STAT 238 - Bayesian Statistics
Lecture Twenty One

Spring 2026, UC Berkeley

Aditya Guntuboyina

13 March 2026

1 Recap: Model from past few lectures

We studied the following model in the past few lectures.

We have a response variable y and a single covariate x. In our example, y denotes weekly earnings and x denotes years of experience. The covariate x takes the values $0, 1, \dots, m$ for some fixed integer m.

Our data is $(x_i, y_i), i = 1, \dots, n$. The model is:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 \text{ReLU}(x_i - 1) + \dots + \beta_m \text{ReLU}(x_i - (m-1)) + \epsilon_i \tag{1}$$

where $\text{ReLU}(u) := \max(u, 0)$, and $\epsilon_i \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. Note we did not include $\text{ReLU}(x_i - m)$ because it always equals 0.

We discussed Bayesian inference in this model using the prior:

$$\beta_0 \sim N(0, C), \beta_1 \sim N(0, C), \beta_2, \dots, \beta_m \overset{\text{i.i.d}}{\sim} N(0, \tau^2) \tag{2}$$

Towards the end of last lecture, we remarked that the resulting model on the regression function:

$$f(x) = \beta_0 + \beta_1 x + \beta_2 \text{ReLU}(x - 1) + \dots + \beta_m \text{ReLU}(x - (m-1))$$

with prior (2) is very similar to:

$$f(x) = \beta_0 + \beta_1 x + \tau I(x) \tag{3}$$

where $I(x)$ is integrated Brownian motion (and $\beta_0, \beta_1 \overset{\text{i.i.d}}{\sim} N(0, C)$). We shall some intuition today as to why the Integrated Brownian Motion is appearing here.

2 Brownian Motion and Integrated Brownian Motion

Before defining Brownian motion, let us recall the definition of Brownian motion. Brownian motion on $[0, M]$ (for some fixed $M > 0$) is a stochastic process $\{B(t), 0 \le t \le M\}$ characterized by the following two properties:

1. Every realization is a continuous function on $[0, M]$, and $B(0) = 0$.

2. For every fixed $t_1, \ldots, t_k \in [0, M]$, the random vector $(B(t_1), \ldots, B(t_k))$ has a multivariate normal distribution with mean zero and covariance matrix

$$\text{Cov}(B(t_i), B(t_j)) = \min(t_i, t_j).$$

This definition suggests the following method for simulating $B(t)$. Construct the dense grid

$$0 = t_0 < t_1 < \cdots < t_N = M$$

on $[0, M]$ with $t_i = iM/N$. One can then simulate $(B(t_1), \ldots, B(t_N))$ from the multivariate normal distribution with mean zero and covariance matrix $\min(t_i, t_j)$.

However, this approach is computationally inefficient for large N. Sampling directly from an $N \times N$ covariance matrix typically requires a matrix factorization (such as a Cholesky decomposition), which has computational cost on the order of $O(N^3)$.

For simulations in particular, the following alternative definition is more appealing:

1. Every realization is a continuous function on $[0, M]$ with $B(0) = 0$.

2. The process has independent increments which means that $B(t_1), B(t_2)-B(t_1), \ldots, B(t_k)-B(t_{k-1})$ are independent whenever $0 \leq t_1 < \cdots < t_k \leq M$.

3. $B(t) - B(s) \sim N(0, t - s)$ whenever $0 \leq s \leq t$.

With this definition, we can simulate $B(t_1), \ldots, B(t_N)$ for $t_i = iM/N$ in the following way:

- First generate $\alpha_1, \ldots, \alpha_N \overset{\text{i.i.d}}{\sim} N(0, M/N)$
- Take $B(t_i) = \alpha_1 + \cdots + \alpha_i$ for $i = 1, \ldots, N$.

Note that in this way $\alpha_i = B(t_i) - B(t_{i-1})$. The equation $B(t_i) = \alpha_1 + \cdots + \alpha_i$ can also be written as

$$B(t) = \sum_{i=1}^{N} \alpha_i I\{t \geq t_i\} \quad \text{for } t \in \{t_1, \ldots, t_N\}.$$

Because of the denseness of the grid (when N is large), we are claiming that

$$B(t) \approx \sum_{i=1}^{N} \alpha_i I\{t \geq t_i\} \quad \text{for all } t \in [0, M]. \tag{4}$$

This means that Brownian motion can be well-approximated (when N is large) by a piecewise constant process which jumps at each grid point $t_i = iM/N$ by a small amount given by $N(0, M/N)$.

We are now ready to define the Integrated Brownian Motion. This is given by:

$$I(t) = \int_0^t B(s) ds. \tag{5}$$

Using the approximation (4), we can write

$$I(t) \approx \int_0^t \left(\sum_{i=1}^{N} \alpha_i I\{t \geq t_i\} \right) ds = \sum_{i=1}^{N} \alpha_i \int_0^t I\{s \geq t_i\} ds = \sum_{i=1}^{N} \alpha_i (t - t_i)_+.$$

Therefore Integrated Brownian Motion can be approximated as:

$$I(t) \approx \sum_{i=1}^{N} \alpha_i (t - t_i)_+ \tag{6}$$

when N is large, $t_i = iM/N$ and $\alpha_i \overset{\text{i.i.d}}{\sim} N(0, M/N)$.

3 The Model (3)

Using (6), the (prior) model (3) can therefore be written as:

$$f(x) \approx \beta_0 + \beta_1 x + \tau \sum_{i=1}^{N} \alpha_i \left(x - \frac{iM}{N} \right)_+$$

Here N is a large number. If we do a further coarse approximation with $N = M$, we get back the model (1) if we make the identification $\beta_i = \tau \alpha_{i-1}$ for $i = 2, \dots, N$ and $M = m$. Therefore, the model (1) can be understood as a coarse discretization of the model (3) which stipulates that f is an Integrated Brownian Motion (scaled by τ) and added to a linear function $\beta_0 + \beta_1 x$ with the uniformative $N(0, C)$ prior on the coefficients β_0, β_1.

The prior (3) is an example of a Gaussian process prior.

4 Gaussian Process Regression

Consider the usual nonparametric regression problem where the goal is to estimate an unknown function $f : \Omega \to \mathbb{R}$ from observations $(x_1, y_1), \dots, (x_n, y_n)$ under the model:

$$y_i = f(x_i) + \epsilon_i \quad \text{where } \epsilon_i \overset{\text{i.i.d}}{\sim} N(0, \sigma^2).$$

Here Ω denotes the domain which is a subset of $\mathbb{R}^d$ for some $d \geq 1$.

We assume that $\{f(x), x \in \Omega\}$ forms a Gaussian process with mean function $\mu(x), x \in \Omega$ and covariance function or *kernel* given by $K(x, x')$ i.e.,

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

The kernel is 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. Often, this matrix will be positive definite, and hence invertible.

The mean function is usually taken to be zero. Here are some standard special cases. 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 $\Omega = [0, \infty)$ and 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. Now

$$K(s, t) = C + \tau^2 \min(s, t).$$

3. **Integrated Brownian Motion**: $\Omega = [0, \infty)$ and

$$f(t) = \int_0^t W_s ds \quad \text{where } W_s \sim BM.$$

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 might be

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

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

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