STAT 238 - Bayesian Statistics
Lecture Twenty Two

Spring 2026, UC Berkeley

Aditya Guntuboyina

16 March 2026

1 Gaussian Processes

The goal is to infer an unknown function $f : \Omega \to \mathbb{R}$ that is defined on a known domain $\Omega \subseteq \mathbb{R}^p$. We assume that $\{f(x), x \in \Omega\}$ forms a Gaussian process with mean zero and covariance function or *kernel* given by $K(x, x')$ i.e.,

$$\text{Cov}(f(x), f(x')) = K(x, x') \text{ for all } x, x' \in \Omega.$$

The kernel needs to be positive semi-definite i.e., for every $N \geq 1$, distinct points $u_1, \dots, u_N \in \Omega$, the $N \times N$ matrix with $(i,j)^{th}$ entry $K(u_i, u_j)$ is positive semi-definite. Sometimes, this matrix will be positive definite, and hence invertible.

Here are some examples of Gaussian processes and kernels.

1. **Brownian Motion**: Here $\Omega = [0, \infty)$ and $K(s,t) = \min(s,t)$.
2. **(scaled) Brownian Motion plus constant**: Here again $\Omega = [0, \infty)$. In the Brownian motion model, $f(0) = 0$ which is generally an unrealistic assumption. To fix this, we assume that $f(t) = \beta_0 + \tau W_t$ where $W_t \sim BM$, $\beta_0 \sim N(0, C)$ and $\tau > 0$ (and independence between β_0 and $\{W_t\}$). C will be taken to be large (potentially $C \to \infty$). Now
$$K(s,t) = C + \tau^2 \min(s,t).$$
3. **Integrated Brownian Motion**: $\Omega = [0, \infty)$ and
$$f(t) = \int_0^t W_s ds \text{ where } W_s \sim BM.$$
The kernel is given by
$$\begin{aligned} K(s,t) &= \text{Cov}\left(\int_0^s W_u du, \int_0^t W_v dv\right) \\ &= \int_0^s \int_0^t \text{Cov}(W_u, W_v) dv du = \int_0^s \int_0^t \min(u,v) dv du \end{aligned}$$
To simplify further, assume that $s \leq t$ so that
$$\begin{aligned} K(s,t) &= \int_0^s \left(\int_0^u v dv + \int_u^t u dv\right) du \\ &= \int_0^s \left(\frac{u^2}{2} + u(t-u)\right) du = \int_0^s \left(ut - \frac{u^2}{2}\right) du = t\frac{s^2}{2} - \frac{s^3}{6}. \end{aligned}$$

For general s and t, we have

$$K(s,t) = \frac{1}{2}\max(s,t)\left(\min(s,t)\right)^2 - \frac{1}{6}\left(\min(s,t)\right)^3.$$

4. **(scaled) Integrated Brownian Motion Plus a Linear Term**: In the IBM model, we have $f(0) = 0$ and $f'(0) = 0$. This might be an unrealistic assumption to make when f is completely unknown. In this case, a better model is:

$$f(t) = \beta_0 + \beta_1 t + \tau \int_0^t W_s ds$$

where $\beta_0, \beta_1, \{W_s\}$ are independent with W_s being Brownian motion and $\beta_0, \beta_1 \overset{\text{i.i.d}}{\sim} N(0, C)$. The kernel now becomes

$$\begin{aligned} K(s,t) &= \mathrm{Cov}\left(\beta_0 + \beta_1 s + \tau \int_0^s W_u du, \beta_0 + \beta_1 t + \tau \int_0^t W_v dv\right) \\ &= C(1+st) + \frac{\tau^2}{2}\max(s,t)\left(\min(s,t)\right)^2 - \frac{\tau^2}{6}\left(\min(s,t)\right)^3. \end{aligned}$$

We will see some more examples of kernels later. Next we see some basic applications of Gaussian Processes.

2 Interpolation

Suppose we are given values of f at $0 \le x_1 < \cdots < x_n \le 1$: $f(x_1), \ldots, f(x_n)$. Let $x \in [0,1]$ be a new point (i.e., distinct from $x_1, \ldots, x_n$). What can we say about $f(x)$?

We will solve this problem assuming a Gaussian process prior for f with mean zero and covariance kernel $K(\cdot,\cdot)$. The answer then will be given by the conditional distribution of $f(x)$ given $f(x_1), \ldots, f(x_n)$. To compute this, we note that:

$$(f(x_1), \ldots, f(x_n), f(x))^T \sim N\left(0, \begin{pmatrix} (K(x_i, x_j))_{n\times n} & (K(x_i, x))_{n\times 1} \\ (K(x, x_i))_{1\times n} & K(x,x) \end{pmatrix}\right).$$

Using the notation $K = (K(x_i, x_j))_{n\times n}$ and $\mathbf{k} = (K(x_i, x))_{n\times 1}$, we can write the conditional distribution of $f(x)$ given $f(x_1), \ldots, f(x_n)$ as

$$f(x) \mid f(x_1), \ldots, f(x_n) \sim N\left(\mathbf{k}^T K^{-1}(f(x_1), \ldots, f(x_n))^T, K(x,x) - \mathbf{k}^T K^{-1}\mathbf{k}\right).$$

Thus the posterior mean (or mode) estimate of $f(x)$ is given by

$$\widehat{f(x)} = \mathbf{k}^T K^{-1}(f(x_1), \ldots, f(x_n))^T. \tag{1}$$

For different kernels K, the above expression behaves differently on $f(x_1), \ldots, f(x_n)$. The simplest example is that of Brownian Motion.

Example 2.1 (Brownian Motion)**.** *Suppose $f(x) = \beta_0 + \tau B(x)$ where $B(x)$ is Brownian Motion, $\beta_0 \sim N(0, C)$ with $C \to \infty$ and τ is a fixed positive constant. Then the kernel is:*

$$K(x_1, x_2) = C + \tau^2 \min(x_1, x_2). \tag{2}$$

In this case, (1) *can be explicitly evaluated (in the limit $C \to \infty$) as*

$$\widehat{f(x)} = \begin{cases} f(x_1), & x \le x_1, \\ \dfrac{x_{i+1} - x}{x_{i+1} - x_i} f(x_i) + \dfrac{x - x_i}{x_{i+1} - x_i} f(x_{i+1}), & x_i \le x \le x_{i+1}, \\ f(x_n), & x \ge x_n. \end{cases} \tag{3}$$

The details behind why (1) *with the kernel* (2) *leads to* (3) *will be seen later.*

The formula (3) *for* $\widehat{f(x)}$ *performs linear interpolation between the observed points, and constant extrapolation outside the observed range.*

Also note that $\widehat{f(x)}$ *in* (3) *does not depend on* τ*. The hyperparameter* τ *only affects the posterior variance of* $f(x)$*, and not the posterior mean.*

Next example is that of the Integrated Brownian Motion prior.

Example 2.2 (Integrated Brownian Motion)**.** *Suppose*

$$f(x) = \beta_0 + \beta_1 x + \tau I(x), \tag{4}$$

where $I(x)$ *is integrated Brownian motion and* $\beta_0, \beta_1 \overset{i.i.d.}{\sim} N(0, C)$ *with* $C \to \infty$*. In this case, the posterior mean in* (1) *can be evaluated explicitly and is the natural cubic spline interpolant of the observations. We will see details behind this later.*

3 Integration

Suppose we are given values of f at $0 \le x_1 < \cdots < x_n \le$: $f(x_1), f \dots, f(x_n)$. What can we say about $\int_0^1 f(x)$?

Again we place a Gaussian process prior on f. By linearity, we can write

$$\begin{aligned} &\mathbb{E}\left[\int_0^1 f(x)dx \mid f(x_1), \dots, f(x_n)\right] \\ &= \int_0^1 \mathbb{E}\left[f(x) \mid f(x_1), \dots, f(x_n)\right] \\ &= \int_0^1 \mathbf{k}^T K^{-1}(f(x_1), \dots, f(x_n))^T \\ &= \left(\int_0^1 k(x_1, x)dx, \dots, \int_0^1 k(x_n, x)dx\right)^T K^{-1}(f(x_1), \dots, f(x_n))^T. \end{aligned}$$

Example 3.1 (Brownian Motion)**.** *For the Brownian motion, we already have the formula* (3) *for* $\mathbb{E}[f(x) \mid f(x_1), \dots, f(x_n)]$*. So the posterior mean for* $\int_0^1 f(x)dx$ *is obtained by simply integrating the right hand side of* (3) *from 0 to 1. This leads to:*

$$\begin{aligned} &\mathbb{E}\left(\int_0^1 f(x)dx \mid f(x_1), \dots, f(x_n)\right) \\ &= x_1 f(x_1) + \sum_{i=1}^{n-1} (x_{i+1} - x_i) \frac{f(x_i) + f(x_{i+1})}{2} + (1 - x_n) f(x_n). \end{aligned}$$

This the trapezoidal integration rule. Therefore, with the Brownian motion prior, the posterior mean estimate of $\int_0^1 f(x)dx$ *coincides with the trapezoidal rule.*

Example 3.2 (Integrated Brownian Motion)**.** *With the prior* (4)*, we get a version of the Simpson integration rule.*

For more on the relation between classical integration (quadrature) rules and Gaussian processes, see the book Hennig et al. [2] or the paper Diaconis [1].

In the next lecture, we shall see how to use Gaussian processes for the regression problem.

References

[1] Diaconis, P. (1988) *Bayesian numerical analysis.* Statistical Decision Theory and Related Topics IV, Vol. 1, 163–175.

[2] Hennig, P., Osborne, M. A., and Kersting, H. P. (2022) *Probabilistic Numerics: Computation as Machine Learning.* Cambridge University Press.