

第 2 章 概率统计基础

在许多仿真过程中，事件的发生是随机的，或者事件属性值的确定具有偶然性，如顾客到达间隔时长、顾客接受服务时长或机器设备故障间隔等。由于这些随机因素的存在，在建模过程中就需要用服从各种分布的随机变量来描述系统中存在的随机和偶然性问题。

随机变量在生产物流系统运作中的存在具有普遍性，因此掌握随机变量的有关知识对理解和分析系统的行为和结果具有重要的作用。本章首先对仿真中的概率统计基本概念做简要说明，然后介绍系统仿真中常见的概率分布形式及其数学表达式，最后对仿真中的随机变量输入模型构建的步骤及过程中所用的方法进行阐述。

2.1 概率统计的基本概念

2.1.1 确定事件和随机事件

确定事件：在给定条件下进行的试验中，一定发生或一定不发生的事件分别称为必然事件和不可能事件，这类事件是确定性的，总称为确定事件。

随机事件：在给定条件下进行的试验中，可能发生也可能不发生，而在大量重复试验中却具有某种规律性的事件，称为随机事件。

2.1.2 随机变量与概率

随机变量：如果试验的每个结果用变量 X 的一个值来表示，即 X 的值根据试验结果来确定，因而它取什么值是随机的，而且对任意实数 x ， $X < x$ 是一个随机事件，这种变量称为随机变量。

频率：如果对某项试验重复进行了 n 次，事件 A 发生了 m 次，则称 A 在这次试验中的频率为 m/n 。在大量重复某项试验时，就会发现事件的频率在试验次数很大而且不断增大的过程中呈现稳定性。这种统计规律性表明：事件发生的可能性大小是事件本身固有的客观属性。

概率：事件 A 发生可能性的大小为事件 A 的概率，记为 $P(A)$ 。当试验次数 n 足够大时，可以用事件的频率作为事件概率的近似值，即

$$P(A) \approx m/n$$

概率是一个在区间 $[0, 1]$ 上取值的实数。

随机变量 X 取值小于实数 x 的可能性大小，即随机事件 $\{X < x\}$ 发生的概率，介于数 0 和 1 之间：

$$0 \leq P(X < x) \leq 1$$

2.1.3 连续型随机变量的概率分布函数

对于随机变量 X , 事件 $\{X < x\}$ 发生的概率 $P(X < x)$ 是实变量 x 的函数, 称之为 X 的概率分布函数, 简称为分布函数, 记为 $F(x)$, 即

$$F(x) = P\{X < x\} \quad (-\infty < x < +\infty)$$

2.1.4 离散型随机变量的概率分布函数

在实际系统中, 有些随机变量取值的数量有限, 在整个区间内呈现离散型特征, 如单位时间内到达某个公交站台的人数、每个订单采购产品的数量等。若离散型随机变量 X 的可能取值为 x_1, x_2, \dots, x_n , 则事件 $\{X = x_k\}$ 为一随机事件, 其发生的概率记作:

$$P_k = P\{X = x_k\}, k = 1, 2, \dots, n$$

P_k 称之为离散型随机变量 X 的概率分布, 且 P_k 满足下列条件表达式:

$$\begin{cases} P_k \geq 0 \\ \sum_{k=1}^n P_k = 1 \end{cases}$$

2.1.5 随机变量概率密度函数

如果随机变量 X 的分布函数 $F(x)$ 能表示成某个函数 $p(x)$ 的积分

$$F(x) = \int_{-\infty}^x p(x) dx$$

就称 X 为连续型随机变量, 并称 $p(x)$ 为 X 的分布密度函数, 简称密度函数。

2.1.6 随机变量的数学特征

随机变量 X 的数学期望记为 $E(X)$, 表示 X 取值的平均值。

若 X 为离散型随机变量, 其可能取值为 x_1, x_2, \dots , 且 $P\{X = x_k\} = p_k$, 则当级数

$$\sum_{k=1}^{+\infty} x_k p_k$$

绝对收敛时, X 的数学期望为:

$$E(X) = \sum_{k=1}^{+\infty} x_k p_k$$

若 X 为连续型随机变量, 其分布密度函数为 $p(x)$, 则当

$$\int_{-\infty}^{+\infty} xp(x) dx$$

绝对收敛时, X 的数学期望为:

$$E(X) = \int_{-\infty}^{+\infty} xp(x) dx$$

随机变量 $[X - E(X)]^2$ 的数学期望 $E[X - E(X)]^2$ 称为 X 的方差, 记为 $D(X)$ 。 $D(X)$ 的平方根称为 X 的均方差(或标准差)。方差和均方差描述随机变量 X 取值相对于其均值 $E(X)$ 的分散程度, 反映了数据波动的大小。

对于离散型随机变量,

$$D(X) = \sum_{k=1}^{+\infty} [x_k - E(X)]^2 p_k$$

对于连续型随机变量,

$$D(X) = \int_{-\infty}^{+\infty} [x_k - E(X)]^2 p_k dx$$

2.2 系统仿真中常用的随机分布

2.2.1 离散分布

1. 0-1 分布

设离散型随机变量 X 只能取 0 和 1 两个值, 其概率分布是:

$$P(X=1) = p, P(X=0) = 1-p, (0 < p < 1)$$

则称 X 服从 (0-1) 分布。

X 的均值和方差分别为:

$$E(X) = 1 \times P(X=1) + 0 \times P(X=0) = 1 \times p + 0 \times (1-p) = p$$

$$D(X) = (1-p)^2 \times p + (0-p)^2 \times (1-p) = p(1-p)$$

2. 贝努利试验及二项分布

令试验 E 的每次结果只能有两种可能的结果, 或者成功 (记为 A) 或者失败 (记为 \tilde{A}), 其中 $P(A) = p, P(\tilde{A}) = 1-p = q (0 < p < 1)$, 将 E 独立地重复进行 n 次, 则称这 n 次重复独立的试验为 n 重贝努利 (Bernoulli) 试验, 简称贝努利试验。

令随机变量 X 表示 n 重贝努利试验中事件 A 发生的次数, 则 A 恰好发生 $k (0 \leq k \leq n)$ 次的概率为:

$$P(X=k) = \begin{cases} C_n^k p^k q^{n-k} & k = 0, 1, 2, \dots, n \\ 0 & \text{else} \end{cases}$$

由于 $C_n^k p^k q^{n-k}$ 刚好是二项式 $(p+q)^n$ 展开式中的第 $k+1$ 项, 因此也称随机变量 X 服从参数为 n, p 的二项分布, 记为 $X \sim B(n, p)$, 二项分布的概率质量函数 (以下简称 pmf) 曲线, 如图 2.1 所示。

其中, 当 $n=1$ 时, 二项分布即为 (0-1) 分布。

内 容 简 介

本书在论述系统建模与仿真基础理论的基础上,以一个完整的车辆生产物流系统案例作为研究对象,阐述系统建模与仿真技术在采购、生产、物流和销售过程中的应用。全书共分为三个部分,第一部分介绍系统建模与仿真,尤其是离散事件系统建模与仿真的一般理论和概率统计基础知识,第二部分介绍仿真平台 Witness 的特点和建模过程,第三部分介绍车辆生产物流系统整个运作流程的建模与仿真的实战技术和方法。

本书可作为高等院校工业工程、物流管理、工商管理、制造工程等专业的本科生或研究生的仿真课程教材,也可作为相关工程技术人员进行仿真学习和研究的参考资料。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究。

图书在版编目(CIP)数据

生产物流系统建模与仿真 / 王建华, 黄贤凤编著. —北京: 电子工业出版社, 2014.9
(华信经管创优系列)

ISBN 978-7-121-24302-8

I. ①生… II. ①王… ②黄… III. ①企业管理—物流—生产管理—系统建模—高等学校—教材
②企业管理—物流—生产管理—系统仿真—高等学校—教材 IV. ①F273.4-39

中国版本图书馆 CIP 数据核字(2014)第 209631 号

策划编辑: 王二华

责任编辑: 王二华

印 刷:

装 订:

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编: 100036

开 本: 787×1092 1/16 印张: 19 字数: 457 千字

版 次: 2014 年 9 月第 1 版

印 次: 2018 年 1 月第 2 次印刷

定 价: 38.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010)88254888, 88258888。

质量投诉请发邮件至 zlbs@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式: (010)88254532。

第 2 章 概率统计基础

在许多仿真过程中，事件的发生是随机的，或者事件属性值的确定具有偶然性，如顾客到达间隔时长、顾客接受服务时长或机器设备故障间隔等。由于这些随机因素的存在，在建模过程中就需要用服从各种分布的随机变量来描述系统中存在的随机和偶然性问题。

随机变量在生产物流系统运作中的存在具有普遍性，因此掌握随机变量的有关知识对理解和分析系统的行为和结果具有重要的作用。本章首先对仿真中的概率统计基本概念做简要说明，然后介绍系统仿真中常见的概率分布形式及其数学表达式，最后对仿真中的随机变量输入模型构建的步骤及过程中所用的方法进行阐述。

2.1 概率统计的基本概念

2.1.1 确定事件和随机事件

确定事件：在给定条件下进行的试验中，一定发生或一定不发生的事件分别称为必然事件和不可能事件，这类事件是确定性的，总称为确定事件。

随机事件：在给定条件下进行的试验中，可能发生也可能不发生，而在大量重复试验中却具有某种规律性的事件，称为随机事件。

2.1.2 随机变量与概率

随机变量：如果试验的每个结果用变量 X 的一个值来表示，即 X 的值根据试验结果来确定，因而它取什么值是随机的，而且对任意实数 x ， $X < x$ 是一个随机事件，这种变量称为随机变量。

频率：如果对某项试验重复进行了 n 次，事件 A 发生了 m 次，则称 A 在这次试验中的频率为 m/n 。在大量重复某项试验时，就会发现事件的频率在试验次数很大而且不断增大的过程中呈现稳定性。这种统计规律性表明：事件发生的可能性大小是事件本身固有的客观属性。

概率：事件 A 发生可能性的大小为事件 A 的概率，记为 $P(A)$ 。当试验次数 n 足够大时，可以用事件的频率作为事件概率的近似值，即

$$P(A) \approx m/n$$

概率是一个在区间 $[0, 1]$ 上取值的实数。

随机变量 X 取值小于实数 x 的可能性大小，即随机事件 $\{X < x\}$ 发生的概率，介于数 0 和 1 之间：

$$0 \leq P(X < x) \leq 1$$

在本例的直方图制作中，分组区间宽度是 1，每个区间内数据的频数是实验结果的频次，可以运用表 2.3 的数据直接生成如图 2.18 所示的直方图。从直方图与常用的随机分布概率质量曲线对比可以初步判断，1000 件产品批中不合格品数量可能服从二项分布或泊松分布。

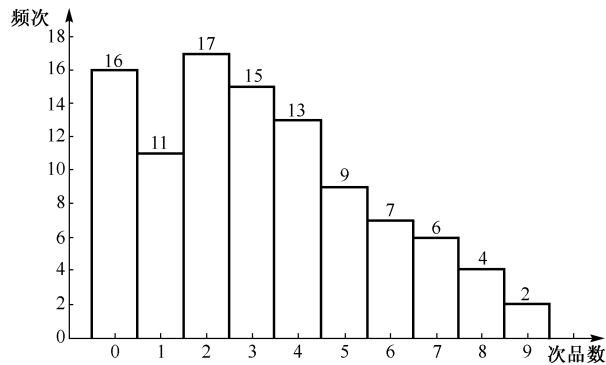


图 2.18 每批 1000 件产品中的次品数直方图



例 2.14 制作连续数据直方图。

电子元器件寿命试验中通常采用恒定压力加速试验，可以通过增加电压、温度、湿度等来进行。在某个电子元件加速试验中，采取试验温度 250℃，获得 81 件产品的寿命记录如下（单位：小时）：



（例题数据表）

2780	38684	16310	5583	58	37076	29919	13677	852
8035	43795	27286	4352	38232	13753	2101	38662	10674
8975	9501	3455	11960	12224	20214	8179	21412	10723
1297	3743	17509	37186	33573	3346	8971	31395	16283
18785	1024	9565	22112	583	12610	28970	25400	1580
4172	30614	26968	1360	80614	1404	2384	21410	14944
4241	8463	7613	28221	3825	25569	7121	13110	32946
14496	11345	20036	2871	3026	19988	15707	45040	3722
8489	66466	6368	9289	2111	28910	10005	1535	14134

由于产品寿命是连续变量，需要进行区间的划分。全部试验数据跨度区间为[58, 80614]，样本总数为 81 件，可以将数据划分成 9 个区间，单个区间跨度为： $(80614-58)/9=8950.7$ 。每个区间的最小值和最大值及试验数据出现的频次如表 2.4 所示。

表 2.4 电子产品加速寿命试验结果分组数据表

序号	1	2	3	4	5	6	7	8	9
min	58	9009	17959	26910	35861	44812	53762	62713	71664
max	9009	17959	26910	35861	44812	53762	62713	71664	80614
频次	33	20	9	10	6	1	0	1	1

(5) 随机分布函数若为连续分布, 利用随机分布函数 $F(x)$ 计算随机变量 X 落在每个分组区间理论概率 $p_i = F(U_i) - F(L_i)$, 并得出了 np_i 值; 随机分布函数若为离散分布, 利用概率质量函数直接计算出每个分组区间的理论概率 $p_i = \sum_{j=L_i}^{U_i} P(X = j)$ 。

(6) 计算 np_i , 如果有些 $np_i < 5$, 将其余相邻的组合并再计算, 获得有效组数 k' ;

(7) 计算 $\frac{(np_i - f_i)^2}{np_i}$, 计算 $\sum_{i=1}^{k'} \frac{(np_i - f_i)^2}{np_i}$ 。

卡方检验临界值 $\chi_{0.05}^2(k' - r - 1)$ 的值可以通过统计学书籍后附录的卡方检验临界值表查询, 也可以通过使用 Excel 中的函数 CHISQ.INV() 来获得, 该函数的使用在例 2.13 和例 2.14 中给出。因为检验中仅有一个随机变量 X , 因此 $r=1$, 再通过查表或 Excel 函数计算获得临界值后, 将临界值同 $\sum_{i=1}^{k'} \frac{(np_i - f_i)^2}{np_i}$ 比较大小, 如果前者大于后者, 则接受随机样本来自于指数分布的假设, 否则拒绝假设。

3. 卡方检验泊松分布

以例 2.13 产品检验数据为例说明如何使用卡方检验进行泊松分布的拟合优度检验 ($\alpha = 0.05$), 主要步骤如下。

- (1) 从数据的直方图来看它可能服从泊松分布。
- (2) 通过计算样本期望, 可以估计参数 $\hat{\lambda} = 3.21$ 。
- (3) 可以设计如下的假设:

H0: 随机变量服从参数 $\hat{\lambda} = 3.21$ 的泊松分布;

H1: 随机变量不服从参数 $\hat{\lambda} = 3.21$ 的泊松分布。

(4) 分组和计算如表 2.6 所示。

其中, 泊松分布的 pmf 如下式:

$$P(X = k) = \begin{cases} \frac{e^{-\lambda} \lambda^k}{k!} & k = 0, 1, 2, \dots \\ 0 & \text{else} \end{cases}$$

表 2.6 卡方检验泊松分布数据处理表

x_i	f_i	p_i	np_i	$f_i - np_i$	$(f_i - np_i)^2 / np_i$
0	16	0.040	4.036	10.010	5.897
1	11	0.130	12.954		
2	17	0.208	20.792	-3.792	0.692
3	15	0.222	22.247	-7.247	2.361
4	13	0.179	17.854	-4.854	1.319
5	9	0.115	11.462	-2.462	0.529
6	7	0.061	6.132	8.345	6.536
7	6	0.028	2.812		
8	4	0.011	1.128		
9	2	0.006	0.583		

其中：因为单项 np_i 小于 5，所以 $x=0$ 和 1 进行合并； $x \geq 6$ 进行合并。最终分组数 $k=6$ ，变量为 1，则该卡方检验的自由度为 4，检验统计量为最后一列值之和 17.334。

(5) 结论：

使用显著性水平 0.05，自由度为 4 的 Excel 卡方函数获得临界值，即在 Excel 任意单元格输入：“=CHISQ.INV(1-0.05,4)”，可以获得临界值 $\chi_{0.05}^2(4)=9.488$ ，远小于 17.334，所以拒绝原假设，即数据不服从参数 $\hat{\lambda}=3.21$ 的泊松分布。因此，分析人员需要寻找一种更合适的模型或使用经验分布来进行拟合。

4. 卡方检验指数分布

以例 2.14 产品寿命数据为例说明如何使用卡方检验进行指数分布的拟合优度检验 ($\alpha=0.05$)，主要步骤如下。

(1) 从数据的直方图来看它可能服从指数分布。

(2) 通过计算样本期望，可以估计参数 $\hat{\lambda}=1/16184$ 。

(3) 可以设计如下的假设：

H0: 随机变量服从参数 $\hat{\lambda}=1/16184$ 的指数分布；

H1: 随机变量不服从参数 $\hat{\lambda}=1/16184$ 的指数分布。

(4) 分组和计算如表 2.7 所示。

(5) 在 Excel 任意单元格输入：“=CHISQ.INV(1-0.05,3)”，获得临界值 $\chi_{0.05}^2(3)=7.815$ ，表 2.7 最后一列求和获得检验统计量为 2.414，由于 $2.414 < 7.815$ ，所以接受假设 H0，即产品寿命服从均值为 1/16184 的指数分布。

表 2.7 卡方检验指数分布数据处理表

序号	L_i	U_i	f_i	p_i	np_i	$f_i - np_i$	$(f_i - np_i)^2 / np_i$
1	$-\infty$	9009	33	0.427	34.557	1.577	0.072
2	9009	17959	20	0.243	19.721	0.279	0.004
3	17959	26910	9	0.140	11.343	2.343	0.484
4	26910	35861	10	0.081	6.524	3.476	1.851
5	35861	44812	6	0.046	3.753	0.166	0.003
6	44812	53762	1	0.027	2.159		
7	53762	62713	0	0.015	1.242		
8	62713	71664	1	0.009	0.714		
9	71664	$+\infty$	1	0.012	0.967		

2.3.5 选择无数据的输入模型

在进行系统输入模型构建过程中，有时候能够获得数据，但是有时候难以获得有效的数据，如实际系统根本不存在、收集数据太昂贵、数据支离破碎或没有相应的协作关系去收集数据。在这种情况下，难以采用前述过程来确定输入模型，这时不得不依靠假设或猜想来设计输入模型。

无数据情况下输入模型选择可以考虑如下方法。

这样，随机数就无法继续产生了。

利用平方取中法的另一个问题是这个方法可能产生退化，即总是得到相同的 x_i 值。



例 3.3 设在产生四位数的随机数过程中，得到了一个 x_i 值为 4500，即

$$x_i = 4500$$

从而， $(4500)^2 = 20250000$; $x_{i+1} = 2500$

$$(2500)^2 = 06250000; \quad x_{i+2} = 2500$$

相继产生的数值总为 2500。

由于这些原因，平方取中法已经被许多新的、能提供更可靠的随机数序列的算法所取代。

3.2.2 线性同余法

线性同余法在 1951 年由莱默尔 (Lehmer) 首先提出。目前大多数随机数发生器都采用这种方法。在这个算法中，随机数序列中的数由如下的递推关系产生：

$$x_{n+1} = (ax_n + c) \bmod m \quad n \geq 0$$

初始值 x_0 称为种子，常数 a 称为乘子，常数 c 称为增量，而常数 m 称为模数。

在上式中，当 $c=0$ 时，该算法称为乘同余法；当 $c \neq 0$ 时，该算法称为混合同余法。从例 3.4 中可以看出利用这种方法产生的序列的重复性，一般来讲任何由此方法产生的序列都存在重复性。但在大多数情况下，合理地选择常数 a 、 c 、 x_0 和 m ，可以使重复周期充分的长。

通过线性同余法获得的随机数序列有：

$$0 \leq x_i \leq m-1 \quad \forall i$$

为了得到 $[0,1)$ 区间或 $[0,1]$ 区间上分布的随机数 R_i ，分别使用下式进行转换：

$$R_i = \frac{x_i}{m} \quad R_i = \frac{x_i}{m-1}$$



例 3.4 设 $a=5$ ， $c=3$ ， $m=16$ ，取 $x_0=7$ ，利用线性同余法产生 $[0,1)$ 区间随机数序列，计算过程如下：

$$\begin{aligned} x_0 &= 7 \\ x_1 &= (5 \times 7 + 3) \bmod 16 = 6 & R_1 &= 6/16 = 0.375 \\ x_2 &= (5 \times 6 + 3) \bmod 16 = 1 & R_2 &= 1/16 = 0.063 \\ x_3 &= (5 \times 1 + 3) \bmod 16 = 8 & R_3 &= 8/16 = 0.500 \\ x_4 &= (5 \times 8 + 3) \bmod 16 = 11 & R_4 &= 11/16 = 0.688 \\ x_5 &= (5 \times 11 + 3) \bmod 16 = 10 & R_5 &= 10/16 = 0.625 \\ x_6 &= (5 \times 10 + 3) \bmod 16 = 5 & R_6 &= 5/16 = 0.313 \end{aligned}$$

表 3.7 科尔莫戈洛夫-斯米尔诺夫法进行均匀随机数检验算例表

i	1	2	3	4	5	6	7	8	9	10
L_i	0.000	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900
U_i	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900	1.000
R_i'	0.123	0.136	0.393	0.417	0.634	0.655	0.663	0.915	0.961	0.987
D_i'	0.123	0.036	0.193	0.117	0.234	0.155	0.063	0.215	0.161	0.087
D_i^+	0.023	0.064	0.093	0.017	0.134	0.055	0.037	0.115	0.061	0.013

3.4.3 科尔莫戈洛夫-斯米尔诺夫检验正态分布

假设通过特定的方法产生了 20 个 $N = (5, 3^2)$ 的正态分布随机数，具体样本数据如下：

8.371 5.789 5.892 5.675 3.174 6.377 5.544 4.506 1.079 7.086
 5.196 7.699 4.934 4.575 3.140 6.415 6.581 6.554 7.377 7.552

试在 $\alpha = 0.10$ 下，检验这些数字是否为来自于 $N = (5, 3^2)$ 的正态分布随机数。

由于科尔莫戈洛夫-斯米尔诺夫检验主要通过检验随机分布的均匀特性来判断样本是否满足某个特定的随机分布，因此需要从正态分布中提取均匀性特征。为了解决这个问题，可以通过观察正态分布概率密度曲线和概率分布曲线（如图 3.7 所示），从图中可以看出，正态分布的概率分布曲线是处于 $[0, 1]$ 区间的，因此，检验过程需要将样本值转换为概率分布函数值。

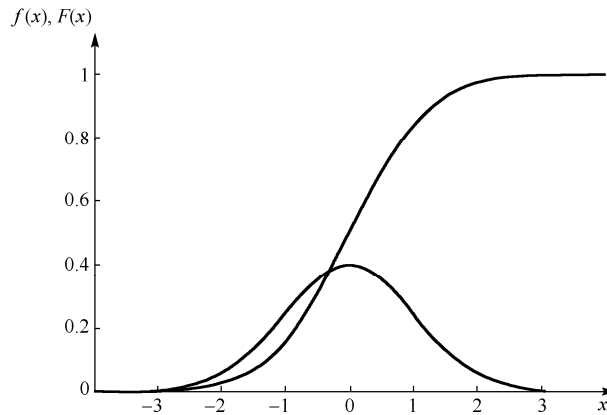


图 3.7 标准正态分布概率密度曲线和概率分布曲线

因此，利用科尔莫戈洛夫-斯米尔诺夫法进行正态分布假设检验，首先将样本由小到大排序，形成排序后的样本 R'_1, R'_2, \dots, R'_N ，然后计算按照下式计算对应样本的 $F(x)$ 值，令其为 W_i ：

$$W_i = \int_{-\infty}^{R'_i} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(R'_i - u)^2}{2\sigma^2}} dx$$

由于总的样本数量为 20 个，因此检验区间需要划分为 20 个，则得到每个区间的上下限值如下：

$$L = \{0.00, 0.05, 0.10, 0.15, \dots, 0.95\}$$

$$U = \{0.05, 0.10, 0.15, 0.20, \dots, 1.00\}$$

通过计算样本值映射至分布函数上的值 W_i 与其理论区间上下限之间的差值，以获取检验统计量 D ，具体计算表格如表 3.8 所示。

表 3.8 科尔莫戈洛夫-斯米尔诺夫法进行正态分布随机数检验算例表

i	1	2	3	4	5	6	7	8	9	10
L_i	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
U_i	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
R_i'	1.079	3.140	3.174	4.506	4.575	4.934	5.196	5.544	5.675	5.789
W_i	0.096	0.268	0.271	0.435	0.444	0.491	0.526	0.572	0.589	0.604
D_i^-	0.096	0.218	0.171	0.285	0.244	0.241	0.226	0.222	0.189	0.154
D_i^+	0.046	0.168	0.121	0.235	0.194	0.191	0.176	0.172	0.139	0.104
i	11	12	13	14	15	16	17	18	19	20
L_i	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
U_i	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1
R_i'	5.892	6.377	6.415	6.554	6.581	7.086	7.377	7.552	7.699	8.371
W_i	0.617	0.677	0.681	0.698	0.701	0.757	0.786	0.803	0.816	0.869
D_i^-	0.117	0.127	0.081	0.048	0.001	0.007	0.014	0.047	0.084	0.081
D_i^+	0.067	0.077	0.031	0.002	0.049	0.043	0.064	0.097	0.134	0.131

根据表 3.8 可以获得检验统计量 D 为 0.285，通过查附录科尔莫戈洛夫-斯米尔诺夫检验临界值表，获得临界值 $D_\alpha (n=20)$ 为 $0.264 < D$ ，所以拒绝该样本服从 $N[5,3^2]$ 的正态分布。



思考题

1. 如何把在区间 $[0, 1]$ 上均匀分布的随机数转换成在区间 $[11, 17]$ 上均匀分布的随机数？
2. 使用线性同余法产生 6 个两位随机整数序列，令 $x_0=37$ ， $a=12$ ， $c=57$ ， $m=100$ 。
3. 在第 2 题中，如果取 $x_0=0$ ，将会出现什么问题？
4. 使用乘同余法产生 6 个三位随机整数序列。令 $x_0=117$ ， $a=53$ ， $m=1100$ 。
5. 使用平方取中法产生 4 位随机整数序列，如果种子数为 7117，该序列什么时候产生蜕化？什么时候产生循环？
6. 产生的随机数列如下：

0.42 0.08 0.80 0.91 0.83 0.56 0.66 0.72 0.04 0.89 0.49

使用科尔莫戈洛夫-斯米尔诺夫检验法检验该序列是否服从 $[0, 1]$ 区间上的均匀分布？取 $\alpha=0.05$ 。

7. 分析下面这些线性同余法能否达到最大周期，并给出达到最大周期的 X_0 的限制：
 - (a) 混合同余法，其中 $a=2\ 814\ 749\ 767\ 109$ ， $c=59\ 482\ 661\ 568\ 307$ ， $m=2^{48}$
 - (b) 乘同余法，其中 $a=69\ 069$ ， $c=0$ ， $m=2^{32}$
 - (c) 混合同余法，其中 $a=4951$ ， $c=247$ ， $m=256$
 - (d) 乘同余法，其中 $a=6507$ ， $c=0$ ， $m=1024$
8. 编写一个计算机程序，使用线性同余法产生四位随机整数序列，允许用户输入 X_0 ， a ， c ， m 的值。
9. 开发一个用于生成具有如下形式 pdf 的随机变量 X 的随机发生器：

续表

A	B	C	D	E	F	G	H	I
		仿真钟		仿真钟		仿真钟		
卡车序号	到达间隔时间 (分钟)	到达时刻	接受服务时长 (分钟)	服务开始时刻	等待时长 (分钟)	服务结束时刻	通过系统时间 (分钟)	叉车空闲时间 (分钟)
17	10.2	94.6	4.6	94.6	0.0	99.2	4.6	0.2
18	0.6	104.9	2.9	104.9	0.0	107.8	2.9	5.6
19	1.9	105.5	7.0	107.8	2.3	114.8	9.4	0.0
20	18.6	107.4	4.5	114.8	7.5	119.3	12.0	0.0
...
100	2.2	0.0	3.1	0.0	0.0	3.1	3.1	0.0
最大	25.4	540.1	7.7	564.3	37.8	567.2	40.3	13.8
平均	5.5	284.3	4.8	295.7	11.4	300.4	16.2	0.9
总和	548.9	28429.6	479.1	29565.7	1136.1	30044.9	1615.3	88.1

C: 到达时刻, 表示当前顾客卡车到达系统的仿真钟, 除了第一个顾客卡车在仿真钟 0 时刻到达外, 其他顾客卡车的到达时间均为前一个顾客卡车到达时间+到达间隔时间, 因此设定该列中第 i 行单元格的公式为 “ $= C_{i-1} + B_{i-1}$ ”。

D: 接受服务时长 (分钟), 表示当前顾客卡车在入库处接受卸载入库服务的时间长度, 需要满足 $U(2,8)$ 的均匀分布, 根据均匀分布随机变量生成公式, 设定该列中各行单元格的公式为 “ $= 2 + 6 * RAND()$ ”。

E: 服务开始时刻, 表示当前顾客卡车接受卸载入库服务的开始时刻, 应该为当前顾客卡车到达系统时间和前一顾客卡车服务结束时间中的最大值, 即接受服务时间只有在当前顾客到达系统之后, 而且前一个顾客已经服务完毕之后, 因此设定该列中第 i 行单元格的公式为 “ $= MAX(C_i, G_{i-1})$ ”。

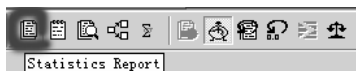
F: 等待时长, 表示当前顾客卡车在队列中等待的时间长度, 等于该卡车开始接受卸载入库服务的时间与其到达系统时间的差值, 因此设定该列中第 i 行单元格的公式为 “ $= E_i - C_i$ ”。

G: 服务结束时刻, 表示当前顾客卡车的服务完毕时间, 等于开始接受服务时间与该顾客接受服务时长之和, 因此设定该列中第 i 行单元格的公式为 “ $= E_i + D_i$ ”。

H: 通过系统时间, 表示当前顾客卡车在系统中滞留的总时间, 为其进入系统到离开系统之间的时间间隔, 即等于服务结束时刻减去其到达时刻, 因此设定该列中第 i 行单元格的公式为 “ $= G_i + C_i$ ”。

I: 叉车空闲时间, 表示入库叉车在对当前顾客卡车服务之前与对前一个顾客卡车服务完毕之后这个时间段的空闲时间, 等于当前顾客服务开始时间与前一顾客服务结束时刻之差, 因此设定该列中第 i 行单元格的公式为 “ $= E_i - G_{i-1}$ ”。

表 4.2 下半部分是对于每一列数据的简单统计, 包括最大值计算、平均值计算和总和的计算。从表 4.2 中的单次仿真实验统计结果可以直接看出, 本次实验顾客卡车到达间隔时间均值为 5.5 分钟, 大于其统计均值; 每位顾客卡车卸载入库服务时间均值为 4.8 分钟, 小于其统计均值; 每位顾客卡车的平均等待时间为 11.4 分钟; 叉车在对每辆顾客卡车进行服务之前平均有 0.8 分钟的空闲时间。从实验结果中还可以间接得出其他有意义的绩效指标。



将鼠标放置在该按钮上，会显示 Statistics Report 文字提示，即可弹出该元素的统计报告窗口。注：如果 Witness 软件界面没有该工具栏，请通过菜单：View→Toolbars→Reporting 打开该工具栏。

2. 模型元素统计分析

下面分别打开本模型中四类（Part、Machine、Conveyor、Variable）元素的统计报告窗口，查看各元素的统计指标，并进行相关分析和说明。

1) 零部件 Widget 的统计指标

使用第一种方法打开 Widget 元素统计报告窗口，如图 6.17 所示。



图 6.17 part 元素统计界面

上图为 Part 类型零件的标准统计报表，如果统计表格不是上图的竖列，而是横向的，使用图上方框区域进行横、竖向转换。Part 统计项目含义如表 6.2 所示。

表 6.2 Part 类型元素统计项目含义

统计项	含义
No. Entered	该零件进入系统的数量
No. Shipped	该零件被 Ship 掉的数量，即出系统的数量
No. Scrapped	该零件作为废料被抛弃的数量
No. Assembled	该零件被组装掉的数量
No. Rejected	该零件被拒绝的数量
W.I.P.	当前该零件的库存量
Avg W.I.P.	在整个仿真时间段内，该零件的平均库存量
Avg Time	所有零件在模型中平均滞留时间
Sigma Rating	废品在总体中所占的比率，以 Sigma 方式表示出来

(5) 交互窗口文本 Interact Box。

系统属性 PEN 经常用于改变零件或劳动者显示图标的颜色，以反映零件或劳动者状态的变化。通常在动作中设置代码改变零件的 PEN 属性数值。Part 类型元素系统属性 PEN 取值不同所对应的 Part 元素显示颜色，如表 7.6 所示。

表 7.6 系统属性 Pen 对应的颜色

PEN 值	前景/背景色	PEN 值	前景/背景色
0	元素图标缺省颜色	13	黑/紫红
1	红/黑	14	黑/青
2	绿/黑	15	黑/白
3	黄/黑	16	白/黑
4	蓝/黑	17	红/红
5	紫红/黑	18	绿/绿
6	青/黑	19	黄/黄
7	白/黑	20	兰/兰
8	黑/黑	21	紫红/紫红
9	白/红	22	青/青
10	黑/绿	23	白/白
11	黑/黄	24	黑/黑
12	白/兰		

注意，要想实现 Pen 控制相关元素的动态图标颜色，这些元素 Style 可视化项目对应的图标必须是在 Witness 中的 Picture Gallery 中设定的单色图标，而不能是 JPG、BMP 等图像，单色图标设定参考图 7.4。



思考题

1. 在原模型基础上，进行如下修改，并进行仿真实验，将结果同原模型的结果进行比较：
 - (1) 将工人 Inspector 的数量设定为 2；
 - (2) 将油漆工位作业时间修改为 1.5 分钟；
 - (3) 将座椅返工比率修改为 20%、5%。
2. 将原模型运行至 48000 仿真时间，查看系统状态，并分析原因。
3. 从网络上查找合适的设备图标并保存下来，将模型中四个设备元素的可视化图标修改为保存下来的图标。
4. 在 Witness 图库中，手工绘制模型中的 part 元素 Seat、Back、Legs 图标，并将其同元素连接起来。
5. 将模型中机器或工位的作业时间都设定为满足 $[0.5t, 1.5t]$ 的均匀分布，其中 t 为模型中的原有时间，仿真至 4800 分钟，查看系统状态，并同原模型进行比较，分析原因。
6. 假设每台机器的日（8 小时/天）作业成本依次为 400 元/天、600 元/天、200 元/天、300 元/天，工人的日工资为 100 元，座椅销售利润为 40 元/把。现在工厂愿意进行系统改善，改善的可能方案为最多增加两台机器和增加 1 名工人，但是管理者不确定增加哪些工位的

workerB 的忙率为 76%，看起来 workerB 的忙率并不高，但是结合 cutter4 的等待工人比率 0.74%和阻塞率 14.3%，可以看出在当前运作方式下，workerB 能力不足，会影响系统的产能；workerB 的空闲比率是 cutter4 等待行车将钢板运送过来的时间比率。Crane 忙率为 74.75%，还有 25.25%的空闲，但是这个空闲时间比率是由于在某些时间切割机均处于工作状态，而不需要行车作业造成的；而有些时候，同时会有两个以上的切割机需要行车进行装、卸作业，这就导致行车虽然看起来忙率不高，但是仍然是系统瓶颈的结果。



思考题

1. 在原模型基础上，修改模型，实现下列功能。

(1) 将每种工件的单位切割时间和单位分拣时间都降为原来的一半，进行仿真实验，对结果进行比较分析。

(2) 增加一台行车用于钢板的装卸载，假设其运作方式同第一台行车完全一样，且不考虑行车在运行过程中的空间冲突，进行仿真实验，并对结果进行比较分析。

(3) 增加一名 workerB，进行仿真实验，并对结果进行比较分析。

2. 在原模型基础上，调整数据文件中日排料计划，以便 A、C、D、E 的产量趋于均衡，并通过仿真实验，验证修改后的效果。

3. 在冲床前增加一个容量为 100 的 buffer，切割机 4 切割下来的物料直接放到该 buffer，而冲床从该 buffer 中提取工件 B 进行冲压，试修改原模型，进行仿真实验，统计结果有何变化。

4. 根据如下的排料原则，试通过编程进行每日的排料计划制定。

(1) 保证 A、C、D 和 E 的下料数量基本均衡。

(2) 保证能够满足当日的总装计划量。

进行仿真实验验证排料效果。



(章节自测题)

第 11 章 流水布局机加焊接作业系统建模与仿真

11.1 模型描述

一个机加焊接车间需要对工件 A 和 E 进行加工作业，最后将这两个部件焊接为最终产品 F，具体设备布局如图 11.1 所示。其中工件 A 和 E 随机到达仓储区 2，然后进行打坡口、切削作业，再在点焊工位将 1 个 A 和 2 个 E 点焊成 1 个 F，并最终由通焊工位进行全面焊接成型，然后在打磨工位将 F 的焊缝进行磨平，从而完成了装配件 F 的加工过程，基本工艺过程和设备布局如图 11.1 所示。

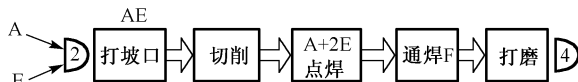


图 11.1 流水布局机加焊接作业工艺过程和设备布局图

各道工序的作业工时如表 11.1 所示。

表 11.1 工序作业工时表

作业	A 打坡口	E 打坡口	A 切削	E 切削	点焊	通焊	打磨
工时: 分钟	2	3	2.5	2.5	3	7	5

打坡口和切削工位换模时间为 20 分钟，工件 A 到达间隔服从均匀分布 $uniform(30,50)$ 随机分布，每次到达批量服从整数均匀分布 $iuniform(3,8)$ ，工件 E 到达间隔服从均匀分布 $uniform(30,60)$ ，每次到达批量服从整数均匀分布 $iuniform(9,15)$ 。为了防止打坡口和切削工位频繁换模，工人对 A 或 E 产品采用整批的方式进行加工，即先加工 10 个 A，然后再加工 20 个 E 的方式进行作业，当某类产品对应批量加工完毕后，另一类产品尚没有存量，则直接停工，直至另一类产品到达。

通过 Witness 仿真，运行 50 天（50 天×8 小时/天×60 分钟/小时=24000 分钟），统计如下数据：

- (1) 各道工序的忙闲状态比率；
- (2) 系统平均库存水平；
- (3) 工作负荷不平衡系数；
- (4) 产线产量；
- (5) 调整打坡口和切削工位对两种产品的加工批量，运行模型，比较上述四项指标的差异。

在学习过程中，主要关注如下功能的实现：

续表

元素名称	元素类型	元素数量	元素作用
BufTyreToCross		1	模拟从 BufTyre 到交叉点的轨道
CrossToBuf1		1	模拟从交叉点到 Buf1 的轨道
CrossToBuf3		1	模拟从交叉点到 Buf3 的轨道
CrossToBufTyre		1	模拟从交叉点到 BufTyre 的轨道
forkIdle		1	模拟车辆空闲时的停车位
fork	Vehicle	1	模拟运输叉车
Belta	Variable(Integer)	3	分别存储 B、C、D、Tyre 运输触发值
forkLoadLimit		2	分别存储叉车运输下限和上限重量
getBatch		1	记录叉车某次运输的工件数量
partWeight		4	分别存储四种工件的重量
toWhere		1	记录特定时点叉车装载目的地编号
decideCDQty	Function	-	进行 C、D 可装载数量的判断函数
goDecision		-	进行叉车调运决策的函数



建立的 Witness 仿真模型最终界面图如 14.2 所示，模型见“chap14 车间物流系统-叉车.mod”。

(示例程序模型下载)

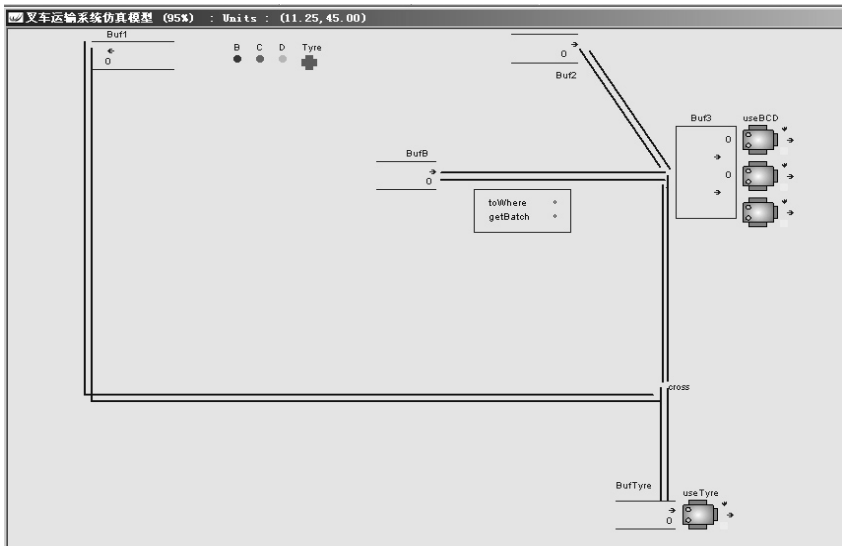


图 14.2 车间物流系统——叉车模型最终界面

14.2.2 模型细节设计

1. Part 类型元素设计

1) B、C、D 和 Tyre 元素细节设计

根据 part 元素 B、C、D 和 Tyre 的到达规律，分别设计这四种工件的细节如表 14.2 所示。

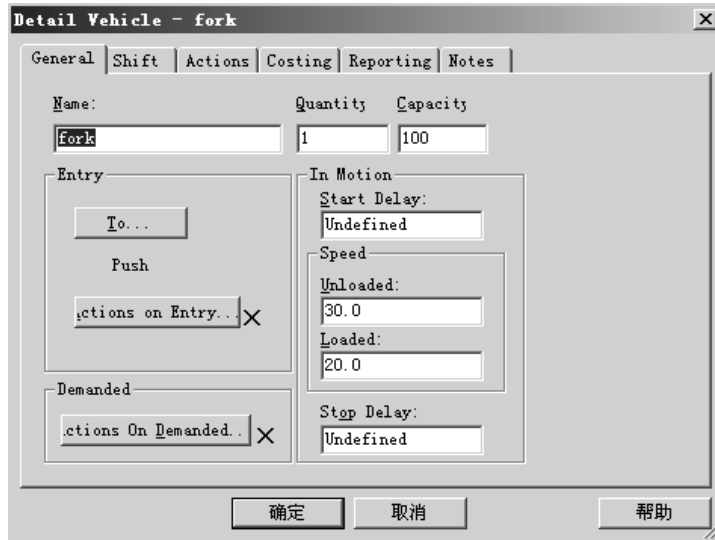


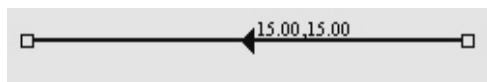
图 14.3 fork 元素细节设计界面

注:

- (1) 根据叉车的最大承载重量 1000 千克和 B、C、D 及 Tyre 的单件重量 40 千克/件、30 千克/件、50 千克/件、100 千克/件，叉车容量 Capacity 只要大于 $1000/30=33.3$ 件即可，在模型中设计其容量为 100;
- (2) Fork 车辆数量 quantity 为 1 辆;
- (3) 空载行驶速度 unloaded 为 30 米/分钟，装载后的行驶速度 loaded 为 20 米/分钟;
- (4) 车辆初始位置 To...为轨道 forkIdle，即 PUSH to forkIdle。

5. Track 类型元素细节设计

Witness 中 Track 元素具有方向性，车辆只能沿着轨道上箭头的方向从一端行驶至另一端，如下图所示的轨道，车辆只能从右侧进入轨道，并行驶至左侧端点 Front，图上轨道上的数字 15.00 表示该段轨道在界面上的像素长度，可以使用函数 PhysicalPathLength (TrackName)获取该轨道的物理长度。



根据图 14.1 的系统布局图上各作业或存储点之间的距离，设计模型中各条轨道的可视化长度，并设计每条轨道的 Physical Length 为 PhysicalPathLength (ELEMENT)，则车辆在每条轨道上的行驶时间将由轨道自身长度、车辆是否装载了工件及车辆的速度决定。

上述设计为模型中 11 条轨道的共性设计，而要实现模型的运行，还需要进行轨道的个性设计。为了便于描述，将模型中的轨道根据功能分为三类:

- (1) 具有装载功能的轨道为 Buf3ToBuf2、Buf3ToBufB、CrossToBuf1;
- (2) 具有卸载功能的轨道为 Buf2ToBuf3、BufBToBuf3、CrossToBufTyre;
- (3) 仅有路由功能的轨道为 Buf1ToCross、CrossToBuf3、Buf3ToCross、forkIdle。

(2) 确定能表示输入过程的概率分布。在获得数据之后,这一步基本上就是从识别数据的频度分布或直方图开始。在已知随机分布 pdf 曲线和输入过程的条件下,选定输入数据服从的分布类型。

(3) 确定分布的参数。在数据和分布类型确定的基础上,估计分布函数的相关参数值。

(4) 评价所选分布及参数与实际数据的拟合优良度。拟合优良度可以通过非正规的方法——图表法或正规的方法——统计检验来评估。卡方检验和科尔莫戈洛夫-斯米尔诺夫检验是最常用的拟合优良度检验方法。如果检验发现所选的分布对数据的近似不能令人满意,则再转到步骤(2),重新选择一个分布类型,重复该过程。如果多次迭代都没能实现假设分布和所收集的数据之间的拟合,则需要考虑使用经验分布作为该数据的输入模型。

下面对上述 4 个步骤进行具体的分析和描述。现在可以使用相关统计软件 Matlab、Minitab、SAS 等来辅助完成上述的步骤(2)、步骤(3)和步骤(4),但是了解构建输入模型的原理和方法仍然非常重要,一方面有助于更好地使用这些统计软件,另一方面有助于开发更合适的随机输入模型。

2.3.1 收集数据

1. 仿真系统输入数据分类

在进行运作系统仿真时,可能的输入数据可以划分为如下两类。

(1) 时间量:加工时间、故障时间间隔、维修时间、准备及换模时间、运输时间。

(2) 结构量:产品——种类、数量、成本、重量或颜色属性。

设备——种类、数量、作业方式(Single、Assembly)、成本。

人员——类别、数量、作业能力、成本。

仓库——类别、数量、容量、存储方式、成本。

方法——作业流程、生产调度方式、应急处理方式。

对于特定运作系统,结构量一般为确定性的数值或流程,只需要将实际系统中对应的数据收集起来,运用到模型中即可,它是建立模型结构的基础,需要在建模之前完成。而时间量通常具有随机性,需要进行特别的统计处理,该类数据的收集处理可以在建模之前、当中,甚至在模型结构建立完成之后完成,但是对其收集和统计处理要花费较多的人力和物力。

2. 收集数据的注意事项

下面以自助洗衣店仿真数据收集过程为例说明相关的注意事项。

为了对自助洗衣店服务效率进行评估而进行仿真建模,需要进行相关数据的收集。自助洗衣店由 10 台洗衣机和 6 台干衣机组成,表面看起来是相当简单的系统。这项工作安排两位仿真的学生去仿真现有系统的运行。

虽然系统看起来很简单,但是很快就发现数据收集量变得非常大。

通过观测发现顾客到达间隔的分布并不相似:它每天随时在变,每周随天在变。洗衣店每周营业 7 天,每天营业 16 小时(或每周 112 小时)。在可利用的资源有限(这两个学

生还有四门课程)和时间很紧(仿真要在四周内完成)的情况下,想观察洗衣店的所有的工
作是不可能的。另外,一周内的到达间隔时间分布在下一周可能不适用。作为一种折中,
选择一种时间样本,并且根据到达速率,将到达间隔时间的分布分类为“高”、“中”、“低”。

通过多次的观察,发现服务时间的分布也存在一个难题:必须要观察和记录需要不同
服务组合的顾客的比例。顾客选择洗衣和干衣的组合方式多样,最简单的情况是顾客需要
一台洗衣机和一台干衣机,或顾客可能选择两台洗衣机和一台干衣机,或顾客仅仅需要一
台干衣机,等等。为了时间收集的简单化,将每台机器都编好序号,然后通过序号来跟踪
顾客,而不需要根据个人特征来记住他们。同时,由于每个顾客对洗衣机的需求和对干衣
机的需求之间具有依赖关系,因此,将洗衣机的服务时间和干衣机的服务时间作为独立变
量分开考虑可能不能准确反映实际情况。

顾客洗衣流程也存在差别。某些顾客耐心地等待衣服的洗衣和干衣周期完成,然后立
刻将它们取走。另一些顾客在衣服放入机器后可能会暂时离开,等他们衣服的洗衣周期完
成之后再回来取。在比较忙的时期,顾客的衣服洗完之后管理人员会将他们的衣服拿出来
装在篮子里。同时规定将机器里的东西清空的那个时间点作为服务终止时间,这个时间
点需要被及时地预测。

机器也会不时地出现故障。故障时间的长度从一会儿(如果管理人员及时修理好机器)
到几天(如果机器是在周五晚上出现的故障,需要的零件洗衣店仓库中没有,要到下一星
期一才能修好)。短期的修理时间可以由学生小组来记录,长期的修理时间需要管理人员来
估计。这样,故障就成为该仿真的一部分。

通过上面的示例可以发现收集数据有时是非常复杂的过程。

下面给出一些可以增强和帮助数据收集的建议。

(1) 编制数据收集计划,设计数据收集表格并安排足够的时间。可以从一次实习或预观
测开始编制计划,开始试着收集数据,并为这一目的的设计表格。在收集实际数据之前,这
些表格很可能被修改几次,要注意异常情况及其数据收集问题,思考如何去处理它们。在可
能的情况下,可以对系统进行录像,然后观看录像分析选取数据。计划编制对于确保采集到
合适的数据是非常重要的,即使在自动数据收集的情况下(如通过计算机收集数据)也是
这样。当数据由其他人收集时,要确保有足够的时间来将收集的数据转化为可用的形式。

(2) 在数据收集的同时,试着分析数据。弄清楚所收集的数据是否够用,以便提供作
为仿真输入所需的分布,看看某些收集到的数据对于仿真来说是否是无用的,没有必要收
集多余的数据。

(3) 试着合并相似的数据集。检查连续的时间周期和连续几天内相同的时间周期的数
据的相似性。例如,要检查从下午 2:00 到下午 3:00 和下午 3:00 到下午 4:00 的数据的相似
性,同时也要检查星期四和星期五下午 2:00 到下午 3:00 的数据的相似性。在检查数据的相
似性时,先用一个双样本的 t 检验看一看两个分布的均值(如平均间隔时间)是否相同。更
彻底的分析需要进行分布的等价检验,这可能要借助于分位点图。

(4) 做散布图来解释两个变量之间是否有关系。有时,用眼观察散布图就可以知道两
个所关心的变量之间是否有关系。

(5) 一系列看似无关的观察实际上却是自相关的,要考虑到这种可能性。自相关可能