# 数学模型 Lecture Notes

Zhennan Zhou

2018 年 6 月 15 日

PRELIMINARY DRAFT. NOT FOR WIDE CIRCULATION.

## 8 随机模拟方法初步

As far as the laws of mathematics refer to reality, they are not certain, and as far as they are certain, they do not refer to reality.

——Albert Einstein

很多自然科学、社会科学现象很适合用含随机性的模型来刻画，很多时候，这些模型是以随机微分方程的形式出现。这些方程的一般形式如下，

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = b(t, X_t) + \sigma(t, X_t)\eta(t). \tag{8.1}$$

这里，$\eta(t)$ 代表某种“噪声”，使得对于特定的时间 $t$，$\eta(t)$ 是一个随机变量。

但是，在进一步了解随机微分方程之前，我们先学习一些随机模拟方法。这可能要回到一个看似简单的问题，如何用电脑按照给定的概率分布生成一个随机数。

### 8.1 蒙特卡洛方法

#### 8.1.1 随机数的生成

我们先考虑可能是一个最基本的任务，生成一个均匀分布的随机变量，即生成一个随机变量 $X \sim U([0,1])$. 当然，给定一个只会进行确定性运算的电脑，我们不能真正的得到一个真随机数。我们至多能期待可以生成一个“伪随机数”，是的虽然它不是真正的随机，但是“看起来随机”。常用的算法叫做线性同余生成器（linear congruential generator，LCG）。这是一个很有趣的话题，但是我们就不深入讨论了，有趣的同学欢迎自己查阅更多的资料。

有了均匀分布的生成器之后，下一个问题是，如何生成一个随机变量 $Y$ 使其（累积）分布函数为 $F$。

常用的算法是如下的逆变换法（或称，积分变换法）：

1. 生成 $X \sim U([0,1])$;
2. 令 $Y = F^{-1}(X)$.

则，生成的随机变量 $Y$ 对应的分布函数为 $F$。

注意到，分布函数 $F$ 是单调递增的，先假设 $F$ 是严格递增的，则所以我们有

$$\Pr(Y \le y) = \Pr(F^{-1}(X) \le y) = \Pr(X \le F(y)) = F(y). \tag{8.2}$$

如果 $F$ 不是严格递增的，我们只需要按如下的方式修改定义

$$F^{-1}(u) = \inf\{x \mid F(x) > u\}, \tag{8.3}$$

容易验证，此时算法仍然适用。

直观上看，这个算法是在纵轴上取样了一个服从均匀分布的随机变量，然后通过 $F$ 去寻找它的原象。因此，在 $F(y)$ 增长更快的地方，概率密度更高，而我们也更容易在这样的地方取样 $Y$. 当 $F^{-1}$ 容易计算的时候，逆变换法非常适用。可惜，能解析算出 $F^{-1}$ 的例子有限。

*Example 1.* 指数分布 Exponential distribution.

- 密度函数。 $p(x;\lambda) = \lambda\exp(-\lambda x), x \geq 0.$
- 分布函数。

$$F(x;\lambda) = \int_0^x \lambda\exp(-\lambda y)dy = 1 - \exp(-\lambda x)$$
$$F^{-1}(x;\lambda) = -\frac{1}{\lambda}\ln(1-x).$$

- 取样方法。 $Y = -\frac{1}{\lambda}\ln(1-X)$ 或者等价的， $Y = -\frac{1}{\lambda}\ln X$，其中 $X \sim U([0,1])$。

*Example 2.* 二维标准正态分布 Box-Muller Method.

- 密度函数。 $p(x,y)dxdy = \frac{1}{2\pi}\exp(-\frac{x^2+y^2}{2})dxdy.$
  在极坐标下， $p(r,\theta)rdrd\theta = \frac{1}{2\pi}\exp(-\frac{r^2}{2})rdrd\theta = \frac{d\theta}{2\pi} \times \exp\left(-\frac{r^2}{2}\right)d\left(\frac{r^2}{2}\right).$
  所以在极坐标下，$\theta$ 和 $r^2/2$独立。
- 取样方法。因为，$\theta$服从 0 到 $2\pi$ 上的均匀分布，$\frac{1}{2}r^2$ 服从指数分布。转换回直角坐标系得

$$\begin{cases} x = \sqrt{-2\log(\xi_1)}\cos(2\pi\xi_2) \\ y = \sqrt{-2\log(\xi_1)}\sin(2\pi\xi_2) \end{cases}$$
$$\xi_1 \sim U(0,1), \quad \xi_2 \sim U(0,1).$$

对于更一般的连续随机变量，我们介绍另一种方法，von Neumann 取舍原则（rejection sampling，acceptance-rejection method）。

- 若密度为 $p_0(x)$ 的随机变量容易生成，现在要生成密度为 $p(x)$ 的随机变量，而且 $p(x) \leq cp_0(x)$。
- 生成密度为 $p_0(x)$ 的随机变量列 $\{\xi_1, \xi_2, \cdots, \xi_n\}$。
- 对 $\{\xi_k\}$ 进行随机筛选，留下的随机变量列就是密度为 $p(x)$ 的随机变量列。
- 筛选方法：独立生成 $U_1, \cdots, U_n \sim U([0,1])$，如果 $U_k \geq \frac{p(\xi_k)}{cp_0(\xi_k)}$，就删除 $\xi_k$，否则保留。
- 注意，留下的随机变量仍然是相互独立同分布的，它们中的每一个 ( 记为 $\eta$ ) 的分布是

$$\begin{aligned} \Pr(\eta < x) &= \Pr(\xi_k < x \mid \xi_k \text{ kept}) \\ &= \Pr\left(\xi_k < x \mid U_k \leq \frac{p(\xi_k)}{cp_0(\xi_k)}\right) \\ &= \frac{\Pr\left(\xi_k < x, U_k \leq \frac{p(\xi_k)}{cp_0(\xi_k)}\right)}{\Pr\left(U_k \leq \frac{p(\xi_k)}{cp_0(\xi_k)}\right)} \\ &= \frac{\int_{-\infty}^{x}\frac{p(y)}{cp_0(y)}p_0(y)dy}{\int_{-\infty}^{+\infty}\frac{p(y)}{cp_0(y)}p_0(y)dy} = \int_{-\infty}^{x} p(y)dy = F(x). \end{aligned}$$

### 8.1.2 蒙特卡罗积分

蒙特卡罗积分是一种常用于高维情况下近似计算定积分的技巧。为了便于说明，我们先考虑一个如下的一维定积分

$$I(f) = \int_0^1 f(x)\,\mathrm{d}x. \tag{8.4}$$

不同于基于格点或者节点的确定性算法（大家可能会在应数导论、或者数值分析中接触到），蒙特卡罗积分通过选取区间里的一些随机点来近似 $I(f)$。我们令 $X_1, \cdots, X_N$ 为独立且同分布的随机变量（independent and identically distributed random variables，常被简写为 iid R.V.），服从 $U([0,1])$。然后，我们做如下近似

$$I(f) \approx I_N(f) = \frac{1}{N}\sum_{i=1}^{N} f(X_i). \tag{8.5}$$

显然这是一个无偏估计。根据大数定律，我们知道

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} f(X_i) = \mathbb{E} f(X) = I(f). \tag{8.6}$$

所以，当 $N$ 很大的时候，$I_N(f)$ 是对 $I(f)$ 的一个很好的估计。为了得到收敛速度，我们计算误差的方差，

$$\mathbb{E}\big(I_N(f) - I(f)\big)^2 = \frac{1}{N^2}\sum_{i,j=1}^{N} \mathbb{E}\big(f(X_i) - I(f)\big)\big(f(X_j) - I(f)\big) = \frac{1}{N}\mathrm{Var}(f), \tag{8.7}$$

这里，

$$\mathrm{Var}(f) = \int_0^1 f^2 dx - \left(\int_0^1 f dx\right)^2.$$

所以，如果我们用误差的标准差来当作误差的估值的话，我们发现误差的阶数为 $O(N^{-1/2})$:

$$\mathrm{Error} \approx \frac{\sqrt{\mathrm{Var}\, f}}{\sqrt{N}}.$$

这的误差其实比很多的确定性算法要大，例如，数值积分法中的梯形公式，误差的阶数为$O(N^{-2})$。

然而，蒙特卡罗积分的优势在高维的时候是明显的。对于一个 $d$ 维的积分，如果我们采用确定性的数值积分法，假设总共有 $N$ 个节点，那么每一维中，节点之间的距离约为 $N^{-1/d}$，则一个二阶的方法所产生的误差约为（$C$为一个常数）

$$\mathrm{Error} \approx C(\Delta x)^2 = \frac{C}{N^{2/d}}.$$

另外一方面，对蒙特卡罗积分做误差估计的时候，所有的推导都与维数没有关系，因此，当维数很高的时候，蒙特卡罗积分是收敛更快的。

我们再回顾一下蒙特卡罗积分的误差估计。虽然误差中的 $N^{-1/2}$ 因子是误差改变的，误差的大小也取决于 $f(x)$ 的方差，而 $f(x)$ 是我们估计 $I(f)$ 的估计量，或称估计子。所以，如果我们能减少估计子的方差，我们就能得到更高效的积分公式。这其实是蒙特卡罗方法的核心问题之一。

重要性采样（importance sampling，也叫g-样本方法）就是这么一种积分法。它的基本思想是，不再在积分区域内均匀的取样，而是在被积函数的值比较大的地方取更多的点。具体来说，对于积分$\int_D f(x)\,\mathrm{d}x$，重要性采样的方法是，

- 选取 iid 的 $\{X_i\}$ 服从 $D$ 中的概率密度函数 $g(x)$;
- 将积分近似为

$$I_n^{(g)} = \frac{1}{N}\sum_{i=1}^{N} \frac{f(X_i)}{g(X_i)}. \tag{8.8}$$

这个算法的数学原理是显然的，(不妨假设 $g$ 处处大于 0)

$$\int f(x)\,\mathrm{d}x = \int \frac{f(x)}{g(x)} g(x)\,\mathrm{d}x = \mathbb{E}^g\Big(\frac{f}{g}\Big) = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \frac{f(X_i)}{g(X_i)}. \tag{8.9}$$

我们还可以进一步的计算误差的方差，

$$\mathbb{E}\big(I_N^{(g)} - I\big)^2 = \frac{1}{N}\mathrm{Var}^g\Big(\frac{f}{g}\Big) = \frac{1}{N}\Big(\int \frac{f^2}{g}\,\mathrm{d}x - \Big(\int f\,\mathrm{d}x\Big)^2\Big). \tag{8.10}$$

我们观察到，假设 $f(x)$ 在 $D$ 上恒正，如果我们可以取$g(x) = f(x)/Z$，$Z = \int_D f\,dx$，那我们上面的方差其实是 0。也就是说，其实没有一点的误差！但是呢，这必然是骗人的，因为 $Z$ 正是我们需要找的量。但是，这个观察至少给我们如下的启示，应该尽可能的贴近 $f(x)$ 来选取 $g(x)$。

## 8.2 马尔科夫蒙特卡洛（MCMC）法

我们一些学习了一些生成随机变量的方法，但是当问题维数特别高的时候，这些方法也不再那么有用了。例如，我们考虑一个由 $n$ 个粒子构成的系统，他们位置分别为 $x_1,\cdots,x_n \in \mathbb{R}^3$，并且有两两的相互作用，

$$U(x) = \sum_{i<j} V(|x_i - x_j|). \tag{8.11}$$

在统计力学里面，（经过一些合理假设）系统的稳态分布是如下的 Boltzmann 分布

$$\pi(x) = Z^{-1} e^{-\beta U(x)}, \tag{8.12}$$

这里，$x \in \mathbb{R}^{3n}$, $\beta = (k_B T)^{-1}$, $k_B$ 是 Boltzmann constant，$T$ 是绝对温度，而 $Z = \int e^{-\beta U(x)}\,\mathrm{d}x$ 是一个未知的归一化常数。如果需要了解这个系统的各种性质，我们就需要对这个 $3n$ 维的 Boltzmann 分布进行取样。

为了取样这么一个分布，有一类方法是，构造一个马尔科夫链，使得目标分布 $\pi$ 是这个马尔科夫链的不变分布。那么，我们可以通过模拟这个马尔科夫链的方式来对这个分布 $\pi$ 进行取样。因而，这类方法被称为马尔科夫蒙特卡洛（Markov Chain Monte Carlo，简称MCMC）法。由于这种方法的问世, 使随机模拟在很多领域的计算中, 相对于确定性算法，显示出它的巨大的优越性。而有时随机模拟与确定性算法的结合使用, 会显出更多的长处。常用的马尔科夫链的构造方法有 Metropolis-Hastings algorithm，Gibbs sampler 等方法，本课程中我们只讨论 Metropolis-Hastings（MH） algorithm。

我们先考虑离散的情况。假设我们有 $N$ 个节点，对应的不变分布为 $\pi$。MH 算法给出了如下的节点之间的取样轨道：

- 假设当前在状态 i，从一个“预选矩阵”中选择状态 j，使得

$$h_{ij} = \Pr\big(\text{propose state } j \mid \text{current state } i\big); \tag{8.13}$$

- 我们以 $a_{ij}$ 的概率接受预选（之后会给出 $a_{ij}$ 具体形式），从状态 i 转移到状态 j；否则，留在状态 i。

注意，第二步可以通过生成一个均匀分布的随机数来实现（类似于von Neumann 取舍原则）。

显然，这个诱导出来的马尔科夫链的有如下的转移概率

$$p_{ji} = \begin{cases} h_{ji}a_{ji}, & i \neq j; \\ 1 - \sum_{i\neq j} h_{ji}a_{ji}, & i = j. \end{cases} \tag{8.14}$$

注意到，这个马尔科夫链的不变测度满足

$$\pi_j = \sum_i p_{ij}\pi_i. \tag{8.15}$$

一般我们通过选择 $a$ 使得它满足如下的 detailed balance

$$\pi_i p_{ij} = \pi_i h_{ij} a_{ij} = \pi_j h_{ji} a_{ji} = \pi_j p_{ji}, \tag{8.16}$$

进而

$$\pi_j = \sum_i \pi_j p_{ji} = \sum_i \pi_i p_{ij}. \tag{8.17}$$

常用的选择是（Hastings algorithm）

$$a_{ij} = \min\Big(1, \frac{\pi_j h_{ji}}{\pi_i h_{ij}}\Big). \tag{8.18}$$

如果预选矩阵是对称的（Metropolis algorithm），则我们有 $a_{ij} = \min\big(1, \frac{\pi_j}{\pi_i}\big)$. 我们观察到，算法中，我们只需要差个常数倍的知道 $\pi$。因此，对于 Boltzmann分布来说，我们不要知道归一化常数 $Z$ 就可以使用 MH算法。事实上，在应用中归一化常数 $Z$ 往往是不知道的。

MH 算法在连续密度函数上的推广是直接的。但是选取一个好的预算概率（矩阵）是一门艺术。我们一方面希望能“跳”得远些以加速空间的探索，另外一方面，我们又希望接受的比例不能太低。

我们用 $\{x_k\}$ 来表示一列由 MCMC 产生的状态，我们可以根据马尔科夫链的遍历性质来估计这个均值

$$\mathbb{E}_\pi f(X) \approx \frac{1}{N}\sum_{k=1}^{N} f(x_k). \tag{8.19}$$

那么，如何分析 MCMC 算法的误差呢？假设样本 $\{x_k\}$ 是 iid $\sim \pi$，那么类似于对蒙特卡罗积分的分析，我们知道误差的估值为 $\sqrt{\mathrm{Var}\, f}/\sqrt{N}$。可是，显然，由于样本是依此生成的，所有样本并不独立！为了考虑进去样本之间的相关性，我们引入 effective sample size （ESS）

$$M = \frac{N}{1 + 2\sum_{k=1}^{\infty}\rho_k}, \tag{8.20}$$

这里，$\rho_k$是自相关函数（Autocorrelation function）

$$\rho_k = \frac{\mathrm{Cov}_\pi(f(x_i), f(x_{i+k}))}{\mathrm{Var}_\pi f(x_i)}. \tag{8.21}$$

显然，当这使得误差的估值被修正为 $\sqrt{\mathrm{Var}\, f}/\sqrt{M}$。

## 8.3 随机分析初步

我们回顾，有“噪声”的常微分方程模型可以写成如下的形式：

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = b(t, X_t) + \sigma(t, X_t)\eta(t).$$

数学上，我们常对噪声做如下的假设

1. $\mathbb{E}\eta(t) = 0$;
2. $\eta(t)$ 是平稳的(这个严格定义超出了本课的范围);
3. $\mathbb{E}\eta(t)\eta(s) = 0$ if $t \neq s$.

这样的噪声被称为**白噪声**。事实上，可以证明，满足这样的$\eta(t)$是不会有连续的路径的。为了理解这类方程，我们来考虑一个离散的形式：在时间点$t_0 < t_1 \cdots$，我们令 $X_k = X(t_k)$:

$$X_{k+1} - X_k = b(t_k, X_k)\Delta t_k + \sigma(t_k, X_k)\int_{t_k}^{t_{k+1}} \eta(s)\,\mathrm{d}s. \tag{8.22}$$

我们不妨引入$V_t = V(t)$，使得

$$\int_{t_k}^{t_{k+1}} \eta(s)\,\mathrm{d}s = V(t_{k+1}) - V(t_k) =: \Delta V_k. \tag{8.23}$$

根据$\eta(t)$的性质，我们知道

1. $\mathbb{E}V_t = 0$;
2. 平稳的增长;
3. 独立的增长.

而且，因为我们希望 $X_t$ 至少是连续的，所以我们要求 $V_t$ 有连续的路径。事实上，唯一可能的满足上述条件的过程是布朗运动（Brownian motion, or Wiener process），而且，他还满足

- $\mathbb{E}W_t W_s = \min(s, t)$;
- for $t > s$, $W_t - W_s \sim N(0, t-s)$.

于是，我们可设，

$$X_k = X_0 + \sum_{j=0}^{k-1} b(t_j, X_j)\Delta t_j + \sum_{j=0}^{k-1} \sigma(t_j, X_j)\Delta W_j, \tag{8.24}$$

然后考虑 $\Delta t_j \to 0$ 的极限。假设极限存在，则我们可写成

$$X_t = X_0 + \int_0^t b(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}W_s, \tag{8.25}$$

或者他的微分形式

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t. \tag{8.26}$$

注意，布朗运动 $W_t$ 其实是关于时间不可微的，因此，我们常避免写 $\frac{\mathrm{d}W_t}{\mathrm{d}t}$。但是，事实上，一些人还是会这么写……

我们需要理解在什么意义下，对于哪些随机过程 $f(s,\omega)$（这里，$\omega$ 是概率空间的参数），$\int_0^t f(s,\omega)\,\mathrm{d}W_s$ 是存在的。因此，我们先考虑如下的积分

$$I = \int_0^t W_s\,\mathrm{d}W_s. \tag{8.27}$$

让我们尝试在 $[k\Delta t, (k+1)\Delta t)$ 等区间上用黎曼积分计算这个积分。我们考虑两种特殊的选择

A. 被积函数在区间的左端点取值

$$I^{(n)} = \sum_{k=0}^{n-1} W_{k\Delta t}\left(W_{(k+1)\Delta t} - W_{k\Delta t}\right), \tag{8.28}$$

于是，我们计算得到

$$\mathbb{E}I^{(n)} = \sum k\Delta t - k\Delta t = 0. \tag{8.29}$$

B. 被积函数在区间的中点取值

$$I^{(n)} = \sum_{k=0}^{n-1} W_{(k+\frac{1}{2})\Delta t}\big(W_{(k+1)\Delta t} - W_{k\Delta t}\big), \tag{8.30}$$

于是我们计算得到

$$\mathbb{E}I^{(n)} = \sum (k+\frac{1}{2})\Delta t - k\Delta t = \frac{t}{2}. \tag{8.31}$$

可见，在不同的逼近选择下，我们得到了不同的极限。所以，黎曼积分是并不存在的。为了解决这个问题，我们需要先决定被积函数在每个区间的逼近方式。这里，有两个常见的选择：

**定义 1** *Itô integral* 是黎曼和中被积函数在左端点取值的极限，记为

$$\int_0^t f(s,\omega)\,\mathrm{d}W_s.$$

**定义 2** *Stratonovich integral* 是黎曼和中被积函数利用梯形公式取值的极限，即

$$\sum_j \frac{1}{2}\big(f(t_j,\omega) + f(t_{j+1},\omega)\big)\Delta W_j.$$

记为

$$\int_0^t f(s,\omega) \circ \mathrm{d}W_s.$$

在下文中，我们只考虑 Itô integral。这里，我们列举 Itô Calculus 的几个常用的性质

- 对于 Itô integral，我们有如下的一般性结果

$$\mathbb{E}\int_0^t f(s,\omega)\,\mathrm{d}W_s = 0. \tag{8.32}$$

- 对于 Itô integral，我们有如下的 Itô isometry

$$\mathbb{E}\Big(\int_0^t f(s,\omega)\,\mathrm{d}W_s\Big)^2 = \mathbb{E}\int_0^t f(s,\omega)^2\,\mathrm{d}s. \tag{8.33}$$

- Itô formula ( Itô lemma):

  Let $X_t$ be the solution to

$$\mathrm{d}X_t = b(t,\omega)\,\mathrm{d}t + \sigma(t,\omega)\,\mathrm{d}W_t$$

  where $b$, $\sigma$ are functions of $(t,\omega)$. Given $g(t,x) \in C^2([0,\infty)\times\mathbb{R})$. Then $Y_t = g(t,X_t)$ satisfies the equation

$$\mathrm{d}Y_t = \frac{\partial g}{\partial t}(t,X_t)\,\mathrm{d}t + \frac{\partial g}{\partial x}(t,X_t)\,\mathrm{d}X_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t,X_t)(\mathrm{d}X_t)^2, \tag{8.34}$$

  where $(\mathrm{d}X_t)^2$ is computed according to the rules

$$\mathrm{d}t\cdot\mathrm{d}t = \mathrm{d}t\cdot\mathrm{d}W_t = \mathrm{d}W_t\cdot\mathrm{d}t = 0, \qquad \mathrm{d}W_t\cdot\mathrm{d}W_t = \mathrm{d}t.$$

**例子1.** 计算 $\mathrm{d}(\frac{1}{2}W_t^2)$。令 $g(t,x) = \frac{1}{2}x^2$，则

$$\mathrm{d}Y_t = W_t\,\mathrm{d}W_t + \frac{1}{2}(\mathrm{d}W_t)^2 = W_t\,\mathrm{d}W_t + \frac{1}{2}\mathrm{d}t.$$

**例子2. Ornstein-Uhlenbeck process.** 考虑方程

$$\mathrm{d}X_t = -aX_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \quad X_0 = \xi. \tag{8.35}$$

注意到，右端的第一项代表指数型衰减，而第二项表示“波动”变化。

我们来利用积分因子法求解此问题。方程两边乘以 $e^{at}$，我们得到

$$\mathrm{d}\big(e^{at}X_t\big) = e^{at}\,\mathrm{d}X_t + ae^{at}X_t\,\mathrm{d}t = \sigma e^{at}\,\mathrm{d}W_t, \tag{8.36}$$

于是，积分得到

$$e^{at}X_t - X_0 = \int_0^t \sigma e^{as}\,\mathrm{d}W_s. \tag{8.37}$$

所以，方程的解是

$$X_t = e^{-at}X_0 + \sigma\int_0^t e^{-a(t-s)}\,\mathrm{d}W_s. \tag{8.38}$$

我们注意到，方程对初值的“记忆”会指数型衰减。并且，我们做如下的计算

- Mean. $\mathbb{E}X_t = e^{-at}\mathbb{E}X_0$ 指数型衰减到 0。
- Covariance. (假设初值和 $W_t$ 无关)

$$\begin{aligned}\mathbb{E}X_tX_s - \mathbb{E}X_t\mathbb{E}X_s &= e^{-at}e^{-as}\big(\mathbb{E}X_0^2 - (\mathbb{E}X_0)^2\big) + \sigma^2\mathbb{E}\left[\int_0^s\int_0^t e^{-a(t-v)}e^{-a(s-u)}\,\mathrm{d}W_u\,\mathrm{d}W_v\right]\\ &= e^{-a(s+t)}\operatorname{Var}X_0 + \sigma^2\int_0^{s\wedge t} e^{-a(s-u)}e^{-a(t-u)}\,\mathrm{d}u\\ &= e^{-a(s+t)}\operatorname{Var}X_0 + \frac{\sigma^2}{2a}\big(e^{-a|s-t|} - e^{-a(s+t)}\big).\end{aligned}$$

特别的，我们注意到 $\lim_{t\to\infty}\operatorname{Var}X_t = \frac{\sigma^2}{2a}$.

**例子3. Geometric Brownian motion.** 考虑方程

$$\mathrm{d}N_t = rN_t\,\mathrm{d}t + \alpha N_t\,\mathrm{d}W_t. \tag{8.39}$$

这是另外一个人口增长模型，这里 $r$ 是（相对）增长系数，$\alpha$ 是波动系数。这个例子也跟金融数学里面的Black-Scholes model 有关。这个意义下， $N_t$ 可以理解称资产的定价，$\alpha$ 可以理解为波动性。

为了求解此方程，我们在两端除以$N_t$，便得到

$$\frac{\mathrm{d}N_t}{N_t} = r\,\mathrm{d}t + \alpha\,\mathrm{d}W_t. \tag{8.40}$$

注意，利用 Itô formula，我们有

$$\mathrm{d}(\log N_t) = \frac{1}{N_t}\,\mathrm{d}N_t - \frac{1}{2N_t^2}(\mathrm{d}N_t)^2 = \frac{1}{N_t}\,\mathrm{d}N_t - \frac{\alpha^2}{2}\,\mathrm{d}t. \tag{8.41}$$

于是我们得到，

$$\mathrm{d}(\log N_t) = \big(r - \frac{\alpha^2}{2}\big)\,\mathrm{d}t + \alpha\,\mathrm{d}W_t, \tag{8.42}$$

容易得出方程的解

$$N_t = N_0e^{(r-\frac{\alpha^2}{2})t + \alpha W_t}. \tag{8.43}$$

接下来，我们来计算一下解的均值（期望）。注意到，问题的关键在于求 $Y_t := e^{\alpha W_t}$ 的均值。利用 Itô formula，我们有

$$\mathrm{d}Y_t = \alpha e^{\alpha W_t}\,\mathrm{d}W_t + \frac{\alpha^2}{2}e^{\alpha W_t}\,\mathrm{d}t = \alpha Y_t\,\mathrm{d}W_t + \frac{\alpha^2}{2}Y_t\,\mathrm{d}t, \tag{8.44}$$

两边积分，并求期望，我们得到

$$\mathbb{E}Y_t = \mathbb{E}Y_0 + \frac{\alpha^2}{2}\int_0^t \mathbb{E}Y_s\,\mathrm{d}s, \tag{8.45}$$

于是有

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}Y_t = \frac{\alpha^2}{2}\mathbb{E}Y_t. \tag{8.46}$$

显然，$\mathbb{E}Y_0 = 1$，于是我们有$\mathbb{E}Y_t = e^{\frac{\alpha^2}{2}t}$。

最终，我们得到，

$$\mathbb{E}N_t = (\mathbb{E}N_0)e^{rt},$$

也就是说，人口数量期望的增长和无随机模型的人口增长是一样的。

本学期的课程内容全部结束了！但是数学模型的学习道路是没有终点的……