

蒙特卡罗法正式出现在 20 世纪 40 年代,但实际上其起源可上溯到 17 世纪。蒙特卡罗法的发展大体上经历了三个阶段。

蒙特卡罗法的第一阶段主要发展直接仿真的方法,也就是对具有随机性质的问题进行直接仿真,没有形成一种普遍的和有效的解决实际问题的方法。这段时间主要的成果是发现和论证随机变量的统计性质和对随机变量进行定量分析的理论与方法。

蒙特卡罗法的第二阶段主要发展了数学仿真方法,是蒙特卡罗法真正确立的阶段。数学仿真方法就是针对具体的随机性质问题,把对随机现象进行直接仿真改为用数学方法进行仿真求得解答。例如,18 世纪用随机投针法计算 π 值,1889 年研究了随机游动问题,1908 年著名统计学家 Gossett 推导了 t 分布和一种 χ^2 分布,1928 年 Courant、Friedrichs 和 Lewy 讨论了二维和三维布朗运动。这种方法显然比直接仿真法有很多优越性。

蒙特卡罗法的第三阶段主要发展了拟仿真方法,该方法使蒙特卡罗法得到迅速发展和广泛应用。为了更加有效地解决实际问题,可以人为地将原来的随机现象变为另一种与原问题答案有一定关系的人为的“随机现象”,然后利用人为随机现象与原问题答案间的关系,再利用计算机仿真方法得到原问题的解答,拟仿真方法的关键就是建立替代的人为随机现象。拟仿真方法使蒙特卡罗法进入了迅速发展的阶段,可以说,没有拟仿真方法就没有蒙特卡罗法的应用和发展。

现在,蒙特卡罗法仍然在进一步的发展中,特别是随着非线性物理学的发展、混沌现象、生命科学和经济学的深入研究,蒙特卡罗法已经成为必备的定量分析研究的工具和仿真模拟的理论基础。蒙特卡罗法学习的意义已经大大超出计算电磁学的领域。

3.1 蒙特卡罗法的基本原理

在自然界有相当多的物理量是具有一定程度不确定性的量。实际上,任何物理现象都是在不断变动中的,环境因素也在不断地变化,就必然具有随机性。随机的物理量可以用概率论和随机过程的理论研究和表达。但是,有些物理量,例如分形、混沌等非线性过程的参量,根本无法用解析式表达而必须用计算机经不断的迭代和产生随机参量才能获得解答。目前,处理这类问题最有效的方法就是计算机模拟,采用的工具就是蒙特卡罗法。从工程设计的角度,如果不考虑物理量的不确定性就不会得到成功的、可靠的工程解决方案。

从科学的角度看,随机现象总是伴随着确定的物理规律,当改换角度看同一物理现象时,

确定的物理现象的背景很可能是随机现象。例如,加热封闭空间中的气体,温度和气压满足热力学定律是确定不变的,但是从分子运动的角度理解加热封闭空间中的气体现象时,我们看到的是大量的随机运动的分子,热力学规律不过是大量分子随机运动的统计结果。从自然哲学的角度看,确定性与随机性是互为依存的同一运动的两方面,只看运动的确定性或者说只看运动的随机性都是不完全的,都曲解了运动的本来面目。因此,本章的内容除了要介绍蒙特卡罗法的基本原理和基本处理方法,还希望读者能够牢固地建立起基于随机性的理解和分析各种现象的能力。如果读者能够主动地站在随机性的角度观察和理解世界,将能够看到同以往完全不同的非常奇妙的图景,将能够发掘出随机现象中的确定性和规律性,将能够解释以前认为是不可理解的、匪夷所思的现象。

在具体介绍蒙特卡罗法之前,笔者首先要强调,进行计算机模拟并不是最终的目的,计算机模拟只是工具,一定要解决实际提出的工程和理论问题。因此,计算机模拟完成之后,还必须回到原来提出的问题去,要对计算机模拟的结果进行分析研究并回到现实世界去,给出对现实问题和现象的物理解释,否则其他人无法使用所得出的数值分析结果。这个过程的难度和重要意义不亚于计算机模拟的全部工作,中国有句古语:行百里者半九十,说的就是这种情形,最后的十里没有完成或没有走对,前面的九十里就毫无意义。在学习蒙特卡罗法时,这一点尤其重要,拟仿真方法实际上是采用已有的随机模型的样板解决实际的问题,得到的解答就必须回到原来的问题对得到的结果作出合理的物理解释。

3.1.1 蒙特卡罗法的基本过程

随机变量的全体称为总体,随机变量总数称为容量,容量可以是无限也可以是有限的。当某个随机变量与存在的另一个随机变量之间有一定的函数关系时,可以定义后一个随机变量为统计量。但随机变量本身并无确定的规律可循,又不可能每次都把所有的随机变量进行研究。因此,为了研究随机变量之间的函数关系,为了在有限范围内研究在无限统计分析时才具有的规律性,就必须首先构成样本空间。

在随机变量总体中抽出一定数量的个体来进行观测的过程称为抽样。抽样要采用科学的方法使抽出来的个体能反映总体的统计或随机性质,反映以总体为单位时参与物理变化的情况。例如,一棵树可以是一片森林的抽样,但是树叶就不是森林的科学抽样,一片森林对调节地球或区域的生态有作用但是却无法讨论一棵树对生态的作用。

设随机变量总体为 Ω ,在保证条件无任何变化的情况下对随机变量 Ω 进行 n 次重复独立观测,得到的结果称为 n 次简单随机抽样, n 次抽样所得的结果记为 $\zeta=(\xi_1, \xi_2, \dots, \xi_n)$,称为来自总体 Ω 的随机样本,抽样次数 n 称为样本容量。

样本应该尽可能多地反映出随机变量总体的特性,即每个 ζ_i 都具有随机变量总体的特征,也就是 ζ_i 与 Ω 具有相同的随机分布。此外,每次抽样都要保证是独立进行的,互不影响,即 $(\xi_1, \xi_2, \dots, \xi_n)=\zeta_i$ 是相互独立的随机变量。

定义 3.1.1 如果随机变量 $(\xi_1, \xi_2, \dots, \xi_n)=\zeta_i$ 相互独立,并且每一个 ζ 与总体 Ω 有相同的概率分布,则随机变量 $(\xi_1, \xi_2, \dots, \xi_n)=\zeta_i$ 称为总体 Ω 的容量为 n 的样本。

抽样之后的量 ζ_i ,仍是独立的随机变量,但 $\xi_1, \xi_2, \dots, \xi_n$ 则具有确切的数值。此时, ζ_i 可以代表 Ω 参加函数运算,并且由此定义出一类由随机变量的样本产生的随机变量,即统计量。

定义 3.1.2 设 $\zeta_i=(\xi_1, \xi_2, \dots, \xi_n)$ 是总体 Ω 的一个样本, $G(\xi_1, \xi_2, \dots, \xi_n)$ 是样本 $\zeta_i=(\xi_1, \xi_2, \dots, \xi_n)$ 的一个函数,则称 $G(\xi_1, \xi_2, \dots, \xi_n)$ 为一个统计量。因为 G 是随机变量 $\xi_1,$

ξ_2, \dots, ξ_n 的函数, 所以统计量也是一个随机变量, 也有自己的概率分布, 称为抽样分布。求抽样分布是数理统计的基本问题之一。

设 $X = (x_1, x_2, \dots, x_n)$ 是来自该统计结果的一个样本, 样本的均值及方差与随机变量的均值及方差有如下的关系。

设 (x_1, x_2, \dots, x_n) 是来自正态分布 $N(\mu, \sigma^2)$ 的一个样本, μ 和 σ^2 是未知的, 现在要用样本来求取 μ 和 σ^2 的近似值。

定义样本均值为

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

样本方差为

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

设 $E(\bar{x}) = \mu$, 即样本均值 \bar{x} 是原随机变量 μ 的无偏估值。由于样本方差的均值为

$$E(S_n^2) = \frac{\sigma^2}{n} E\left(\frac{nS_n^2}{\sigma^2}\right) = \frac{n-1}{n} \sigma^2 \quad (3.1.1)$$

所以, S_n^2 并不是原随机变量方差 σ^2 的无偏估值。但是对 S_n^2 略加修正即可写为

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3.1.2)$$

S^2 就是原随机变量方差 σ^2 的无偏估值。当 n 趋于无穷大时 S_n^2 趋于 S^2 , 因此称 S_n^2 称为 σ^2 的渐进无偏估计。通常使用时, 由于样本容量不够大, 使用 S_n^2 会使得过小估计 σ^2 , 因此实际中采用 S^2 的比较多。

蒙特卡罗法应用时首先要构成一个概率空间, 然后在该概率空间中确定一个依赖随机变量 X , 其分布函数为 $F(x)$ 的统计量 $g(x)$, 使得 $g(x)$ 的数学期望

$$E(g) = \int g(x) dF(x) \quad (3.1.3)$$

正好等于所要求取的值 G , 式(3.1.3)中 $F(x)$ 是 x 的分布函数。最后, 产生随机变量 x 的简单子样 x_1, x_2, \dots, x_N , 取其相应的统计量 $g(x_1), g(x_2), \dots, g(x_N)$ 的算术平均值 \hat{G}_N 为统计量 $g(x)$ 的无偏估值, 作为 $G(\xi_i)$ 的近似估计。令

$$\hat{G}_N = \frac{1}{N} \sum_{n=1}^N g(x_n) \quad (3.1.4)$$

若用 $\hat{G}_N(X)$ 表示对 $G(\xi_i)$ 的估计量, 则 $E_\xi(\hat{G}_N(X)) = G(X)$ 称为 $G(\xi_i)$ 的无偏估计。实际上不可能进行无穷次实验获得渐进无偏估计 $G(X)$, 因此一般使用渐进无偏估计作为 $G(X)$ 的近似估计。

显然, 蒙特卡罗法应用的关键是要人为地确定一个统计量, 使其数学期望正好等于所要求的值, 称这样的统计量为无偏估计量。

蒙特卡罗法的最低要求: 能确定一个与计算步数 N 有关的无偏估计量 \hat{G}_N 使得当 $N \rightarrow \infty$ 时, $\hat{G}_N(X)$ 依概率收敛于所要求的值 $G(X)$, 即对任意小的 $\epsilon > 0$, 有

$$\lim_{N \rightarrow \infty} P(|\hat{G}_N - G| < \epsilon) = 1 \quad (3.1.5)$$

式中, $P(*)$ 表示事件 $*$ 的概率。

蒙特卡罗法可以用来解决两类问题。

第一类是确定性问题,例如,计算重积分、求逆矩阵、解线性代数方程组、解积分方程、解偏微分方程的边值问题、计算微分算子特征值等。

第二类是随机性问题,例如,中子扩散问题、运筹学的库存问题、随机服务系统中的排队问题、信息和传输的算法、随机游动问题、动物生态竞争问题、传染病蔓延问题等。

解第一类问题时要假设一个概率事件作为拟仿真建模,使概率事件的统计结果就是要求的值。首先,构成一个概率空间,也就是建立一个相应于所要求解问题的概率模型,使所要求的解就是建立概率模型的某一随机变量的数学期望值。然后,在该概率空间中确定一个随机变量 x 和这个随机变量的统计量 $g(x)$,并对 $g(x)$ 进行抽样和试验。通常在计算机上试验时,要按 x 的分布函数 $F(x)$ 抽取随机变量 x 的简单子样 x_1, x_2, \dots, x_N 。最后,这个统计量的数学期望值 $E(g)$ 的近似值或估计量 $g(x_1), g(x_2), \dots, g(x_N)$ 的算术平均值

$$\hat{G}_N = \frac{1}{n} \sum_{n=1}^N g(x_n) \quad (3.1.6)$$

就是要求的解 G 。

求解第二类问题时,通常没有或很难求得确定的表达式,一般都采用直接模拟的方法,即根据实际问题的随机性用计算机进行抽样,再对试验得到的随机变量值进行统计平均,求得所要的结果。

从以上两种问题可看出,蒙特卡罗法最关键的一步是确定一个概率空间中的统计量,即使相应的无偏估计量的数学期望恰好等于所求的值。

3.1.2 蒙特卡罗法的基本问题

使用蒙特卡罗法一般要讨论如下一些基本问题,这些是不同于其他计算方法之处。

1. 蒙特卡罗法的收敛性

蒙特卡罗法尽管可以进行相当多次的抽样,但是总不能达到无穷多次,因此通常用 G 的近似估计值 \hat{G}_N 代替数学期望值 G 。这就有个收敛的问题,因此蒙特卡罗法的最基本要求是能确定这样一个与计算步数 N 有关的统计估计值 \hat{G}_N ,使当 $N \rightarrow \infty$ 时, $\hat{G}_N \rightarrow$ 期望值 G 。

蒙特卡罗法的近似估计值 \hat{G}_N 依概率 1 收敛于 G ,即

$$P(\lim_{n \rightarrow \infty} \hat{G}_N = G) = 1 \quad (3.1.7)$$

要满足这一点,无偏估计量 $g(x)$ 的均值应满足

$$E(|g|) = \int |g(x)| dF(x) < +\infty \quad (3.1.8)$$

的条件,也就是要求这个无偏估计量的绝对数学期望值存在,式中 $F(x)$ 为 x 的分布函数。

收敛速度:如果无偏估计量满足条件 $E(|g|^r) = \int |g(x)|^r dF(x) < +\infty$,其中 $1 \leq r \leq 2$,则有

$$P(\lim_{N \rightarrow \infty} N^{\frac{r-1}{r}} (\hat{G}_N - G) = 0) = 1 \quad (3.1.9)$$

\hat{G}_N 依概率 1 收敛于 G 的速度为 $N^{\frac{1-r}{r}}$,因为使分式 $(1-r)/r$ 取得最大值的 r 为 2,所以收敛速度不会超过 $N^{-0.5}$ 。

2. 蒙特卡罗法的误差

根据中心极限定理,只要所确定的无偏估计量具有有限的不等于零的方差 σ^2 ,则对于任意非负的 X 值均有

$$\lim_{N \rightarrow \infty} P\left(\frac{\sqrt{N}}{\sigma} |\hat{G}_N - G| < X\right) = \frac{1}{\sqrt{2\pi}} \int_{-X}^X e^{-\frac{t^2}{2}} dt$$

当 N 足够大时,有

$$P\left(|\hat{G}_N - G| < \frac{X\sigma}{\sqrt{N}}\right) \approx \frac{1}{\sqrt{2\pi}} \int_{-X}^X e^{-\frac{t^2}{2}} dt = 1 - \alpha \quad (3.1.10)$$

$1 - \alpha$ 称为置信水平, α 称为置信度, 式(3.1.10)说明发生事件 $|\hat{G}_N - G| < (X\sigma/\sqrt{N})$ 的概率由问题要求确定的置信水平决定。实际上可以由问题的要求得到 $1 - \alpha$, 再按正态分布确定 X , 则可以由 X 求出真实值与估值的偏差为 $|\hat{G}_N - G| < (X\sigma/\sqrt{N})$ 。为计算方便, 通常取置信水平为 0.5, 0.95, 0.997 对应的 X 为 0.6745, 1.96, 3。置信度与正态分布取值区间关系如表 3.1 所示。

表 3.1 置信度与正态分布取值区间关系

| | | | | | | | | | | |
|----------|--------|--------|--------|-------|--------|-------|--------|--------|--------|--------|
| α | 0.0063 | 0.0027 | 0.0455 | 0.134 | 0.3173 | 0.617 | 0.50 | 0.05 | 0.02 | 0.01 |
| X | 4 | 3 | 2 | 1.5 | 1 | 0.5 | 0.6745 | 1.9600 | 2.3263 | 2.5758 |

3. 蒙特卡罗法的费用

根据上述讨论, 设问题要求误差为 ϵ , 置信水平为 $1 - \alpha$, 则可以利用正态积分表确定 X , 再由置信水平和 $(X\sigma/\sqrt{N}) \leq \epsilon$ 确定 N 。也就是要求误差小于 ϵ 时, 就必须要求样本容量 N 为

$$N \geq \left(\frac{X\sigma}{\epsilon}\right)^2 \quad (3.1.11)$$

设每计算一次无偏估计量所需的费用为 C , 则误差为 ϵ 的蒙特卡罗法总花费为

$$N_C \approx \left(\frac{X\sigma}{\epsilon}\right)^2 C = \left(\frac{X}{\epsilon}\right)^2 \sigma^2 C \propto \sigma^2 C \quad (3.1.12)$$

即蒙特卡罗法的费用与无偏估计量方差 σ^2 、费用 C 的乘积 $\sigma^2 C$ 成正比。在同一置信水平下, $\sigma^2 C$ 值越小就表示该方法越好。

3.1.3 蒙特卡罗法的特点

蒙特卡罗法有如下几个特点。

(1) 收敛速度与问题维数无关。由式(3.1.12)可以看出, 当置信水平确定之后, 蒙特卡罗法的误差由无偏估计量的方差 σ^2 和样本容量 N 唯一确定。因此, 蒙特卡罗法的收敛速度与问题的维数无关, 维数变化只是引起一次观察所需要的花费 C 的大小变化。

(2) 受问题条件限制的影响不大。采用蒙特卡罗法求积分问题或边值问题时, 一旦确定了拟仿真的概率模型边界条件和区域形状只影响概率事件的判决, 问题的区域形状的限制对蒙特卡罗法影响不大, 不会明显增加问题的难度和计算花费。

(3) 不必进行离散化处理。一般计算方法都首先要将问题进行离散化, 进行区域切分要确定网格, 网格一旦确定后, 问题只能针对网格结点上的值展开, 网格以外点的值只能采用插值方法近似求得。蒙特卡罗法完全不同, 蒙特卡罗法不需要进行任何离散化处理, 其计算针对连续变化的量。这一优点可使蒙特卡罗法更加适用于高维问题, 使精度进一步增加, 可以节省大量存储单元, 对建立蒙特卡罗通用程序很有利。

(4) 蒙特卡罗法是一种直接解决问题的方法。特别是那些具有随机性质的问题, 蒙特卡

罗法可以直接建立随机模型并在计算机上实现针对求解问题的随机过程的实验,因此是一种试验方法,是一种可靠的客观的方法,其结果常作为标准去衡量其他方法的正确性。

(5) 误差容易确定。一般计算方法要估计计算结果与真值的误差是十分困难的事,计算结果的误差与离散方法、计算机的有效位等因素都有关。而蒙特卡罗法的误差则只与 X 、 N 、 σ 有关, X 由置信度确定, N 为试验次数, σ 可以通过 \hat{G}_N 来给出:

$$\hat{\sigma}_N \approx \left[\frac{1}{N} \sum_{n=1}^N g^2(x_n) - \hat{G}_N^2 \right]^{\frac{1}{2}} \quad (3.1.13)$$

由式(3.1.13)可知,蒙特卡罗法的误差很容易确定。

(6) 蒙特卡罗法的缺点是对小维数的问题不如其他方法好,求得的结果一般都比真实值低。

3.1.4 蒙特卡罗法待研究的若干问题

1. 随机数

具有均匀分布的总体中所产生的简单子样称为随机数序列,每一个体称为随机数。伪随机数则是由数学递推公式所产生的随机数。

产生伪随机数的方法有平方取中法、乘同余方法和乘加同余方法等。后两者很好地解决了出现周期性的问题,但是仍然有许多需要研究的问题。例如,20世纪60年代初人们提出了伪随机数序列的独立性问题。

2. 已知分布的随机抽样

冯·诺依曼早在1951年就完成了随机抽样的基本理论,1954年Kahn对冯·诺依曼的理论系统进行了补充研究并扩充了部分方法,但是此后在随机抽样的基本原理方面几乎没有有什么进展,特别是如何将原理用于若干重要分布的随机抽样得到理想的随机抽样方法的研究。

3. 非归一问题的随机抽样

有些问题分布是已知的但不是归一的,其抽样方法是1953年由Metropolis确定的,但这种方法有很大的缺点,特别对连续型分布的情况,需要提出更加理想的抽样方法。

4. 蒙特卡罗法的基本技巧

蒙特卡罗法的基本技巧主要是讨论如何确定合适的无偏估计量使它相应的 σ^2 尽量小的技巧。蒙特卡罗法的基本技巧随问题不同而变化,随着具体问题的应用的迅速增长,需要应用者进一步发展各种不同类型的应用方法。

5. 蒙特卡罗法的并行化计算方法

略。

3.1.5 随机变量的基本规律

1. 随机变量

如果每次试验的结果都可以用 ξ 来表示, ξ 为实数而且对任意实数 X ，“ $\xi < X$ ”事件的发生有着确定的概率,则称 ξ 为随机变量。

事件 $\xi < X$ 对应的概率分布为 $P\{\xi < X\} = F(x)$,其中的随机变量可以是离散随机变量,也可以是连续随机变量。连续随机变量 ξ 可以用概率分布函数或概率分布密度 $p(z)$ 来描写,概率分布函数 $F(x)$ 可以表示为

$$F(x) = \int_{-\infty}^x p(z) dz \quad (3.1.14)$$

表示随机变量 ξ 出现在 z 到 $z+dz$ 范围内的概率为 $p(z)dz$, 其中 $p(z)$ 称为随机变量 ξ 的分布密度函数, 有如下性质:

$$\begin{cases} p(z) \geq 0 \\ \int_{-\infty}^{\infty} p(z) dz = 1 \\ P\{a \leq \xi \leq b\} = F(b) - F(a) = \int_a^b p(z) dz \end{cases} \quad (3.1.15)$$

2. 数学期望值

数学期望值定义为随机变量出现概率最大的值, 用 E 表示, 即

$$E(\xi) = \int_{-\infty}^{\infty} X dF(x) = \int_{-\infty}^{\infty} X p(x) dx \quad (3.1.16)$$

若 $F_{\xi}(x)$ 是 ξ 的分布函数, $f(x)$ 是连续函数, 则有

$$E(f(\xi)) = \int_{-\infty}^{\infty} f(x) dF_{\xi}(x) = \int_{-\infty}^{\infty} f(x) p(x) dx$$

3. 方差

方差表示随机变量值偏离数学期望值的特性, 定义为

$$D(\xi) = E(\xi - E(\xi))^2 = \int_{-\infty}^{\infty} (\xi - E(\xi))^2 dF_{\xi}(x) \quad (3.1.17)$$

$$D(\xi) = E(\xi^2) - (E(\xi))^2 \quad (3.1.18)$$

4. 特征函数

设随机变量 ξ 的分布函数为 $F(x)$, 则称 $e^{it\xi}$ 的数学期望 $f(t) = E(e^{it\xi}) = \int e^{itx} dF(x)$ 为 ξ 的特征函数。

唯一性定理: 分布函数由特征函数唯一决定, 两者有一一对应关系。

5. 中心极限定理

无论单个随机变量的分布如何, 大量的独立随机变量之和总是满足正态分布的, 称为高斯分布, 该分布可以由给定的数学期望值和方差完全确定下来, 其分布密度函数为

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (3.1.19)$$

6. 分布函数的基本性质

(1) 单调不减: 若 $x_1 < x_2$ 则 $F(x_1) \leq F(x_2)$ 。

(2) 定义 $F(-\infty) = \lim_{x \rightarrow -\infty} F(x)$, $F(+\infty) = \lim_{x \rightarrow +\infty} F(x)$, 则有 $F(-\infty) = 0$, $F(+\infty) = 1$ 。

(3) 左连续: 对于 $-\infty < x_0 < +\infty$, 有 $\lim_{x \rightarrow x_0-0} F(x) = F(x_0)$ 。

这三条基本性质不但是分布函数要满足的必要条件, 也构成随机变量分布函数的充要条件。下面用符号 $N(\mu, \sigma^2)$ 表示正态分布的概率函数, 其概率密度函数为

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

式中, $\sigma > 0$ 。

7. 随机变量序列的收敛性

1) 依分布收敛

设随机变量 $\xi_n (n = 1, 2, \dots)$ 和随机变量 ξ 的分布函数分别为 $F_n(x) (n = 1, 2, \dots)$ 和 $F(x)$, 若 $F(x)$ 的所有连续点 x 上都有 $\lim_{n \rightarrow \infty} F_n(x) = F(x)$, 则称随机变量序列 $\{\xi_n\}$ 依分布收敛于随机变量 ξ , 简记为 $\xi_n \xrightarrow{L} \xi$ 。

2) 依概率收敛

若对任意给定的 $\epsilon > 0$ 有 $\lim_{n \rightarrow \infty} P\{|\xi_n - \xi| < \epsilon\} = 1$ 或 $\lim_{n \rightarrow \infty} P\{|\xi_n - \xi| \geq \epsilon\} = 0$, 则称随机变量序列 $\{\xi_n\}$ 依概率收敛于随机变量 ξ , 简记为 $\xi_n \xrightarrow{P} \xi$ 。

3) r 阶收敛

若有 $E|\xi|^r < \infty$ 和 $E|\xi_n|^r < \infty$, 且 $\lim_{n \rightarrow \infty} E|\xi_n - \xi|^r = 0 (r > 0, n = 1, 2, \dots)$, 则称随机变量序列 $\{\xi_n\}$ 为依 r 阶收敛于随机变量 ξ , 简记为 $\xi_n \xrightarrow{r} \xi$, $r = 1$ 时称 $\{\xi_n\}$ 为平均收敛, $r = 2$ 时称 $\{\xi_n\}$ 为均方收敛。

4) 依概率 1 收敛

若有 $P\{\omega : \lim_{n \rightarrow \infty} \xi_n(\omega) = \xi(\omega)\} = 1$, 简记为 $P\{\lim_{n \rightarrow \infty} \xi_n = \xi\} = 1$, 则称随机变量序列 $\{\xi_n\}$ 以概率 1 (或几乎处处) 收敛于随机变量 ξ , 记为 $\xi_n \xrightarrow{a.e} \xi$ 。

上述的几种收敛之间的关系可以用图 3.1 形象化表示, 其中由左向右依次满足包含关系。



图 3.1 几种收敛的关系

5) 切比雪夫(Чебыщев)不等式

设随机变量 ξ 的方差 $D\xi$ 为有限值, 则对任意的 $\epsilon > 0$, 有

$$P\{|\xi - E\xi| \geq \epsilon\} \leq \frac{D\xi}{\epsilon^2} \quad \text{或} \quad P\{|\xi - E\xi| < \epsilon\} \geq 1 - \frac{D\xi}{\epsilon^2}$$

3.1.6 大数定律及中心极限定理的一般形式

1. 大数定律

设 $\{\xi_n\}$ 是随机变量序列, 若存在常数序列 $\{a_n\}$ 对任给的 $\epsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{k=1}^n \xi_k - a_n\right| < \epsilon\right\} = 1$$

则称随机变量序列 $\{\xi_n\}$ 服从大数定律。

3 个常用的大数定律形式如下所述。

1) 伯努利(Bernoulli)大数定律

n 重独立试验中事件 A 出现的频率 $\frac{\nu_n}{n}$ 依概率收敛于事件 A 在每次试验中出现的概率 p

($0 < p < 1$), 即对任给的 $\epsilon > 0$, 有 $\lim_{n \rightarrow \infty} P\left\{\left|\frac{\nu_n}{n} - p\right| < \epsilon\right\} = 1$ 。

伯努利大数定律以数学形式表达了随机事件出现的频率的稳定性, 即当 n 很大时, 事件 A 发生的频率以很大的概率趋向于每次试验中出现的概率, 即 $\frac{\nu_n}{n} \rightarrow p(A) = p$ 。

2) 切比雪夫大数定律

设 $\{\xi_n\}$ 是相互独立的随机变量序列, 又有 $D\xi_n \leq C$, C 为常数, $n = 1, 2, \dots$, 则对任意的 $\epsilon > 0$, 有 $\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n \xi_k - \frac{1}{n} \sum_{k=1}^n E\xi_k \right| < \epsilon \right\} = 1$ 。

3) 辛钦(Хинчин)大数定律

设 $\{\xi_n\}$ 是相互独立同分布的随机变量序列, 且有有限的数学期望 $E\xi_n = \mu$, $n = 1, 2, \dots$, 则对任意的 $\epsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n \xi_k - \mu \right| < \epsilon \right\} = 1 \quad \text{即} \quad \frac{1}{n} \sum_{k=1}^n \xi_k \xrightarrow{p} \mu$$

辛钦大数定律可理解为: 对一个随机变量 ξ 在进行 n 次重复独立观测时, 得到的算术平均 $\frac{1}{n} \sum_{k=1}^n \xi_k$ 在 n 很大时会以很大的概率接近一个常数, 即接近数学期望值。

2. 中心极限定理

设相互独立的随机变量序列 $\{\xi_n\}$ 有有限的数学期望 $E\xi_n$ 与方差 $D\xi_n > 0$ ($n = 1, 2, \dots$), 并

设 $\eta_n = \frac{\sum_{k=1}^n \xi_k - \sum_{k=1}^n E\xi_k}{\sqrt{\sum_{k=1}^n D\xi_k}}$, 则 η_n 称为 $\{\xi_n\}$ 前 n 项和的标准化(规范化), 中心极限定理就是要寻找 $\{\xi_n\}$ 满足什么条件时, η_n 的渐近分布是标准正态分布, 即

$$\lim_{n \rightarrow \infty} P \{ \eta_n < x \} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (3.1.20)$$

3.1.7 4个常见的中心极限定理

1. 勒维-林德伯格(Lévy-Lindeberg)中心极限定理

设 $\{\xi_n\}$ 是相互独立、同分布的随机变量序列, 且有有限的数学期望与方差 $E\xi_n = \mu$, $D\xi_n = \sigma^2 > 0$ ($n = 1, 2, \dots$), 则对任意的实数 x , 有

$$\lim_{n \rightarrow \infty} P \left\{ \frac{\sum_{k=1}^n \xi_k - n\mu}{\sigma/\sqrt{n}} < x \right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (3.1.21)$$

或

$$\frac{\sum_{k=1}^n \xi_k - E\left(\sum_{k=1}^n \xi_k\right)}{\sqrt{D\left(\sum_{k=1}^n \xi_k\right)}} \xrightarrow{L} N(0, 1), \quad n \rightarrow \infty \quad (3.1.22)$$

或

$$\frac{\frac{1}{n} \sum_{k=1}^n \xi_k - \mu}{\sigma/\sqrt{n}} \xrightarrow{L} N(0, 1), \quad n \rightarrow \infty \quad (3.1.23)$$

这个定理表明: 当 n 充分大时, $\frac{1}{n} \sum_{k=1}^n \xi_k$ 近似服从正态分布 $N\left(\mu, \frac{\sigma^2}{n}\right)$ 。

2. 棣莫弗-拉普拉斯(De Moivre-Laplace)中心极限定理

设 $\{\xi_n\}$ 是相互独立同分布的随机变量序列, 且 $\xi_n \sim B(1, p)$ ($0 < p < 1, n = 1, 2, \dots$), 则对任意实数 x , 有

$$\lim_{n \rightarrow \infty} P \left\{ \frac{\sum_{k=1}^n \xi_k - np}{\sqrt{np(1-p)}} < x \right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (3.1.24)$$

式(3.1.24)常常等价地表达如下: 设 ν_n 表示 n 重独立试验中事件 A 出现的次数, p ($0 < p < 1$) 是事件 A 在每次试验中出现的概率, 则对任意区间 $[a, b)$, 有

$$\lim_{n \rightarrow \infty} P \left\{ a \leq \frac{V_n - np}{\sqrt{np(1-p)}} < b \right\} = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (3.1.25)$$

此处, $B(n, p)$ 为二项分布, $E\xi = np$, $D\xi = np(1-p)$ 。

3. 李雅普诺夫中心极限定理

设 $\{\xi_n\}$ 是相互独立的随机变量序列, 如果对某个 $\delta > 0$ 有 $0 < E|\xi_k - E\xi_k|^{2+\delta} < \infty, k = 1, 2, \dots$, 且满足条件 $\lim_{n \rightarrow \infty} \frac{1}{B_n^{2+\delta}} \sum_{k=1}^n E|\xi_k - E\xi_k|^{2+\delta} = 0$, 其中 $B_n = \sqrt{\sum_{k=1}^n D\xi_k}$, 则对任意实数 x , 有

$$\lim_{n \rightarrow \infty} P \left\{ \frac{\sum_{k=1}^n \xi_k - \sum_{k=1}^n E\xi_k}{B_n} < x \right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (3.1.26)$$

4. 林德伯格中心极限定理

设 $\{\xi_n\}$ 是独立随机变量序列, 且有有限的方差 $D\xi_n > 0, n = 1, 2, \dots$, 如果对任意实数 $\tau > 0$ 满足条件

$$\lim_{n \rightarrow \infty} \frac{1}{B_n^2} \sum_{k=1}^n \int_{|x - E\xi_k| > \tau B_n} |x - E\xi_k|^2 dF_k(x) = 0$$

式中, $B_n = \sqrt{\sum_{k=1}^n D\xi_k}$, $F_k(x)$ 是 ξ_k 的分布函数, 则对任意函数 x , 有

$$\lim_{n \rightarrow \infty} P \left\{ \frac{\sum_{k=1}^n \xi_k - \sum_{k=1}^n E\xi_k}{B_n} < x \right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (3.1.27)$$

此定理可解释为: 若一个随机变量是由大量相互独立的随机因素的影响所造成的, 而每一个别因素在总影响中所起的作用都不是很大, 则这种随机变量通常都服从或近似服从正态分布。

大数法则和中心极限定理是蒙特卡罗法的基础, 应该通过各种形式的表述和应用来加深理解。

3.1.8 几种常见的概率模型和分布

常见的概率模型(以下简称为概型)和分布有以下 15 种。

1. 伯努利概型——二项分布

$$F(x) = P(\mu < x) = \sum_{K < x} C_n^K p^K q^{n-K}, \quad q = 1 - p \quad (3.1.28)$$

2. 泊松 (Poisson) 分布

$$F(x) = \sum_{K < x} \frac{(\lambda t)^K}{K!} e^{-\lambda t} \quad (3.1.29)$$

式中, λ 为常数, 在电信科学中泊松分布可以用来表示单位时间内 K 次呼叫的占线率。

3. 均匀分布

令 Beta 分布中 $a=b=1$, 即得均匀分布

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases} \quad (3.1.30)$$

4. 正态分布

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad (3.1.31)$$

5. 指数分布

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \leq 0 \end{cases} \quad (3.1.32)$$

6. Gamma 分布

Gamma 分布是用如下概率密度表示的概率分布, 记为 $\text{Ga}(\alpha, \lambda)$ 。

$$p(x; \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad (3.1.33)$$

7. Beta 分布

Beta 分布是用如下概率密度表示的概率分布, 记为 $\text{Be}(a, b)$, 其中 a, b 是两个正值参数。

$$p(a; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \quad (3.1.34)$$

8. t 分布

t 分布是用如下概率密度表示的概率分布, 记为 $t(\alpha)$, 其中 α 是正实数。

$$p(x, \alpha) = \frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{\sqrt{\alpha\pi}\Gamma\left(\frac{\alpha}{2}\right)} \left(1 + \frac{x^2}{\alpha}\right)^{-\frac{(\alpha+1)}{2}} \quad (3.1.35)$$

9. z 分布

z 分布是用如下概率密度表示的概率分布, 记为 $z(a, b)$, 其中 a, b 是两个正值参数。

$$p(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{x^{a-1}}{(1+x)^{a+b}} \quad (3.1.36)$$

10. χ^2 分布

在 Gamma 分布中, 令 $\alpha = n/2, \lambda = 1/2$, 即得自由度为 n 的 χ^2 分布, 记为 $\chi^2(n)$ 。

$$p(x; \alpha, \lambda) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, \quad x > 0 \quad (3.1.37)$$

11. 指数分布

在 Gamma 分布中, 令 $\alpha=1$, 即得指数分布, 记为 $\exp(\lambda)$ 。

$$p(x; \lambda) = \lambda e^{-\lambda x}, \quad x > 0 \quad (3.1.38)$$

12. 反余弦分布

在 Beta 分布中,令 $a=b=1/2$,即得反余弦分布。

13. 多项分布

$$P(X_1 = n_1, \dots, X_r = n_r) = \frac{n!}{n_1! \cdots n_r!} p_1^{n_1} \cdots p_r^{n_r} \quad (3.1.39)$$

14. 非中心 Gamma 分布

非中心 Gamma 分布是如下概率密度表示的概率分布,记为 $\text{Ga}(\alpha, \lambda, \gamma)$ 。

$$p(x; \alpha, \lambda, \gamma) = \frac{e^{-\gamma} \gamma^m}{m!} P_G(x; \alpha + m, \lambda) \quad (3.1.40)$$

15. 非中心 t 分布

非中心 t 分布是用如下概率密度表示的概率分布,记为 $t(n, \gamma)$,其中 γ 是非负实数。

$$p(x, n, \gamma) = \frac{n^{\frac{n}{2}} e^{-\frac{\gamma^2}{2}}}{\sqrt{\pi} \Gamma\left(\frac{n}{2}\right) (n + x^2)^{\frac{n+1}{2}}} \sum_{m=0}^{\infty} \Gamma\left(\frac{m+n+1}{2}\right) \frac{\lambda^m}{m!!} \left(\frac{\sqrt{2}x}{\sqrt{n+x^2}}\right)^m \quad (3.1.41)$$

3.1.9 蒙特卡罗法简单应用举例

为了有一个直观的基本概念,举例如下。

例 3.1 Buffon 投针试验求 π 的近似值(1777 年)。

平面上有相距 $2a$ 的两条平行线,向这一平面随机投放一个长为 $2l$ 的针,设 $a > l > 0$,针的中点用 M 表示, y 是中点 M 与最近一条平行线的距离, ϕ 是针与平行线的交角。显然有如下的条件(参见图 3.2)。

(1) $0 \leq y \leq a, 0 \leq \phi \leq \pi$;

(2) 针与平行线相交的充要条件为 $y \leq l \sin \phi$,于是,针与平行线相交概率为

$$p = \frac{1}{\pi a} \int_0^{\pi} \int_0^{l \sin \phi} dy d\phi = \frac{1}{\pi a} \int_0^{\pi} l \sin \phi d\phi = \frac{2l}{\pi a} \quad (3.1.42)$$

此处, y 的变化范围是由 0 到 a , ϕ 则只能由 0 变到 π , ϕ 和 y 变化范围扫过的总面积为 πa ,总的概率为 1,参见图 3.3,则所有可能发生的事件的总值为 πa ,而投针与平行线相交的概率面积为 $\int_0^{\pi} \int_0^{l \sin \phi} dy d\phi$ 。用 N 表示投针的次数, ν 是针与平行线相交的次数。当试验次数足够多时,相交

概率为 $p = \nu/N$,由式(3.1.42)有 $\frac{\nu}{N} \approx \frac{2l}{\pi a}$,于是有 $\pi \approx \frac{2lN}{a\nu}$ 的关系。

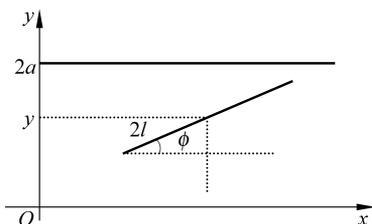


图 3.2 Buffon 投针试验示意图

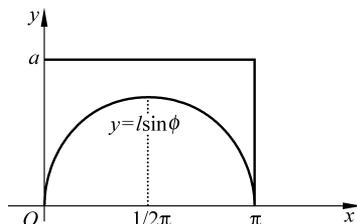


图 3.3 投针试验中针与线相交概率

Buffon 投针试验可以编计算机程序进行计算。Wolf 在 1853 年取 $a = 45, l = 36, N =$

5000 进行试验时,得到 $v=2532$, 计算出的 π 值为 3.1596。表 3.2 列出一些试验结果。

表 3.2 历史上投针试验获得的 π 的近似值

| 试验者 | 时间 | 针长(l/a) | 投针次数 | 相交次数 | π 的估计值 |
|-----------|--------|-------------|------|------|--------------|
| Wolf | 1850 年 | 0.80 | 5000 | 2532 | 3.159 56 |
| Smith | 1855 年 | 0.60 | 3204 | 1218 | 3.156 65 |
| Fox | 1884 年 | 0.75 | 1030 | 489 | 3.159 51 |
| Lazzarini | 1925 年 | 0.83 | 3408 | 1808 | 3.141 592 92 |

实际上,该试验的精度并不高。设 $l=a$, 则 $p=2/\pi$, 由中心极限定理可知: 当试验次数为 n 时, 如果要求置信度为 0.95, 同时要求 π 的估计值达到小数点后三位有效数字, 则试验次数 n 应该达到 8.87×10^5 , 显然表 3.2 中 Lazzarini 的结果纯属巧合(该结果接近祖冲之的密率 355/113)。

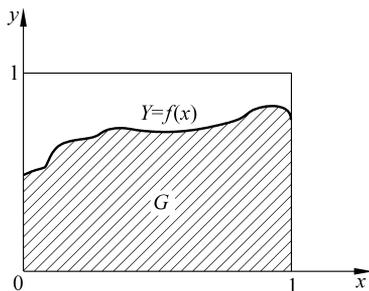


图 3.4 随机投点求积分值

例 3.2 见图 3.4, 求定积分值 $I = \int_0^1 f(x) dx$ 。

设 $f(x)$ 是 $[0, 1]$ 上的连续函数且 $0 \leq f(x) \leq 1$, 那么当向正方形 $[1, 1]$ 内均匀投点 (ξ, η) 时, 随机点落入曲线 $y=f(x)$ 下面的概率就是要求的积分值。

$$P_r \{y \leq f(x)\} = \int_0^1 \int_0^{f(x)} dy dx = \int_0^1 f(x) dx = I$$

例 3.3 应用泊松过程估算地面移动设备雷击概率。

(1) 泊松分布与二项分布。

泊松分布源于二项分布, 因为泊松分布是某段时间间隔或某个给定区域内发生的结果是否达到了某个数量, 本质上还是达到与没达到的选择。二项分布为

$$P \{y = k\} = \binom{n}{k} p^k q^{n-k} \quad p + q = 1 \quad k = 0, 1, 2, \dots, n$$

实际情况中, 当 p 或 q 很小, 而试验次数 n 很大时, $n \rightarrow \infty$, $p \rightarrow 0$, 两者是同级的无限大/小, 所以 $np \rightarrow \lambda$, 一定趋于一个常数 λ 。设试验时间无限长, 每次试验的间隔时间为 Δ , 在区间 Δ 中发生 k 次呼叫的概率满足如下的二项分布:

$$p_n(k) = \frac{n!}{(n-k)! k!} p^k (1-p)^{n-k} \quad k = 0, 1, 2, \dots, n$$

泊松定理可推导出二项分布的极限分布。

泊松定理如下: 如果 $n \rightarrow \infty$, $p \rightarrow 0$, $np \rightarrow \lambda$, 则有

$$\frac{n!}{(n-k)! k!} p^k (1-p)^{n-k} \xrightarrow{n \rightarrow \infty} e^{-\lambda} \frac{\lambda^k}{k!} \quad k = 0, 1, 2, \dots$$

(2) 首次发生的小概率事件的泊松分布。

把泊松定理应用于 p 很小, 而试验次数 n 很大时的二项分布问题中, 也就是用于随机泊松点的问题: 随机放置 n 个点到 $(-T/2, T/2)$ 区间并且用 $P\{k \text{ 在 } t_a \text{ 内}\}$ 表示这些点中的 k 个将落在 (t_1, t_2) 内的概率。设问题要考虑的区间长度为 $t_a = t_2 - t_1$, $p = t_a/T$, 则在 $n \gg 1$ 并且 $t_a = t_2 - t_1 \ll T$ 时, 有

$$P\{k \text{ 在 } t_a \text{ 内}\} \approx e^{-\frac{nt_a}{T}} \frac{\left(\frac{nt_a}{T}\right)^k}{k!} = e^{-\lambda t_a} \frac{(\lambda t_a)^k}{k!}$$

用 X 表示在一个泊松试验中得到的某个随机变量结果的数量,称为泊松随机变量,在公式中表示为 $X > x$,其概率分布为泊松分布,结果的均值为 $\mu = \lambda t$, t 是具体的时间、距离、面积或体积; λ 是结果的发生率,表示在单位时间、长度、面积或体积内得到结果的平均数量。某段时间间隔或某个给定区域内某事件发生的次数满足泊松分布:

$$p(x; \lambda t) = \frac{e^{-\lambda t} (\lambda t)^x}{x!} \quad x = 0, 1, 2, \dots$$

其中, $X(x)$ 表示在给定的时间间隔或指定的区域 t 内结果的发生数量; X 为泊松随机变量; λ 是在单位时间、长度、面积或体积内得到结果的平均数量。

例如,一个试验中, $1\mu\text{s}$ 内通过计数器的平均辐射粒子数为 4 个。在某个 $1\mu\text{s}$ 中有 6 个辐射粒子通过计数器的概率是多少?

按所给条件知道, $x=6, \mu=\lambda t=4$, 有

$$p(6; 4) = \frac{e^{-4} 4^6}{6!} = \sum_{x=0}^6 p(x; 4) - \sum_{x=0}^5 p(x; 4) = 0.8893 - 0.7851 = 0.1042$$

又例如,在 12 小时内,有 180 个电话随机打来,问 4 小时中打入电话数介于 50~70 个的概率有多大? 此时 $p=4/12, x=180, k=50\sim 70$ 。

泊松分布为离散分布,用来计算在某时间间隔或空间范围内一定数目的泊松事件发生的概率,就变成了连续分布的泊松问题。在很多应用中,时间或空间可以是随机变量,而泊松分布是一个单参数的分布,参数 λ 可以理解为单位时间内事件发生的平均数目。

在许多情况下,更关心的是问题中的事件发生没发生,可以等价地考虑首次发生该事件的时间是否落在我们关心的时间或区间段内。当我们考虑描述事件首次发生所需要的等待时间的随机变量时,可以研究在时间 t 时,事件还没有发生时的概率,于是互补的概率问题是只要发生事件的概率。于是,有

$$(\text{首次发生事件所需要的等待时间的概率}) = 1 - (\text{只要发生事件的概率})$$

泊松事件首达时刻,即在某段时间间隔或某个给定区域内还没有发生概率事件的概率等价于发生所有事件的概率,即

$$P(X \geq 1) = 1 - p(0; \lambda t) = 1 - \frac{e^{-\lambda t} (\lambda t)^0}{0!} = 1 - e^{-\lambda t} = e^{-\lambda t}$$

(3) 根据问题条件构建概率密度函数。

为了构造移动设备雷击概率函数,首先要回顾一下构建概率函数的理论。针对随便选取或构造的一些函数,怎么判断它是否为概率函数? 从数学上讲,应该满足随机函数的存在性定理。

存在性定理: 给定一个函数 $f(x)$ 或它的积分 $F(x) = \int_{-\infty}^x f(u) du$, 如果要求能够构造一个随机变量 x 和一个随机试验,具有分布密度函数 $f(x)$ 或分布函数 $F(x)$, 则要求该函数必须具有如下性质: 函数 $f(x)$ 或分布函数 $F(x)$ 必须是非负的并且在 $(-\infty, +\infty)$ 上的积分 $F(x) = 1$; $F(x)$ 必须是右连续的,并且当 x 从 $-\infty$ 增加到 ∞ 时,函数 $f(x)$ 必须单调地从 0 增加到 1。

在构建关于随机变量的概率分布函数时,其均值是最重要的信息,因为这是无限次概率试

验得到的必然性的结果。对于泊松概率分布来讲,其均值为 $\mu = \lambda t$,

没有事件发生时 $x=0$,泊松分布变成指数分布,则 x 的累积分布为

$$P = 1 - e^{-\lambda x} = e^{-\lambda x}$$

指数分布函数:如果随机变量 x 的概率密度函数为 $f_x(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{其他} \end{cases}$,则称随机变量 x 是服从具有参数 λ 的指数分布。如果在互不相交的区间上事件发生是相互独立的,例如电话呼叫的到达时间、公共汽车到达一个车站的时间或类似的事件的等待时间,都可以用指数分布来描述。

因此,在构建移动设备雷击概率函数时,首先要合理地假设其数学期望值或均值。在我们的问题中,讨论的是近地面发生的雷击事件,尽管描述车辆、公路、雷云的参数复杂各异,但是有意义的最主要的是地面物体的引雷面积。因此,取地面移动目标的等效静止区域面积与雷云区行车路线决定的公路的面积之比作为发生雷击概率的均值 $\mu = \lambda t$ 。定义概率 p_{em} 主要由处于雷云区的公路的总的引雷面积和移动目标的等效静止引雷面积决定。

设移动目标的等效静止区域面积为 σ_{em} ;雷云区的行车路线的公路总面积的下确界,也就是公路的总建筑面积,仍然用 σ_{cm} 表示;用 σ'_{cm} 表示由移动出发点到发生雷击点的公路的总建筑面积;公路的总长度用 L_{cm} 表示;由移动出发点到发生雷击点的公路的总长度用 L'_{cm} 表示;移动目标的等效静止引雷面积用 σ_{em} 表示,而移动目标的等效静止引雷面的长度用则 L_{em} 表示。

从实际情况考虑,公路的宽度与移动车辆的宽度有一定的比例关系,例如两车道、四车道、六车道等。假设最简单公路情况,也就是两车道的情况,就可以把公路宽度 D 设为移动车辆的 2 倍。根据地面物体引雷面与高度的关系,考虑车辆比地面有一定的高度,所以可以忽略这个差异,则有

$$\sigma_{em} = L_{em} \times d_{车宽} \cong L_{em} \times D \quad (3.1.43)$$

$$\sigma_{cm} = L_{cm} \times D \quad (3.1.44)$$

在我们的问题中, λ 是事件的发生率或均值,可以认为 λ 等于移动目标的等效静止区域面积与雷云区的行车路线的公路总面积之比,也就是公路的总长度与移动目标的等效静止引雷长度之比,即 $\lambda = \frac{L_{em}}{L_{cm}}$ 。

问题中,首次发生雷击事件的位置是不确定的随机数,就是近地面发生的雷击地面移动设备事件中的随机变量 x 。不管雷击发生在何处,都属于地面移动目标的“首次发生雷击”这个事件,设 L'_{cm} 是“首次发生雷击”的随机位置,就可以把随机变量 x 用 $x = \frac{L_{em}}{L_{cm}}$ 表示。

经过推导得到

$$p_{em} = \left(\frac{L_{em}}{L_{cm}} \right) e^{-\frac{L_{em}}{L_{cm}} x} = \lambda e^{-\frac{L_{em}}{L_{cm} \left(1 - \frac{L_{em}}{L_{cm}} \right)}} \rightarrow \lambda e^{-\lambda \frac{1}{1-x}} = \lambda e^{-\lambda (x+x^2+x^3+\dots)} \approx \lambda e^{-\lambda x}$$

在这个概率表达式中,选取 $e^{-\frac{L_{em}}{L_{cm} \left(1 - \frac{L_{em}}{L_{cm}} \right)}}$ 是为了使概率函数满足存在性定理的要求: $F(x)$ 必须是右连续的,并且当 x 从 $-\infty$ 增加到 ∞ 时,函数 $f(x)$ 必须单调地从 0 增加到 1。

当 $L'_{cm} = L_{cm}$ 时,由于公路非常短,不可能发生雷击,分母为 0,则 $e^{-\infty} = 0$;当 L_{cm} 为无穷长时,表示公路非常长,其上移动的车辆一定会遭受雷击,也就是 $e^{-\frac{L_{em}}{L_{cm}}} \rightarrow e^{-\frac{L_{em}}{\infty}} = e^0 = 1$ 。综上所述

述,我们设计的首次发生雷击事件概率函数符合存在性定理的要求条件。

设 $\lambda = \frac{L_{em}}{L_{cm}}$, 而首次发生雷击事件的位置 L'_{cm} 是不确定的随机数, 但是不管发生在何处, 都

属于地面移动目标的“首次发生雷击”这个事件, 设 $x = \frac{L_{em}}{L_{cm}}$, 有

$$e^{-\frac{L_{em}}{L_{cm}(1-\frac{L_{em}}{L_{cm}})}} \rightarrow e^{-\left(\frac{L_{em}}{L_{cm}}\right)\frac{1}{1-x}} = e^{-\left(\frac{L_{em}}{L_{cm}}\right)(x+x^2+x^3+\dots)} \rightarrow e^{-\left(\frac{L_{em}}{L_{cm}}\right)x} = e^{-\lambda x}$$

以典型数值为例, 设地面移动目标长度的线度 $L_{\text{移动目标}}$ 为 10~20m, 再加上地面移动目标运动产生的等效长度 40~60m, 设移动目标等效静止引雷长度 L_{em} 为 50~80m。选取行驶的公路长度为 1km, 按此数据计算可得到 p_{em} 为 0.047 5~0.078 4, p_{em} 随着等效静止引雷长度和公路长度取值的不同而变化。

根据上述演算, 大型运输车运输中被雷击是个小概率事件。但是, 这个举例只是配合教学的简单情况, 实际情况中, 雷云高度、厚度, 云-地温度差决定的雷云带电量, 以及运输货物的易燃易爆性质等因素都对结果产生很大影响, 并不能得出大型运输车运输中被雷击是小概率事件的结论。

3.2 伪随机数

3.2.1 简单子样

产生随机数是执行蒙特卡罗法的基本条件, 通常用随机数序列的方法实现。具有单位均匀分布的总体中所产生的简单子样称为随机数序列, 其中每一个体称为随机数。

简单子样指的是这样的随机子样 $\{x_n\}_{n=1}^N = \{x_1, \dots, x_N\}$, 它们之间相互独立, 对于任意的 x 和所有的 $n=1, 2, \dots, N$, 满足

$$P(x < x_n < x + dx) = f(x)dx \quad (3.2.1)$$

简单子样的性质如下。

(1) 对于任意的统计量 $g(x)$, $\{g(x_n)\}_{n=1}^N$ 具有同一分布且相互独立, 数学期望均等于

$$G = E(g) = \int g(x)dF(x) = \int g(x)f(x)dx \quad (3.2.2)$$

$F(x)$ 与 $f(x)$ 分别为随机变量 ξ 的分布函数和分布密度函数。

(2) 只要 $E(|g|) = \int |g(x)|f(x)dx$ 存在, 则蒙特卡罗近似估计值 $\hat{G}_N^S = \frac{1}{N} \sum_{n=1}^N g(x_n)$ 便依概率 1 收敛于 G 。

(3) 只要统计量 $g(x)$ 的方差 σ^2 存在, 且异于 0, 则 $\sqrt{N}|\hat{G}_N^S - G|/\sigma$ 就渐进于正态分布。

3.2.2 随机数与伪随机数

前面提到, 由具有已知分布的总体中产生的简单子样在蒙特卡罗法的问题中占有非常重要的地位。形式多样的简单子样可以按照一定的规则利用基础的随机变量产生, 显然最合适的基础随机变量是单位均匀分布的随机变量, 就如同可以用黑点在白纸上构造出任意形状的模式的情形一样。在蒙特卡罗法中, 把这种分布的简单子样作为基础量来看待, 而针对其他分布产生简单子样时, 就把这种具有单位均匀分布的随机变量看成已知条件。通常人们采用随

机数作为计算机上的单位均匀分布的随机变量。因此,生产随机数成为应用蒙特卡罗法的大前提。

1. 随机数

均匀分布也称为矩形分布,其中最基本的是单位均匀分布,其密度函数为 $f(x)=1(0\leq x\leq 1)$,而从具有单位均匀分布的总体中产生的简单子样 x_1, x_2, \dots, x_N 称为随机数序列,其中的每一个体称为随机数,并且用 $\xi_1, \xi_2, \dots, \xi_N$ 等符号表示。

2. 随机数的性质

设随机数 ξ_1, ξ_2, \dots 是相互独立的具有相同单位均匀分布的随机变量序列,则对任意自然数 S ,由 S 个随机数所组成的 S 维空间上的点 $(\xi_{n+1}, \xi_{n+2}, \dots, \xi_{n+s})$ 也在 S 维空间的单位立方体 G_S 上均匀分布,也就是对于任意小的 $a_i, 0\leq a_i\leq 1, i=1, 2, \dots, s$,有下式成立:

$$P(\xi_{n+i} \leq a_i, i=1, 2, \dots, s) = \prod_{i=1}^s a_i \quad (3.2.3)$$

客观上存在许多真正的随机变量,因而用物理的方法产生随机数应当不成问题。例如,可以用放射性同位素的衰变来产生随机数。1942年,RAND公司就是以随机脉冲源,用电子旋转轮产生随机数表的,但要产生又快又准的随机数是很困难的。例如,1975—1978年,Frigerio做了一个高分率的计数器,每小时能产生6000个31位的真随机数。但是,随机过程一去不复返,而且设备的维修费用昂贵,显然产生的随机数的数量少而且慢。

3. 伪随机数

由于采用真随机数遇到困难,人们普遍用数学方法产生随机数,但这种随机数是由递推公式和初值推出的,也就不满足随机数相互独立的要求。由于计算机所能表示的 $[0, 1]$ 上的数量是有限多个,而随机数又由递推公式产生,所以就会出现周期现象。因此,一般称用数学方法产生的随机数为伪随机数。

尽管存在这些问题,但在一定范围的应用中,只要这些伪随机数序列能通过一系列的统计检验,如均匀性和独立性的检验,就可以把它当作随机数来使用。伪随机数的应用范围与伪随机数的总体大小有直接关系,在应用中伪随机数的数量应能满足所设概率模型和问题精度的要求。实际上,随机数无边无际变化的性质总是体现在有限逼近的过程中。

4. 伪随机数的最大容量

由伪随机数组成的子样有一个最大限度,超出这个限度将出现周期循环现象,则从伪随机数的初始值开始到出现循环现象为止所产生的伪随机数的个数称为伪随机数的最大容量。

3.2.3 产生伪随机数的几种方法

1. 平方取中法

这是最早的一种方法,由冯·诺依曼提出,是历史上第一个产生伪随机数的方法。

以十进制为例,平方取中法是把一个 $2s$ 位的十进制数自乘,去头截尾保留中间的 $2s$ 位数字,再用 10^{2s} 来除使结果落在 $[0, 1]$ 区间,得到的结果就是 $[0, 1]$ 上的伪随机数,平方取中法可用公式表示为

$$X_{n+1} \equiv [10^{-s} X_n^2] \pmod{10^{2s}}$$

十进制:

$$\xi_{n+1} = \frac{X_{n+1}}{10^{2s}} \quad (3.2.4)$$

$$X_{n+1} \equiv [2^{-s} X_n^2] \pmod{2^{2s}}$$

二进制:

$$\xi_{n+1} = \frac{X_{n+1}}{2^{2s}} \quad (3.2.5)$$

式中,用 10^{-s} 乘 X_n^2 表示“截尾”, $(\text{mod } 10^{2s})$ 运算表示“去头”, $A \text{ mod } (N)$ 表示取余运算,即 A 除以 N 所剩的余数, $[*]$ 符号(方括号)表示取整运算,取不超过“*”的最大整数。

与此类似,Forsythe 提出类似的方法如下:

$$X_{n+2} \equiv [2^{-s} X_n X_{n+1}] \pmod{2^{2s}}$$

$$\xi_{n+2} = \frac{X_{n+2}}{2^{2s}} \quad (3.2.6)$$

平方取中法的最大容量随初值的不同而变化,情形非常复杂。

练习题: 取十进制数,用平方取中法产生伪随机数序列并比较序列长度,找到周期发生部位(建议数字: $2s=4$, $X_0=1111$ 和 $X_0=6400$)。

当用数学方法产生伪随机数时,如果计算的伪随机数产生了零值或产生了重复的元素,则伪随机数序列中断,并决定了此随机数最大序列长度。平方取中法缺点很多,现已很少采用,目前使用比较多的是乘同余法和乘加同余法。

2. 加同余法

任取两个数作为初值,用下面的公式产生后续随机数:

$$X_{n+2} \equiv X_n + X_{n+1} \pmod{M}$$

$$\xi_{n+2} = \frac{X_{n+2}}{M} \quad (3.2.7)$$

当采用 $X_1=X_2=1$ 时,所产生的序列就是著名的斐波那契(Fibonacci)数列,该数列在生物等多种学科中有广泛的应用。一般性地讨论加同余法的最大容量很困难,但是当取 $M=2^s$ 时,该随机序列的最大容量为 $3 \times 2^{s-1}$ 。

3. 乘同余法

乘同余法是 Lehmer 首先提出的。其方法为:对任意取定的初值再用式(3.2.8)决定后续随机数。

$$\begin{cases} X_{n+1} \equiv aX_n \pmod{M} \\ \xi_{n+1} = \frac{X_{n+1}}{M} \end{cases} \quad (3.2.8)$$

式中, a 为选定的常数,在式中最好取 X_1 为 $4q+1$ 型的数, q 为任意整数, a 取 5^{2K+1} 型的数, K 是使 5^{2K+1} 达到计算机允许的最大奇数时所确定的数。

其最大容量主要取决于产生周期的时刻,与 a 的取值有关,通常与初值 X_1 的选取无关,与 M 的具体值关系不大,但一般要取 X_1 与 M 为互素数。

4. 乘加同余法

乘加同余法由 Rotenberg 提出,其方法如下:对任意初值 X_1 ,由下式确定伪随机数。

$$X_{n+1} \equiv aX_n + c \pmod{M}$$

$$\xi_{n+1} = \frac{X_{n+1}}{M}, \quad a, c \text{ 为常数} \quad (3.2.9)$$

乘子 a 为小于模 M 的正整数, c 为非负整数,常选为 0。当 $c=0$ 、 M 取素数、 a 是 M 的一个原根时用乘同余法产生的随机数的周期为 $M-1$,而 M 通常取为 $M=2^{31}-1$, $a=16\ 087$ 或

630 360 016。当取 $M = 2^K$ 、 K 为整数、 a 为模 8 余 3 或 5 的整数时所产生的随机数周期为 2^{K-2} 。

1968 年, Marsaglia 从理论上指出乘同余法的缺点是: 给定长度 K , 设 $(\xi_{i1}, \xi_{i2}, \dots, \xi_{ik})$ 为乘同余法所产生的随机数空间上的一点, 这些点的分布具有栅格结构, 也就是, 这些点落在具有确定数量的等距离的平行超平面上, 也就是说是有规律的, 破坏了应有的随机性。

5. 移位寄存器法——Tausworthe 法

一般形式为

$$\begin{cases} \beta_j = \beta_0 \cdot x T_j \\ \xi_j = \sum_{k=1}^n b_{(j-1) \times n + k} x 2^{k-1}, \quad j = 1, 2, \dots \end{cases} \quad (3.2.10)$$

式中, $b_j \in Z$ (Z 表示只含数 0 和 1 的二元有限域), $\beta_j = (\beta_{1,j}, \beta_{2,j}, \dots, \beta_{n,j})$ 是 $1 \times n$ 的二进制矩阵。

$$T = (I + R^s)(I + L^t)$$

$$I = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}_{n \times n} \quad R = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}_{n \times n}$$

$$L = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}_{n \times n}$$

t 为整数, I, R, L 为 $n \times n$ 矩阵。

移位寄存器法产生的伪随机数周期为 $2^n - 1$, 但这种伪随机数发生器可能具有严重的高位序列相关性。

6. 斐波那契法

该法的一般形式为 $X_i = (X_{i-p} \otimes X_{i-q}) \bmod 2^m$, 其中 \otimes 是一种运算操作, 可以定义为加、减、乘、逻辑异或 (p, q 为延迟步, 且 $p > q$)。

斐波那契法最大的优点是可以直接进行浮点运算, 而不像其他算法需要经整数转换。

7. 混合法

混合法于 1965 年由 Donald 和 Marsaglia 提出, 该方法将两个或三个乘同余方法简单组合。步骤如下。

(1) 给出两个单独的乘同余发生器:

$$\begin{cases} X_i = (\alpha X_{i-1}) \bmod M_x \\ Y_i = (\beta Y_{i-1}) \bmod M_y \end{cases}$$

(2) 组合规则:

$$\begin{cases} Z_i = X_i \otimes Y_i = \begin{cases} X_i - Y_i + M_x - 1, & X_i - Y_i \leq 1 \\ X_i - Y_i, & \text{其他} \end{cases} \\ \xi_i = Z_i / M_x, \quad i = 1, 2, 3, \dots \end{cases}$$

如果取 $\alpha=40\ 014, \beta=40\ 692, M_x=2\ 147\ 483\ 563, M_y=2\ 147\ 483\ 399$, 产生的随机数周期为 $(M_x-1)(M_y-1)/2$ 。

8. 复杂组合法

复杂组合法于 1990 年由 Marsaglia、Zaman 和 Tsang 设计。

基本思想：采用斐波那契序列和乘同余方法设置初值，再利用一个简单的算术序列和斐波那契序列给出随机数。步骤如下。

(1) 产生初始种子序列 $\{Y_n\}\{Z_n\}\{b_n\}$ ：

$$Y_n = Y_{n-3} * Y_{n-2} * Y_{n-1} \text{ mod } 179$$

$$Z_n = 53 * I_{n-1} \text{ mod } 169$$

$$b_n = \begin{cases} 0, & Y_n * Z_n \text{ mod } 64 < 32 \\ 1, & \text{其他} \end{cases}$$

初值 Y_0, Y_1, Y_2 为 $[1, 178]$ 区间内的任意整数，不能同时全为 1；初值 Z_0 为 $[0, 168]$ 区间的任意整数。由此产生 $M \times N$ 个 b_i ，即 $i = 1, 2, \dots, M \times N$ 。 M 表示所需的种子数， N 为二进制小数的位数。把种子转换为实型数：

$$X_i = \sum_{j=1}^N b_{(i-1)*N+j} * 2^{-j}, \quad i = 1, 2, \dots, M$$

(2) 产生简单算术序列 $\{C_n\}$ ：

$$C_n = C_{n-1} \otimes d = \begin{cases} C_{n-1} - d, & C_{n-1} \geq d \\ C_{n-1} - d + e, & \text{其他} \end{cases}$$

式中， d 等于任意整数除以 2^N 所得的数， e 等于对小于 2^N 的最大素数除以 2^N 所得的数， C_0 等于任意整数除以 2^N 所得的数，序列 $\{C_n\}$ 的周期为小于 2^N 的最大素数。

(3) 斐波那契序列及组合规则。

斐波那契序列为

$$X_n = X_{n-i_p} \otimes X_{n-j_p}$$

定义运算 \otimes 为

$$X \otimes Y = \begin{cases} x - y, & x \geq y \\ x - y + 1, & \text{其他} \end{cases}$$

(4) 组合产生随机数序列 $\{U_n\}$ ：

$$U_n = X_n \otimes C_n$$

这种方法产生的伪随机数序列有很长的周期且能通过各种统计检验，如果对(3)的斐波那契序列稍做改进即可得到周期更长的伪随机数序列。

3.2.4 伪随机数的检验

只有经过一系列检验，所产生的伪随机数序列才能作为随机数使用。重要的检验内容简述如下。

1. 均匀性检验

一般伪随机数序列可表示为 $\{\xi_n\}_{n=1}^N$ ，均匀偏度定义为

$$\Delta(N) = \sup_{0 \leq x \leq 1} \left| \frac{1}{N} \sum_{n=1}^N \eta(\xi_n < X) - X \right| \quad (3.2.11)$$