

随机过程的生成

王璐

随机向量是有限个随机变量的集合，随机过程涉及无限个随机变量的集合。比如研究一个粒子随机运动时位置随时间的变化，我们既可以研究粒子在离散时间点 $t \in \{0, 1, 2, \dots\}$ 的位置，也可以研究粒子在一段时间 $[0, T]$ 内连续变化的位置，这两种方式都涉及无限个随机变量。很多非参数 Bayesian 模型的计算都涉及对随机过程的抽样，本章我们将介绍如何对一些常见的随机过程进行抽样，比如随机游走 (random walk), Gaussian process, Poisson process 以及 Dirichlet process.

1 随机过程的一些基本概念

一个随机过程一般记为 $\{X(t) \mid t \in \mathcal{T}\}$ ，指标集 (index set) \mathcal{T} 可以是离散的集合，如 $\mathcal{T} = \{1, 2, \dots\}$ ，或者连续的集合，如 $\mathcal{T} = [0, \infty)$ 。对于离散的随机过程，有时将 $X(t)$ 简记为 X_t 。如果有两个随机过程，一般记为 $X_1(t)$ 和 $X_2(t)$ 。

有些涉及空间的随机过程，指标集 \mathcal{T} 可以是 \mathbb{R}^d 上的一个区域，比如 $X(t)$ 可能表示某个地点 t 的温度。这种定义在 \mathbb{R}^d ($d > 1$) 子集上的随机过程也被称为 **random field**。

随机过程 $\{X(t) \mid t \in \mathcal{T}\}$ 的一次实现定义了一个从 \mathcal{T} 到 \mathbb{R} 的随机函数 $f(\cdot)$ 。随机函数 $f(\cdot)$ 也被称为该过程的一条**样本路径** (sample path)。

实际应用中，虽然对随机过程的一次抽样只会产生有限个值，但会遇到与随机向量抽样不同的问题。比如，不断生成一个粒子在新时刻 t_j 的位置直到粒子离开某一特定区域。假设粒子的一条样本路径为 $(X(t_1), \dots, X(t_m))$ ，终止时刻对应的 m 可以看作一个随机整数 M 的样本，即该向量的维度是随机的。虽然 $P(M < \infty) = 1$ ，但在抽样前我们对维度 M 并没有有界的预期。在有些随机过程中，我们选择抽样的时刻 t_j 还与之前抽到的某个 $X(t_k)$ 的取值有关。因此随机过程抽样的挑战在于如何用高效的方法前后一致地产生各部分的值。

随机过程主要通过它的任意有限维分布来描述。从指标集 \mathcal{T} 选取任意有限个点 $t_1, \dots, t_m \in \mathcal{T}$ ，观察随机过程在这些点的分布，称 $(X(t_1), \dots, X(t_m))$ 的联合分布为随机过程 $X(t)$ 的一个**有限维分布**。Kolmogorov's extension theorem 告诉我们如果一组有限维分布是相容的 (compatible)，即没有矛盾 (contradictions)，则一定存在一个随机过程具有这样的有限维分布。

但是有限维分布不能唯一确定一个随机过程，即两个不同的随机过程可能有完全相同的有限维分布，比如，令 $X(t)$ 为定义在指标集 $\mathcal{T} = [0, 1]$ 上的一个随机过程，随机抽取 $s \sim U[0, 1]$ ，定义另一个随机过程 $Y(t)$ 如下：

$$Y(t) = \begin{cases} X(s) + 1, & t = s \\ X(t), & t \neq s \end{cases}$$

由于 $Y(t)$ 与 $X(t)$ 只在一个零测集上不同，因此

$$P(X(t_1) \leq x_1, \dots, X(t_m) \leq x_m) \equiv P(Y(t_1) \leq x_1, \dots, Y(t_m) \leq x_m), \quad \forall t_1, \dots, t_m \in \mathcal{T}$$

即它们的任意有限维分布都相同。

如果需要研究的性质涉及随机过程 $X(t)$ 在无限个点 t 处的值，比如计算 $\mu = E[g(X(\cdot))]$ ，可以使用 Monte Carlo 方法做近似估计。此时需要从随机过程中生成 n 条样本路径 $X_i(t_{ij}), i = 1, \dots, n$ 。假设第 i 条路径有 M_i 个点 (Monte Carlo 方法只能产生样本路径上有限个点)，则 μ 可估计为

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n g(X_i(t_{i1}), \dots, X_i(t_{iM_i})).$$

2 随机游走

随机游走一般具有以下形式

$$X_t = X_{t-1} + Z_t, \quad t = 1, 2, \dots \quad (1)$$

其中 Z_t 是 iid 的随机向量。初始点 X_0 通常取为 $\mathbf{0}$ 。如果我们知道如何对 Z_t 抽样，就很容易根据(1)生成 X_t 的路径。图1展示了离散和连续随机游走的一些样本路径。在这两个例子中， $E(Z_t) = \mathbf{0}$ 。如果 $E(Z_t) = \mu$ ，称随机游走具有 drift μ 。如果 Z_t 的协方差矩阵 Σ 有限，根据中心极限定理，

$$\frac{1}{\sqrt{t}}(X_t - t\mu) \rightarrow N(\mathbf{0}, \Sigma), \quad t \rightarrow \infty$$

我们可以对随机游走进行扩展，使 Z_t 的分布随 t 变化，比如 Pólya's urn process.

- **Pólya's urn process.** 在过程开始的时候，桶 (urn) 里有一个黑球和一个红球。在随后的每一步，我们从桶里随机取出一个球，将该球和一个与它同色的球放回桶中。令 $X_t = (R_t, B_t)$ ，其中 R_t 代表 t 时刻的红球数， B_t 代表黑球数。初始时刻 $X_0 = (1, 1)$ ，每一步的增量 Z_t 服从如下分布：

$$Z_t = \begin{cases} (1, 0), & \text{概率} = R_t / (R_t + B_t) \\ (0, 1), & \text{概率} = B_t / (R_t + B_t) \end{cases}$$

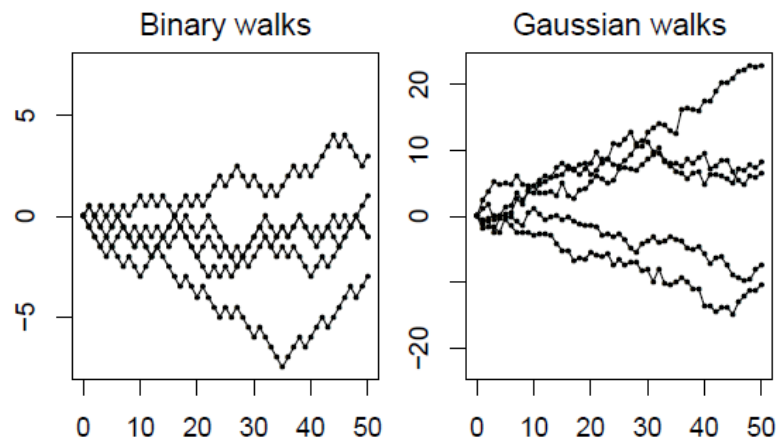


Figure 1: 离散和连续随机游走的 5 条样本路径。每条路径从 $X_0 = 0$ 开始, 持续 50 步。左图中的随机游走, 增量 Z_t 取 ± 1 的概率各为 0.5; 右图 $Z_t \sim N(0, 1)$. Picture source: Art B. Owen

在这一过程中, 我们感兴趣的变量是 $Y_t = R_t / (R_t + B_t)$, $t \rightarrow \infty$. 即足够长时间后桶中红球所占的比例。数学家 Pólya 证明上述过程 Y_t 的每条样本路径都会收敛到一个值 Y_∞ , 但 Y_∞ 本身也是随机的, 服从 $U(0, 1)$. 我们可以用 Monte carlo 方法检验该结论及 Y_t 的收敛速度。图2展示了 Y_t 的 25 条样本路径, 每条路径持续 1000 步。可以看到每条路径都收敛了, 但是收敛到不同的值。

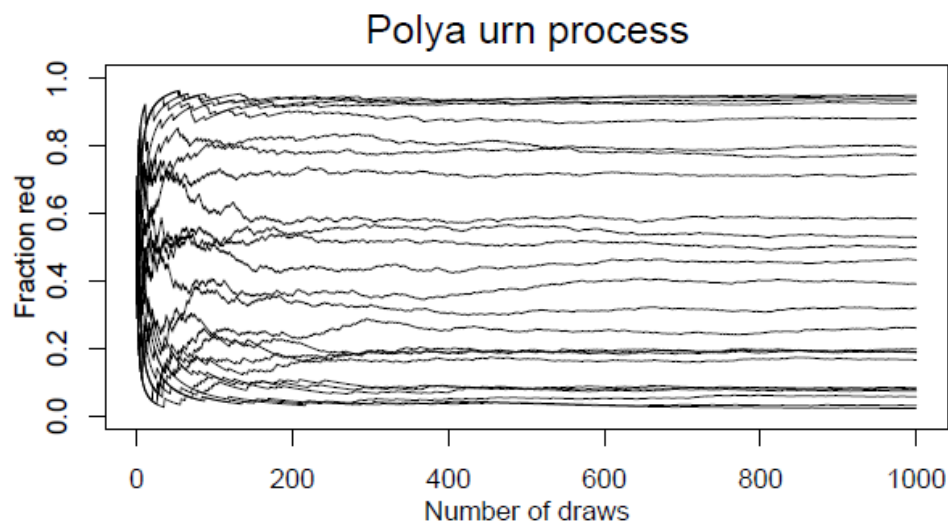


Figure 2: Pólya urn process 的 25 条样本路径。Picture source: Art B. Owen

- **作业.** 对 Pólya's urn process 稍做修改可以用来描述市场竞争中赢家通吃 (winner-take-all) 的现象。比如用 (R_t, B_t) 代表两家公司的用户数, 即使它们提供的产品完全相同, 如果新用户

倾向于购买他们朋友购买的产品，公司的用户增量 Z_t 可能服从如下分布：

$$Z_t = \begin{cases} (1, 0), & \text{概率} = R_t^\alpha / (R_t^\alpha + B_t^\alpha) \\ (0, 1), & \text{概率} = B_t^\alpha / (R_t^\alpha + B_t^\alpha) \end{cases} \quad (2)$$

其中 $\alpha > 1$. 这种情况下两家公司最终不会平分市场份额，而是由一家公司占领全部市场。最终的结果与早期的一些优势或运气有很大关系。选择不同的 $\alpha > 1$ 的值，基于(2)生成若干条 $Y_t = R_t / (R_t + B_t)$ 的样本路径（初始时刻 $R_0 = 1, B_0 = 1$ ），观察是否出现赢家通吃的现象以及 α 的取值对路径收敛速度的影响。

3 Gaussian processes

Gaussian process 的任意有限维分布都是一个多元正态分布。由于多元正态分布只取决于期望和协方差矩阵，因此定义一个 Gaussian process $\{X(t) \mid t \in \mathcal{T}\}$ 只需要确定一个**期望函数**

$$\mu(t) = E[X(t)], \quad t \in \mathcal{T}$$

和一个**协方差函数**

$$\Sigma(t, s) = \text{Cov}(X(t), X(s)), \quad \forall t, s \in \mathcal{T}.$$

显然协方差函数 $\Sigma(\cdot, \cdot)$ 需满足对称性 $\Sigma(t, s) = \Sigma(s, t)$. 此时 Gaussian process 的任意有限维分布可写为

$$\begin{pmatrix} X(t_1) \\ \vdots \\ X(t_m) \end{pmatrix} \sim N_m \left(\begin{pmatrix} \mu(t_1) \\ \vdots \\ \mu(t_m) \end{pmatrix}, \begin{pmatrix} \Sigma(t_1, t_1) & \cdots & \Sigma(t_1, t_m) \\ \vdots & \ddots & \vdots \\ \Sigma(t_m, t_1) & \cdots & \Sigma(t_m, t_m) \end{pmatrix} \right)$$

Gaussian process 的期望函数可以是任意函数 $\mu: \mathcal{T} \rightarrow \mathbb{R}$ ，而协方差函数还需要再满足一个限制条件：由于多元正态分布的协方差矩阵是（半）正定的，一个有效的协方差函数 $\Sigma(\cdot, \cdot)$ 需满足

$$\sum_{i=1}^m \sum_{j=1}^m x_i x_j \Sigma(t_i, t_j) \geq 0, \quad \forall m \geq 1, t_i \in \mathcal{T}, x_i \in \mathbb{R}.$$

Gaussian process 的一个重要应用是为函数插值提供不确定性估计 (uncertainty quantification)，它也是 nonparametric Bayesian model 常用的 prior. 假设 $f(\cdot)$ 是 Gaussian process 的一条样本路径，且已知该 Gaussian process 的期望函数 $\mu(\cdot)$ 和协方差函数 $\Sigma(\cdot, \cdot)$ ，我们对 $f(\cdot)$ 的任意有限维分布就有了先验信息 (prior information). 当观察到该样本路径上 k 个点的值 $f(t_1), \dots, f(t_k)$ 后，利用 $(k+1)$ 维正态分布的条件分布公式，我们可以计算样本路径上任一点 $f(t)$ 的条件期望和方

差: $E(f(t) | f(t_1), \dots, f(t_k))$, $Var(f(t) | f(t_1), \dots, f(t_k))$, 即 $f(t)$ 的 posterior mean 和 posterior variance.

用每一点的条件期望 (posterior mean) 定义一个预测函数:

$$\hat{f}(t) = E(f(t) | f(t_1), \dots, f(t_k)), \quad t \in \mathcal{T}.$$

显然 $\hat{f}(t_j) = f(t_j)$, $j = 1, \dots, k$. 即 \hat{f} 是对观察值的一个插值函数。同时我们知道 $f(t)$ 在每一点的条件方差, 因此可以给出 f 的置信区间 (由每一点的置信区间组成)。由于给定 $f(t_1), \dots, f(t_k)$ 后, $f(t)$ 的条件分布 (posterior distribution) 也是一个正态分布, 因此可以在每一点对 $f(t)$ 抽样, 生成一条通过已知点的样本路径。

如果对任意间隔 Δ , $\forall t \in \mathcal{T}$, $X(t)$ 和 $X(t + \Delta)$ 都是同分布, 称随机过程 $X(t)$ 是平稳的 (stationary). 对于 Gaussian process, stationarity 等价于

$$\mu(t + \Delta) = \mu(t), \quad \Sigma(t + \Delta, s + \Delta) = \Sigma(t, s), \quad \forall \Delta, \forall t, s \in \mathcal{T}.$$

通常 \mathcal{T} 包含 0, 因此对于 Gaussian process, stationarity 意味着

$$\mu(t) \equiv \mu(0), \quad \Sigma(t, s) = \Sigma(t - s, 0), \quad \forall t, s \in \mathcal{T}.$$

下面列举一些常见的 stationary Gaussian processes.

- **Exponential covariance.** Gaussian process with exponential covariance 的期望函数 $\mu(t) \equiv 0$, 协方差函数为

$$\Sigma(t, s) = \sigma^2 \exp(-\theta|t - s|), \quad \theta > 0.$$

该过程的样本路径是连续的但不可导。

- **Gaussian covariance.** Gaussian process with Gaussian covariance 的期望函数 $\mu(t) \equiv 0$, 协方差函数为

$$\Sigma(t, s) = \sigma^2 \exp(-\theta(t - s)^2), \quad \theta > 0.$$

Gaussian covariance 也被称为 squared exponential covariance. 它的样本路径是任意阶可导的。

图3展示了分别使用 Gaussian process with exponential covariance 和 Gaussian covariance 对三个已知点 $f(0) = 1$, $f(0.4) = 3$ 及 $f(1) = 2$ 插值的结果。在画预测函数 \hat{f} 时, 从指标集 \mathcal{T}

选取的节点为 -0.25 到 1.25 之间间隔为 0.01 的一系列点。上述插值得到的预测函数 \hat{f} 与 σ 的取值无关，因为 σ 在计算条件期望时被消掉了。但是 \hat{f} 与 θ 有关：当 θ 很大时，节点之间的相关性随着距离 $|t - s|$ 增加迅速下降，观察点附近节点的预测值与观察值的相关性变得很小，节点的预测值被迅速拉向整体的期望函数 $\mu(t) \equiv 0$ ；当 θ 较小时，节点之间的相关性随距离 $|t - s|$ 增加下降地较慢，观察点附近节点的预测值与观察值的相关性很高。可以看到 exponential covariance 在 θ 较小时的插值预测函数似乎是分段线性的 (piecewise linear)，而 Gaussian covariance 在 θ 较小时的预测函数非常平滑。

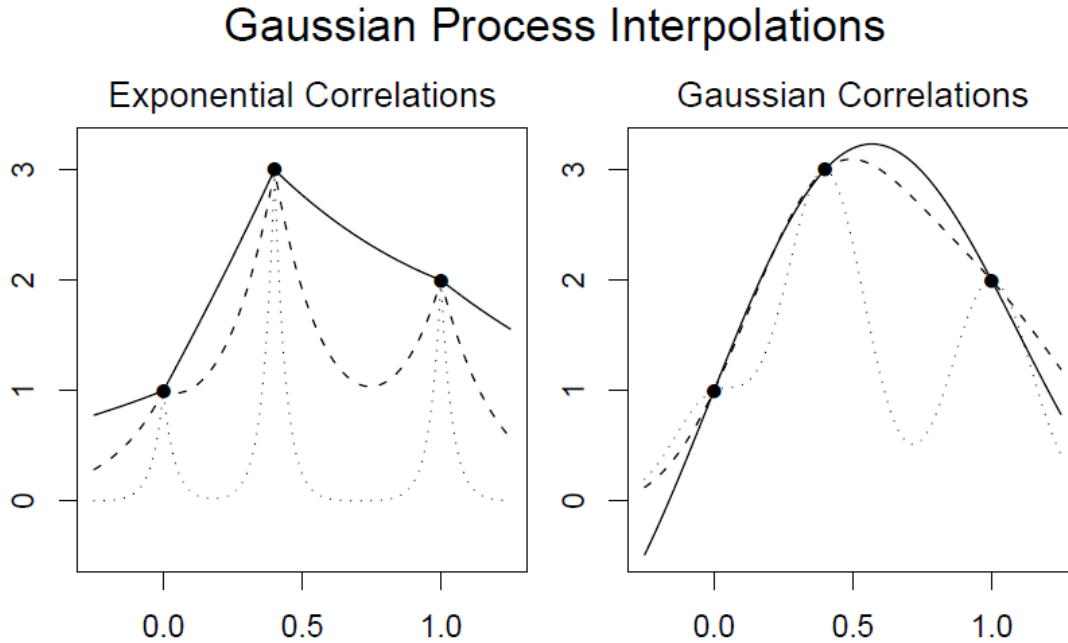


Figure 3: 使用 Gaussian process 对三个观察点插值。左图使用 exponential covariance, 右图使用 Gaussian covariance; 图中的实线, 虚线, 点线分别对应 $\theta = 1, 5, 25$. Picture source: Art B. Owen

图4展示了 Gaussian process with $\mu(t) = 0$, $\Sigma(t, s) = \exp(-(t - s)^2)$ 生成的若干条通过已知点 $f(0) = 1$, $f(0.4) = 3$ 和 $f(1) = 2$ 的样本路径。从这些模拟中，我们可以近似得到 $f(\cdot)$ 的最大值点 t^* 的 (posterior) 分布。

Gaussian covariance 产生的样本路径有时过于平滑，Matérn covariances 可以提供介于 exponential 和 Gaussian covariance 之间的平滑度。

- **Matérn covariances.** Matérn class of covariances 由一个平滑度 (smoothness) 系数 ν 控制。对一般的 $\nu > 0$, 协方差函数 $\Sigma(t, s; \nu)$ 通过 Bessel function 定义。当 $\nu = m + 1/2$ 且 m 是非

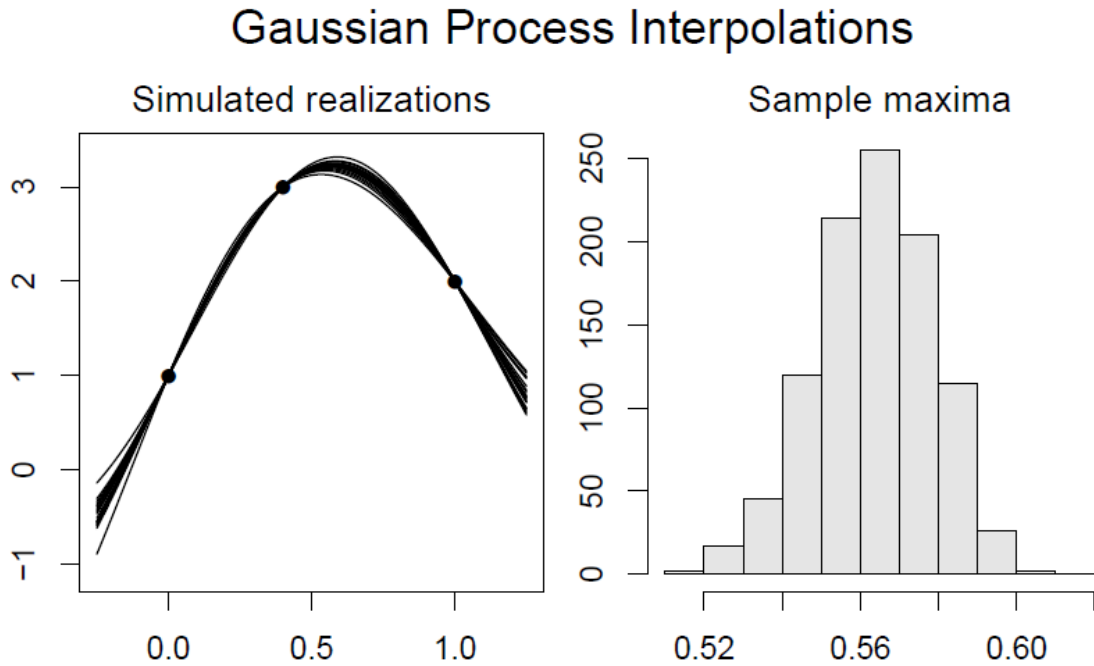


Figure 4: Gaussian process with Gaussian covariance 生成的 20 条通过 3 个已知点的样本路径, $\theta = 1$, $\sigma^2 = 1$ (左图). 1000 条通过左图方式生成的样本路径上最大值点的分布 (右图)。Picture source: Art B. Owen

负整数时, 协方差函数 $\Sigma(t, s; \nu)$ 可以极大简化, 比如前 4 个特例为

$$\begin{aligned}\Sigma\left(t, s; \frac{1}{2}\right) &= \sigma^2 \exp(-\theta|t-s|) \\ \Sigma\left(t, s; \frac{3}{2}\right) &= \sigma^2 (1 + \theta|t-s|) \exp(-\theta|t-s|) \\ \Sigma\left(t, s; \frac{5}{2}\right) &= \sigma^2 \left(1 + \theta|t-s| + \frac{1}{3}\theta^2|t-s|^2\right) \exp(-\theta|t-s|) \\ \Sigma\left(t, s; \frac{7}{2}\right) &= \sigma^2 \left(1 + \theta|t-s| + \frac{2}{5}\theta^2|t-s|^2 + \frac{1}{15}\theta^3|t-s|^3\right) \exp(-\theta|t-s|)\end{aligned}$$

其中 $\theta > 0$.

显然 exponential covariance 是 Matérn class 的一个特例 ($\nu = 1/2$). Matérn covariances 在 $\nu \rightarrow \infty$ 时收敛到 Gaussian covariance. Matérn covariance with $\nu = m + 1/2$ 生成的样本路径有 m 阶导数。图5展示了 Matérn process 生成的一些样本路径, 可以看到 ν 越大, 对应的样本路径越光滑; θ 越大, 样本路径的振荡越多。

从 Gaussian process 生成一条有 m 个点的样本路径 f , 等价于对一个 m 维的多元正态分

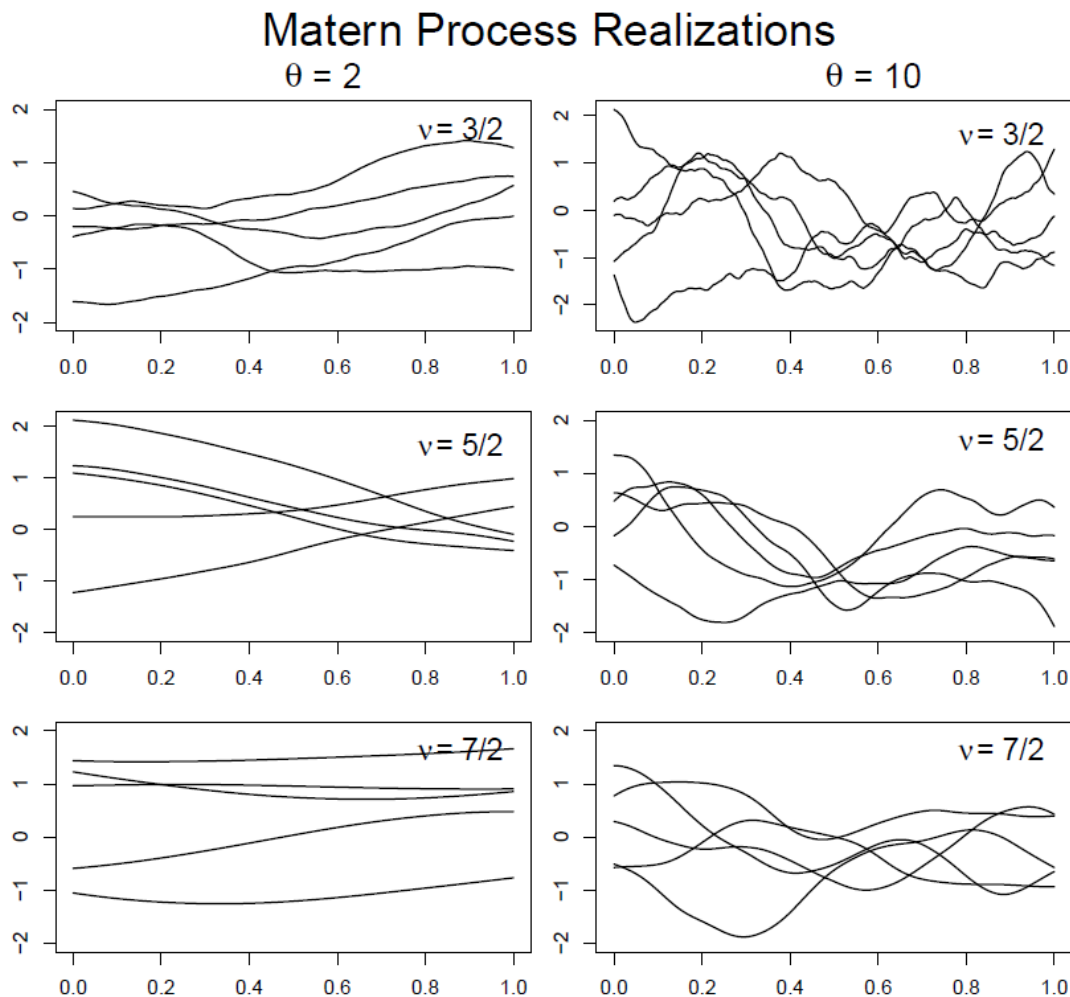


Figure 5: Matérn process 在 $\nu = 3/2, 5/2, 7/2$, $\theta = 2, 10$ 下分别产生的 5 条样本路径 ($\sigma^2 = 1$).
Picture source: Art B. Owen

布进行一次抽样。取定 t_1, \dots, t_m , 通过期望函数可以得到随机向量 $(f(t_1), \dots, f(t_m))$ 的期望 $\boldsymbol{\mu}$, 通过协方差函数可以计算出它的协方差矩阵 Σ . 上一章我们介绍过从 $N_m(\boldsymbol{\mu}, \Sigma)$ 抽样需要计算一个矩阵 C 使得 $\Sigma = CC^\top$, 这一过程的计算量一般为 $O(m^3)$. 抽样时先抽 $\mathbf{Z} \sim N_m(\mathbf{0}, I_m)$, 则 $\boldsymbol{\mu} + C\mathbf{Z} \sim N_m(\boldsymbol{\mu}, \Sigma)$. 如果想得到非常平滑的样本路径, 比如使用 Gaussian covariance (θ 很小), 对应的协方差矩阵 Σ 可能非常接近 singular (节点间相关系数几乎为 1), 此时推荐使用特征值分解计算 C , 即做分解 $\Sigma = P\Lambda P^\top$, 然后令 $C = P\Lambda^{1/2}$. 另一个办法是给 Σ 加一个 nugget effect, 即用 $\Sigma_\epsilon = \Sigma + \epsilon I_m$ 替代 Σ , 其中 ϵ 是很小的正数. 如果 Σ 是一个有效的协方差矩阵 (半正定), Σ_ϵ 也是有效的协方差矩阵且是可逆的 (正定). 此时相当于将模型修改为 $\Sigma_\epsilon(t, s) = \text{Cov}(X(t) + \varepsilon_t, X(s) + \varepsilon_s)$, 其中 ε_t 's iid 服从 $N(0, \epsilon)$, 它们可以看作加在原过程 $X(t)$ 上的一些“扰动”(jitter) 或测量误差。

3.1 Brownian motion

Brownian motion 可能是最重要的一个 Gaussian process, 本节我们讨论如何对 Brownian motion 抽样。Standard Brownian motion 是定义在 $\mathcal{T} = [0, \infty)$ 上的 Gaussian process, 记为 $B(t)$, 它有三条性质:

1. $B(0) = 0$.
2. 对于任意的 $0 = t_0 < t_1 < \cdots < t_m$, $B(t_i) - B(t_{i-1}) \stackrel{ind}{\sim} N(0, t_i - t_{i-1})$, $i = 1, \dots, m$.
3. $B(t)$ 的样本路径在 $[0, \infty)$ 上以概率 1 连续。

Standard Brownian motion 也被称为 Wiener process, 以纪念数学家 Norbert Wiener, 他在 1923 年证明了满足上述 3 条性质的随机过程是存在的。尽管 $B(t)$ 的样本路径是连续的, 但它也以概率 1 处处不可导。易证 $B(t)$ 的期望函数 $\mu(t) = 0$, 协方差函数 $\Sigma(t, s) = \min(t, s)$, 因此 Brownian motion 不是平稳的。

在 standard Brownian motion 的基础上, 可以生成更复杂的 Brownian motions. 将 standard Brownian motion $B(t)$ 记为 $B(\cdot) \sim \text{BM}(0, 1)$. 定义一个新的随机过程

$$X(t) = \delta t + \sigma B(t)$$

容易证明 $X(t)$ 的期望函数 $\mu(t) = \delta t$, 协方差函数 $\Sigma(t, s) = \sigma^2 \min(t, s)$. 我们称 $X(t)$ 是 drift δ , 方差 σ^2 的 Brownian motion, 记为 $X(\cdot) \sim \text{BM}(\delta, \sigma^2)$.

如果想得到 $X(\cdot) \sim \text{BM}(\delta, \sigma^2)$ 在 $[0, T]$ 上的样本路径, 只需先抽取 $B(\cdot) \sim \text{BM}(0, 1)$ 在 $[0, 1]$ 上的样本路径, 然后令 $X(t) = \delta t + \sigma \sqrt{T} B(t/T)$ 即可. 因此我们只需关注如何对 standard Brownian motion 在 $[0, 1]$ 上抽样。对于 $[0, 1]$ 上的任意一系列点, $0 < t_1 < t_2 < \cdots < t_m \leq 1$, 根据定义可以如下得到 $B(\cdot)$ 在这些点的样本:

$$\begin{aligned} B(t_1) &= \sqrt{t_1} Z_1, \\ B(t_j) &= B(t_{j-1}) + \sqrt{t_j - t_{j-1}} Z_j, \quad j = 2, \dots, m \end{aligned} \tag{3}$$

其中 $Z_j \stackrel{ind}{\sim} N(0, 1)$, $j = 1, \dots, m$. 也可以将上述过程写为矩阵形式

$$\begin{pmatrix} B(t_1) \\ B(t_2) \\ \vdots \\ B(t_m) \end{pmatrix} = \begin{pmatrix} \sqrt{t_1} & 0 & \cdots & 0 \\ \sqrt{t_1} & \sqrt{t_2 - t_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{t_1} & \sqrt{t_2 - t_1} & \cdots & \sqrt{t_m - t_{m-1}} \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_m \end{pmatrix}. \tag{4}$$

等式(4)中的系数矩阵恰好是该随机向量协方差矩阵的 Cholesky 分解矩阵:

$$\text{Var} \left(\begin{pmatrix} B(t_1) \\ B(t_2) \\ \vdots \\ B(t_m) \end{pmatrix} \right) = (\min(t_j, t_k))_{1 \leq j, k \leq m} = \begin{pmatrix} t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & \cdots & t_2 \\ \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & \cdots & t_m \end{pmatrix}.$$

References