Ch 9 Input Modeling

投入資料是模擬模式的重要驅力，在等候系統，投入資料是到達時間與服務時間之分配。在存貨系統投入資料是前置時間與需求分配。在 Ch2 和 Ch3 的範例中，分配均為已知，在現實生活中，收集投入資料是一項重要工作。

✓ 建構有用模式中之收集投入資料有四個步驟：

1. 由現實系統中收集資料：通常十分耗時與浪費資源，有些情況下收集資料不可能，則詢諸專家意見。
2. 辨認一機率分配代表投入過程：資料備妥，首先建構直方圖，Ch5 的分配提供一些實務上常見之分配
3. 選擇參數代表分配。常以資料中之 μ、σ 代表
4. 評估選擇之分配的適合度：可以不正式的圖形方法或是以正式的統計方法，χ^2 test 與 K-S test。

9.1 資料收集

資料收集是解決一真實問題最大工作之一，通常最重要與最困難的問題在模擬。如果

資料無法正確收集，即使模式正確建構，結果深入分析，仍然會導致決策者錯誤判斷。

e.g. 9.1 自助洗衣店中有10台洗衣机，6台乾衣机，如何收集投入資料，顧客到達時間（一日中、与一星期中有離尖峰），顧客使用洗衣机与乾衣机時间（每位顧客之洗衣量与習慣不同），机器經常故障，這是一個簡單現實生活中的範例，如何在有限時間內，收集足以代表系統特性。下列建議可加強與便利資料收集：

1. 有效支配時间：收集資料前事先觀察系統，並設計欲收集資料之型式，如可能以錄影方式收集，如是他人給予之資料，設法將其轉換成可用格式。
2. 嘗試分析資料於收集过程中：已合資料是否足以使用，不需收集过多的資料。
3. 嘗試合併同質性的資料：2:00～3:00 Thursday 與 3:00～4:00 Friday 資料，藉由平时数是否相等。
4. 注意資料普查而部分想要的資料不被包含在內，因收集資料过程是坐在想收集資料的時间前後。

5. 決定兩變數間是否有關係，建立一散佈圖。

6. 考慮一序列的觀察值是否獨立與自我相關。

7. 分清楚投入資料與產出資料，投入資料非系統所能控制，產出資料代表系統之績效受投入資料影響。

9.2 使用收集之資料辨認，一机率分配

9.2.1 直方圖

直方圖之建構步驟如下：

1. 依據資料全距分成若干個區間（區間通常是等距）

$K = 1 + 3.322 \log n$

組數(區間個數)

2. 在 x 軸標上各個區間

3. 決定產生在各區間的次數

4. 在 y 軸標上刻度

5. 將各區間產生的次數寫在 y 軸上

圖 9.1 中 (a) 代表原始資料 — 凹凸不平的

(b) 合併相鄰區間 — 太粗糙

(c) 合併相鄰區間 — 適合的

直方圖宜平滑，避免凹凸不平

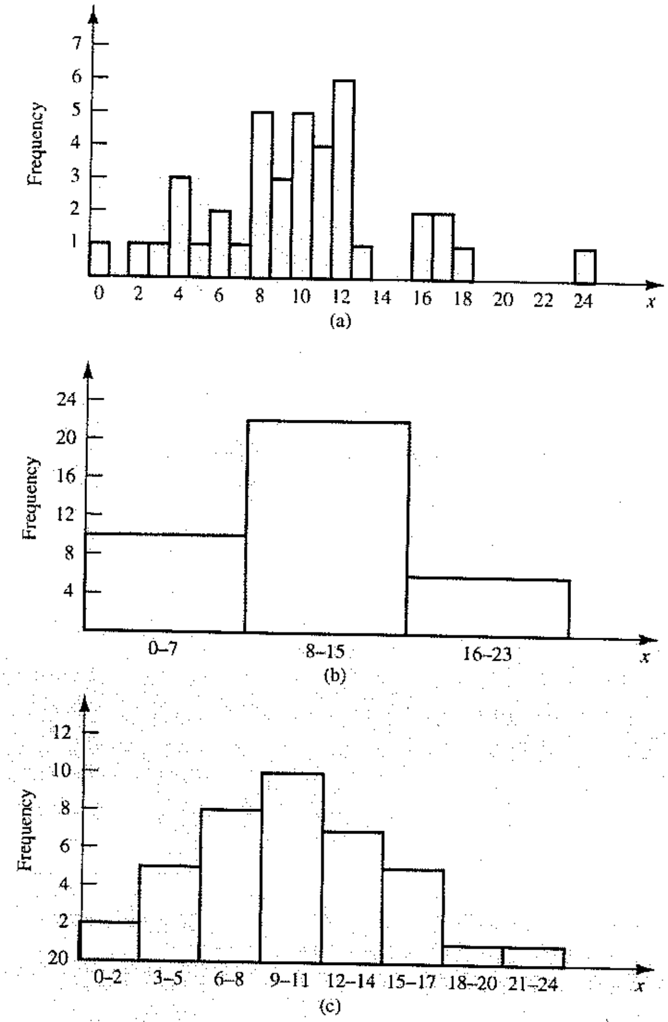


Figure 9.1. Ragged, coarse, and appropriate histograms: (a) original data — too ragged; (b) combining adjacent cells — too coarse; (c) combining adjacent cells — appropriate.

example 9.2 離散資料

表 9.1 顯示，早上 7:00 ~ 7:05 分間汽車到達數量資料取得是觀察平日 (five workdays)，經過 20 星期而得。表中第一列代表中前面代表有 12 個觀察期間是沒有汽車到達，第二列有 10 個觀察期間是有 1 輛汽車到達，依此類推。

到達汽車數目為一離散變數，且有足夠資料其結果顯示於表 9.2。

Table 9.1. Number of Arrivals in a 5-Minute Period

Arrivals per Period	*Frequency*	*Arrivals per Period*	*Frequency*
0	12	6	7
1	10	7	5
2	19	8	5
3	17	9	3
4	10	10	3
5	8	11	1

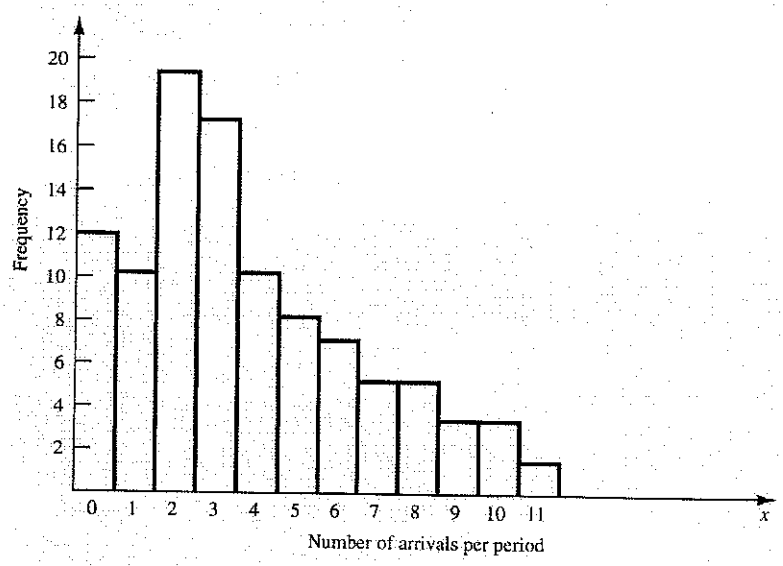


Figure 9.2. Histogram of number of arrivals per period.

example 9.3 連續型資料

電子晶片之生命測試資料如下.

79.919	3.081	0.062	1.961	5.845
3.027	6.505	0.021	0.013	0.123
6.769	59.899	1.192	34.760	5.009
18.387	0.141	43.565	24.420	0.433
144.695	2.663	17.967	0.091	9.003
0.941	0.878	3.371	2.157	7.579
0.624	5.380	3.148	7.078	23.960
0.590	1.928	0.300	0.002	0.543
7.004	31.764	1.005	1.147	0.219
3.217	14.382	1.008	2.336	4.562

生命時間(壽命)通常視為一連續變數,資料中記錄至小數點後三位。資料範圍從 0.002 至 144.695。大部分的壽命是落在 0 與 5 之間,使用表 9.2 的區間,其壽命之直方圖顯示於圖 9.3

Table 9.2. Electronic Chip Data

Chip Life (Days)	*Frequency*
$0 \le x_j < 3$	23
$3 \le x_j < 6$	10
$6 \le x_j < 9$	5
$9 \le x_j < 12$	1
$12 \le x_j < 15$	1
$15 \le x_j < 18$	2
$18 \le x_j < 21$	0
$21 \le x_j < 24$	1
$24 \le x_j < 27$	1
$27 \le x_j < 30$	0
$30 \le x_j < 33$	1
$33 \le x_j < 36$	1
.	.
.	.
.	.
$42 \le x_j < 45$	1
.	.
.	.
.	.
$57 \le x_j < 60$	1
.	.
.	.
.	.
$78 \le x_j < 81$	1
.	.
.	.
.	.
$144 \le x_j < 147$	1

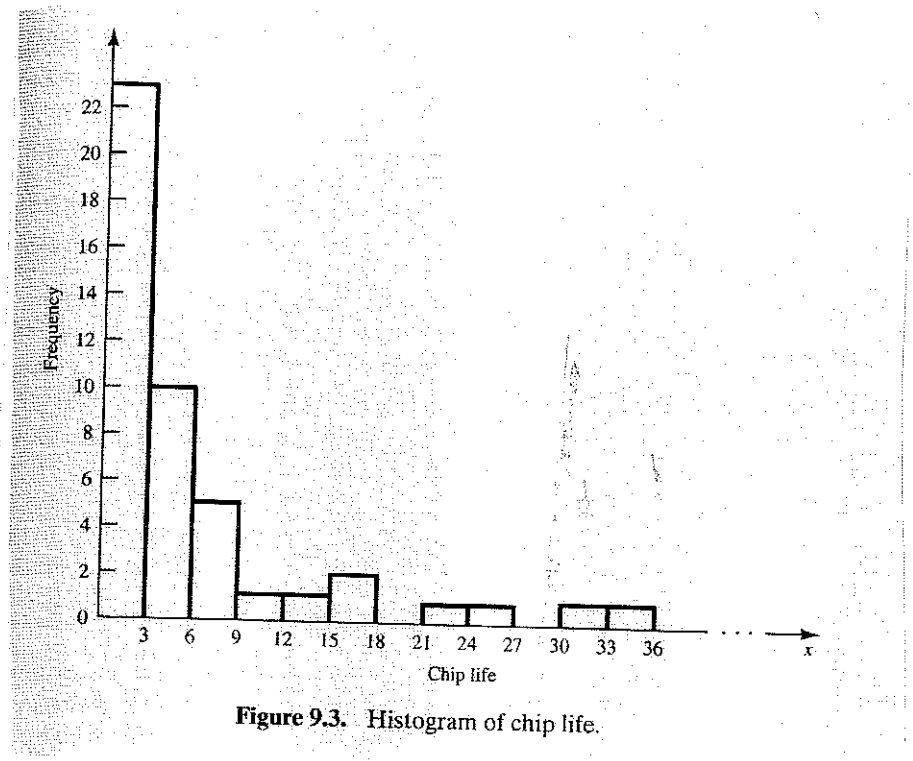


Figure 9.3. Histogram of chip life.

9.2.2 選擇分配

準備直方圖的目的是為了推論資料為一已知分配，所以對Ch5所介紹之分配應有所瞭解

FIGURE 2–2 Two useful discrete distributions.

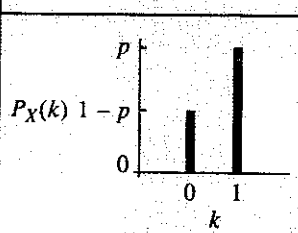
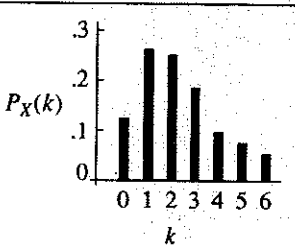
Distribution	Parameters	Mean	Variance	Probability Mass Function		Cumulative Distribution Function
Bernoulli	$0 < p < 1$	p	$p(1-p)$		$P_X(k) = \begin{cases} 1-p, & k=0 \\ p, & k=1 \\ 0 & \text{otherwise} \end{cases}$	$F_X(b) = \begin{cases} 0, & b<0 \\ 1-p, & 0 \le b < 1 \\ 1, & b>1 \end{cases}$
Poisson	$\lambda > 0$	λ	λ		$P_X(k) = \dfrac{\lambda^k e^{-\lambda}}{k!} \quad k = 0, 1, 2, \ldots$	$F_X(b) = \sum_{k=0}^{b} \dfrac{\lambda^k e^{-\lambda}}{k!} \quad b = 0, 1, 2, \ldots$

FIGURE 2–3 Some useful continuous distributions.

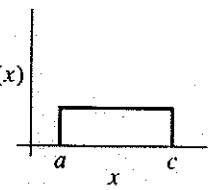
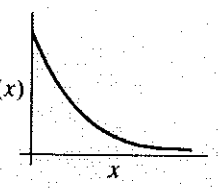
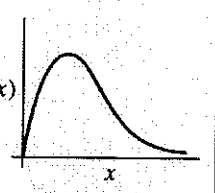
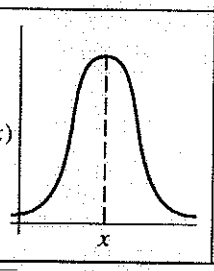
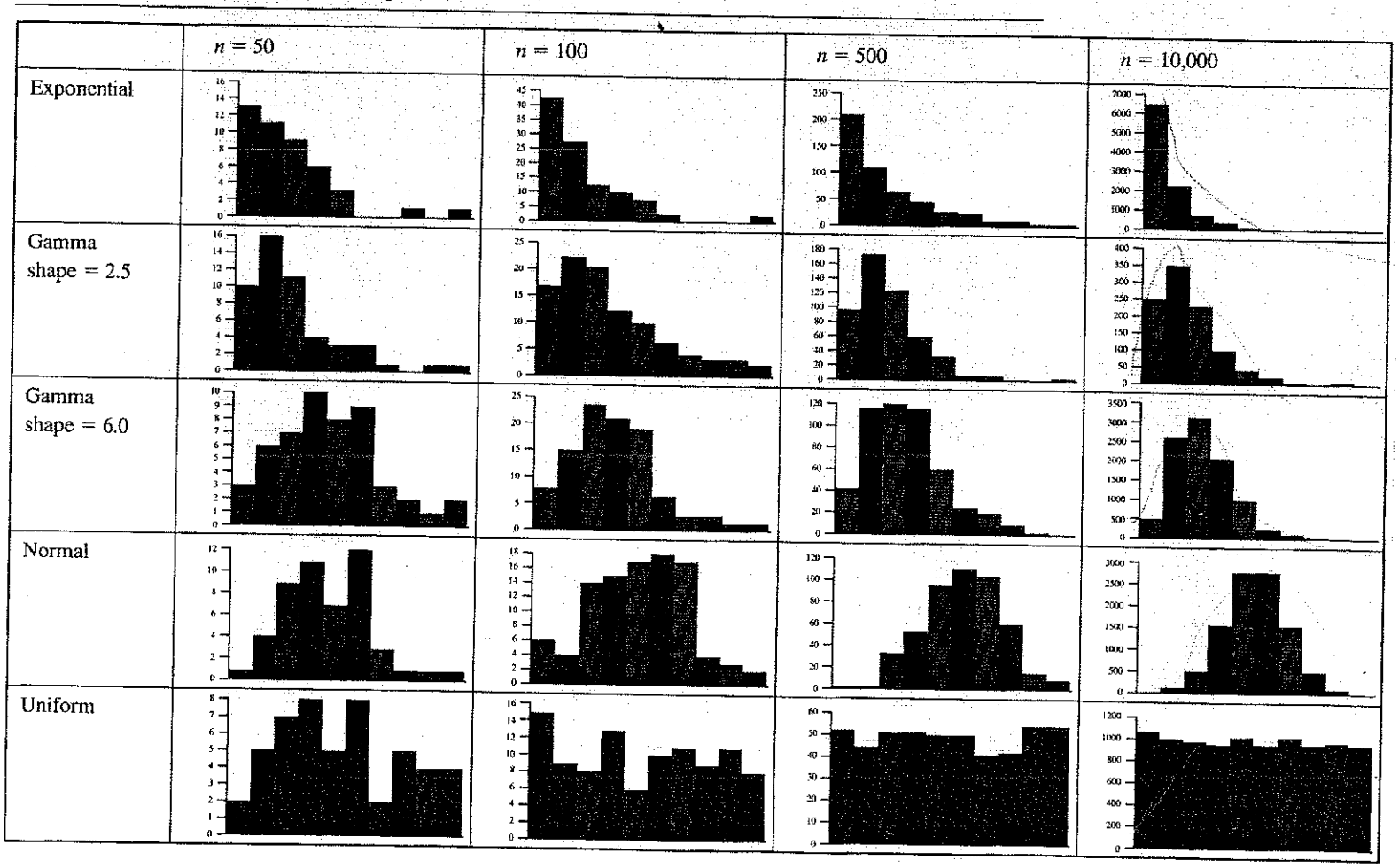
Distribution	Parameters	Mean	Variance	Density Function	Cumulative Distribution Function
Uniform	$a < c$	$\frac{a+c}{2}$	$\frac{(c-a)^2}{12}$	 $f_X(x) = \begin{cases} \frac{1}{c-a} & a \leq x \leq c \\ 0 & \text{otherwise} \end{cases}$	$F_X(b) = \begin{cases} 0 & b < a \\ \frac{b-a}{c-a} & a \leq b \leq c \\ 1 & b > c \end{cases}$
Exponential	$\lambda > 0$ λ is the *rate*.	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	 $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$	$F_X(b) = \begin{cases} 0 & b < 0 \\ 1 - e^{-\lambda b} & b \geq 0 \end{cases}$
Erlang	$k = 1, 2, \ldots$ $\lambda > 0$ k is the *shape*; λ is the *scale*.	$\frac{k}{\lambda}$	$\frac{k}{\lambda^2}$	 $f_X(x) = \begin{cases} \frac{x^{k-1}\lambda^k e^{-\lambda x}}{(k-1)!} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$	$F_X(b) = \begin{cases} 1 - \sum_{i=0}^{k-1} \frac{(b\lambda)^i e^{-\lambda b}}{i!} & b \geq 0 \\ 0 & \text{otherwise} \end{cases}$
Normal	$-\infty < \mu < \infty$ $\sigma > 0$	μ	σ^2	 $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]$ $-\infty < x < \infty$	Difficult to compute. Values can be found in statistical tables such as the one in Appendix A.

FIGURE 5–4 Histograms of variates generated from five different distributions. Notice that the histograms of small samples are ragged and difficult to identify, whereas the histograms of large samples look very much like the distributions from which they were generated.



詳細之分配與分配之型態，請參閱 Law and Kelton 書中 Selecting Input Probability Distributions 這章。如果無法找出適合之理論分配，則以原始資料產生 Empirical distribution 實證分配。在模擬實務中，實證分配有其重要性，一般簡單(化)的產生方式是以實際資料產生之直方圖與其相對次數線合產生。

7.2.3 Quantile - Quantile Plots

當資料過少(30或更小)直方圖會凹凸不平，且直方圖區間的間距亦會影響直方圖圖形。Q-Q plot 是一有用的工具，不受上述問題影響，大部分的軟体均可產生 Q-Q plot (SAS)(IMSL statistical Library)

$$x_1 \le x_2 \le x_3, \ldots, \le x_n \quad \Leftarrow \quad \frac{j-0.5}{n}$$

$$y_1 \le y_2 \le y_3, \ldots, \le y_n \qquad (x_i, y_i)$$

99.79	99.56	100.17	100.33
100.26	100.41	99.98	99.83
100.23	100.27	100.02	100.47
99.55	99.62	99.65	99.82
99.96	99.90	100.06	99.85

j	Value	j	Value	j	Value	j	Value
1	99.55	6	99.82	11	99.98	16	100.26
2	99.56	7	99.83	12	100.02	17	100.27
3	99.62	8	99.85	13	100.06	18	100.33
4	99.65	9	99.90	14	100.17	19	100.41
5	99.79	10	99.96	15	100.23	20	100.47

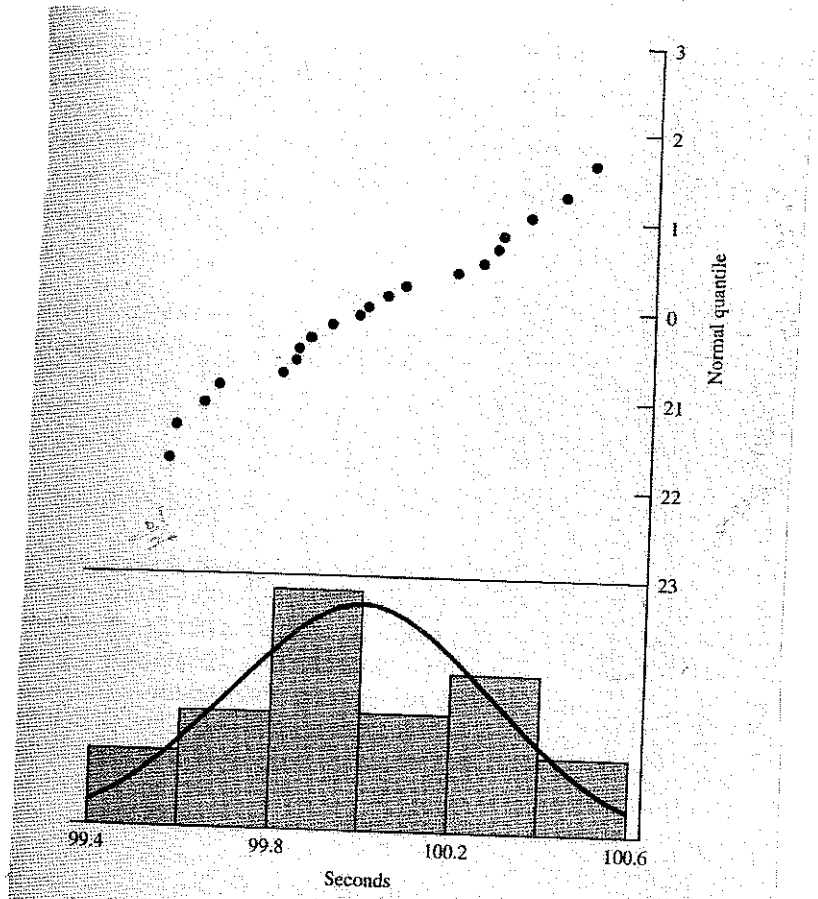


Figure 9.4. Histogram and q-q plot of the installation times.

圖 9.4 顯示直方圖與 q-q plot，q-q plot 看來近似直線，所以其常態性成立(從直方圖不易判斷)。使用 q-q plot 應注意事項

1. 觀察值不全落在一直線上

2. 排序後之相值不是獨立的

3. 最大與最小值差異可能極大，注意 plot 的中間是否為直線遠較兩端重要。

9.3 參數估計

當分配確定後，下一步驟便是計算分配的參數，樣本平均數與樣本變異數常用來估計分配的參數。

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \quad (9.1)\ \text{樣本平均數}$$

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1} \quad (9.2)\ \text{樣本變異數}$$

} 資料未分組

$$\bar{X} = \frac{\sum_{j=1}^{k} f_j X_j}{n} \quad (9.3)\ \text{樣本平均數}$$

$$S^2 = \frac{\sum_{j=1}^{k} f_j X_j^2 - n\bar{X}^2}{n-1} \quad (9.4)\ \text{樣本變異數}$$

} 資料已分組

其中 k 是不同 X 值的數目，f_j 是 X_j 的次數

Example 9.5 分組資料

由 Table 9.1 資料

X_j	f_j
0	12
1	10
2	19
3	17
4	10
5	8
6	2
7	5
8	5
9	3
10	3
11	1

$n = 100$ $f_1 = 12$ $X_1 = 0$

$f_2 = 10$ $X_2 = 1$

$$\sum_{j=1}^{k} f_j X_j = 364$$

$$\sum_{j=1}^{k} f_j X_j^2 = 2050$$

由 (9.3) $\bar{X} = \frac{364}{100} = 3.64$

(9.4) $S^2 = \frac{2050 - 100(3.64)^2}{99} = 2.63$

儘可能使用原始資料，但資料若已經轉換成直方圖，則使用下列式子去近似

$$\bar{X} \doteq \frac{\sum_{j=1}^{c} f_j m_j}{n} \quad (9.5)$$

$$S^2 \doteq \frac{\sum_{j=1}^{c} f_j m_j^2 - n\bar{X}^2}{n-1} \quad (9.6)$$

其中 f_j 是第 j 組之觀察次數，

m_j 為第 j 組之組中點。

Example 9.6 連續資料在組距中 (Continuous Data in Class Intervals)

以 Example 9.3 為例 $f_1=23, m_1=1.5, f_2=10, m_2=4.5, \ldots, \sum_{j=1}^{49} f_j m_j = 614, \sum_{j=1}^{49} f_j m_j^2 = 37266.5, n=50$

$$\bar{X} \doteq \frac{614}{50} = 12.28$$

$$S^2 \doteq \frac{37266.5 - 50(12.28)^2}{49} = 605.849$$

如果使用原始資料得出 $\bar{X} = 11.894$

$S = 24.953$

所以當原始資料不在，以 (9.5) 與 (9.6) 式近似仍會有些許之誤差。[discards data 不全面捨棄]

9.3.2 建議的估計量 Suggested Estimator

表 9.3 包含在模擬中常用的估計量

parameter 參數係一未知常數，estimator 估計量為一隨機變數，取決於樣本的數值。

Table 9.3. Suggested Estimators for Distributions Often Used in Simulation

Distribution	*Parameter(s)*	*Suggested Estimator(s)*
Poisson	α	$\hat{\alpha} = \bar{X}$
Exponential	λ	$\hat{\lambda} = \frac{1}{\bar{X}}$
Gamma	β, θ	$\hat{\beta}$ (see Table A.9)
		$\hat{\theta} = \frac{1}{\bar{X}}$
Normal	μ, σ^2	$\hat{\mu} = \bar{X}$
		$\hat{\sigma}^2 = S^2$ (unbiased)
Lognormal	μ, σ^2	$\hat{\mu} = \bar{X}$ (after taking ℓn of the data)
		$\hat{\sigma}^2 = S^2$ (after taking ℓn of the data)
Weibull with $\nu = 0$	α, β	$\hat{\beta}_0 = \frac{\bar{X}}{S}$
		$\hat{\beta}_j = \hat{\beta}_{j-1} - \frac{f(\hat{\beta}_{j-1})}{f'(\hat{\beta}_{j-1})}$
		See Equations (9.12) and (9.15) for $f(\hat{\beta})$ and $f'(\hat{\beta})$.
		Iterate until convergence:
		$\hat{\alpha} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^{\hat{\beta}}\right)^{1/\hat{\beta}}$

example 9.7 (Poisson Distribution)

假設表 9.1 中資料需要分析，圖形類似 Poisson 分配，表 9.3 建議參數 α，而 α 的估計量 $\hat{\alpha} = \bar{X}$ 由計算而得 $\hat{\alpha} = 3.64$, $S^2 = 7.63$

example 9.8 (Lognormal Distribution)

10 項投資組合之報酬率分別為 18.8, 29.9, 21.0, 6.1, 32.4, 5, 22.9, 1, 3.1 及 8.3 為估計 lognormal 的參數首先對上述資料取 (ln) 對數得 2.9, 3.3, 3, 1.8, 3.6, 1.6, 3.1, 0, 1.1 及 2.1 令 $\hat{\mu} = \bar{X} = 2.3$ $\hat{\sigma}^2 = S^2 = 1.3$

example 9.9 (Normal Distribution)

常態分配的參數為 μ 與 σ^2，以 $\bar{X}$ 和 S^2 估計，使用 example 9.4 的資料，可以算出 $\hat{\mu} = \bar{X} = 99.9865$

$\hat{\sigma}^2 = S^2 = (0.2832)^2$

example 9.10 (Gamma Distribution)

其參數為 β，其估計量 $\hat{\beta}$，需由 A.9 查表而得，為得到 $\hat{\beta}$，需計算 $\frac{1}{M}$，其中

$$M = \ln \bar{X} - \frac{1}{n} \sum_{i=1}^{n} \ln X_i \qquad \hat{\theta} = \frac{1}{\bar{X}}$$

前置時間20筆資料如下

Order	Lead Time (Days)	Order	Lead Time (Days)
1	70.292	11	30.215
2	10.107	12	17.137
3	48.386	13	44.024
4	20.480	14	10.552
5	13.053	15	37.298
6	25.292	16	16.314
7	14.713	17	28.073
8	39.166	18	39.019
9	17.421	19	32.330
10	13.905	20	36.547

To determine $\hat{\beta}$ and $\hat{\theta}$, it is first necessary to determine M using Equation (9.7). Here, $\bar{X}$ is determined from Equation (9.1) to be

$$\bar{X} = \frac{564.32}{20} = 28.22$$

Then,

$$\ell n\bar{X} = 3.34$$

Next,

$$\sum_{i=1}^{20} \ell n X_i = 63.99$$

Then,

$$M = 3.34 - \frac{63.99}{20} = 0.14$$

and

$$1/M = 7.14$$

By interpolation in Table A.9, $\hat{\beta} = 3.728$. Finally, Equation (9.8) results in

$$\hat{\theta} = \frac{1}{28.22} = 0.035$$

$\frac{1}{M}$ β

7.0 3.658

7.14 x

7.3 3.808

插由內插

Example 9.11 (Exponential Distribution)

假設 Example 9.3 的資料來自 exponential distribution

$$\hat{\lambda} = \frac{1}{\bar{X}} = \frac{1}{11.894} = 0.084 \text{ 每天}$$

Example 9.12 (Weibull Distribution)

Weibull distribution 的參數 α、β. 欲計算其估計量 $\hat{\alpha}$、$\hat{\beta}$ 需藉助電腦計算

$$\hat{\beta}_0 = \frac{\bar{X}}{S} \qquad (9.14)$$

$$\hat{\beta}_j = \hat{\beta}_{j-1} - \frac{f(\hat{\beta}_{j-1})}{f'(\hat{\beta}_{j-1})} \qquad (9.13)$$

$$f(\beta) = \frac{n}{\beta} + \sum_{i=1}^{n} \ln X_i - \frac{n\sum_{i=1}^{n} X_i^{\beta} \ln X_i}{\sum_{i=1}^{n} X_i^{\beta}} \qquad (9.12)$$

$$f'(\beta) = -\frac{n}{\beta^2} - \frac{n\sum_{i=1}^{n} X_i^{\beta} (\ln X_i)^2}{\sum_{i=1}^{n} X_i^{\beta}} + \frac{n\left(\sum_{i=1}^{n} X_i^{\beta} \ln X_i\right)^2}{\left(\sum_{i=1}^{n} X_i^{\beta}\right)^2} \qquad (9.15)$$

運算直到收斂.

$$\hat{\alpha} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^{\hat{\beta}}\right)^{\frac{1}{\hat{\beta}}}$$

以 Example 9.3 的資料為例 $n=50$ $\bar{X}=11.894$ $\bar{X}^2=141.467$

$$\sum_{i=1}^{50} X_i^2 = 37575.85 \qquad S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1} = \frac{37575.85 - 50(141.467)}{49} = 622.65$$

$\therefore S = 24.953$

於是 $\hat{\beta}_0 = \frac{\bar{X}}{S} = \frac{11.894}{24.953} = 0.477$

為計算 $\hat{\beta}_1$ 由 (9.13) 得知

$$\hat{\beta}_1 = \hat{\beta}_0 - \frac{f(\hat{\beta}_0)}{f'(\hat{\beta}_0)}$$

$$\sum_{i=1}^{50} X_i^{\hat{\beta}_0} = 115.125 \qquad \sum_{i=1}^{50} \ln X_i = 38.394 \qquad \sum_{i=1}^{50} X_i^{\hat{\beta}_0} \ln X_i = 393.629$$

$$\sum_{i=1}^{50} X_i^{\hat{\beta}_0} (\ln X_i)^2 = 1057.781$$

$$f(\hat{\beta}_0) = \frac{50}{0.477} + 38.394 - \frac{50(393.629)}{115.125} = 16.004$$

$$f'(\hat{\beta}_0) = \frac{-50}{(0.477)^2} - \frac{50(1057.781)}{115.125} + \frac{50(292.629)^2}{(115.125)^2} = -356.110$$

$$\hat{\beta}_1 = 0.477 - \frac{16.024}{-356.110} = 0.522$$

Table 9.4. Iterative Estimation of Parameters of the Weibull Distribution

j	$\hat{\beta}_j$	$\sum_{i=1}^{50} X_i^{\hat{\beta}_j}$	$\sum_{i=1}^{50} X_i^{\hat{\beta}_j} \ln X_i$	$\sum_{i=1}^{50} X_i^{\hat{\beta}_j} (\ln X_i)^2$	$f(\hat{\beta}_j)$	$f'(\hat{\beta}_j)$	$\hat{\beta}_{j+1}$
0	0.477	115.125	292.629	1057.781	16.024	−356.110	0.522
1	0.522	129.489	344.713	1254.111	1.008	−313.540	0.525
2	0.525	130.603	348.769	1269.547	0.004	−310.853	0.525
3	0.525	130.608	348.786	1269.614	0.000	−310.841	0.525

經過4次運算 $|f(\hat{\beta}_3)| \leq 0.001$.

$\hat{\beta} = \hat{\beta}_4 = 0.525$ 收斂

$$\hat{\alpha} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^{\hat{\beta}}\right)^{\frac{1}{\hat{\beta}}} = \left[\frac{130.608}{50}\right]^{\frac{1}{0.525}} = 6.227$$

9.4 Goodness-of-Fit Test 適合度檢定

適合度檢定提供有效的引導評估適合的投入模式，無論如何沒有單一正確的分配在真實世界，瞭解樣本起的影响是十分重要的。如果僅有很少的資料，適合度檢定不可能拒絕所有分配，有非常多的資料，適合度檢定可能拒絕所有分配。

9.4.1 Chi-Square Test

卡方檢定

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \qquad (9.16)$$

其中 O_i 是在第 i 組中的觀察次數，E_i 是在第 i 組的期望次數。$E_i = np_i$ 其中 n 為所有樣本數目；p_i 是伴隨第 i 組的理論上之機率。

統計量 χ_0^2 在於服從 χ^2 分配，有 $k-s-1$ 的自由度，其中 k 為分組數目，s 為欲檢查之分配的參數數目。

H_0：隨機變數 X 符合某分配假設並有參數對應該分配，且其參數由估計量而得

H_1：隨機變數 X 不符合某分配。

我們拒絕 H_0，如果 $\chi_0^2 > \chi_{\alpha, k-s-1}^2$ 其中 $\chi_{\alpha, k-s-1}^2$ 可經由查表而得。

期望次數 E_i 會影響 χ_0^2，雖然沒有關於最小期望次數 E_i 的一致結論，一般 3, 4, 5 是廣泛使用的數字。如果期望次數太小，則合併其相鄰的組別。

Tabl 9.5 連續資料建議分組數

樣本數目	分組數目
20	不要用 χ^2
50	5 ~ 10
100	10 ~ 20
大於 100	$\sqrt{n}$ ~ $n/5$

Example 9.13 (Chi-Square Test Applied to Poisson Assumption)

在 example 9.7 分析 example 9.2 車輛到達資料, 直方圖看起來像 Poisson 分配。估計量 $\hat{\alpha} = 3.64$

H_0: 隨机变数是 Poisson 分配

H_1: 隨机变数不是 Poisson 分配

$$p(x) = \begin{cases} \dfrac{e^{-\alpha}\alpha^x}{x!}, & x = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases} \qquad (9.17)$$

$\alpha = 3.64$ 代入 (9.17)

$p(0) = 0.026$	$p(6) = 0.085$
$p(1) = 0.096$	$p(7) = 0.044$
$p(2) = 0.174$	$p(8) = 0.02$
$p(3) = 0.211$	$p(9) = 0.008$
$p(4) = 0.192$	$p(10) = 0.003$
$p(5) = 0.14$	$p(11) = 0.001$

以上述資料, Tabl 9.6 被建構

Table 9.6. Chi-Square Goodness-of-Fit Test for Example 9.2

x_i	Observed Frequency, O_i		Expected Frequency, E_i		$\frac{(O_i - E_i)^2}{E_i}$
0	12	22	2.6	12.2	7.87
1	10		9.6		
2	19		17.4		0.15
3	17		21.1		0.80
4	10		19.2		4.41
5	8		14.0		2.57
6	7		8.5		0.26
7	5	17	4.4	7.6	11.62
8	5		2.0		
9	3		0.8		
10	3		0.3		
11	1		0.1		
	100		100.0		27.68

$E_1 = np_0 = 100(0.026) = 2.6$

$E_2 = np_1 = 100(0.096) = 9.6$

$\because E_1 < 5 \quad \therefore$ 合併 E_1 與 E_2

同理最後 5 組也被合併

χ_0^2 是 27.68 $\quad \chi^2 = k - s - 1 = 9 - 1 - 1 = 5$

$\chi^2_{.05,5} = 11.1$

$\chi_0^2 = 27.68 > 11.1$ 所以拒絕 Poisson 分配，

必需尋求其他分配或使用實證分配。

9.4.2 χ^2 test with equal probabilities

許多學者建議如果檢定連續分配，應使用

組距產生机率相同而不是使用相同的組距

使用相同机率，則 $p_i = \frac{1}{k}$，既然建議 $E_i = np_i \geq 5$

以 $p_i = \frac{1}{k}$ 代入 $E_i = np_i \geq 5$

$$\frac{n}{k} \geq 5 \quad \therefore k \leq \frac{n}{5} \quad (9.18)$$

(9.8)式就是用以建議 Table 9.5 中的最大組距。本節介紹的方法可用於 normal, exponential, 或 Weibull 分配。如果 Gamma 或其他分配計算組端点的方法十分複雜需藉助統計分析軟体。

Example 9.14 (Chi-Square Test for Exponential Distribution)

在 example 9.11 分析 example 9.3 資料，由資料建構的直方圖看起來像指数分配。估計量 $\hat{\lambda} = \frac{1}{\bar{x}} = 0.084$

H_0：隨机变数是指数分配

H_1：隨机变数不是指数分配

為執行 χ^2 檢定且各組卷生机率相同，需將各組組界計算出來。首先決定分組数目，在 Table 9.5 中建議 $k \leq 10$ ∵ $n = 50$，令 $k = 8$ 組則各組的机率為 $\frac{1}{8} = 0.125$。各組的端点、

可由指数分配的 cdf 計算出來，亦即

$$F(a_i) = 1 - e^{-\lambda a_i} \qquad (9.19)$$

其中 a_i 代表第 i 組的組上界（endpoint）

既然 $F(a_i)$ 是從 0 至 a_i 的累積面積

$F(a_i) = ip$ 於是 (9.19) 可改寫成

$ip = 1 - e^{-\lambda a_i}$ 或 $e^{-\lambda a_i} = 1 - ip$ 兩邊取對數

則得 $$a_i = \frac{-1}{\lambda} \ln(1 - ip), \quad i = 0, 1, \ldots, k \qquad (9.20)$$

以 $\hat{\lambda} = 0.084$ 與 $k = 8$

$$a_1 = \frac{-1}{0.084} \ln(1 - 0.125) = 1.59$$

依續使用 (9.20) 計算 $a_2, a_3, \ldots, a_7$ 可得

3.425, 5.595, 8.252, 11.677, 16.503 和 24.755 — 各組的組界

可依序建構如 Tab 9.7

Table 9.7. Chi-Square Goodness-of-Fit Test for Example 9.14

Class Interval	*Observed Frequency, O_i*	*Expected Frequency, E_i*	$\frac{(O_i - E_i)^2}{E_i}$
[0, 1.590)	19	6.25	26.01
[1.590, 3.425)	10	6.25	2.25
[3.425, 5.595)	3	6.25	0.81
[5.595, 8.252)	6	6.25	0.01
[8.252, 11.677)	1	6.25	4.41
[11.677, 16.503)	1	6.25	4.41
[16.503, 24.755)	4	6.25	0.81
[24.755, ∞)	6	6.25	0.01
	50	50	39.6

χ_0^2 是 39.6，$k = 8$，$s = 1$，$\chi^2_{.05,\,8-1-1}$ 是 12.6

$\chi_0^2 > \chi^2_{.05,6}$ 所以拒絕 H_0

9.4.3 Kolmogorov-Smirnov Goodness-of-Fit Test

χ^2_{test} 需要將資料分組，在連續變數分組組數是可任意選擇，而分組的組數會影響 χ^2 test 的結果。當 χ^2_0 統計量接近 $\chi^2_{\alpha, k-s-1}$ 臨界值時，這種分組可能接受 H_0，另一種分組則可能拒絕 H_0。

在第章 9.4.1 介紹過的與 9.4.4 介紹過的 gap test 均是 Kolmogorov-Smirnov，任何連續分配的適合度檢定可使用 9.4.1 介紹的方法。任何離散分配的檢定可使用 9.4.4 的方法。

K-S 檢定在樣本組目少和參數的估計量未被計算時特別有用。

Example 9.15 (K-S Test for Exponential Distribution)

在100分鐘的正商內，50個到達時商商隔資料收集如下：

0.0044	0.0097	0.0301	0.0575	0.0775	0.0805	0.1059	0.1111	0.1313	0.1502
0.1655	0.1676	0.1956	0.1960	0.2095	0.2927	0.3161	0.3356	0.3366	0.3508
0.3553	0.3561	0.3670	0.3746	0.4300	0.4694	0.4796	0.5027	0.5315	0.5382
0.5494	0.5520	0.5977	0.6514	0.6526	0.6845	0.7008	0.7154	0.7262	0.7468
0.7553	0.7636	0.7880	0.7982	0.8206	0.8417	0.8732	0.9022	0.9680	0.9744

H_0：到達時間間隔是指數分配

H_1：到達時間間隔不是指數分配

使用 K-S test 於 2.4.1 中介紹過，在 (0,1) 區間這些点將是 $\left[\frac{T_1}{T}, \frac{T_1+T_2}{T}, \ldots, \frac{T_1+T_2+\cdots+T_{50}}{T}\right]$ 其資料如下：

0.44	0.53	2.04	2.74	2.00	0.30	2.54	0.52	2.02	1.89	1.53	0.21
2.80	0.04	1.35	8.32	2.34	1.95	0.10	1.42	0.46	0.07	1.09	0.76
5.55	3.93	1.07	2.26	2.88	0.67	1.12	0.26	4.57	5.37	0.12	3.19
1.63	1.46	1.08	2.06	0.85	0.83	2.44	2.11	3.15	2.90	6.58	0.64

依據 example 2.6 的程序，計算 $D^+ = 0.1054$

$D^- = 0.008$

K-S 統計量 $D = \max(0.1054, 0.008) = 0.1054$

D 的臨界值由表 Table A.8 在 $\alpha = .05$ 下：

$n = 50$ $D_{0.05} = \frac{1.36}{\sqrt{n}} = 0.1923$

$D = 0.1054 < D_{0.05} = 0.1923$

H_0：到達時間間隔是指數分配不可以被拒絕 (i.e. 接受)