第 5 章模拟方法建模

韩建伟

2020/11/25

信息学院
hanjianwei@zjgsu.edu.cn

为什么要用模拟方法进行建模？

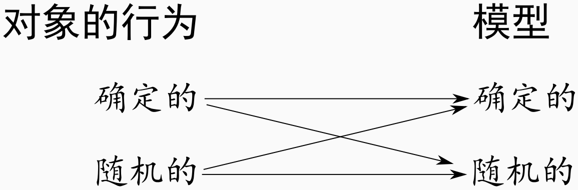
- 在某些情况下，对对象的行为进行直接观测或重复实验可能是不可行的，如早高峰时电梯系统提供的服务、大城市交通控制系统。
- 还有另一些情况，需要对可供选择的模式做试验的系统甚至可能不存在，如一座办公大楼的通讯网络选择、新工厂各台机器的布局、核电站事故防护和疏散方案选择。

模拟方法建模
当对象的行为不能做分析性的解释，或数据无法直接收集的情况下，建模者可以用某种方式间接地模拟其行为，试验所研究的供选择的各种方案，以估计它们怎样影响对象的行为，然后收集数据来确定哪种方案是最好的。

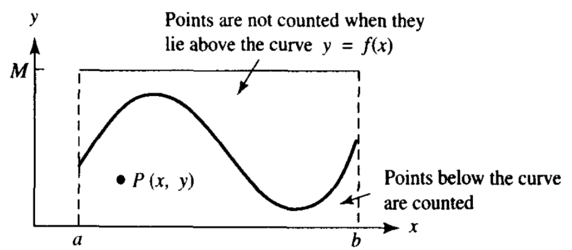
蒙特卡罗模拟

Wikipedia.org

也称统计模拟方法，是二十世纪四十年代中期由于科学技术的发展和电子计算机的发明，而被提出的一种以概率统计理论为指导的一类非常重要的数值计算方法。是指使用随机数（或更常见的伪随机数）来解决很多计算问题的方法。



确定行为的模拟：曲线下的面积



$$\frac{\text{曲线下的面积}}{\text{矩形面积}} \approx \frac{\text{曲线下的点数}}{\text{随机点的总数}}$$

计算面积的蒙特卡罗算法

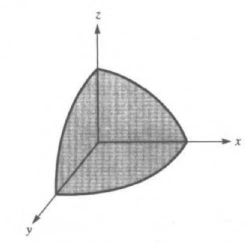
输入	模拟中产生的随机点总数 n
输出	$AREA =$ 给定区间 $a \le x \le b$ 上曲线 $y = f(x)$ 下的近似面积，其中 $0 \le f(x) \le M$.
第 1 步	初始化：$COUNTER = 0$
第 2 步	对 $i = 1, 2, ..., n$, 进行第 $3 \sim 5$ 步
第 3 步	计算随机坐标 x_i 和 y_i，满足 $a \le x_i \le b$，$0 \le y_i \le M$.
第 4 步	对随机坐标 x_i 计算 $f(x_i)$.
第 5 步	若 $y_i \le f(x_i)$，则 COUNTER 加 1；否则 COUNTER 不变
第 6 步	计算 $AREA = M(b-a)COUNTER/n$
第 7 步	输出 $(AREA)$
	停止

$y = cos(x)$ 的蒙特卡罗近似

Number of points	Approximation to area	Number of points	Approximation to area
100	2.07345	2000	1.94465
200	2.13628	3000	1.97711
300	2.01064	4000	1.99962
400	2.12058	5000	2.01429
500	2.04832	6000	2.02319
600	2.09440	8000	2.00669
700	2.02857	10000	2.00873
800	1.99491	15000	2.00978
900	1.99666	20000	2.01093
1000	1.96664	30000	2.01186

图 1: 区间 $-\pi/2 \leq x \leq \pi/2$ 上曲线 $y = cos(x)$ 下面积的蒙特卡罗近似

曲面下的体积



$$\frac{\text{曲面下的体积}}{\text{盒子体积}} \approx \frac{\text{第 1 卦限内曲面下的点数}}{\text{随机点的总数}}$$

随机数的生成

- 平方取中方法
- 线性同余
- C 语言中产生 0 到 1 之间的随机数

```c
#include <time.h>
#include <stdio.h>
#include <stdlib.h>

int main() {
  float rand_num;
  srand((unsigned)time(NULL));
  rand_num = (float)rand()/RAND_MAX;
  return 0;
}
```

随机行为的模拟

$$\text{概率} = \frac{\text{有效事件数}}{\text{事件的总数}}$$

- 抛一枚正规的硬币
- 掷一个正规的骰子
- 掷一个或一对不正规的骰子

一枚正规的硬币

设 x 是 $[0,1]$ 内的随机数，$f(x)$ 的定义如下：

$$f(x) = \begin{cases} \text{正面} & 0 \leq x \leq 0.5 \\ \text{反面} & 0.5 < x \leq 1 \end{cases}$$

抛正规硬币的蒙特卡罗算法

输入	模拟中产生的随机抛硬币的总次数 n
输出	抛硬币时得到正面的概率
第 1 步	初始化: $COUNTER = 0$
第 2 步	对 $i = 1, 2, ..., n$, 进行第 3 4 步
第 3 步	得到 $[0, 1]$ 内的随机数.
第 4 步	若 $0 \leq x_i \leq 0.5$，则 $COUNTER = COUNTER + 1$. 否则，$COUNTER$ 不变.
第 5 步	计算 $P(\text{正面}) = COUNTER/n$
第 6 步	输出正面的概率 $P(\text{正面})$
	停止

结果

Number of flips	Number of heads	Percent heads
100	49	0.49
200	102	0.51
500	252	0.504
1,000	492	0.492
5,000	2469	0.4930
10,000	4993	0.4993

扩展：如何用蒙特卡罗方法模拟掷一个正规的骰子？

掷一个不正规的骰子

x_i 的值	骰子的值	出现概率
$[0, 0.1]$	1	0.1
$(0.1, 0.2]$	2	0.1
$(0.2, 0.4]$	3	0.2
$(0.4, 0.7]$	4	0.3
$(0.7, 0.9]$	5	0.2
$(0.9, 1]$	6	0.1

存储模型：汽油与消费需求

问题

确定每隔多长时间及把多少汽油运送到各个加油站。每次运送汽油的费用 d 和运送量无关。加油站的需求量可看作常数。如何最小化每天的平均运费，并且每个加油站存储足够的汽油以满足消费需求。

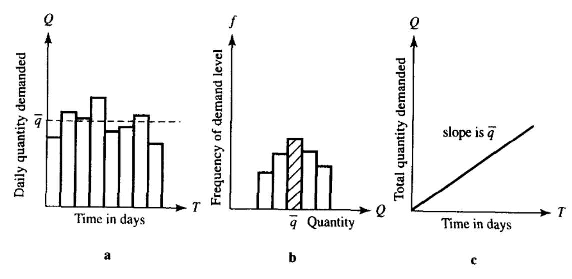
$$\text{日平均费用} = f(\text{存储费}, \text{运费}, \text{需求率})$$

运费 每次的运费 d 和运送量无关.

存储费 单位存储量的费用是常数.

需求量 每天的需求量可合理地看作常数.

常数需求率



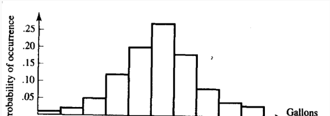
在第 13 章可以得到 (符号含义见课本):

$$T^* = \sqrt{\frac{2d}{sr}}$$

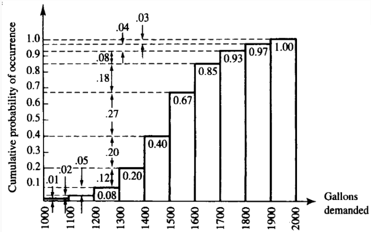
$$Q^* = rT^*$$

需求量的概率统计

需求量（加仑）	出现天数	出现概率
$1000 \sim 1099$	10	0.01
$1100 \sim 1199$	20	0.02
$1200 \sim 1299$	50	0.05
$1300 \sim 1399$	120	0.12
$1400 \sim 1499$	200	0.20
$1500 \sim 1599$	270	0.27
$1600 \sim 1699$	180	0.18
$1700 \sim 1799$	80	0.08
$1800 \sim 1899$	40	0.04
$1900 \sim 1999$	30	0.03
总计	1000	1.00



需求量的累计直方图



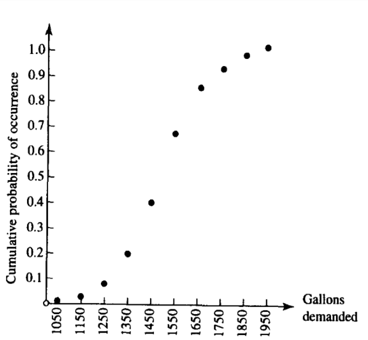
Random number	Corresponding demand	Percent occurrence
$0 \le x < 0.01$	1000–1099	0.01
$0.01 \le x < 0.03$	1100–1199	0.02
$0.03 \le x < 0.08$	1200–1299	0.05
$0.08 \le x < 0.20$	1300–1399	0.12
$0.20 \le x < 0.40$	1400–1499	0.20
$0.40 \le x < 0.67$	1500–1599	0.27
$0.67 \le x < 0.85$	1600–1699	0.18
$0.85 \le x < 0.93$	1700–1799	0.08
$0.93 \le x < 0.97$	1800–1899	0.04
$0.97 \le x \le 1.00$	1900–1999	0.03

(a) 需求子模型的累计直方图　　(b) 随机数与需求量的对应关系

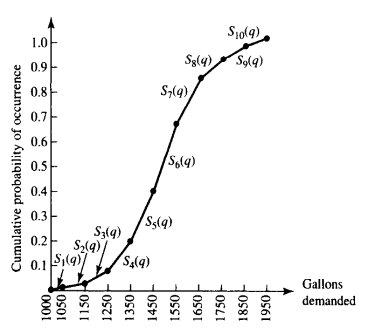
需求子模型的蒙特卡罗近似

Interval	Number of occurrences/expected no. of occurrences 1000 trials	 10,000 trials
1000–1099	8/10	91/100
1100–1199	16/20	198/200
1200–1299	46/50	487/500
1300–1399	118/120	1205/1200
1400–1499	194/200	2008/2000
1500–1599	275/270	2681/2700
1600–1699	187/180	1812/1800
1700–1799	83/80	857/800
1800–1899	34/40	377/400
1900–1999	39/30	284/300
	1000/1000	10,000/10,000

需求子模型的累计图



(a) 点表示



(b) 线性样条

蒙特卡罗存储算法术语一览

Q	汽油运量（加仑）
T	运送的时间间隔（天）
I	当前的存储量（加仑）
d	每次运送的费用（元）
s	每加仑汽油每天的存储费（元）
C	总费用
c	每天平均费用
N	模拟运行的天数
K	模拟的天数
x_i	$[0,1]$ 内的随机数
q_i	日需求量
$Flag$	用于中止计算的指标

蒙特卡罗存储算法

输入	Q, T, d, s, N
输出	c
第 1 步	初始化: $K = N$ $I = 0$ $C = 0$ $Flag = 0$
第 2 步	以一次运送开始下一存储周期: $I = I + Q$ $C = C + d$
第 3 步	确定该存储期的模拟是否中止: 若 $T \geq K$，置 $T = K, Flag = 1.$

蒙特卡罗存储算法

第 4 步	模拟这个存储期（或剩余部分）的每一天 对于 $i=1,2,...,T$，执行第 5 ~ 9 步.
第 5 步	生成随机数 x_i
第 6 步	利用需求子模型计算 q_i
第 7 步	修正当前的存储量 $I=I-q_i$
第 8 步	若存储量用完，即 $I\leq 0$，置 $I=0$，转第 9 步； 否则，计算日存储费和总费用： $C=C+I*s$
第 9 步	减少模拟中剩下的天数 $K=K-1$

蒙特卡罗存储算法

第 10 步	若 $Flag = 0$，转第 2 步；否则，转第 11 步
第 11 步	计算每天平均费用 $c = C/N$
第 12 步	输出 c
	停止

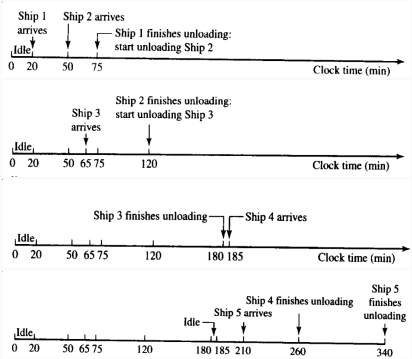
排队模型

港口系统
一个只能同时为 1 只船卸货的港口，相邻两艘船到达的时间间隔在 15 分钟到 145 分钟之间变化，每艘船的卸货时间在 45 分钟到 90 分钟之间变化，需要回答以下问题：

- 每艘船只在港口的平均时间和最长时间是多少？
- 平均等待时间和最长等待时间是多少？
- 卸货设备空闲的时间百分比是多少？
- 船只排队最长的长度是多少？

5 艘船只的例子

	Ship 1	Ship 2	Ship 3	Ship 4	Ship 5
Time between successive ships	20	30	15	120	25
Unload time	55	45	60	75	80



模拟概要

Figure 5.9
Idle and unloading times for the ships and docking facilities

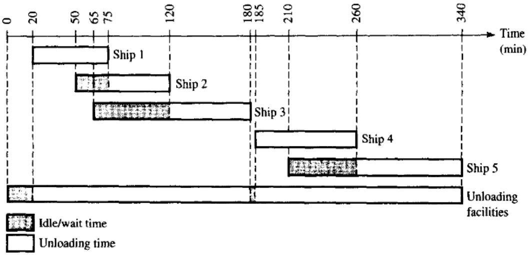


Table 5.14 Summary of the harbor system simulation

Ship no.	Random time between ship arrivals	Arrival time	Start service	Queue length at arrival	Wait time	Random unload time	Time in harbor	Dock idle time
1	20	20	20	0	0	55	55	20
2	30	50	75	1	25	45	70	0
3	15	65	120	2	55	60	115	0
4	120	185	185	0	0	75	75	5
5	25	210	260	1	50	80	130	0
Totals (if appropriate):					130			25
Averages (if appropriate):					26	63	89	

Note: All times are given in minutes after the start of the clock at time $t = 0$.

港口系统算法术语一览

$between_i$	船 i 与 $i-1$ 的到达时间间隔 (15-145 之间的随机整数)
$arrive_i$	从时钟 t=0 分开始计时，船 i 到达港口的时间
$unload_i$	船 i 在港口卸货所需的时间 (45 和 90 之间的随机整数)
$start_i$	船 i 开始卸货的时间
$idle_i$	恰在船 i 开始卸货之前码头设备空闲的时间
$wait_i$	船 i 到达后开始卸货前在码头的等待时间
$finish_i$	船 i 卸货完毕的时间
$harbor_i$	船 i 呆在港口的总的时间

港口系统算法术语一览

$HARTIME$	每艘船呆在港口的平均时间
$MAXHAR$	一艘船呆在港口的最长时间
$WAITIME$	每艘船卸货之前的平均等待时间
$MAXWAIT$	一艘船的最长等待时间
$IDLETIME$	卸货设备空闲时间占总模拟时间的百分比

港口系统模拟算法

输入	模拟中的船只总数 n
输出	$HARTIME$, $MAXHAR$, $WAITIME$, $MAXWAIT$, $IDLET$
第 1 步	随机生成 $between_1$ 和 $unload_1$，令 $arrive_1 = between_1$
第 2 步	全部输出初始化： $HARTIME = unload_1, MAXHAR = unload_1$ $WAITME = 0, MAXWAIT = 0, IDLETIME = arrive_1$
第 3 步	计算船 1 卸货完毕的时间： $finish_1 = arrive_1 + unload_1$
第 4 步	对于 $i = 2, 3, ..., n$, 执行第 5-16 步
第 5 步	分别在各自的区间上生成一对随机整数 $between_i$ 和 $unlo$

港口系统模拟算法

第 6 步	假定时钟从 $t = 0$ 分钟开始计时，计算船 i 的到达时间 $arrive_i = arrive_{i-1} + between_i$
第 7 步	计算船 i 到达与船 $i-1$ 卸货完毕的时间之差： $timediff = arrive_i - finish_{i-1}$
第 8 步	若 timediff 非负，则卸货设备空闲： $idlei = timediff$ 且 $waiti = 0$ 若 timediff 为负，则船 i 在卸货前必须等待： $waiti = -timediff$ 且 $idlei = 0$
第 9 步	计算船 i 开始卸货的时间 $starti = arrive_i + waiti$

港口系统模拟算法

第 10 步	计算船 i 卸货完毕的时间 $finish_i = start_i + unload_i$
第 11 步	计算船只呆在港口的时间 $harbor_i = wait_i + unload_i$
第 12 步	将 $harbor_i$ 加入总的港口时间 $HARTIME$，平均时用到
第 13 步	若 $harbor_i > MAXHAR$，则令 $MAXHAR = harbor_i$; 否则 $MAXHAR$ 不变
第 14 步	将 $wait_i$ 加入总的等待时间 $WAITIME$，平均时用到
第 15 步	将 $idle_i$ 加入总的空闲时间 $IDLETIME$
第 16 步	若 $wait_i > MAXWAIT$，则令 $MAXWAIT = wait_i$; 否则 $MAXWAIT$ 不变

港口系统模拟算法

第 17 步	令 $HARTIME = HATIME/n, WAITIME = WAITIME/n$ 且 $IDLETIME = IDLETIME/finish_n$
第 18 步	输出 $(HARTIE, MAXHAR, WAITIME,$ $MAXWAIT, IDLETIME)$
	停止

模拟结果

Average time of a ship in the harbor	106	85	101	116	112	94
Maximum time of a ship in the harbor	287	180	233	280	234	264
Average waiting time of a ship	39	20	35	50	44	27
Maximum waiting time of a ship	213	118	172	203	167	184
Percentage of time dock facilities are idle	0.18	0.17	0.15	0.20	0.14	0.21

Note: All times are given in minutes. Time between successive ships is 15–145 min. Unloading time per ship varies from 45 to 90 min.

Average time of a ship in the harbor	74	62	64	67	67	73
Maximum time of a ship in the harbor	161	116	167	178	173	190
Average waiting time of a ship	19	6	10	12	12	16
Maximum waiting time of a ship	102	58	102	110	104	131
Percentage of time dock facilities are idle	0.25	0.33	0.32	0.30	0.31	0.27

Note: All times are given in minutes. Time between successive ships is 15–145 min. Unloading time per ship varies from 35 to 75 min.

Average time of a ship in the harbor	114	79	96	88	126	115
Maximum time of a ship in the harbor	248	224	205	171	371	223
Average waiting time of a ship	57	24	41	35	71	61
Maximum waiting time of a ship	175	152	155	122	309	173
Percentage of time dock facilities are idle	0.15	0.19	0.12	0.14	0.17	0.06

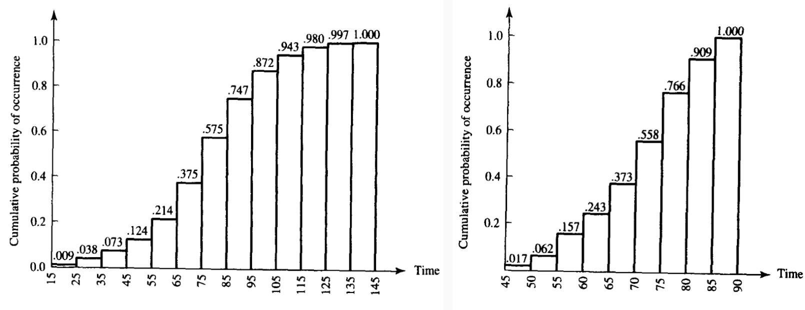
Note: All times are given in minutes. Time between successive ships is 10–120 min. Unloading time per ship varies from 35 to 75 min.

搜集港口数据

Time between arrivals	Number of occurrences	Probability of occurrence	Unloading time	Number of occurrences	Probability of occurrence
15–24	11	0.009			
25–34	35	0.029			
35–44	42	0.035	45–49	20	0.017
45–54	61	0.051	50–54	54	0.045
55–64	108	0.090	55–59	114	0.095
65–74	193	0.161	60–64	103	0.086
75–84	240	0.200	65–69	156	0.130
85–94	207	0.172	70–74	223	0.185
95–104	150	0.125	75–79	250	0.208
105–114	85	0.071	80–84	171	0.143
115–124	44	0.037	85–90	109	0.091
125–134	21	0.017		1200	1.000
135–145	3	0.003			
	1200	1.000			

Note: All times are given in minutes.

累计直方图



a. Time between arrivals

b. Unloading time

基于直方图的模拟结果

Average time of a ship in the harbor	108	95	125	78	123	101
Maximum time of a ship in the harbor	237	188	218	133	250	191
Average waiting time of a ship	38	25	54	9	53	31
Maximum waiting time of a ship	156	118	137	65	167	124
Percentage of time dock facilities are idle	0.09	0.09	0.08	0.12	0.06	0.10

Note: Based on the data exhibited in Table 5.18. All times are given in minutes.

扩展

早高峰时间的电梯调度算法.