

觅食算法

李昱辰 张恒 时侠圣 李崇慧 刘子旋 潘镇涛 朱春林 谷超明

2018 年 4 月 19 日

目录

觅食算法之布朗运动

随机游走

Heavy-tailed Distribution And Light-tailed Distribution

莱维过程

分解定理

随机微分方程

觅食算法实例

觅食算法之布朗运动

随机过程与布朗运动

我们周边充满了各种各样的数据，所谓大数据时代，这些数据最基本的特点就是含有巨量的噪音，而随机过程就是尝试从这些噪音里提取出信息。

随机过程描述的是一个量随时间可能的变化，在这个过程中，每一个时刻变化的方向都是不确定的，或者说随机过程就是由一系列随机变量组成，每一个时刻系统的状态都由一个随机变量表述，而整个过程则构成态空间的一个轨迹（随机过程的实现）。

布朗运动的简史

最初由英国生物学家布朗 (Brown) 于 1827 年提出这种物理现象。1905 年爱因斯坦首次对这一现象的物理规律给出数学描述。1918 年维纳 (Wiener) 运用数学理论严格描述这种无规则运动。并用随机过程理论和概率理论建立了数学模型。因此又称为维纳过程。是具有连续时间参数和连续状态空间的一类随机过程。

自 1860 年以来,许多科学家都在研究此种现象,后来发现布朗运动有下列的主要特性:

- ▶ 粒子的运动由平移及转移所构成, 显得非常没规则而且其轨迹几乎是处处没有切线。
- ▶ 粒子之移动显然互不相关, 甚至于当粒子互相接近至比其直径小的距离时也是如此。
- ▶ 粒子的成分及密度对其运动没有影响。
- ▶ 粒子越小或液体粘性越低或温度越高时, 粒子的运动越活泼。
- ▶ 粒子的运动永不停止。

随机过程与布朗运动

布朗运动是一种正态分布的独立增量连续随机过程。其基本性质为：布朗运动 $W(t)$ 是期望为 0，方差为 t （时间）的正态随机变量。对于任意的 $r \leq s$, $w(t) - W(s)$ 独立于 $W(r)$, 且是期望为 0 方差为 $t - s$ 的正态随机变量。

定义：若随机过程 $\{X(t), t \geq 0\}$ 满足：

(1) $X(t)$ 关于 t 是连续函数；且几乎处处不可微。

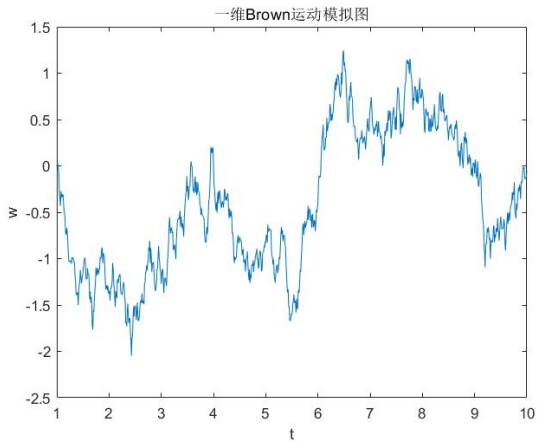
(2) $\{X(t), t \geq 0\}$ 具有平稳独立增量：对任意的有限正数 $0 = t_0 < t_1 < \dots < t_n$, 随机变量 $X(t_0), X(t_1) - X(t_0), \dots, X(t_n) - X(t_{n-1})$ 相互独立。

(3) $\forall s, t > 0$, s.t. $X(s+t) - X(s) \sim N(0, \sigma^2 t)$

则称随机过程 $\{X(t), t \geq 0\}$ 为布朗运动（或维纳过程）。当 $\sigma = 1$ 时，称随机过程 $\{X(t), t \geq 0\}$ 为标准布朗运动。记为 $\{B(t), t \geq 0\}$ 。几何布朗运动是连续时间情况下的随机过程，其中随机变量的对数遵循布朗运动。

布朗运动

布朗运动连续性如下仿真效果：



布朗运动

定义：概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 上的随机过程 $\{X(t), t \geq 0\}$ 称为高斯过程，若对任意的 $n \geq 1, t_1, t_2, \dots, t_n, (X(t_1), \dots, X(t_n))$ 服从高斯分布。

定理：设 $B = \{B(t), t \geq 0\}$ 是零初值的实值随机过程。则它是布朗运动的充要条件是它是一个高斯过程，并且 $\mathbb{E}(B(t)) = 0, \mathbb{E}(B(t)B(s)) = t \wedge s = \min(s, t)$ 。

证明：必要性。由独立增量性 $n \geq 1, t_1, t_2, \dots, t_n, (X(t_1), \dots, X(t_n))$ 服从高斯分布。又因为

$$\begin{pmatrix} B(t_1) \\ B(t_2) \\ \vdots \\ B(t_n) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} B(t_1) \\ B(t_2) - B(t_1) \\ \vdots \\ B(t_n) - B(t_{n-1}) \end{pmatrix} \quad (1)$$

布朗运动

所以有 $(B(t_1), \dots, B(t_n))$ 服从高斯分布。显然有 $\mathbb{E}(B(t)) = 0$, 对任意 $t \geq s \geq 0$, 有

$$\mathbb{E}(B(t)B(s)) = \mathbb{E}(B(t) - B(s))B(s) + \mathbb{E}B^2(s) = s. \quad (2)$$

充分性。先证独立增量性。由于它们服从正态分布，所以只需要证明不相关性即可。事实上，对任意的 $i < j$, 有

$$\begin{aligned} \mathbb{E}(B(t_i) - B(t_{i-1}))(B(t_j) - B(t_{j-1})) &= \mathbb{E}B(t_i)B(t_j) - \mathbb{E}B(t_{i-1})B(t_j) \\ &\quad + \mathbb{E}B(t_{i-1})B(t_{j-1}) \\ &= t_i - t_i - t_{i-1} + t_{i-1} = 0. \end{aligned} \quad (3)$$

显然，对任意的 $t > s$, $B(t) - B(s)$ 服从正态分布，均值为 0，方差为

$$\mathbb{E}(B(t) - B(s))^2 = \mathbb{E}B^2(t) - 2\mathbb{E}B(t)B(s) + \mathbb{E}B^2(s) = t - s. \quad (4)$$

布朗运动的构造

下面利用 Fourier 级数直接构造布朗运动。假设 B 是一个标准布朗运动。定义过程 $W(t) = B(t) - tB(1)$, $t \in [0, 1]$. 称为 $[0, 1]$ 上的布朗桥。注意到 $W(0) = W(1) = 0$. 我们将 W 在 $[0, 1]$ 上的 Fourier 展开 (L^2 意义下)

$$W(t) = \sum_{n=1}^{\infty} X_n \sin(n\pi t) \quad (5)$$

其中系数 X_n 是随机变量，表达式为

$$X_n = 2 \int_0^1 W(t) \sin(n\pi t) dt. \quad (6)$$

则 X_n 服从正态分布，可以计算

$$\mathbb{E}X_n = 0, \mathbb{E}[X_n X_m] = \frac{2}{\pi^2 n^2} \delta_{mn}. \quad (7)$$

令 $Z_0 = B_1$, $Z_n = n\pi X_n / \sqrt{2}$, $n \geq 1$. 则 $\{Z_n\}$ i.i.d $N(0, 1)$ 。我们可以将布朗运动写成

$$B(t) = tZ_0 + \frac{\sqrt{2}}{\pi} \sum_{n=1}^{\infty} \frac{Z_n}{n} \sin(n\pi t), \quad t \in [0, 1] \quad (8)$$

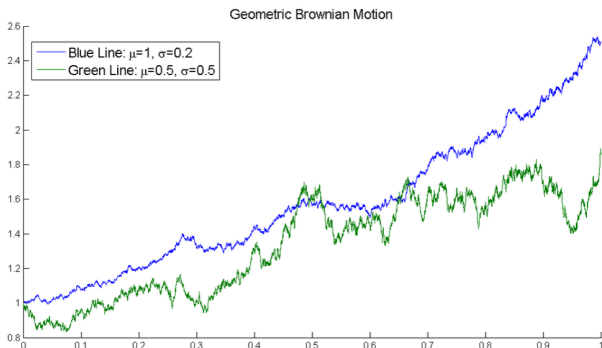
布朗运动的逼近

三种逼近方式：

- (1) 基于随机游动的逼近；
- (2) 布朗运动的帕里-维纳表示；
- (3) 基于小波函数的逼近方法。

布朗运动的应用

金融数学：几何布朗运动是连续时间情况下的随机过程，其中随机变量的对数遵循布朗运动。用来在布莱克-舒尔斯定价模型中模仿股票价格。



图：

布朗运动的应用

使用几何布朗运动来描述股票价格的理由：

- ▶ 几何布朗运动的期望与随机过程的价格（股票价格）是独立的，这与我们对现实市场的期望是相符的。
- ▶ 几何布朗运动过程与我们在股票市场观察到的价格轨迹呈现了同样的“roughness”。
- ▶ 几何布朗运动过程计算相对简单。

使用几何布朗运动来描述股票价格的缺陷：

- ▶ 在真实股票价格中波动随时间变化，但是在几何布朗运动中，波动是不随时间变化的。
- ▶ 在真实股票价格中，收益通常不服从正态分布。

布朗运动的其他应用

- ▶ 研究仪器的灵敏度：在近代无线电技术 (如卫星通讯) 中, 由于放大倍数很高, 电涨落现象表现得特别显著, 引起热噪声, 这个问题也需要用布朗运动理论来研究。
- ▶ 研究各类扩散现象：扩散现象的本质是布朗运动产生的位移, 因此布朗运动理论可用于各类扩散现象。例如半导体中载流子 (电子或空穴) 的扩散, 原子核反应堆中中子的扩散等, 均可用布朗运动理论来研究。
- ▶ 分形布朗运动模型在地形分析、模式识别、数字图像处理、信息讯号的处理中的应用。
- ▶ 1990 年以来, 量子布朗运动理论进一步发展, 转向对纳米结构中粒子运动和生物细胞中分子机器的研究。

随机游走

定义

随机游走 (random walk) 也称随机漫步, 随机行走等是指基于过去的表现, 无法预测将来的发展步骤和方向。核心概念是指任何无规则行走者所带的守恒量都各自对应着一个扩散运输定律, 接近于布朗运动, 是布朗运动理想的数学状态, 现阶段主要应用于互联网链接分析及金融股票市场中。

全局最优化，操作简单，不易陷入局部极小值

- ▶ 全局最优化：目前还没有一个通用的办法可以对任意复杂函数求解全局最优值。

- ▶ 全局最优化：目前还没有一个通用的办法可以对任意复杂函数求解全局最优值。
- ▶ 操作简单：易于依托代码实现该算法。

- ▶ 全局最优化：目前还没有一个通用的办法可以对任意复杂函数求解全局最优值。
- ▶ 操作简单：易于依托代码实现该算法。
- ▶ 不易陷入局部极小值：源于算法中的不断迭代过程。

算法流程

- ▶ 设 $f(x)$ 是一个含有 n 个变量的多元函数, $x=(x_1, x_2, \dots, x_n)$ 为 n 维向量。

算法流程

- ▶ 设 $f(x)$ 是一个含有 n 个变量的多元函数, $x=(x_1, x_2, \dots, x_n)$ 为 n 维向量。
- ▶ 给定初始迭代点 x , 初次行走步长 λ , 控制精度 ϵ 。

算法流程

- ▶ 设 $f(x)$ 是一个含有 n 个变量的多元函数, $x=(x_1, x_2, \dots, x_n)$ 为 n 维向量。
- ▶ 给定初始迭代点 x , 初次行走步长 λ , 控制精度 ϵ 。
- ▶ 给定迭代控制次数 N , k 为当前迭代次数, 置 $k=1$ 。

算法流程

- ▶ 设 $f(x)$ 是一个含有 n 个变量的多元函数, $x=(x_1, x_2, \dots, x_n)$ 为 n 维向量。
- ▶ 给定初始迭代点 x , 初次行走步长 λ , 控制精度 ϵ 。
- ▶ 给定迭代控制次数 N , k 为当前迭代次数, 置 $k=1$ 。
- ▶ 当 $k < N$ 时, 随机生成一个 $(-1, 1)$ 之间的 n 维向量 $u=(u_1, u_2, \dots, u_n), (-1 < u_i < 1, i=1, 2, \dots, n)$ 。

算法流程

- ▶ 设 $f(x)$ 是一个含有 n 个变量的多元函数, $x=(x_1, x_2, \dots, x_n)$ 为 n 维向量。
- ▶ 给定初始迭代点 x , 初次行走步长 λ , 控制精度 ϵ 。
- ▶ 给定迭代控制次数 N , k 为当前迭代次数, 置 $k=1$ 。
- ▶ 当 $k < N$ 时, 随机生成一个 $(-1, 1)$ 之间的 n 维向量 $u=(u_1, u_2, \dots, u_n), (-1 < u_i < 1, i=1, 2, \dots, n)$ 。
- ▶ 将 n 维向量标准化得到

$$u' = \frac{u}{\sqrt{\sum_{i=1}^n u_i^2}} \quad (9)$$

算法流程

- ▶ 设 $f(x)$ 是一个含有 n 个变量的多元函数, $x=(x_1, x_2, \dots, x_n)$ 为 n 维向量。
- ▶ 给定初始迭代点 x , 初次行走步长 λ , 控制精度 ϵ 。
- ▶ 给定迭代控制次数 N , k 为当前迭代次数, 置 $k=1$ 。
- ▶ 当 $k < N$ 时, 随机生成一个 $(-1, 1)$ 之间的 n 维向量 $u=(u_1, u_2, \dots, u_n), (-1 < u_i < 1, i=1, 2, \dots, n)$ 。
- ▶ 将 n 维向量标准化得到

$$u' = \frac{u}{\sqrt{\sum_{i=1}^n u_i^2}} \quad (9)$$

- ▶ 令

$$x_1 = x + \lambda u' \quad (10)$$

从而完成第一步游走。

算法流程

- ▶ 计算函数值，如果 $f(x_1) < f(x)$ ，即找到了一个比初始值好的点，那么 k 重新置为 1，将 x_1 变为 x ，回到第 2 步；否则 $k=k+1$ ，回到第 3 步。

算法流程

- ▶ 计算函数值，如果 $f(x_1) < f(x)$ ，即找到了一个比初始值好的点，那么 k 重新置为 1，将 x_1 变为 x ，回到第 2 步；否则 $k=k+1$ ，回到第 3 步。
- ▶ 如果连续 N 次都找不到更优的值，则认为，最优解就在以当前最优解为中心，当前步长为半径的 N 维球内。

算法流程

- ▶ 计算函数值，如果 $f(x_1) < f(x)$ ，即找到了一个比初始值好的点，那么 k 重新置为 1，将 x_1 变为 x ，回到第 2 步；否则 $k=k+1$ ，回到第 3 步。
- ▶ 如果连续 N 次都找不到更优的值，则认为，最优解就在以当前最优解为中心，当前步长为半径的 N 维球内。
- ▶ 此时，如果 $\lambda < \epsilon$ ，则结束算法；否则，令 λ 减半，回到第 1 步，开始新一轮游走。

Heavy-tailed Distribution And Light-tailed Distribution

承前启后

如果捕食者像无头苍蝇一样漫无目的的做布朗运动，真的能很有效的地寻找到猎物吗？在数学上可以证明，布朗运动和分子自由扩散一样，单位速度的分子在时间 t 内平均只有 \sqrt{t} 的位移量。若捕食者采用这种策略，需要很久才能够成功。

那么有没有比布朗运动更高效的搜索方法呢？一组巴西物理学家于 1999 年提出了一个设想，认为“莱维飞行”比布朗运动有更高的搜索效率，因此自然会偏向与采用“莱维飞行”捕食的生物。为了理解什么是“莱维飞行”，我们首先需要对概率分布的轻尾性和重尾性有一个认识。

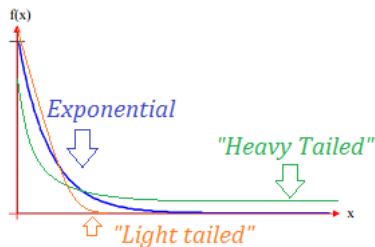
指数分布

指数分布是描述泊松过程中的事件之间的时间的概率分布，即事件以恒定平均速率连续且独立地发生地过程。它是几何分布的连续模拟，它具有无记忆的关键性质。例如，如果 T 是某一元件的寿命，已知元件使用了 t 小时，它总共使用至少 $s + t$ 小时的条件概率与一开始使用时算起它使用至少 s 小时的概率相等。其概率密度函数：

$$F(x; \lambda) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

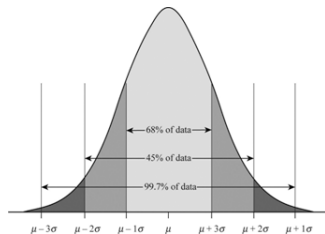
轻尾分布

比指数分布尾部更薄的概率分布是轻尾分布。它们下降到 0 的速度比指数分布的速度快得多，因此在尾部质量更小。Gumbel 分布是轻尾分布的一个例子。轻尾分布并不能很好地反映“真实世界”的数据。



概率分布的轻尾性

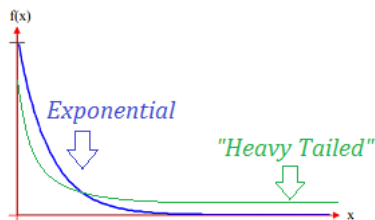
前面说过，布朗运动的步长服从正态分布。正态分布是一种典型的**轻尾分布**，也就是说，不太可能取得极端值的分布。



表达式 $y = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ 。其中 μ 是均值， σ 是标准差。当 x 增大时，函数下滑趋势非常快，因此数学家又将正态分布函数纳入速降函数空间范畴。速降函数是专门为傅里叶分析而量身定做的。

概率分布的重尾性

而在现实生活中，大多数的事情不能用轻尾分布来描述，例如保险领域的保险金-事故可以看作是稀有事件，但一旦发生并通过审核，保险公司就必须大量金额，这样一来保险金的分布就很可能取到很大的值。为了描述这样容易取得极值的随机变量，我们需要引入重尾分布。



典型特点：大头短 + 小尾长。“莱维飞行”的平均位移是
 $t^\gamma, \gamma > \frac{1}{2}$.

重尾分布

当 $x \rightarrow \infty$ 时, 重尾分布下降到 0 的速度慢于指数分布。因而尾部质量较大。其数学定义为:

$$\lim_{x \rightarrow \infty} e^{\lambda x} F(x) = \infty$$

其中, $\lambda > 0$, $F(x)$ 是尾分布函数。

重尾分布更适用于对那些离峰值较远的稀有事件也会有相当的概率发生的情况。重尾分布作为一个大的类别, 还包含三个重要的子类别, 分别是肥尾分布 (Fat-tailed distribution), 长尾分布 (Long-tailed distribution) 和次指数分布 (Subexponential distribution)

概率分布的重尾性

那么，如何验证“莱维飞行”的有效性呢？一个方案就是前面提到的大肠杆菌的**趋化行为**。尽管最初的“莱维飞行”假说是针对动物而言的，但从大肠杆菌的游击战策略可以看出，该假说对微生物同样适用！这样一来，对微生物的研究也能反过来帮助人们更深入地理解动物行为，这便是微观的细胞生物学在宏观的生态学中的应用。

莱维过程

莱维过程 $\{X(t), t \geq 0\}$ 是一种随机过程，它满足的条件比布朗运动宽松：

- ▶ X_0 几乎处处为 0;
- ▶ 独立增量性;
- ▶ 稳定增量性;
- ▶ 样本轨道右连续.

连续的布朗运动和离散的泊松过程都是莱维过程的特例。因此可以大胆猜测，莱维过程就是带“跳跃”的布朗运动。正是这些不连续性的“跳跃”给予莱维过程“重尾”的特性。

莱维过程

定义

一个随机过程 $X = \{X_t : t \geq 0\}$ 如果符合以下条件:

- (1) $X_0 = 0$.
 - (2) 独立增量: 对于任何 $0 \leq t_1 < t_2 < \cdots < t_n < \infty$, $X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, \dots, X_{t_n} - X_{t_{n-1}}$ 相互独立.
 - (3) 稳定增量: 对任何 $s < t$, $X_t - X_s$ 与 X_{t-s} 有相同分布.
 - (4) X 几乎确定是右连左极.
- 我们把 X 称作 levy 过程.

无限可分分布 (infinitely divisible distribution)

- ▶ 定义：对于一个随机变量 U , 如果存在一系列的独立同分布的随机变量 $U_{1,n}, \dots, U_{n,n}$, 使得 $U \stackrel{d}{=} U_{1,n} + \dots + U_{n,n}$, 那么随机变量 U 拥有无限可分分布, $\stackrel{d}{=}$ 意味着同分布.
- ▶ 如果 U 是一个随机变量, 那么它的特征函数 $h: \mathbb{R} \rightarrow \mathbb{C}$ 为:
 $h(\theta) = E[e^{i\theta U}], \theta \in \mathbb{R}$.
- ▶ U 是一个无限可分随机变量, h 是特征函数, 那么:
 - (1) h 是连续并且非 0 的, $h(0)=1$.
 - (2) 存在一个独特的连续函数, 对于所有 $\theta \in \mathbb{R}$ 满足 $e^{f(\theta)} = h(\theta)$, f 是 U 的特征指数

定理：如果 X 是一个 levy 过程，对于任何 $t \geq 0$, 随机变量 X_{t_1} 拥有无限可分分布. 证明：如果 X 是 levy 过程，并且 $t \geq 0$. 那么对于 $n=1,2,\dots$,

$$X_t = X_{t/n} + (X_{2t/n} - X_{t/n}) + \cdots + (X_t - X_{(n-1)t/n})$$

右边的每项都是独立同分布，并且有独立稳定增量属性，因此， X_t 拥有无限可分分布. 接下来，我们看下一维 levy 过程的特征函数，对于 levy 过程 X, X_t 的特征指数为 ψ_t ,

$$e^{\psi_t(\theta)} = E[e^{i\theta X_t}], \theta \in R$$

通过上面的公式，我们可以得到

$$m\psi_1(\theta) = \psi_m(\theta) = n\psi_{m/n}(\theta)$$

也就是说对于 $t \geq 0$,

$$\psi_t(\theta) = t\psi_1(\theta)$$

因此，我们可以得到

$$E[e^{i\theta X_t}] = e^{t\psi(\theta)}, \theta \in R$$

其中， $\psi := \psi_1$ 是 X_1 的特征指数.

莱维-辛钦定理

莱维-辛钦定理: X 为 levy 过程, 特征指数为 ψ . 存在唯一的 $a \in R, \sigma \geq 0$ 以及测度 Π , 满足 $\int_R 1 \wedge x^2 \Pi(dx) < \infty$

$$\psi(\theta) = ia\theta - \frac{1}{2}\sigma^2\theta^2 + \left(\int_R e^{i\theta x} - 1 - i\theta x I_{[-1,1]}(x) \right) \Pi(dx)$$

对于上述三个参数 (a, σ, Π) , a 包含了任何确定的漂移项, σ 是高斯系数, 描述了布朗运动的波动性, 而 Π 描述了 X 跳跃的大小和强度

poission 过程

poission 过程的概率密度函数为 $\mu_{\lambda}(k) = e^{-\lambda} \lambda^k / k!$, 因此 poission 分布的特征函数为

$$\sum_{k=1}^n e^{i\theta k} \mu_{\lambda}(k) = e^{\lambda(e^{i\theta}-1)} = [e^{\frac{\lambda}{n}(e^{i\theta}-1)}]^n$$

因此对于一个 poission 过程 $N_t : t \geq 0$ 如果参数为 λt , 则是 levy 过程, 特征函数为 $E(e^{i\theta N_t}) = e^{\lambda t(e^{i\theta}-1)}$, 特征指数为 $\psi(\theta) = \lambda(e^{i\theta} - 1)$

复合 poisson 过程

N 为 poisson 过程, 对于 $\theta \in R$, 特征函数为

$$E(e^{i\theta \sum_{i=1}^N \xi_i}) = \sum_{n \geq 0} E(e^{i\theta \sum_{i=1}^n \xi_i}) e^{-\lambda} \frac{\lambda^n}{n!}$$

$$= \sum_{n \geq 0} \left(\int_R e^{i\theta x} F(dx) \right)^n e^{-\lambda} = e^{\lambda \int_R e^{i\theta x} F(dx)}$$

其中三个参数分别为

$$a = \lambda \int_{0 < |x| < 1} x F(dx), \sigma = 0, \Pi(dx) = \lambda F(dx).$$

布朗运动

布朗运动的概率密度函数为：

$$\mu_{s,\gamma}(dx) := \frac{1}{\sqrt{2\pi s^2}} e^{-(x-\gamma)^2/2s^2} dx, x \in R$$

我们可以得到他的特征函数和特征指数：

$$E[e^{i\theta U}] = e^{-\frac{1}{2}s^2\theta^2 + i\theta\gamma}$$

三个参数为 $a = -\gamma, \sigma = s, \Pi = 0$

莱维-伊藤分解

根据莱维-辛钦定理，我们可以得到如下公式

$$\begin{aligned}\Psi(\theta) = & \left\{ ia\theta - \frac{1}{2}\sigma^2\theta^2 \right\} + \left\{ \prod(R \setminus (-1, 1)) \int_{|x| \geq 1} (1 - e^{i\theta x}) \frac{\prod dx}{\prod(R \setminus (-1, 1))} \right\} \\ & + \left\{ \int_{0 < |x| < 1} (1 - e^{i\theta x} + i\theta x) \prod(dx) \right\}\end{aligned}$$

第一项是线性布朗运动，第二项是复合泊松过程，第三项是平方可积鞅。因此 levy 过程可由这三个过程组合而成。

分解定理

背景

细菌觅食算法 (BFA) 由 K. M. Passino 于 2002 年基于 Ecoli 大肠杆菌在人体肠道内吞噬食物的行为, 提出一种新型仿真算法, 该算法属于随机搜索算法。该算法因具有群体智能算法并行搜索、易于跳出局部极小值、对初值和参数选择要求低、鲁棒性好、全局搜索等优点, 成为生物启发式计算研究领域的又一热点。



图 1.1 鸟群在被捕食时的反应
Fig. 1.1 The reaction when birds are attacked



图 1.2 鱼群在遭到捕食者袭击时的反应
Fig. 1.2 The reaction when fish are attacked



图 1.3 狼群在围捕一头野牛
Fig. 1.3 A bison is hunted by wolves



图 1.4 蚁群在协作搭桥
Fig. 1.4 Ants are assisting in building a bridge

图: 群体智能

大肠杆菌特点

基于大肠杆菌生物模型理论，算法主要融合了大肠杆菌的趋化行为、复制行为、迁徙行为和描述生物群体感应机制的聚集行为。大肠杆菌是一种常见的普通原核生物，由细胞膜、细胞壁、细胞质和细胞核等构成，其两端钝圆呈杆状。表面布满纤毛和鞭毛。纤毛是一种呈凸起状且可运动的细胞器，通过纤毛向彼此传递消息。



图 2.1 大肠杆菌的外观
Fig. 2.1 The appearance of E.coli

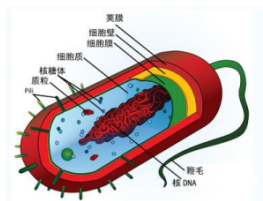


图 2.2 大肠杆菌结构
Fig. 2.2 The structure of E.coli

图: 大肠杆菌模型

大肠杆菌特点

大肠杆菌依靠其表面的鞭毛进行每秒大约 100 200 圈的转动以此来实现其自身的运动，当鞭毛逆时针摆动时，会给细菌一个向前的推力保证其向前游走；而当鞭毛进行顺时针摆动时，会给予其一个拉力其细菌翻转。大肠杆菌正是通过游动和翻转这两个基本动作组合来实现其在空间区域中的移动。

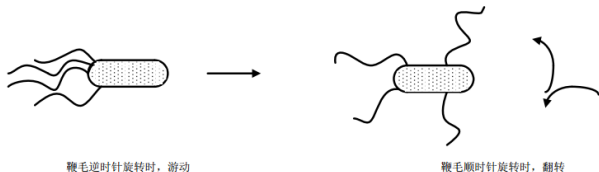


图 2.3 大肠杆菌的游动动作和翻转动作
Fig. 2.3 The swim and tumble of E.coli

图: 大肠杆菌运动模型

大肠杆菌特点

大肠杆菌虽是一种简单的生物，但对所经历的周围环境具有记忆其状态的能力，这种能力帮助大肠杆菌在搜索食物的过程中避开有毒物质向着存在食物的方向运动。同时，大肠杆菌也会对所经历过的所有状态都行评价据此反馈信息给后面的状态改变。包括周围环境的物理、化学因素变动所引起的温度过高、营养匮乏、环境 PH 值过偏等；也包括细菌本身的生理代谢过程也会使得所处环境发生变化。区域细菌浓度也是信号之一。

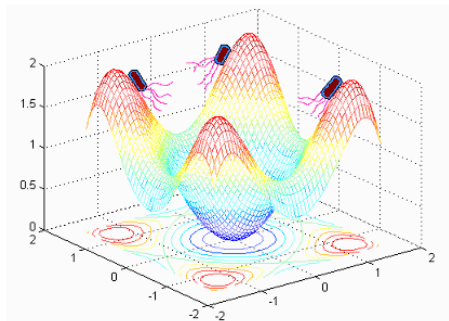


图: 应用示意图

大肠杆菌特点

通常食物是分布在多个区域之中的。首先，大肠杆菌利用自己的鞭毛进行游动来搜索食物的潜在区域。然后大肠杆菌决定是否进入选择的区域，若不进入，则继续搜索。当进入一个目标区域后，如果食物丰富，则停留至食物耗尽，之后沿着上次前进的方向继续搜索食物；如若该区域食物匮乏，那么细菌将会改变前进方向，朝着可能食物丰富区域搜索。

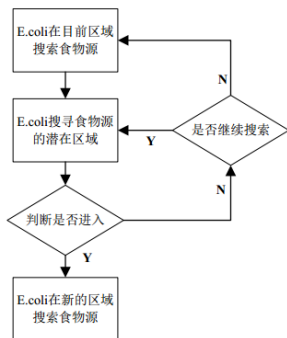


图 2.4 E.coli 搜寻食物源的过程
Fig. 2.4 The process of E.coli searching food

图: 大肠杆菌搜索食物流程图

算法基本原理

大肠杆菌自身的引导控制系统来趋利避害。算法中待求解优化问题的解空间对应为细菌觅食的整个搜索空间，各个优化问题的解则对应为细菌在空间中的位置，那么最优解总是被由更好食物源位置所对应的更好的优化问题解所不断靠近。

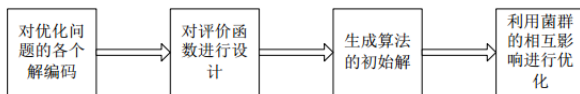


图 2.5 BFO 算法求解优化问题的一般过程

Fig. 2.5 The general process of the BFO for solving optimization problems

图: 求解过程

算法基本原理

细菌觅食算法是随机搜索的智能优化算法，其数学模型主要有三个基本步骤：趋向、复制、迁移。几何解释为三个循环层操作，最里层是趋向，中间层是复制，最外层是迁移。细菌觅食主要就是依据这三层循环操作来寻找食物。

算法基本原理

其觅食特点依靠鞭毛的旋转方向，逆时针，快速前进；顺时针，则停留原地。相应的数学模型解释就是当计算出两个点位置的函数的适应度值，通过比较结果决定大肠杆菌是继续前进还是寻找随机方向移动，当尝试次数达到最大限制数时选择下一个大肠杆菌进行趋向性操作。其数学表达式为

$$\theta'(j+1, k, l) = \theta'(j, k, l) + C(i) \frac{\Delta(i)}{\sqrt{\Delta(i)^T \Delta(i)}} \quad (11)$$

其中 $\theta'(j, k, l)$ 表示大肠杆菌 i 在第 j 次趋向第 k 次复制第 l 次迁移操作之后所在的位置， $C(i)$ 为趋向步长， $\frac{\Delta(i)}{\sqrt{\Delta(i)^T \Delta(i)}}$ 为移动的一个随机前进方向。

算法基本原理

自然界生物进化的准则是优胜劣汰，因此性能差的细菌会被淘汰。在 BFA 模型中, $J(i, j, k, l)$ 表示细菌 i 在第 j 次趋向第 k 次复制第 l 次迁移操作之后的适应值。 J_{health}^i 是细菌 i 的能量函数，其大小决定细菌觅食能力的强弱。将觅食后的细菌的能量函数进行大小排序，淘汰掉 $\frac{S}{2}$ 个能量值较小的细菌，将剩下的进行赋值操作，新生成的复制细菌与原细菌具有相同的觅食能力。

算法基本原理

在生物体中存在的细菌，其环境对其觅食具有很大的影响，比如局部区域温度升高或者食物被消耗掉，这些必然导致细菌被迫迁移到新的区域去寻找食物。这种迁移未必会导致细菌寻找不到新的食物，反而会对种群的良好生长起到促进作用，在经过复制操作之后细菌若按照一定的概率进行迁移，新的种群具有随机特性，与原种群相比可能有不同的觅食能力，这种随机性能使群体跳出局部极值而更好的靠近全局最优解区域。

算法基本原理

在细菌觅食过程中，并不是独立行为，即在模型中体现为引力和斥力，当引力作用是，细菌会快速的聚集在一起，围绕最优食物觅食，形成细菌群。当斥力起作用时，表示细菌能具有更好的独立寻找食物的能力。其数据模型为

$$\begin{aligned} J_{cc}(\theta, P(j, k, l)) &= \sum_{i=1}^S J_{cc}(\theta, \theta^i(j, k, l)) \\ &= \sum_{i=1}^S [-d_{att} \exp(-\omega_{att} \sum_{m=1}^P (\theta - \theta_m^i)^2)] \\ &\quad + \sum_{i=1}^S [h_{rep} \exp(-\omega_{rep} \sum_{m=1}^P (\theta - \theta_m^i)^2)] \end{aligned} \quad (12)$$

其中 d_{att} 是引力深度， ω_{att} 是引力宽度， h_{rep} 是斥力高度， ω_{rep} 是斥力宽度， θ_m^i 是细菌 i 的第 m 个分量。通常取 $d_{att} = h_{rep}$ ，趋向性操作的适应度值的计算公式可表示为：

$$J(i, j+1, k, l) = J(i, j, k, l) + J_{cc}(\theta^i(j+1, k, l), P(j+1, k, l)) \quad (13)$$

算法主要步骤

Step1: 初始化参数, 种群大小 S , 空间维数 P , 趋向性行为的次数 N_c , 趋向性操作前进最大步数 N_s , 复制性操作的次数 N_{re} , 迁移操作的次数 N_{ed} , 迁移概率 P_{ed} , 细菌 i 的信息用 D 维向量表示为 $\theta^i = [\theta_1^i, \theta_2^i, \dots, \theta_D^i]$, $i = 1, 2, \dots, S$, 趋向性步长 $C(i)$.

Step2: 迁移循环操作;

Step3: 复制循环操作;

Step4: 趋向性循环操作;

(1) 细菌 i 进行趋向性一步;

(2) 计算 $J(i, j, k, l)$, 储存最优值 J_{best} ;

(3) 随机生成向量 $\Delta(i)$, 细菌按照步长 $C(i)$ 进行移动;

(4) 计算此时的适应度值 $J(i, j+1, k, l)$;

(5) 旋转判定条件: 若 $m < N_s$, $m = m + 1$, 若 $J(i, j+1, k, l) < J_{best}$, 则 $J_{best} = J(i, j+1, k, l)$, 此时 $\theta^i(j+1, k, l) = \theta^i(j, k, l) + C(i) \frac{\Delta(i)}{\sqrt{\Delta(i)^T \Delta(i)}}$.

根据细菌的信息量 $\theta^i(j+1, k, l)$ 计算新的 $J(j+1, k, l)$ 否则 $m = N_s$.

(6) 重新处理下一个细菌。

算法主要步骤

Step5: 若 $j < N_c$ 返回上一步进行细菌趋向性操作。

Step6: 复制，淘汰 $\frac{S}{2}$ 个能量值小的细菌，对剩下的 $\frac{S}{2}$ 个细菌进行复制。

Step7: 若 $k < N_{re}$ ，返回 Step3.

Step8: 迁移，达到一定条件后细菌以概率 P_{ed} 重新觅食。若 $l < N_d$ ，则返回 Step2，否则寻优结束。

Step9: 循环结束条件判断，条件满足则结束，输出结果。

参数选取

模型中种群的规模 S 太大导致计算量增大, 太小会降低种群的多样性; 趋向性步长 C 太大时, 优化值易陷入局部极值; 当太小时, 算法的计算复杂度会大幅度增加, 不利于算法收敛; 引力参数 d_{att} 和 ω_{att} 大小决定算法的群聚性, 如果值较大会导致不按照自己的信息搜索寻找食物而过度的向中心靠拢, 反之太小会完全按照自身信息搜索食物导致群聚性能很弱, 斥力参数 h_{rep}, ω_{rep} 与引力参数相反; 趋向性操作因子 N_c 决定算法的寻优能力和收敛速度; 趋向性参数 N_s 最大步长对算法的收敛速度有很大的影响; 赋值操作因子 N_{re} 决定算法的计算复杂度和收敛速度, 并且复制操作所采取的原则计算出的适应度值只与当前位置有关, 导致最优解被淘汰掉容易陷入局部最优; 迁移操作因子执行次数 N_{ed} 太小会使算法陷入局部最优, 反之计算量和复杂度增加, 迁移概率 P_{ed} 太大会使算法进入循环搜索, 以至远离最优解。

案例

对优化问题 $\min \sum_{i=1}^n x_i^2 + x_i$ 做测试。

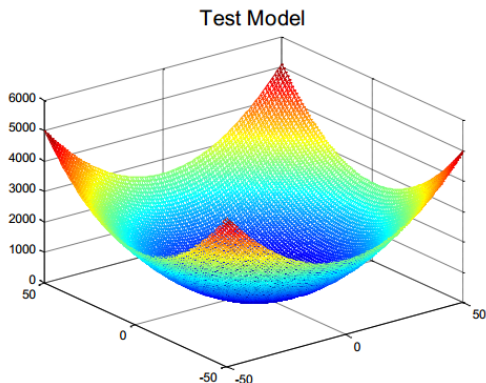


图 3.10 测试函数是 $f(x) = \sum_{i=1}^n x_i^2 + x_i$ 的三维图像

Fig. 3.10 The three-dimensional image of the test function $f(x) = \sum_{i=1}^n x_i^2 + x_i$

图：测试案例

仿真结果

对算法进行 10000 次仿真。对如下四个标志数据：最小值、最大值、平均值、标准差等做评价。

表 3.1 C28-BFO 算法的仿真结果
Table. 3.1 The simulation results of C28-BFO algorithm

仿真次数	min value	max value	average value	stardartd diviation
第一次	9.8535	6.8647e+004	1.3817e+003	7.4929e+003
第二次	9.0499	6.9979e+004	1.3145e+003	7.1791e+003
第三次	10.2501	6.5632e+004	1.3877e+003	7.1940e+003
第四次	12.8886	6.4345e+004	1.3878e+003	7.1507e+003
第五次	9.7737	6.6766e+004	1.2079e+003	6.8080e+003
第六次	7.8143	7.3082e+004	1.3745e+003	7.6817e+003
第七次	9.6852	7.0552e+004	1.3794e+003	7.5952e+003
第八次	10.3507	7.3194e+004	1.4279e+003	7.5933e+003

图：测试结果

算法应用

在控制领域主要是控制器设计，可以用细菌觅食算法求解 PID 控制器参数优化问题。同样，BFA 在图像处理领域中得到很好的应用，主要应用于图像分割、边缘检测、图像压缩等。亦可将细菌觅食算法应用到神经网络学习机制中，应用到负荷预测。也可通过 BFA 来优化模糊系统的参数，优化后的滤波器优于传统的卡尔曼滤波器。在盲信号分离也可应用 BFA，在模式识别领域得到广泛探索。此外，细菌觅食算法在调度问题、股票经济预测、组合优化等问题中也有广泛应用。

觅食策略

多说站在食物链顶端的动物们，例如狮子、老虎等主要靠视觉来搜寻猎物，和靠气味觅食（大白鲨）不同，视觉不存在所谓的梯度差，看不见就是看不见，没有动物具有透视眼。所以觅食策略对于狮子老虎这样的陆地动物是很重要的。在不知道猎物位置的情况下，**布朗运动**或许是最自然的方法了。布朗运动是一个随机过程，它的步长服从**正态分布**。

离散版的布朗运动叫做**随机游走**（Random walk），它的步长和方向都是离散的。尽管布朗运动和随机游走是两个概念，也各自发展出了一套不同的理论，但两者的数学本质是一样的。

布朗运动轨迹

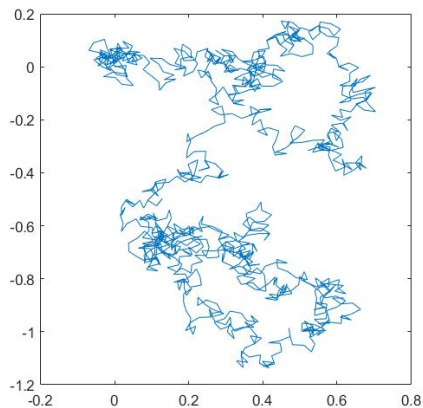


图: 布朗运动轨迹图

现在提供一个布朗运动的轨迹图。

布朗运动

随机过程是一系列随机变量的集合。比如说，每丢一次硬币，便产生一个随机变量，那么一次次的丢下去，便产生一系列的随机变量 X_1, \dots, X_i, \dots 。随机变量序列集合起来，便是随机过程。

布朗运动是悬浮在液体或气体中的微粒所作的永不停息的无规则运动。它是一种正态分布的独立增量连续随机过程。其基本性质为：布朗运动 $W(t)$ 是期望为 0，方差为 t （时间）的正态随机变量。对于任意的 $r \leq s$, $w(t) - W(s)$ 独立于 $W(r)$ ，且是期望为 0 方差为 $t - s$ 的正态随机变量。

觅食策略

在物理学家眼里，布朗运动可以用郎之万方程（Langevin Equation。提出人保罗·郎之万是法国物理学家，居里夫人丈夫皮埃尔·居里的博士生，并且后来给皮埃尔·居里戴了绿帽子）来描述：

$$m \frac{d^2 x}{dt^2} = f(t) + F'(t). \quad (14)$$

其中 $f(t)$ 分成两部分，一部分为阻力 $-\alpha v$ ；一部分为随机作用力 $F(t)$ 。粘滞阻力仍来自介质分子对颗粒的碰撞，将颗粒看作半径为 a 的小球，在粘滞系数为 η 的流体中运动，则有 $\alpha = 6\pi\eta a$ 。

$$m\ddot{x} = -\mu\dot{x}(t) + \eta(t). \quad (15)$$

其中 $\eta(t)$ 是均值为 0，方差为常数的随机过程。更确切地讲，是白噪声过程。

觅食策略

因此郎之万方程本质上就是考虑了**随机误差**的牛顿第二定律，这样的方程又叫做**随机微分方程**（Stochastic Differential Equation）。尽管郎之万是第一个提出随机过程的人，但作为物理学家，他更关心这些方程是否符合实际物理现象，而不关心数学上的严谨性。因此这种模型并没有立即受到数据家的重视。

到了二十世纪中叶，日本数学家伊藤清和苏联数学家 Stratonovich 先后使用概率论的方法，把随机微分方程发展成严谨的数学概念。尽管如此，但两人对随机微分方程定义各有不同，这也显示了**随机性**和**确定性**的本质差异。此外，随机微分方程多用于对一些多样化现象进行建模，比如不停变动的股票价格、部分物理现象如热扰动等。

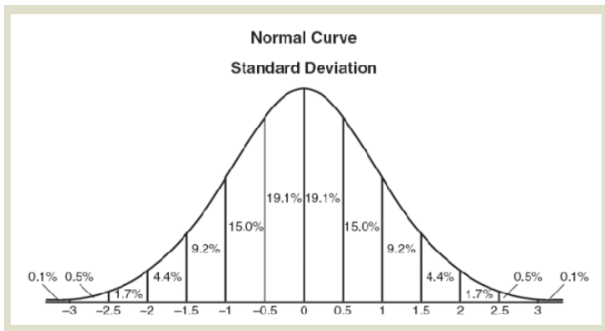
觅食策略

如果捕食者像无头苍蝇一样漫无目的的做布朗运动，真的能很有效的地寻到到猎物吗？在数学上可以证明，布朗运动和分子自由扩散一样，单位速度的分子在时间 t 内平均只有 \sqrt{t} 的位移量。捕食者若采用此种策略，可能需要踏遍千山万水才能成功了。

那么有没有比布朗运动更高效的搜索方法呢？一组巴西物理学家于1999年提出了一个设想，认为“莱维飞行”比布朗运动有更高的搜索效率，因此自然会偏向与采用“莱维飞行”捕食的生物。为了理解什么是“莱维飞行”，我们首先需要对概率分布的重尾性有一个认识。

概率分布的轻尾性

前面说过，布朗运动的步长服从正态分布。正态分布是一种典型的轻尾分布，也就是说，不太可能取得极端值的分布。



图：正态分布图

表达式： $y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ 。其中 μ 是均值， σ 是标准差。当 x 增大时，函数下滑趋势可谓是猛如雪崩，因此数学家又将正态分布函数纳入速降函数空间范畴。速降函数是专门为傅里叶分析而量身定做的。

概率分布的重尾性

而在现实生活中，太多的事件不能用轻尾分布来描述，例如保险领域的保险金——事故可以看作是稀有事件，但一旦发生并通过审核，保险公司就必须大量金额，这样一来保险金的分布就很可能取到很大的值。为了描述这样容易取得极值的随机变量，我们需要引入**重尾分布**。

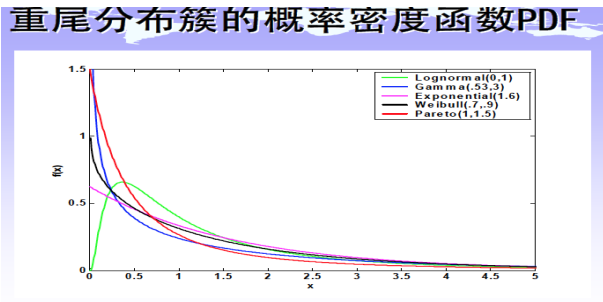


图: 重尾分布图

典型特点：大头短 + 小尾长。“莱维飞行”的平均位移是 $t^\gamma, \gamma > \frac{1}{2}$ 。

概率分布的重尾性

那么，如何验证“莱维飞行”的有效性呢？一个方案就是前面提到的大肠杆菌的**趋化行为**。尽管最初的“莱维飞行”假说是针对动物而言的，但从大肠杆菌的游击战策略可以看出，该假说对微生物同样适用！这样一来，对微生物的研究也能反过来帮助人们更深入地理解动物行为，这便是微观的细胞生物学在宏观的生态学中的应用。

概率分布的重尾性

而在现实生活中，太多的事件不能用轻尾分布来描述，例如保险领域的保险金——事故可以看作是稀有事件，但一旦发生并通过审核，保险公司就必须大量金额，这样一来保险金的分布就很可能取到很大的值。为了描述这样容易取得极值的随机变量，我们需要引入**重尾分布**。

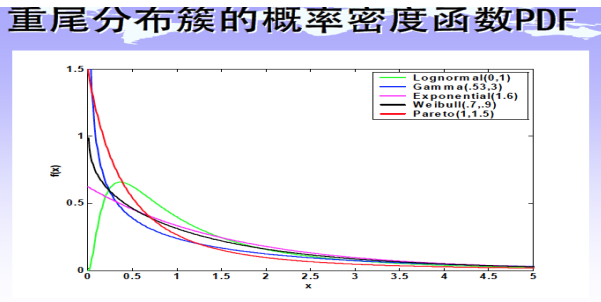


图: 重尾分布图

典型特点：大头短 + 小尾长。“莱维飞行”的平均位移是 $t^\gamma, \gamma > \frac{1}{2}$ 。

莱维飞行

莱维飞行时服从莱维分布的随机搜索路径，是一种短距离的搜索与偶尔长距离的行走相间的行走方式，它能够解释很多自然现象，如布朗运动、随机行走等。目前莱维飞行行为已被用于优化领域，比如布谷鸟算法中就采用了莱维飞行进行位置更新。莱维飞行能增加种群多样性和扩大搜索范围，因此采用莱维飞行的智能算法更容易跳出局部最优解。

莱维过程

下面对莱维过程 (Lévy process) 做一个介绍。



图: 哈哈

莱维分布

莱维分布 (Levy Distribution) 是由 P.Levy 在 19 世纪 30 年代提出的一类分布, 这种分布有两个参数: α 和 γ 。参数 $\gamma > 0$, 参数 α 用于控制分布的形状, 且满足 $0 < \alpha \leq 2$ 。事实上当 $\gamma = 1$ 时, 莱维分布就转换为柯西分布, 而当 $\gamma = 2$ 时, 莱维分布则为正态分布。莱维分布的概率密度函数为

$$L_{\alpha, \gamma} = \frac{1}{\pi} \int_0^{\infty} e^{-\gamma q^{\alpha}} \cos(qy) dq, y \in \mathcal{R}. \quad (16)$$

上述积分很难积, 因此现有的莱维分布基本上使用数值方法计算。设 x, y 是两个独立同分布的随机变量, 且均为标准正态分布, 令随机变量 v 满足:

$$v = \frac{x}{\sqrt{y}} \quad (17)$$

则随机变量 $\{z_n\}_{n=1}^{+\infty}$ 为

$$z_n = \frac{1}{\alpha \sqrt{n}} \sum_{j=1}^n v_j \quad (18)$$

收敛于莱维分布。

莱维分布

现在提供一个 matlab 例子。轨迹如下：

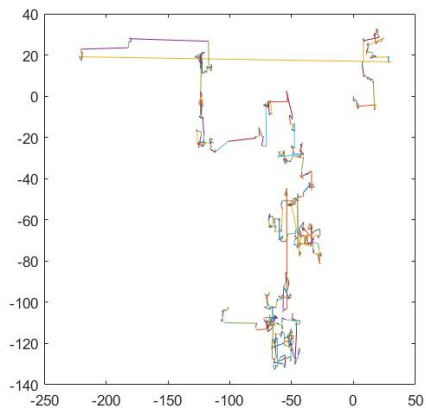


图: Levy 分布轨迹图

莱维过程

莱维过程 $\{X(t), t \geq 0\}$ 是一种随机过程，它满足的条件比布朗运动宽松：

1. $X(0)$ 几乎处处为 0；
2. 独立增量性；
3. 稳定增量性；
4. 样本轨道右连续。

连续的布朗运动和离散的泊松过程都是莱维过程的特例。因此可以大胆猜测，莱维过程就是带“跳跃”的布朗运动。正是这些不连续性的“跳跃”给予莱维过程“重尾”的特性。

莱维过程

独立增量:

设 $X(t)$ 是一个连续时间上的随机过程。也就是说, 对于任何固定的 $t \geq 0$, $X(t)$ 是一个随机变量。过程的增量为差值 $X(s) - X(t)$ 。独立增量意味着对于任意时间 $s > t > u > v$, $X(s) - X(t)$ 与 $X(u) - X(v)$ 相互独立。

稳定增量:

如果增量 $X(s) - X(t)$ 的分布只依赖于时间间隔 $s - t$, 则称增量是稳定的。例如对于维纳过程, 增量 $X(s) - X(t)$ 服从均值为 0, 方差为 $s - t$ 的正态分布。对于泊松过程, 增量 $X(s) - X(t)$ 服从指数为 $s - t$ 的泊松分布。

莱维过程

定理 1(莱维-辛钦公式): 莱维过程 $\{X(t), t \geq 0\}$ 的特征函数 (傅里叶变换) 表达如下:

$$\phi_X(\theta)(t) := E[e^{i\theta X(t)}] = \exp\left(t\left(ai\theta - \frac{1}{2}\sigma^2\theta^2 + \int_{\mathcal{R} \setminus \{0\}} (e^{i\theta x} - 1 - i\theta x I_{|x|<1}) \prod(dx)\right)\right).$$

(19)

这个定理的证明比较复杂, 依赖于测度论中的一系列结果。

莱维过程

定理 2(莱维-伊藤分解): 每个莱维过程 $\{X(t), t \geq 0\}$ 都可以分解为 $\{S(t), t \geq 0\}, \{Y(t), t \geq 0\}$ 和 $\{Z(t), t \geq 0\}$ 三个子过程, 其中:

1. $S(t)$ 是维纳过程 (就是布朗运动, 莱维过程的连续部分);
2. $Y(t)$ 是复合泊松过程 (刻画了较极端的“跳跃”现象);
3. $Z(t)$ 是平方可积的离散鞅 (刻画了较小的“跳跃”现象)。

Basic Definition

随机变量 (r.v.) $X \sim \mu$ 的特征函数是映射 $\Phi : \mathbf{R}^d \rightarrow \mathbb{C}$, 定义如下

$$\Phi(u) = \mathbb{E}(e^{iu \cdot X}) = \int_{\mathbf{R}^d} e^{iu \cdot y} \mu(dy). \quad (20)$$

随机变量 X 是无限可分的, 除非它的该率分布 p_x 是无限可分的, 例如, $X = Y_1^{(n)} + \dots + Y_n^{(n)}$, 其中 $Y_1^{(n)}, \dots, Y_n^{(n)}$ 是独立同分布。那么 X 的特征函数可以写成 $\Phi_X(u) = (\Phi_{Y_1^{(n)}}(u))^n$.

定义 1:

莱维可测是满足如下条件的 $\mathbf{R}^d \setminus \{0\}$ 上的测度 ν , 使得

$$\int (|y|^2 \wedge 1) \nu(dy) < \infty. \quad (21)$$

其中 $|y|^2 \wedge 1 = \begin{cases} 1, & |x| > 1 \\ y^2, & |x| \leq 1 \end{cases}$

Levy-Ito decomposition

定义函数 $X: \mathbf{R}^+ \rightarrow \mathbf{R}$ 是 Cadlag 过程, 即 X 是右连续左极限存在。令 $\Delta X(t) = X(t) - X(t-)$ (由于左极限的存在) 且定义如下泊松随机测度

$$N(t, A) = \#\{\Delta X(s) \in A : s \in [0, t]\}. \quad (22)$$

那么有如下结论:

(1) $N(1, B_\varepsilon^c(0)) < \infty$.

(2) $N(1, \mathbf{R} \setminus \{0\})$ 是可数的。

令 A 是有下界的, i.e., $0 \notin \bar{A}$. 那么 $N(t, A), t \geq 0$ 是泊松过程且强度为 $\mu(A) = \mathbf{E}[N(1, A)]$. 很明显有结论 $\mu(A) < \infty$ 不管 A 是否有下界。所以测度 μ 是 σ -有限的。

Levy-Ito decomposition

推论 1: (1) 对任意的 $t > 0, \omega \in \Omega$, $N(t, \cdot)(\omega)$ 在 $\mathcal{B}(\mathbf{R}^d \setminus \{0\})$ 上是计数可测的。

(2) 对于任意的有下界的 $A, N(t, A), t \geq 0$ 是泊松过程且具有如下强度 $\mu(A) = \mathbf{E}[N(1, A)]$.

(3) 补 $\tilde{N}(t, A) = N(t, A) - t\mu(A)$ 是值鞅可测的。对于有下界的 A , $\tilde{N}(t, A)$ 是一个鞅。

令 $f: \mathbf{R}^d \rightarrow \mathbf{R}$ 是波莱尔可测函数, A 有下界, 那么对任意 $t > 0, \omega \in \Omega$, 我们定义如下关于 f 泊松积分通过一个随机有限和。

$$\int_A f(x) N(t, dx)(\omega) = \sum_{x \in A} f(x) N(t, \{x\})(\omega). \quad (23)$$

Levy-Ito decomposition

定理 2: 令 A 有下界, 则

(1) 对任意的 $\int_A f(x)N(t, dx)$, $t \geq 0$ 是复合泊松过程, 带有如下特征函数:

$$\mathbf{E}[\exp\{iu \cdot \int_A f(x)N(t, dx)\}] = \exp\{t \int_A (e^{iu \cdot x} - 1)\mu_f(dx)\}. \quad (24)$$

其中 $u \in \mathbf{R}^d$, $\mu_f = \mu \circ f^{-1}$.

(2) 如果有 $f \in L^1(A, \mu_A)$, 那么有

$$\mathbf{E}[\int_A f(x)N(t, dx)] = t \int_A f(x)\mu(dx). \quad (25)$$

(3) 如果有 $f \in L^2(A, \mu_A)$, 那么有

$$\text{Var}[\int_A f(x)N(t, dx)] = t \int_A |f(x)|^2 \mu(dx). \quad (26)$$

Levy-Ito decomposition

从定理 2 可以看出, 如果 $f \in L^1(A, \mu_A)$, 一个泊松积分也许不一定全都有有限期望。对于此, 我们定义如下补泊松积分:

$$\int_A f(x) \tilde{N}(t, dx) = \int_A f(x) N(t, dx) - t \int_A f(x) \mu(dx). \quad (27)$$

则有如下结论:

(1) $\int_A f(x) \tilde{N}(t, dx), t \geq 0$ 是一个鞅。

(2) 特征函数:

$$\mathbf{E}[\exp\{iu \cdot \int_A f(x) \tilde{N}(t, dx)\}] = \exp\{t \int_A (e^{iu \cdot x} - 1 - iu \cdot x) \mu_f(dx)\}. \quad (28)$$

其中 $u \in \mathbf{R}^d, \mu_f = \mu \circ f^{-1}$.

(3) 如果有 $f \in L^2(A, \mu_A)$, 那么有

$$\text{Var}[\int_A f(x) \tilde{N}(t, dx)] = t \int_A |f(x)|^2 \mu(dx). \quad (29)$$

Levy-Ito decomposition

对于有下界的 A 和任意的 $t > 0$, $\int_A xN(t, dx) \sum_{0 \leq u \leq t} \Delta X(u) 1_A(\Delta X(u))$ 是集合 A 中不超过时间 t 的所有跳变值和。由于轨迹 X 是 Cadlag, 所以上述是个有限随机和。特别地, $\int_{|x| \geq 1} xN(t, dx)$ 是所有大于 1 的跳变之和。也就是说是一个带有有限次扰动的复合泊松过程。相反地, 也可以证明 $X(t) - \int_{|x| \geq 1} xN(t, dx)$ 带有有限次有次序动作的莱维过程。但是也许有无界扰动。所以我们可以定义

$$b = \mathbf{E}[X(1) - \int_{|x| \geq 1} xN(1, dx)]. \quad (30)$$

Levy-Ito decomposition

现在把关注度放在小跳变上来。引入 $M(t, A) = \int_A f(x) \tilde{N}(t, dx)$. 令 $A_m = \{x : \frac{1}{m+1} < |x| \leq 1\}$. 可以证明在 L^2 里, 当 $m \rightarrow \infty$ 时, 我们有 $M(t, A_m) \rightarrow \int_{|x|<1} x \tilde{N}(t, dx)$. 所以 $\int_{|x|<1} x \tilde{N}(t, dx)$ 是一个鞅。取极限得

$$\mathbf{E}[\exp\{iu \cdot \int_A x \tilde{N}(t, dx)\}] = \exp\{t \int_A (e^{iu \cdot x} - 1 - iu \cdot x) \mu(dx)\}. \quad (31)$$

最后考虑如下随机过程

$$W_A(t) = X(t) - bt - \int_{|x|<1} x \tilde{N}(t, dx) - \int_{|x|\geq 1} x N(t, dx). \quad (32)$$

那么随机过程 $W_A(t)$ 是带有连续采样路径的中心鞅。利用布朗运动的莱维特征化, 我们有 $W_A(t)$ 就是协方差为 A 的布朗运动。

Levy-Ito decomposition

Levy-Ito Decomposition X 是莱维过程。那么存在 $b \in \mathbf{R}^d$, 协方差为 A 的布朗运动 $W_A(t), \mathbf{R}^+ \times \{\mathbf{R}^d \setminus \{0\}\}$ 独立的泊松随机测度 N 使得下式成立:

$$X(t) = bt + W_A(t) + \int_{|x|<1} x \tilde{N}(t, dx) + \int_{|x|\geq 1} x N(t, dx). \quad (33)$$

其中平方可积鞅 (L^2 -鞅) $\int_{|x|<1} x \tilde{N}(t, dx)$ 是所有小跳变的补和。上述跳变是以 1 为界的, 可以推广至任意实数 $R > 0$, 我们有

$$X(t) = b_R t + W_A(t) + \int_{|x|<R} x \tilde{N}(t, dx) + \int_{|x|\geq R} x N(t, dx). \quad (34)$$

其中 $b_R = \mathbf{E}[X(1) - \int_{|x|\geq R} x N(1, dx)]$. 可以计算如下:

- (1) 如果 $1 < R < \infty$, 有 $b_R = b + \int_{1 \leq |x| < R} x \mu(dx)$.
- (2) 如果 $0 < R < 1$, 有 $b_R = b - \int_{R \leq |x| < 1} x \mu(dx)$.

Levy-Ito decomposition

是否可以去除上届约束 R 呢? 此时我们有

$$X(t) = b_{\infty}t + W_A(t) + \int_{|x| \geq 0} x \tilde{N}(t, dx). \quad (35)$$

结论是可以的, 如果我们有 $\mathbf{E}[X(1)] < \infty$ 。此时, 我们有 $b_{\infty} = \mathbf{E}[X(1)] < \infty$ 。

随机微分方程

大纲

- ▶ 基本概念：
随机过程、布朗运动、Ito 积分、SDE...
- ▶ SDE 的数值解
- ▶ SDE 的参数估计

随机过程

例：我们在某一时段对某一地区成人的身高和体重 (X, Y) 进行随机抽样，可得出

$$Z = (X, Y) \sim N(\mu_1, \mu_2, \rho, \sigma_1^2, \sigma_1^2)$$

$$Z(\omega) = (X(\omega), Y(\omega))$$

这是某时段所考查随机变量的概率分布，如果每隔 10 年在同一地区做同样的随机抽样，得：

$$Z(\omega, 0) = (X(\omega, 0), Y(\omega, 0)) \sim N(\mu_{01}, \mu_{02}, \rho_0, \sigma_{01}^2, \sigma_{02}^2)$$

$$\text{第 1 个 10 年: } Z(\omega, 1) = (X(\omega, 1), Y(\omega, 1)) \sim N(\mu_{11}, \mu_{12}, \rho_1, \sigma_{11}^2, \sigma_{12}^2)$$

$$\text{第 2 个 10 年: } Z(\omega, 2) = (X(\omega, 2), Y(\omega, 2)) \sim N(\mu_{21}, \mu_{22}, \rho_2, \sigma_{21}^2, \sigma_{22}^2)$$

.....

$$\text{第 } t \text{ 个 10 年: } Z(\omega, t) = (X(\omega, t), Y(\omega, t)) \sim N(\mu_{t1}, \mu_{t2}, \rho_t, \sigma_{t1}^2, \sigma_{t2}^2)$$

$$t = 0, 1, 2, \dots$$

因此 $\{Z(\omega, t) = (X(t), Y(t)) : t \geq 0\}$ 表示的就是身高和体重这两个随机变量在不同时段的情况。

随机过程

定义 (随机过程)

设 (Ω, \mathcal{F}, P) 为一概率空间，另设集合 T 为一指标集合，如果对于所有 $t \in T$ ，均有一随机变量 $X(t, \omega)$ 定义于概率空间 (Ω, \mathcal{F}, P) ，则集合 $\{X(t, \omega) | t \in T\}$ 为一随机过程

- ▶ 对于固定的 ω ，比如 $\bar{\omega}$ ， $\{X(t, \bar{\omega}), t \geq 0\}$ 被称为路径 (**path**) 或轨迹 (**trajectory**)
- ▶ 对于固定的 t ，比如 \bar{t} ，集合 $\{X(\bar{t}, \omega), \omega \in \Omega\}$ 是时刻 \bar{t} 在该随机过程中的状态集， $X(\bar{t}, \omega)$ 也就是时刻 \bar{t} 的随机变量

布朗运动

标准布朗运动的数学性质.

设连续时间随机过程 $W_t : 0 \leq t < T$ 是 $[0, T)$ 上的标准布朗运动,

- ▶ $W_0 = 0$
- ▶ **独立增量性:** 对于有限个时刻 $0 \leq t_1 < t_2 < \dots < t_n < T$, 随机变量

$$W_{t_2} - W_{t_1}, W_{t_3} - W_{t_2}, \dots, W_{t_n} - W_{t_{n-1}}$$

是独立的

- ▶ **正态性:** 对任意的 $0 \leq s < t < T$, $W_t - W_s$ 服从均值为 0, 方差为 $t-s$ 的正态分布



随机微分方程

一般的随机微分方程

$$\frac{dX(t)}{dt} = h[X(t), t] + g[X(t), t]R(t)$$

形式上

$$R(t) = \frac{dB}{dt}$$

其中记号 $dW(t) = R(t)dt$, $B(\omega, t)$ 是维纳过程
对上式求积分, 得:

$$X(t) - X(0) = \int_0^t h[s, X(s)]ds + \int_0^t g[s, X(s)]dB$$

第一个积分是普通微积分的, 第二个积分是维纳过程的随机函数的积分

布朗运动: $dX_t = \mu dt + \sigma dW_t$

几何布朗运动: $dX_t = \mu X_t dt + \sigma X_t dW_t$

随机积分的求解

与普通微积分的牛顿莱布尼茨公式采用分区间近似求和相同，随机微积分中也是用分片的常数函数来近似 $g[s, X(s)]$ ，即

$$\int_0^t g[s, X(s)] dB \approx \sum_{i=0}^{n-1} g_i[s, X(s)] dB_i = \sum_{i=0}^{n-1} g_i[s, X(s)] [B(t_{i+1}) - B(t_i)], s \in [t_{i+1}, t_i]$$

其中 s 的取值有两种：

- ▶ 在区间左端点 t_i 取值

$$g_i[s, X(s)] = G[t_i, X(t_i)]$$

相应的随机积分

$$\int_0^t g[s, X(s, \omega)] dB(\omega, t)$$

称为 Ito 积分

- ▶ 在区间端点取值求平均

$$g_i[s, X(s)] = \frac{g[t_i, X(t_i)] + g[t_{i+1}, X(t_{i+1})]}{2}$$

相应的随机积分

$$\int_0^t g[s, X(\omega, s)] dB(\omega, t)$$

称为 Stratonovich 积分

随机积分的求解：例

设 $g[t, B(t)] = B(t)$, 则在 Ito 积分中:

$$I_1 = \int_0^t B(s) dB \quad (36)$$

$$\approx - \sum_{i=0}^{n-1} B_i (B_i - B_{i+1}) \quad (37)$$

$$= -[B_0^2 - B_0 B_1 + B_0^2 - B_0 B_1 + \dots + B_{n-1}^2 - B_{n-1} B_n] \quad (38)$$

$$= -\frac{1}{2}[B_0^2 + \sum_{i=1}^{n-1} (B_{i+1} - B_i)^2 - B_n^2] = \frac{1}{2}(B_t^2 - B_0^2) - \frac{1}{2} \sum_{i=1}^{n-1} \Delta B_i^2 \quad (39)$$

在 Stratonovich 积分中:

$$I_2 = \int_0^t B(s) dB \approx \sum_{i=0}^{n-1} \frac{1}{2} [B(t_{i+1}) + B(t_i)] [B(t_{i+1}) - B(t_i)] \quad (40)$$

$$= \frac{1}{2} \sum_{i=0}^{n-1} [B^2(t_{i+1}) - B^2(t_i)] \quad (41)$$

$$= \frac{1}{2} [B^2(t_1) - B^2(t_0) + B^2(t_2) - B^2(t_1) + \dots + B^2(t_n) - B^2(t_{n-1})] \quad (42)$$

$$= \frac{1}{2} [B^2(t_n) - B^2(t_0)] = \frac{1}{2} B^2(t) \quad (43)$$

Ito 公式

普通微积分中，若 X 连续可微，则

$$X(t) - X(0) = \sum_{i=1}^n [X(t_i) - X(t_{i-1})] = \sum_{i=1}^n X'(\xi_i)(t_i - t_{i-1}) = \int_0^t X'(s)ds$$

将 t 替换成 B_t ，此时 ξ_i 是介于 B_{i-1} 和 B_i 之间的点， ξ_i 在区间不同位置的取值会影响积分值，Ito 将取 ξ_i 为左端点进行求和，这就需要将泰勒公式多展开一阶 $X(B_{t_j}) - X(B_{t_{j-1}}) = X'(B_{t_{j-1}})(B_{t_j} - B_{t_{j-1}}) + \frac{1}{2}X''(\xi_j)(B_{t_j} - B_{t_{j-1}})^2$
右端第一项求和，得

$$\sum_{j=1}^n X'(B_{t_{j-1}})(B_{t_j} - B_{t_{j-1}}) \rightarrow \int_0^t X'(B_s)dB_s$$

右端第二项求和，

$$\sum_{j=1}^n \frac{1}{2} X''(\xi_j)(B_{t_j} - B_{t_{j-1}})^2 \rightarrow \frac{1}{2} \int_0^t X''(B_s)ds$$

即 $X(B_t) = X(B_0) + \int_0^t X'(B_s)dB_s + \frac{1}{2} \int_0^t X''(B_s)ds$
微分形式： $dX(B_t) = X'(B_t)dB_t + \frac{1}{2}X''(B_t)dt$

Ito 公式的应用

假设 $X(t)$ 满足几何布朗运动的 SDE:

$$dX_t = \mu X_t dt + \sigma X_t dW_t$$

如何解出 $X(t)$?

设 $Y(t) = \ln X(t)$, 则 $\frac{\partial Y}{\partial X} = \frac{1}{X}$, $\frac{\partial^2 Y}{\partial X^2} = -\frac{1}{X^2}$, 由 Ito 公式得:

$$dY = \frac{\partial Y}{\partial X} dX + \frac{1}{2} \frac{\partial^2 Y}{\partial X^2} (X_t)^2 \quad (44)$$

$$= \frac{1}{X} (\mu X dW) + \frac{1}{2} \left(-\frac{1}{X^2}\right) \sigma^2 X^2 dt \quad (45)$$

$$= \mu dt + \sigma dW - \frac{1}{2} \sigma^2 dt \quad (46)$$

$$= \left(\mu - \frac{1}{2} \sigma^2\right) dt + \sigma dW \quad (47)$$

$$(48)$$

Ito 公式的应用

则 $Y(t)$ 是布朗运动, 因此

$$Y(t) = Y(t_0) + (\mu - \frac{1}{2}\sigma^2)(t - t_0) + \sigma(W(t) - W(t_0)) \quad (49)$$

$$X(t) = \exp(Y(t)) \quad (50)$$

$$= X(t_0)\exp[(\mu - \frac{1}{2}\sigma^2)(t - t_0) + \sigma(W(t) - W(t_0))] \quad (51)$$

SDE 的数值解

并不是所有的 SDE 都能解出显式解, 更多的 SDE 只能通过迭代式求数值解, 求 SDE 数值解的过程也就是模拟出解的路径。

Euler 格式.

$$X_{i+1} = X_i + \mu(t_i, X_i)(t_{i+1} - t_i) + \sigma(t_i, X_i)(W_{i+1} - W_i)$$

□

Milstein 格式.

$$X_{i+1} = X_i + \mu(t_i, X_i)(t_{i+1} - t_i) + \sigma(t_i, X_i)(W_{i+1} - W_i) \quad (52)$$

$$\frac{1}{2} \sigma(t_i, X_i) \sigma_y(t_i, X_i) \{ (W_{i+1} - W_i)^2 - (t_{i+1} - t_i) \} \quad (53)$$

□

其中

$$W_{i+1} - W_i = \sqrt{t_{i+1} - t_i} Z_{i+1}, i = 0, 1, \dots, n-1$$

而 Z_1, \dots, Z_n 是相互独立的标准正态随机变量

SDE 的参数估计

► 极大似然估计

以布朗运动 $dX_t = \mu(X_t; \theta)dt + \sigma(X_t; \theta)dW_t$ 为例假设 x_0, \dots, x_N 是 $X(t)$ 在均匀离散时刻 $t_i = \Delta t$ 的观测, 其中 $i = 0, 1, \dots, N, \Delta t = T/N$.

令 $p(t_k, x_k | t_{k-1}, x_{k-1}; \theta)$ 是从 (t_{k-1}, x_{k-1}) 到 (t_k, x_k) 的传递概率密度, 假设初始状态的概率密度为 $p_0(x_0 | \theta)$, 似然函数为:

$$f(\theta) = p_0(x_0 | \theta) \prod_{k=1}^N p(t_k, x_k | t_{k-1}, x_{k-1}; \theta)$$

考虑 SDE 的 Euler 近似, 有

$$X(t_k) = x_{k-1} + \mu(t_{k-1}, x_{k-1}; \theta)\Delta t + \sigma(t_{k-1}, x_{k-1}; \theta)\sqrt{\Delta t}\eta_k$$

其中 $\eta_k \sim N(0, 1)$ 。因此

$$p(t_k, x_k | t_{k-1}, x_{k-1}; \theta) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(x_k - \mu_k)^2}{2\sigma_k^2}\right)$$

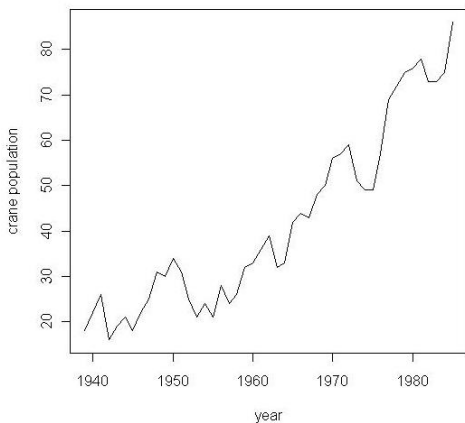
其中 $\mu_k = x_{k-1} + \mu(t_{k-1}, x_{k-1}; \theta)\Delta t, \sigma_k = \sigma(t_{k-1}, x_{k-1}; \theta)\sqrt{\Delta t}$

SDE 的参数估计

► 实例：种群动力学

考虑美洲鹤 1939-1985 年的种群数据，假设 t 时刻种群大小 $X(t)$ 满足 SDE:

$$dX(t) = \theta_1 X(t)dt + \theta_2 \sqrt{X(t)}dW(t), X(0) = 18$$



觅食算法实例

- ▶ 资源分配和调度
- ▶ 信号处理
- ▶ 预测控制
- ▶ 图形图像

资源分配和调度

- ▶ 对计算机中的多区域温度调度平台 (*Multizone Temperature Experimentation Platform*) 进行资源分配，以低温区域为“食物”，细菌的觅食过程就是将数据的获取与计算工作分配至低温区域以获得更高的能效
- ▶ 车间作业中利用觅食算法进行空闲时间片段的调度，将未完成的工序合理地分配到不同的空闲时间中。在工序顺序和机器占用的约束下，有效降低调度的规模和复杂性

独立组分分析 (ICA) 用来对统计信号进行处理，找到非高斯分布数据的线性表示。利用觅食算法与 ICA 的结合，可以提高信号处理的收敛速度，降低均方误差

预测控制

- ▶ 利用觅食算法对股票市场指数趋势进行预测
- ▶ 利用觅食算法对控制系统进行局部优化

利用细菌觅食算法，在图像识别中，提高识别的速度与准确率。通常会与基因算法等其他算法进行结合，在面部识别等领域提高特征点的搜索速度。