. Moreover, K_ν is the modified Bessel function of order ν and $\Gamma(\nu)$ is the gamma function evaluated at ν . Importantly, a complicated covariance function can be defined as a linear combination of other simpler covariance functions in order to incorporate different insights about the data-set at hand.

The inferential results are dependent on the values of the hyperparameters θ (e.g. ℓ and σ) defining the model's behaviour. A popular choice for θ is to provide *maximum a posteriori* (MAP) estimates of it with some chosen prior. If the prior is very near uniform, this is the same as maximizing the *marginal likelihood* of the process; the marginalization being done over the observed process values \mathbf{y} .^[7] This approach is also known as *maximum likelihood II*, *evidence maximization*, or *empirical Bayes*.^[9]

Continuity

For a Gaussian process, *continuity in probability* is equivalent to *mean-square continuity*,^{[10]:145} and *continuity with probability one* is equivalent to *sample continuity*.^{[11]:91} "Gaussian processes are discontinuous at fixed points." The latter implies, but is not implied by, continuity in

probability. Continuity in probability holds if and only if the mean and autocovariance are continuous functions. In contrast, sample continuity was challenging even for stationary Gaussian processes (as probably noted first by Andrey Kolmogorov), and more challenging for more general processes. [12]:Sect. 2.8 [13]:69,81 [14]:80 [15] As usual, by a sample continuous process one means a process that admits a sample continuous modification. [16]:292 [17]:424

Stationary case

For a stationary Gaussian process $\mathbf{X} = (X_t)_{t \in \mathbb{R}}$, some conditions on its spectrum are sufficient for sample continuity, but fail to be necessary. A necessary and sufficient condition, sometimes called Dudley–Fernique theorem, involves the function σ defined by

$$\sigma(h) = \sqrt{\mathbb{E}[X(t+h) - X(t)]^2}$$

(the right-hand side does not depend on t due to stationarity). Continuity of \mathbf{X} in probability is equivalent to continuity of σ at $\mathbf{0}$. When convergence of $\sigma(h)$ to $\mathbf{0}$ (as $h \rightarrow \mathbf{0}$) is too slow, sample continuity of \mathbf{X} may fail. Convergence of the following integrals matters:

$$I(\sigma) = \int_0^1 \frac{\sigma(h)}{h\sqrt{\log(1/h)}} dh = \int_0^\infty 2\sigma(e^{-x^2}) dx,$$

these two integrals being equal according to integration by substitution $h = e^{-x^2}$, $x = \sqrt{\log(1/h)}$. The first integrand need not be bounded as $h \rightarrow \mathbf{0}+$, thus the integral may converge ($I(\sigma) < \infty$) or diverge ($I(\sigma) = \infty$). Taking for example $\sigma(e^{-x^2}) = \frac{1}{x^a}$ for large x , that is, $\sigma(h) = (\log(1/h))^{-a/2}$ for small h , one obtains $I(\sigma) < \infty$ when $a > 1$, and $I(\sigma) = \infty$ when $0 < a \leq 1$. In these two cases the function σ is increasing on $[0, \infty)$, but generally it is not. Moreover, the condition

(*) there exists $\varepsilon > \mathbf{0}$ such that σ is monotone on $[0, \varepsilon]$

does not follow from continuity of σ and the evident relations $\sigma(h) \geq \mathbf{0}$ (for all h) and $\sigma(\mathbf{0}) = \mathbf{0}$.

Theorem 1 — Let σ be continuous and satisfy (*). Then the condition $I(\sigma) < \infty$ is necessary and sufficient for sample continuity of \mathbf{X} .

Some history.^{[17]:424} Sufficiency was announced by Xavier Fernique in 1964, but the first proof was published by Richard M. Dudley in 1967.^{[16]:Theorem 7.1} Necessity was proved by Michael B. Marcus and Lawrence Shepp in 1970.^{[18]:380}

There exist sample continuous processes \mathbf{X} such that $I(\sigma) = \infty$; they violate condition (*). An example found by Marcus and Shepp^{[18]:387} is a random lacunary Fourier series

$$X_t = \sum_{n=1}^{\infty} c_n (\xi_n \cos \lambda_n t + \eta_n \sin \lambda_n t),$$

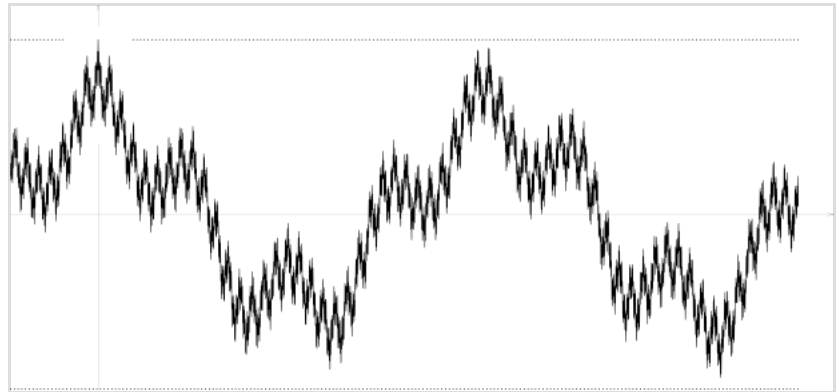
where $\xi_1, \eta_1, \xi_2, \eta_2, \dots$ are independent random variables with standard normal distribution;

frequencies $0 < \lambda_1 < \lambda_2 < \dots$ are a fast growing sequence; and coefficients $c_n > 0$ satisfy $\sum_n c_n < \infty$. The latter relation implies

$$\mathbb{E} \sum_n c_n (|\xi_n| + |\eta_n|) = \sum_n c_n \mathbb{E}[|\xi_n| + |\eta_n|] = \text{const} \cdot \sum_n c_n < \infty,$$

whence $\sum_n c_n (|\xi_n| + |\eta_n|) < \infty$ almost surely, which ensures uniform convergence of the Fourier series almost surely, and sample continuity of \mathbf{X} .

Its autocovariation function



Autocorrelation of a random lacunary Fourier series

$$\mathbb{E}[X_t X_{t+h}] = \sum_{n=1}^{\infty} c_n^2 \cos \lambda_n h$$

is nowhere monotone (see the picture), as well as the corresponding function σ ,

$$\sigma(h) = \sqrt{2\mathbb{E}[X_t X_t] - 2\mathbb{E}[X_t X_{t+h}]} = 2 \sqrt{\sum_{n=1}^{\infty} c_n^2 \sin^2 \frac{\lambda_n h}{2}}.$$

Brownian motion as the integral of Gaussian processes

A Wiener process (also known as Brownian motion) is the integral of a white noise generalized Gaussian process. It is not stationary, but it has stationary increments.

The Ornstein–Uhlenbeck process is a stationary Gaussian process.

The Brownian bridge is (like the Ornstein–Uhlenbeck process) an example of a Gaussian process whose increments are not independent.

The fractional Brownian motion is a Gaussian process whose covariance function is a generalisation of that of the Wiener process.

RKHS structure and Gaussian process

Let \mathbf{f} be a mean-zero Gaussian process $\{X_t; t \in T\}$ with a non-negative definite covariance function \mathbf{K} and let \mathbf{R} be a symmetric and positive semidefinite function. Then, there exists a Gaussian process \mathbf{X} which has the covariance \mathbf{R} . Moreover, the Reproducing kernel Hilbert space associated to \mathbf{R} coincides

with the Cameron–Martin theorem associated space $R(H)$ of X , and all the spaces $R(H)$, H_X , and $\mathcal{H}(K)$ are isometric^[19]. From now on, let $\mathcal{H}(R)$ be a Reproducing kernel Hilbert space with positive definite kernel R .

Driscoll's zero-one law is a result characterizing the sample functions generated by a Gaussian process:

$$\lim_{n \rightarrow \infty} \text{tr}[K_n R_n^{-1}] < \infty,$$

where K_n and R_n are the covariance matrices of all possible pairs of n points, implies

$$\Pr[f \in \mathcal{H}(R)] = 1.$$

Moreover,

$$\lim_{n \rightarrow \infty} \text{tr}[K_n R_n^{-1}] = \infty$$

implies ^[20]

$$\Pr[f \in \mathcal{H}(R)] = 0.$$

This has significant implications when $K = R$, as

$$\lim_{n \rightarrow \infty} \text{tr}[R_n R_n^{-1}] = \lim_{n \rightarrow \infty} \text{tr}[I] = \lim_{n \rightarrow \infty} n = \infty.$$

As such, almost all sample paths of a mean-zero Gaussian process with positive definite kernel K will lie outside of the Hilbert space $\mathcal{H}(K)$.

Linearly constrained Gaussian processes

For many applications of interest some pre-existing knowledge about the system at hand is already given. Consider e.g. the case where the output of the Gaussian process corresponds to a magnetic field; here, the real magnetic field is bound by Maxwell's equations and a way to incorporate this constraint into the Gaussian process formalism would be desirable as this would likely improve the accuracy of the algorithm.

A method on how to incorporate linear constraints into Gaussian processes already exists:^[21]

Consider the (vector valued) output function $f(x)$ which is known to obey the linear constraint (i.e. \mathcal{F}_X is a linear operator)

$$\mathcal{F}_X(f(x)) = 0.$$

Then the constraint \mathcal{F}_X can be fulfilled by choosing $f(x) = \mathcal{G}_X(g(x))$, where $g(x) \sim \mathcal{GP}(\mu_g, K_g)$ is modelled as a Gaussian process, and finding \mathcal{G}_X such that

$$\mathcal{F}_X(\mathcal{G}_X(g)) = 0 \quad \forall g.$$

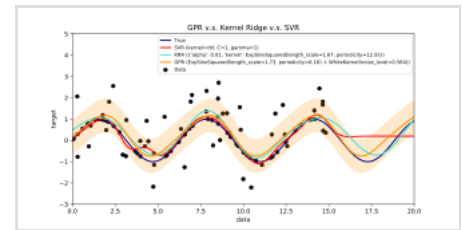
Given \mathcal{G}_X and using the fact that Gaussian processes are closed under linear transformations, the Gaussian process for f obeying constraint \mathcal{F}_X becomes

$$f(x) = \mathcal{G}_X g \sim \mathcal{GP}(\mathcal{G}_X \mu_g, \mathcal{G}_X K_g \mathcal{G}_X^T).$$

Hence, linear constraints can be encoded into the mean and covariance function of a Gaussian process.

Applications

A Gaussian process can be used as a prior probability distribution over functions in Bayesian inference.^{[7][23]} Given any set of N points in the desired domain of your functions, take a multivariate Gaussian whose covariance matrix parameter is the Gram matrix of your N points with some desired kernel, and sample from that Gaussian. For solution of the multi-output prediction problem, Gaussian process regression for vector-valued function was developed. In this method, a 'big' covariance is constructed, which describes the correlations between all the input and output variables taken in N points in the desired domain.^[24] This approach was elaborated in detail for the matrix-valued Gaussian processes and generalised to processes with 'heavier tails' like Student-t processes.^[25]



An example of Gaussian Process Regression (prediction) compared with other regression models.^[22]

Inference of continuous values with a Gaussian process prior is known as Gaussian process regression, or kriging; extending Gaussian process regression to multiple target variables is known as *cokriging*.^[26] Gaussian processes are thus useful as a powerful non-linear multivariate interpolation tool. Kriging is also used to extend Gaussian process in the case of mixed integer inputs.^[27]

Gaussian processes are also commonly used to tackle numerical analysis problems such as numerical integration, solving differential equations, or optimisation in the field of probabilistic numerics.

Gaussian processes can also be used in the context of mixture of experts models, for example.^{[28][29]} The underlying rationale of such a learning framework consists in the assumption that a given mapping cannot be well captured by a single Gaussian process model. Instead, the observation space is divided into subsets, each of which is characterized by a different mapping function; each of these is learned via a different Gaussian process component in the postulated mixture.

In the natural sciences, Gaussian processes have found use as probabilistic models of astronomical time series and as predictors of molecular properties.^[30]

Gaussian process prediction, or Kriging

When concerned with a general Gaussian process regression problem (Kriging), it is assumed that for a Gaussian process \mathbf{f} observed at coordinates \mathbf{x} , the vector of values $\mathbf{f}(\mathbf{x})$ is just one sample from a multivariate Gaussian distribution of dimension equal to number of observed coordinates n . Therefore, under the assumption of a zero-mean distribution, $\mathbf{f}(\mathbf{x}') \sim N(0, K(\theta, \mathbf{x}, \mathbf{x}'))$, where $K(\theta, \mathbf{x}, \mathbf{x}')$ is the covariance matrix between all possible pairs $(\mathbf{x}, \mathbf{x}')$ for a given set of hyperparameters θ .^[7] As such the log marginal likelihood is:

$$\log p(\mathbf{f}(\mathbf{x}') \mid \theta, \mathbf{x}) = -\frac{1}{2} (\mathbf{f}(\mathbf{x})^\top K(\theta, \mathbf{x}, \mathbf{x}')^{-1} \mathbf{f}(\mathbf{x}') + \log \det(K(\theta, \mathbf{x}, \mathbf{x}')) + n \log 2\pi)$$

and maximizing this marginal likelihood towards θ provides the complete specification of the Gaussian process f . One can briefly note at this point that the first term corresponds to a penalty term for a model's failure to fit observed values and the second term to a penalty term that increases proportionally to a model's complexity. Having specified θ , making predictions about unobserved values $f(\mathbf{x}^*)$ at coordinates \mathbf{x}^* is then only a matter of drawing samples from the predictive distribution $p(\mathbf{y}^* | \mathbf{x}^*, f(\mathbf{x}), \mathbf{x}) = N(\mathbf{y}^* | \mathbf{A}, \mathbf{B})$ where the posterior mean estimate \mathbf{A} is defined as

$$\mathbf{A} = \mathbf{K}(\theta, \mathbf{x}^*, \mathbf{x})\mathbf{K}(\theta, \mathbf{x}, \mathbf{x}')^{-1} f(\mathbf{x})$$

and the posterior variance estimate \mathbf{B} is defined as:

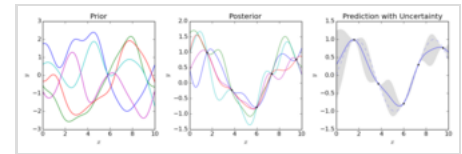
$$\mathbf{B} = \mathbf{K}(\theta, \mathbf{x}^*, \mathbf{x}^*) - \mathbf{K}(\theta, \mathbf{x}^*, \mathbf{x})\mathbf{K}(\theta, \mathbf{x}, \mathbf{x}')^{-1} \mathbf{K}(\theta, \mathbf{x}, \mathbf{x}^*)^T$$

where $\mathbf{K}(\theta, \mathbf{x}^*, \mathbf{x})$ is the covariance between the new coordinate of estimation \mathbf{x}^* and all other observed coordinates \mathbf{x} for a given hyperparameter vector θ , $\mathbf{K}(\theta, \mathbf{x}, \mathbf{x}')$ and $f(\mathbf{x})$ are defined as before and $\mathbf{K}(\theta, \mathbf{x}^*, \mathbf{x}^*)$ is the variance at point \mathbf{x}^* as dictated by θ . It is important to note that practically the posterior mean estimate of $f(\mathbf{x}^*)$ (the "point estimate") is just a linear combination of the observations $f(\mathbf{x})$; in a similar manner the variance of $f(\mathbf{x}^*)$ is actually independent of the observations $f(\mathbf{x})$. A known bottleneck in Gaussian process prediction is that the computational complexity of inference and likelihood evaluation is cubic in the number of points $|\mathbf{x}|$, and as such can become unfeasible for larger data sets.^[6] Works on sparse Gaussian processes, that usually are based on the idea of building a *representative set* for the given process f , try to circumvent this issue.^{[31][32]} The kriging method can be used in the latent level of a nonlinear mixed-effects model for a spatial functional prediction: this technique is called the latent kriging.^[33]

Often, the covariance has the form $\mathbf{K}(\theta, \mathbf{x}, \mathbf{x}') = \frac{1}{\sigma^2} \tilde{\mathbf{K}}(\theta, \mathbf{x}, \mathbf{x}')$, where σ^2 is a scaling parameter. Examples are the Matérn class covariance functions. If this scaling parameter σ^2 is either known or unknown (i.e. must be marginalized), then the posterior probability, $p(\theta | \mathbf{D})$, i.e. the probability for the hyperparameters θ given a set of data pairs \mathbf{D} of observations of \mathbf{x} and $f(\mathbf{x})$, admits an analytical expression.^[34]

Bayesian neural networks as Gaussian processes

Bayesian neural networks are a particular type of Bayesian network that results from treating deep learning and artificial neural network models probabilistically, and assigning a prior distribution to their parameters. Computation in artificial neural networks is usually organized into sequential layers of artificial neurons. The number of neurons in a layer is called the layer width. As layer width grows large, many Bayesian neural networks reduce to a Gaussian process with a closed form compositional kernel. This Gaussian process is called the Neural Network Gaussian Process (NNGP).^{[7][35][36]} It allows predictions from Bayesian neural networks to be more efficiently evaluated, and provides an analytic tool to understand deep learning models.



Gaussian Process Regression (prediction) with a squared exponential kernel. Left plot are draws from the prior function distribution. Middle are draws from the posterior. Right is mean prediction with one standard deviation shaded.

Computational issues

In practical applications, Gaussian process models are often evaluated on a grid leading to multivariate normal distributions. Using these models for prediction or parameter estimation using maximum likelihood requires evaluating a multivariate Gaussian density, which involves calculating the determinant and the inverse of the covariance matrix. Both of these operations have cubic computational complexity which means that even for grids of modest sizes, both operations can have a prohibitive computational cost. This drawback led to the development of multiple approximation methods.

See also

- Bayes linear statistics
- Bayesian interpretation of regularization
- Kriging
- Gaussian free field
- Gauss–Markov process
- Gradient-enhanced kriging (GEK)
- Student's t-process

References

1. MacKay, David, J.C. (2003). *Information Theory, Inference, and Learning Algorithms* (<http://www.inference.phy.cam.ac.uk/itprnn/book.pdf>) (PDF). Cambridge University Press. p. 540. ISBN 9780521642989. "The probability distribution of a function $y(\mathbf{x})$ is a Gaussian processes if for any finite selection of points $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}$, the density $P(y(\mathbf{x}^{(1)}), y(\mathbf{x}^{(2)}), \dots, y(\mathbf{x}^{(N)}))$ is a Gaussian"
2. Dudley, R.M. (1989). *Real Analysis and Probability*. Wadsworth and Brooks/Cole. ISBN 0-534-10050-3.
3. Amos Lapidoth (8 February 2017). *A Foundation in Digital Communication* (<https://books.google.com/books?id=6oTuDQAAQBAJ&q=independence>). Cambridge University Press. ISBN 978-1-107-17732-1.
4. Kac, M.; Siegert, A.J.F (1947). "An Explicit Representation of a Stationary Gaussian Process" (<https://doi.org/10.1214%2Faoms%2F1177730391>). *The Annals of Mathematical Statistics*. **18** (3): 438–442. doi:10.1214/aoms/1177730391 (<https://doi.org/10.1214%2Faoms%2F1177730391>).
5. Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. Springer. ISBN 978-0-387-31073-2.
6. Barber, David (2012). *Bayesian Reasoning and Machine Learning* (<http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.HomePage>). Cambridge University Press. ISBN 978-0-521-51814-7.

7. Rasmussen, C.E.; Williams, C.K.I (2006). *Gaussian Processes for Machine Learning* (<http://www.gaussianprocess.org/gpml/>). MIT Press. ISBN 978-0-262-18253-9.
8. Grimmett, Geoffrey; David Stirzaker (2001). *Probability and Random Processes*. Oxford University Press. ISBN 978-0198572220.
9. Seeger, Matthias (2004). "Gaussian Processes for Machine Learning". *International Journal of Neural Systems*. **14** (2): 69–104. CiteSeerX 10.1.1.71.1079 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.71.1079>). doi:10.1142/s0129065704001899 (<https://doi.org/10.1142/s0129065704001899>). PMID 15112367 (<https://pubmed.ncbi.nlm.nih.gov/15112367>). S2CID 52807317 (<https://api.semanticscholar.org/CorpusID:52807317>).
10. Dudley, R. M. (1975). "The Gaussian process and how to approach it" (<https://www.mathunion.org/fileadmin/ICM/Proceedings/ICM1974.2/ICM1974.2.ocr.pdf>) (PDF). *Proceedings of the International Congress of Mathematicians*. Vol. 2. pp. 143–146.
11. Dudley, R. M. (2010). "Sample Functions of the Gaussian Process" (<http://projecteuclid.org/euclid.aop/1176997026>). *Selected Works of R.M. Dudley*. Vol. 1. pp. 66–103. doi:10.1007/978-1-4419-5821-1_13 (https://doi.org/10.1007/978-1-4419-5821-1_13). ISBN 978-1-4419-5820-4. {{cite book}}: |journal= ignored (help)
12. Talagrand, Michel (2014). *Upper and lower bounds for stochastic processes: modern methods and classical problems* (<https://www.springer.com/gp/book/9783642540745>). *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics*. Springer, Heidelberg. ISBN 978-3-642-54074-5.
13. Ledoux, Michel (1996), "Isoperimetry and Gaussian analysis", in Dobrushin, Roland; Groeneboom, Piet; Ledoux, Michel (eds.), *Lectures on Probability Theory and Statistics: Ecole d'Été de Probabilités de Saint-Flour XXIV–1994*, *Lecture Notes in Mathematics*, vol. 1648, Berlin: Springer, pp. 165–294, doi:10.1007/BFb0095676 (<https://doi.org/10.1007/BFb0095676>), ISBN 978-3-540-62055-6, MR 1600888 (<https://mathscinet.ams.org/mathscinet-getitem?mr=1600888>)
14. Adler, Robert J. (1990). *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. Vol. 12. Hayward, California: Institute of Mathematical Statistics. ISBN 0-940600-17-X. JSTOR 4355563 (<https://www.jstor.org/stable/4355563>). MR 1088478 (<https://mathscinet.ams.org/mathscinet-getitem?mr=1088478>). {{cite book}}: |journal= ignored (help)
15. Berman, Simeon M. (1992). "Review of: Adler 1990 'An introduction to continuity...'". *Mathematical Reviews*. MR 1088478 (<https://mathscinet.ams.org/mathscinet-getitem?mr=1088478>).
16. Dudley, R. M. (1967). "The sizes of compact subsets of Hilbert space and continuity of Gaussian processes" ([https://doi.org/10.1016/0022-1236\(67\)90017-1](https://doi.org/10.1016/0022-1236(67)90017-1)). *Journal of Functional Analysis*. **1** (3): 290–330. doi:10.1016/0022-1236(67)90017-1 ([https://doi.org/10.1016/0022-1236\(67\)90017-1](https://doi.org/10.1016/0022-1236(67)90017-1)).
17. Marcus, M.B.; Shepp, Lawrence A. (1972). "Sample behavior of Gaussian processes" (<https://projecteuclid.org/euclid.bsmsp/1200514231>). *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability, vol. II: probability theory*. Vol. 6. Univ. California, Berkeley. pp. 423–441.

18. Marcus, Michael B.; Shepp, Lawrence A. (1970). "Continuity of Gaussian processes" (<https://doi.org/10.1090%2Fs0002-9947-1970-0264749-1>). *Transactions of the American Mathematical Society*. **151** (2): 377–391. doi:10.1090/s0002-9947-1970-0264749-1 (<https://doi.org/10.1090%2Fs0002-9947-1970-0264749-1>). JSTOR 1995502 (<https://www.jstor.org/stable/1995502>).
19. Azmoodeh, Ehsan; Sottinen, Tommi; Viitasaari, Lauri; Yazigi, Adil (2014). "Necessary and sufficient conditions for Hölder continuity of Gaussian processes". *Statistics & Probability Letters*. **94**: 230–235. arXiv:1403.2215 (<https://arxiv.org/abs/1403.2215>). doi:10.1016/j.spl.2014.07.030 (<https://doi.org/10.1016%2Fj.spl.2014.07.030>).
20. Driscoll, Michael F. (1973). "The reproducing kernel Hilbert space structure of the sample paths of a Gaussian process" (<https://doi.org/10.1007%2FBF00534894>). *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*. **26** (4): 309–316. doi:10.1007/BF00534894 (<https://doi.org/10.1007%2FBF00534894>). ISSN 0044-3719 (<https://search.worldcat.org/issn/0044-3719>). S2CID 123348980 (<https://api.semanticscholar.org/CorpusID:123348980>).
21. Jidling, Carl; Wahlström, Niklas; Wills, Adrian; Schön, Thomas B. (2017-09-19). "Linearly constrained Gaussian processes". arXiv:1703.00787 (<https://arxiv.org/abs/1703.00787>) [stat.ML (<https://arxiv.org/archive/stat.ML>)].
22. The documentation for scikit-learn also has similar examples (http://scikit-learn.org/stable/auto_examples/gaussian_process/plot_compare_gpr_krr.html).
23. Liu, W.; Principe, J.C.; Haykin, S. (2010). *Kernel Adaptive Filtering: A Comprehensive Introduction* (<https://web.archive.org/web/20160304042652/http://www.cnel.ufl.edu/~weifeng/publication.htm>). John Wiley. ISBN 978-0-470-44753-6. Archived from the original (<http://www.cnel.ufl.edu/~weifeng/publication.htm>) on 2016-03-04. Retrieved 2010-03-26.
24. Álvarez, Mauricio A.; Rosasco, Lorenzo; Lawrence, Neil D. (2012). "Kernels for vector-valued functions: A review" (<http://eprints.whiterose.ac.uk/114503/1/1106.6251v2.pdf>) (PDF). *Foundations and Trends in Machine Learning*. **4** (3): 195–266. doi:10.1561/22000000036 (<https://doi.org/10.1561%2F22000000036>). S2CID 456491 (<https://api.semanticscholar.org/CorpusID:456491>).
25. Chen, Zexun; Wang, Bo; Gorban, Alexander N. (2019). "Multivariate Gaussian and Student-t process regression for multi-output prediction" (<https://doi.org/10.1007%2Fs00521-019-04687-8>). *Neural Computing and Applications*. **32** (8): 3005–3028. arXiv:1703.04455 (<https://arxiv.org/abs/1703.04455>). doi:10.1007/s00521-019-04687-8 (<https://doi.org/10.1007%2Fs00521-019-04687-8>).
26. Stein, M.L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer.
27. Saves, Paul; Diouane, Youssef; Bartoli, Nathalie; Lefebvre, Thierry; Morlier, Joseph (2023). "A mixed-categorical correlation kernel for Gaussian process". *Neurocomputing*. **550**: 126472. arXiv:2211.08262 (<https://arxiv.org/abs/2211.08262>). doi:10.1016/j.neucom.2023.126472 (<https://doi.org/10.1016%2Fj.neucom.2023.126472>).

28. Platanios, Emmanouil A.; Chatzis, Sotirios P. (2014). "Gaussian Process-Mixture Conditional Heteroscedasticity". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **36** (5): 888–900. doi:10.1109/TPAMI.2013.183 (<https://doi.org/10.1109%2FTPAMI.2013.183>). PMID 26353224 (<https://pubmed.ncbi.nlm.nih.gov/26353224>). S2CID 10424638 (<https://api.semanticscholar.org/CorpusID:10424638>).
29. Chatzis, Sotirios P. (2013). "A latent variable Gaussian process model with Pitman–Yor process priors for multiclass classification". *Neurocomputing*. **120**: 482–489. doi:10.1016/j.neucom.2013.04.029 (<https://doi.org/10.1016%2Fj.neucom.2013.04.029>)
.
30. Griffiths, Ryan-Rhys (2022). *Applications of Gaussian Processes at Extreme Lengthscales: From Molecules to Black Holes* (PhD thesis). University of Cambridge. arXiv:2303.14291 (<https://arxiv.org/abs/2303.14291>). doi:10.17863/CAM.93643 (<https://doi.org/10.17863%2FCAM.93643>).
31. Smola, A.J.; Schoellkopf, B. (2000). "Sparse greedy matrix approximation for machine learning". *Proceedings of the Seventeenth International Conference on Machine Learning*: 911–918. CiteSeerX 10.1.1.43.3153 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.43.3153>).
32. Csato, L.; Opper, M. (2002). "Sparse on-line Gaussian processes". *Neural Computation*. **14** (3): 641–668. CiteSeerX 10.1.1.335.9713 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.335.9713>). doi:10.1162/089976602317250933 (<https://doi.org/10.1162%2F089976602317250933>). PMID 11860686 (<https://pubmed.ncbi.nlm.nih.gov/11860686>). S2CID 11375333 (<https://api.semanticscholar.org/CorpusID:11375333>).
33. Lee, Se Yoon; Mallick, Bani (2021). "Bayesian Hierarchical Modeling: Application Towards Production Results in the Eagle Ford Shale of South Texas" (<https://doi.org/10.1007%2Fs13571-020-00245-8>). *Sankhya B*. **84**: 1–43. doi:10.1007/s13571-020-00245-8 (<https://doi.org/10.1007%2Fs13571-020-00245-8>).
34. Ranftl, Sascha; Melito, Gian Marco; Badeli, Vahid; Reinbacher-Köstinger, Alice; Ellermann, Katrin; von der Linden, Wolfgang (2019-12-31). "Bayesian Uncertainty Quantification with Multi-Fidelity Data and Gaussian Processes for Impedance Cardiography of Aortic Dissection" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7516489>). *Entropy*. **22** (1): 58. Bibcode:2019Entrp..22...58R (<https://ui.adsabs.harvard.edu/abs/2019Entrp..22...58R>). doi:10.3390/e22010058 (<https://doi.org/10.3390%2Fe22010058>). ISSN 1099-4300 (<https://search.worldcat.org/issn/1099-4300>). PMC 7516489 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7516489>). PMID 33285833 (<https://pubmed.ncbi.nlm.nih.gov/33285833>).
35. Novak, Roman; Xiao, Lechao; Hron, Jiri; Lee, Jaehoon; Alemi, Alexander A.; Sohl-Dickstein, Jascha; Schoenholz, Samuel S. (2020). "Neural Tangents: Fast and Easy Infinite Neural Networks in Python". *International Conference on Learning Representations*. arXiv:1912.02803 (<https://arxiv.org/abs/1912.02803>).
36. Neal, Radford M. (2012). *Bayesian Learning for Neural Networks*. Springer Science and Business Media.

External links

Literature

- The Gaussian Processes Web Site, including the text of Rasmussen and Williams' Gaussian Processes for Machine Learning (<https://gaussianprocess.org>)
- Ebden, Mark (2015). "Gaussian Processes: A Quick Introduction". [arXiv:1505.02965](https://arxiv.org/abs/1505.02965) (<https://arxiv.org/abs/1505.02965>) [math.ST (<https://arxiv.org/archive/math>.ST)].
- A Review of Gaussian Random Fields and Correlation Functions (http://publications.nlr.no/917_Rapport.pdf)
- Efficient Reinforcement Learning using Gaussian Processes (<https://web.archive.org/web/20180826005000/https://pdfs.semanticscholar.org/c9f2/1b84149991f4d547b3f0f625f710750ad8d9.pdf>)

Software

- GPML: A comprehensive Matlab toolbox for GP regression and classification (<http://www.gaussianprocess.org/gpml/code/matlab/doc/>)
- STK: a Small (Matlab/Octave) Toolbox for Kriging and GP modeling (<http://sourceforge.net/projects/kriging>)
- Kriging module in UQLab framework (Matlab) (<http://www.uqlab.com/>)
- CODES Toolbox: implementations of Kriging, variational kriging and multi-fidelity models (Matlab) (<http://codes.arizona.edu/toolbox/>)
- Matlab/Octave function for stationary Gaussian fields (<http://au.mathworks.com/matlabcentral/fileexchange/38880>;)
- Yelp MOE – A black box optimization engine using Gaussian process learning (<https://github.com/Yelp/MOE>)
- ooDACE (<http://www.sumo.intec.ugent.be/ooDACE>) Archived (<https://web.archive.org/web/20200809021046/http://sumo.intec.ugent.be/ooDACE>) 2020-08-09 at the Wayback Machine – A flexible object-oriented Kriging Matlab toolbox.
- GPstuff – Gaussian process toolbox for Matlab and Octave (<https://web.archive.org/web/20141009045756/http://becs.aalto.fi/en/research/bayes/gpstuff/>)
- GPy – A Gaussian processes framework in Python (<https://github.com/SheffieldML/GPy>)
- GSTools - A geostatistical toolbox, including Gaussian process regression, written in Python (<https://github.com/GeoStat-Framework/GSTools>)
- Interactive Gaussian process regression demo (<http://www.tmpl.fi/gp/>)
- Basic Gaussian process library written in C++11 (<https://github.com/ChristophJud/GPR>)
- scikit-learn (<http://scikit-learn.org>) – A machine learning library for Python which

includes Gaussian process regression and classification

- [1] (<https://github.com/modsim/KriKit>) - The Kriging toolKit (KriKit) is developed at the Institute of Bio- and Geosciences 1 (IBG-1) of Forschungszentrum Jülich (FZJ)

Video tutorials

- Gaussian Process Basics by David MacKay (http://videolectures.net/gpip06_mackay_gpb)
- Learning with Gaussian Processes by Carl Edward Rasmussen (http://videolectures.net/epsrcws08_rasmussen_lgp)
- Bayesian inference and Gaussian processes by Carl Edward Rasmussen (http://videolectures.net/mlss07_rasmussen_bigp)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Gaussian_process&oldid=1248409134"