Title	Statistical Analyses of grouped observations from power-normal distribution
Author(s)	濱崎, 俊光
Citation	大阪大学, 1998, 博士論文
Version Type	VoR
URL	https://doi.org/10.11501/3151131
rights	
Note	

Statistical Analyses of Grouped Observations from Power-normal Distribution

Toshimitsu Hamasaki

1998

Abstract

In medical science and biology, observations may be specified only by two certain observed points (each observation being defined by upper and lower bounds) because of some restrictions in the observation system. This grouped observation (or interval-censored observation) arises naturally whenever individuals or experimental units are observed only occasionally and where the failure event of interest dosed not preclude continued follow-up. Thus, the midpoint of the interval can be used as a representative value, and some corrections such as Sheppard, may be used to calculate some statistics or quantities such as mean and variance (Stuart and Ord, 1986). However, this procedure may not be appropriate in some cases because observations are often found to have skewed distribution, the numbers of intervals are often small or the lengths of intervals are wide.

In this paper, we have proposed an exploratory approach to the inference of grouped observations based on the power-normal distribution (PND) proposed by Goto *et al.* (1983) when the underlying distribution is unknown or there is no strong knowledge of process generating the data.

We have firstly developed the procedure for fitting the PND to univariate grouped observations. We have considered the most elementary case, in which there was only one variable of interest and its observations were given in a grouped form such as a frequency table. The explicit expressions for the maximum likelihood estimates of parameters were not available in ungrouped observation case, as shown in Goto *et al.* (1983), and the present case was no exception. However, it was possible to gain some insight into our proposed procedure using the criterion of maximization of likelihood, by investigating the asymptotic properties of them. In deriving these properties, we have followed the approach to discrete distributions in Hernandez and Johnson (1981). In their approach, the criterion of minimization of Kullback-Leibler information was used to obtain the parameter estimates. Our procedure showed that the maximum likelihood estimates of parameters had the strong consistency and the asymptotic normality under certain conditions. After having concrete and practical grasp of our procedure through several numerical examples, medium-sized simulation experiment was performed to evaluate (i) the precision of the maximum likelihood estimates of parameters and (ii) the effect of grouping or categorizing on them. For (i), the results of simulation showed that, the precision of maximum likelihood estimate of transforming parameter was not so much influenced by the shape of the PND as the precision of the estimate in ungrouped observation case. Nevertheless, the precision of maximum likelihood estimates of mean and variance was as much influenced by the shape of the PND as the precision of those estimates in ungrouped observation case. However, it was shown from some numerical examples that these effects could be removed by using the normalizing power-transformation, which was the adjusted power-transformation by Jacobian of transformation formula. For (ii), the results of simulation showed that the number of intervals had a great influence on the precision of maximum likelihood estimates and the precision decreased as the number of intervals increased.

The PND was also applied in the mixed observation case, where observation involved ungrouped (exactly specified) and grouped ones. We have particularly focused on the two case of the right/or left censored observations and ungrouped observations available in the tail. Furthermore, it was applied in the observations subjected to an upper constraint such as examination score and percentage or proportion, and in discrete observations.

Next, we have focused on two variables situations and considered two variables grouped in (i) a correlation table, (ii) grouped bivariate regression and (iii) simple linear regression when one or both variables were given in a grouped form and their observations were generated form the bivariate power-normal distribution (BPND). In all of (i), (ii) and (iii), the analogy of the procedure used in univariate grouped observation case showed that the maximum likelihood estimates of parameters had the strong consistency and asymptotic normality under certain conditions. Correlation table and regression were equivalent situation in our viewpoint. Thus, the solution provided for correlation tables was easily modified to include bivariate regression. The same sets of observations were then used to illustrate both procedures.

Acknowledgements

The author has much suggestions, helps, and encouragement from many people throughout his work on the preparation and completion of his dissertation. He owes greatly enjoying in his work to them. He has learnt much from them, and so his knowledge and experiment of the field has greatly increased. His deep thanks are given to Professor Masashi Goto, of Osaka University, who provided a basis for his investigations and gave the adequate suggestions throughout the various stages of the preparation of his dissertation. Professor Shingo Shirahata, of Osaka University reviewed the first draft of his dissertation carefully and provided useful comments. Professor Nobuo Inagaki, of Osaka University provided helpful comments. The author is grateful for their assistance.

He received valuable help from the members of Biostatistiacal Research Association (BRA); in particular, he would like to thank Mr. Fumihiko Goto, Dr. Yoshihiro Matsubara, Mrs. Seishi Yamamoto, Mitsugu Kanba, Takao Soda, and Kazuhiko Kuribayashi.

That the author has been able to devote so much time to his dissertation is a tribute most of all to the patience and understanding cooperation of his superiors in Pfizer Pharmaceuticals Inc., Dr. Yukiharu Yamaguchi, Mr. Masahiko Ohishi and Dr. Yoichi Ii. He thanks also Ms. Keiko Aoyama and Yuko Ogata, of his colleagues in Pfizer, whose encouraging cry filled him with redoubled courage.

Finally, the author express his special thanks to his parents and sister for their concern and support throughout his formal education. He thanks also to Dr. Martha M Yaoita who gave mental supports.

Contents

Notations

Notation	Definition/Example	Content
General		
o, O	$o(1)$, $O(1/n)$	Stochastic order
Γ	$\Gamma(\nu) = \int_0^\infty e^{-x} x^{\nu-1} dx$	Gamma function
$\prod$	$\prod_{i=1}^{n} x_i$	Product
Vector and Matrix		
A	$A = (a_{ij}; i = 1, \cdots p; j = 1, \cdots, q)$	Matrix with p rows and q columns
$(A)_{ij}$		The ij element of matrix A
x	$x = (x_1, x_2, \cdots, x_n)^T$	Column vector with components $x_1, \cdots, x_n$
^{T}	x^T, A^T	Transposed matrix of matrix A
$^{-1}$	A^{-1}	Inverse matrix of matrix A
$\mathbf{1}_n$	$\mathbf{1}_n = (1, 1, \cdots, 1)^T$	$n \times 1$ column vector with elements all equal to unity
$\mathbf{0}_n$	$\mathbf{0}_n = (0, 0, \cdots, 0)^T$	$n \times 1$ column vector with elements all equal to zero
$\mathbf{I}_n$	$\mathbf{I}_n = (\delta_{ij}; i, j = 1, \cdots, n)$ $\delta_{ij} = 1(i = j)$	$n \times n$ matrix with all diagonal elements are unity and zeros elsewhere: Identity matrix
∇f	$\nabla f = \frac{\partial f(x)}{\partial x} = \left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_p} \right)$	Gradient vector
$\nabla^2 f$	$\nabla^2 f = \left(\partial f / \partial x_i \partial x_j ; i, j = 1, \cdots, p \right)$	Hessian matrix
$\Re$	$\Re(A)$	Space spanned by matrix A
Random variables and Distribution		
X		Random variable

Notation	Definition/Example	Content
X	$X = (X_1, \cdots, X_p)^{\mathrm{T}}$	p -dimensional random variable
$\xrightarrow{a.s.}$ $a.s.$	$X_n \xrightarrow{a.s.} X (n \to \infty)$	Almost surely convergence Almost surely
$\xrightarrow{p}$	$X_n \xrightarrow{p} X (n \to \infty)$	Convergence in probability
$\xrightarrow{d}$	$X_n \xrightarrow{d} X$	Convergence in distribution
Pr	$\Pr[A]$	Probability of event A
$\mathrm{E}[\cdot]$	$\mathrm{E}[X]$, $\mathrm{E}[X]$	Expectation
$\mathrm{E}[\|]$	$\mathrm{E}[Y\|x]$	Conditional expectation
$\mathrm{var}[\cdot]$	$\mathrm{var}[X] = \mathrm{E}[X - \mathrm{E}[X]]^2$ $\mathrm{var}[X] = \mathrm{E}[X - \mathrm{E}[X]]^2$	Variance
$\mathrm{cov}[\cdot]$	$\mathrm{cov}[X, Y]$	Covariance
Σ	$\Sigma = (\sigma_{ij}) = \mathrm{var}(X)$	Variance-covariance matrix
$\mathrm{N}(\mu, \sigma^2)$		Normal distribution
$\mathrm{N}_n(\mu, \Sigma)$		n -dimensional normal distribution
$\mathrm{PND}(\lambda, \mu, \sigma)$		Power-normal distribution
Others		
λ		Transforming parameter
$\hat{\lambda}$		Maximum likelihood estimate of λ
$\hat{\mu}$		Maximum likelihood estimate of μ
$\hat{\sigma}$		Maximum likelihood estimate of σ
$\tilde{\mu}$		Weighted least square estimate of μ
$\tilde{\sigma}$		Weighted least square estimate of σ

Abbreviations

Abbreviation	Definition
AIC	Akaike's information criterion
ANOVA	Analysis of variance
APT	Asymmetric power-transformation
AOPT	Asymmetric odds-ratio power-transformation
BPND	Bivariate power-normal distribution
DI	Data investigation
FPT	Folded power-transformation
GLIM	Generalized linear model
NPT	Normalizing power-transformation
OPT	Ordinary power-transformation
PND	Power-normal distribution
PT	Power-transformation
PVE	Proportion of variation explained
SPT	Symmetric power-transformation

Lists of Data

Example 1 (Stuart and Ord, 1986): The data given below are 8,585 adult males in the United Kingdom (including, at the time of collection of the data, the whole of Ireland), distributed according to height.

Height	No. of Men	Height	No. of Men
57 -	2	68 -	1,230
58 -	4	69 -	1,063
59 -	14	70 -	646
60 -	41	71 -	392
61 -	82	72 -	202
62 -	169	73 -	79
63 -	394	74 -	32
64 -	669	75 -	16
65 -	990	76 -	5
66 -	1,223	77 -	2
67 -	1,329	**Total**	8,585

Example 2 (Stuart and Ord, 1986): The data given below are 7,749 adult males in the United Kingdom (including, at the time of collection of the data, the whole of Ireland), distributed according to weight.

Weight(lb.)	No. of Men	Weight(lb.)	No. of Men
90 -	2	190 -	263
100 -	34	200 -	107
110 -	152	210 -	85
120 -	390	220 -	41
130 -	867	230 -	16
140 -	1,623	240 -	11
150 -	1,559	250 -	8
160 -	1,326	260 -	1
170 -	787	270 -	-
180 -	476	280 -	1
		Total	7749

Example 3 (Daniel, 1987): The data given below are the age of 75 cases of a certain disease reported during a year in a particular state.

Age		No. of Cases
5 -	14	5
15 -	24	10
25 -	34	20
35 -	44	22
45 -	54	13
55 -	64	5
Total		75

Example 4 (Pagano and Gauvreau, 1993): The data given below are the birth weight of 3,751,275 infants.

Birth Weight(grams)		No. of Infants
0 -	499	4,843
500 -	999	17,487
1,000 -	1,499	23,139
1,500 -	1,999	49,112
2,000 -	2,499	160,919
2,500 -	2,999	597,738
3,000 -	3,499	1,376,008
3,500 -	3,999	1,106,634
4,000 -	4,499	344,390
4,500 -	4,999	62,769
5,000 -	5,499	8,236
Total		3,751,275

Example 5 (Siegel and Morgan, 1996): The data given below are the 36 women's professional golf tournaments in 1979 paid the following prizes.

Thousands of Dollars		No. of Tour-naments
10 -	14	3
15 -	19	20
20 -	24	6
25 -	29	0
30 -	34	7
Total		36

Example 6 (Anderson *et al.*, 1985: Bland, 1995): The data given below are the age of death associated with volatile substance abuse (VSA) mortality for Great Britain.

Age		VSA Death
0 -	9	0
10 -	14	44
15 -	19	150
20 -	24	45
25 -	29	15
30 -	39	8
40 -	49	2
50 -	59	7
60 -		4
Total		260

Example 7 (Brown and Hollander, 1977): The data given below are the results of the measurement of serum cholesterol on the 500 adult males in a very small city.

Cholesterol Measurement		No. of Males
75 -	99	3
100 -	124	6
125 -	149	14
150 -	174	26
175 -	199	36
200 -	224	49
225 -	249	63
250 -	274	65
275 -	299	57
300 -	324	48
325 -	349	46
350 -	374	33
375 -	399	24
400 -	424	16
425 -	449	4
450 -	549	10
Total		500

Example 8 (Silverman, 1986): The data given below are the lengths of 86 spells of psychiatric treatment undergone by patients used as controls in a study if suicide risks. Silverman (1986) divides the domain [0,800] into 20 intervals of length 40 to applying the density smoothing to the data set.

1	1	1	5	7	8	8	13	14	14
17	18	21	21	22	25	27	27	30	30
31	31	32	34	35	36	37	38	39	39
40	49	49	54	56	56	62	63	65	65
67	75	76	79	82	83	84	84	84	90
91	92	93	93	103	103	111	112	119	122
123	126	129	134	144	147	153	163	167	175
228	231	235	242	256	256	257	311	314	322
369	415	573	609	640	737				

Example 9 (Daniel, 1987): The data given below are the weights in ounces of malignant tumors removed from the abdomens of 57 subjects. Daniel (1987) divides the data into the 7 interval of length 10 by using Sturges' rule.

68	65	12	22
63	43	32	43
42	25	49	27
27	74	38	49
30	51	42	28
36	36	27	23
28	42	31	19
32	28	50	46
79	31	38	30
27	28	21	43
22	25	16	49
23	45	24	12
24	12	69	
25	57	47	
44	51	23	

Example 10 (Chen, 1995): The data given below are the numbers of cycles to failure for a group of 60 electrical appliance in a life test.

14	34	59	61	69	80	123	142	165	210
381	464	479	556	574	839	917	969	991	1,064
1,088	1,091	1,174	1,270	1,275	1,355	1,397	1,477	1,578	1,649
1,702	1,893	1,932	2,001	2,161	2,292	2,326	2,337	2,628	2,785
2,811	2,886	2,993	3,122	3,248	3,715	3,790	3,857	3,912	4,100
4,106	4,116	4,315	4,510	4,586	5,267	5,299	5,583	6,065	9,701

Example 11 (Industry Wage Surveys, 1976): The data given below are the weekly earnings of secretaries. The observations outside the range [130,250) are given ungrouped. So five observations falling in [120,130) are 121,123,125,127,129 respectively, and the last three observations are 255, 285 and 325 respectively.

Weekly Earning		No. of Secretaries	Ungrouped
120 -	130	5	121,123,125,127,129
130 -	140	45	
140 -	150	125	
150 -	160	126	
160 -	170	141	
170 -	180	100	
180 -	190	69	
190 -	200	70	
200 -	210	48	
210 -	220	43	
220 -	230	11	
230 -	240	24	
240 -	250	4	
250 -	260	1	255
260 -	270	0	
270 -	280	0	
280 -	290	1	285
290 -	300	0	
300 -	310	0	
310 -	320	0	
320 -	330	1	325
Total		814	

Example 12 (Fisher and Belle, 1993): The data given below are the aflatoxin levels of raw peanut kernels. The observations outside [20,50) are given grouped. Therefore, the twelve observations falling in [20,50) are 26,26,22,27,23,28,30,36,31,35,37 and 48 respectively.

Aflatoxin levels		No. of peanut	Ungrouped
10 -	20	1	
20 -	50	12	26,26,22,27,23,28,30,36,31,35,37,48
50 -		2	
Total		15	

Example 13 (Sugiura, 1980,1981): The data given below are the total score of the common first-stage examination for university entrance conducted in fiscal 1979, 1980 and 1981.

Score	Year		
	1979	1980	1981
0 - 300	2,622	2,400	4,500
301 - 350	4,972	5,700	7,800
351 - 400	9,166	11,400	13,500
401 - 450	14,996	17,400	21,600
451 - 500	22,351	25,500	30,600
501 - 550	31,015	34,800	37,500
551 - 600	38,808	44,100	43,800
601 - 650	45,079	49,800	45,600
651 - 700	46,363	48,600	43,200
701 - 750	42,061	41,700	36,300
751 - 800	33,429	29,700	27,900
801 - 850	22,632	16,200	17,700
851 - 900	11,160	5,700	8,100
901 - 950	2,486		1,910
Total	327,140	333,000	340,010

Example 14 (Stuart and Ord, 1986): The data given below are the number of major labour strikes in the U.K., 1948-59, commencing in each week. Stuart and Ord (1986) give the data as an example to describe the Poisson distribution.

No. of commencing in week	No. of weeks with this no. commencing
0	252
1	229
2	109
3	28
4 or over	8
Total	626

Example 15 (Domae and Miyahara, 1984): The data given below are body weight of 1,084 students who entered Saitama University in fiscal 1982.

All			Male			Female		
Weight		No. of Students	Weight		No. of Students	Weight		No. of Students
-	33.3	0	-	39.5	0	-	31.9	0
34.9 -	36.6	1	41.0 -	42.4	0	31.9 -	34.5	0
38.3 -	40.0	3	43.9 -	43.5	1	34.5 -	37.2	1
41.6 -	43.3	22	46.8 -	48.3	12	37.2 -	39.8	2
45.0 -	46.7	53	49.7 -	51.2	36	39.8 -	42.4	14
48.3 -	50.0	89	52.6 -	54.1	62	42.4 -	45.0	27
51.7 -	53.4	142	55.5 -	57.0	105	45.0 -	47.7	48
55.0 -	56.7	177	58.4 -	59.9	112	47.7 -	50.3	55
58.4 -	60.1	155	61.4 -	62.8	136	50.3 -	52.9	57
61.7 -	63.4	166	64.3 -	65.7	112	52.9 -	55.6	51
65.1 -	66.8	112	67.2 -	68.6	66	55.6 -	58.2	31
68.4 -	70.1	80	70.1 -	71.5	44	58.2 -	60.8	22
71.8 -	73.5	38	73.0 -	74.5	26	60.8 -	63.4	8
75.1 -	76.8	22	85.9 -	77.4	21	63.4 -	66.1	7
78.5 -	80.2	11	88.8 -	80.3	8	66.1 -	68.7	2
81.8 -	83.5	4	81.7 -	83.2	4	68.7 -	71.3	2
83.5 -		9	83.2 -		8	71.3 -		4
Total		1,084	**Total**		753	**Total**		331

Example 16(Cramér, 1974): The frequencies in several groups of age of parents for 475,322 boy children in Norway during nineteen years period 1871-1900.

Age of Father	Age of mother - 20	20 - 25	25 - 30	30 - 35	35 - 40	40 - 45	45 -	Total
0 - 20	377	974	555	187	93	25	6	2,217
20 - 25	2,173	18,043	11,173	3,448	1,022	258	30	36,147
25 - 30	1,814	26,956	43,082	16,760	4,564	973	123	94,272
30 - 35	700	14,252	38,505	41,208	14,475	3,243	287	112,670
35 - 40	238	4,738	17,914	32,240	31,573	8,426	836	95,965
40 - 45	103	1,791	6,586	16,214	24,770	18,079	2,171	69,714
45 - 50	47	695	2,593	5,952	12,453	13,170	4,006	38,916
50 - 55	21	311	995	2,503	4,492	6,322	2,574	17,218
55 - 60	5	133	412	925	1,790	2,141	1,086	6,492
60 - 65	10	57	190	408	736	822	348	2,571
65 - 70	6	25	68	173	266	283	131	952
70 -	2	12	46	59	119	113	48	399
Total	5,496	67,987	122,119	120,077	96,353	53,855	11,646	477,533

Example 17(Holmes, 1974): The frequencies in several groups of total miles driven and family income for 4012 families (car owners) during 1973.

Miles driven	Family income 0 - 5	5 - 10	10 - 15	15 - 20	20 - 25	25 - 30	30 - 35	35 -	Total
0 - 3,050	123	42	31	15	7	2	1	6	227
3,050 - 4,900	141	91	50	16	18	6	6	9	337
4,900 - 6,550	99	109	102	32	20	7	10	13	392
6,550 - 8,700	89	134	109	46	44	12	16	16	466
8,700 - 10,850	82	102	130	60	51	19	18	25	487
10,850 - 12,900	41	69	146	75	53	30	25	36	475
12,900 - 15,350	46	68	116	80	51	42	23	33	459
15,350 - 18,500	18	44	100	97	60	33	28	39	419
18,500 - 23,500	18	51	76	68	62	44	19	44	382
23,500	6	39	61	58	60	37	43	64	368
Total	663	749	921	547	426	232	189	285	4,012

Example 18(Winkelmann, 1994): The frequencies of the number of employers and the number of unemployment spells during the ten years period 1974-1984 for 1,962 individuals, provided by German Socio-economic panel (SOEP).

Unem-	Direct job change												
ployment	0	1	2	3	4	5	6	7	8	9	10	12	Total
0	1,102	301	105	25	20	5	1	2	1		2		1,564
1	146	79	21	10	1	3	2	1		1			264
2	34	16	6	6	2	2	1	1				1	69
3	20	4	1		2								27
4	7	2		2									11
5	6	2											8
6	2									1			3
7	2												2
8	3												3
9	3												3
10	7												7
15	1												1
Total	1,333	404	133	43	25	10	4	4	1	2	2	1	1,962

Example 19 (Mardia, 1970: Stuart and Ord, 1986): The frequency of the number of 9,440 beans according to both length and breadth.

Length	Breath (midpoint)												Total
	6.375	6.625	6.875	7.125	7.375	7.625	7.875	8.125	8.375	8.625	8.875	9.125	
9.5	1												1
10.0	1	3	1	1	1								7
10.5	1	4	7	6									18
11.0		1	13	11	11								36
11.5			12	32	22	4							70
12.0		2	21	25	37	27	3						115
12.5	1		8	35	78	55	19	3					199
13.0			9	28	124	175	89	12					437
13.5			1	18	137	361	330	73	9				929
14.0				13	91	469	794	362	56	2			1,787
14.5				1	23	236	871	913	227	23			2,294
15.0					6	65	385	956	574	93		3	2,082
15.5						4	81	375	494	156	19		1,129
16.0						1	7	44	105	101	17		275
16.5								4	18	23	8	2	55
17.0										2	4		6
Total	4	10	72	170	530	1,397	2,579	2,742	1,483	400	48	5	9,440

Example 20 (Stuart and Ord, 1986): The frequencies of the number of students in the University of London, 1995, classified by the number of newspapers read and the number of newspapers looked at only, on a particular day.

Number read	Number looked at only						**Total**
	0	1	2	3	4	5 or more	
0	77	75	19	10	1	2	184
1	179	136	65	20	15	2	417
2	86	70	45	18	1	3	223
3	17	21	13	3	4		58
4	4	2	2	2			10
5 or more	2	2		2			6
Total	365	306	144	55	21	7	898

1 Introduction

1.1 Outline

In medical science and biology, observations may be specified only by two certain observed points (each observations being defined by upper and lower bounds) because of some restrictions during process of phenomenon occurring, or an observation system. Observations are also grouped into the interval in advance by some criteria. This grouped observation (or interval censored observation) arises naturally whenever individuals or experimental units are observed only occasionally and where the failure event of interest dosed not preclude continued follow-up. Also, observations may be some property connected with the health of the patient in clinical trials, so that patients are classified as "absent", "mild", "moderate" and " severe". Furthermore, it is desirable for ease of calculation or graphical representation to divide the observed range of variation into classes (Cox, 1957: Borkowf *et al.*, 1997: Altman, 1998). On the other hands, it is thought that variables are observed and recorded in finite precision. In a fundamental sense, all variables are eventually rounded or coarsened, i.e., grouped (Heitjan, 1989). In these case, the midpoint of the interval can used as a representative value, and some corrections such as Sheppard, may be used to calculate some statistics or quantities such as mean and variance (Stuart and Ord, 1986). However, this procedure may not often be appropriate in some cases because observations are often found to have skewed distribution, the numbers of intervals are often small or the lengths of intervals are wide (Heitjan, 1989).

So, the probability, which the observations lie within the interval, can be obtained by supposing some underlying distribution, which are appropriate for describing the phenomenon. Accordingly, the likelihood can be constructed and the inferences about the parameters of distribution can be derived based on the likelihood. In the inferences, the estimation of the parameter, which describe the shape of distribution, have received particularly focus (Gjeddebaek, 1949, 1956, 1957, 1959a, 1959b, 1961, 1968: Kulldorff, 1962: Kempthorne, 1966: Barnard, 1976: Kempthorne and Folks, 1971: Copas, 1972: Giesbrecht and Kempthorne, 1976: Nabeya, 1983: Stuart and Ord, 1986: Heitjan, 1989: Atkinson *et al.*, 1991). Then, goodness of fit test on chi-square statistics, or information criteria such as Akaike's information criterion (AIC, Akaike, 1972) have be used to select the underlying distribution (Kariya, 1975, 1979: Kariya and Akasaka, 1976: Sugiura, 1980, 1981: Domae and Miyahara, 1984). Since such measures are used under these underlying distributions, only the gap between the real and the ideal (hypothesis) can be quantitatively represented and evaluated. However, it is difficult to try the flexible fitting of the distributions on date-adaptive as long as one persist in these underlying distributions. In practice, it is desirable that the inference can be carried out on the family of the distribution or data-adaptive distribution which include the underlying distribution, if it were possible, normal distribution with good "property" in the statistical inferences (Matsubara and Goto, 1979: Goto, 1986).

In this paper, we propose an exploratory approach to the inference of grouped observations based on the power-normal distribution (PND) proposed by Goto *et al.* (1983) when the underlying distribution is unknown or when there is no strong knowledge of process generating the data. Here, the PND is a theoretical distribution supporting the underlying concept of power-transformation and includes normal distribution.

In Chapter 2, we firstly proceed to develop the procedure for fitting of the PND to univariate grouped observations. We consider the most elementary case, which is when there is only one variable of interest and its observations are given in a grouped form such as a frequency table. This case is investigated in Section 2.2 and serves to introduce the general methodology which will be used in subsequent section. In all case, the maximum likelihood estimates are not obtainable analytically and iterative procedures have to be employed to solve the likelihood equations. This procedure is considered in Section 2.2 and 2.3. In order to have a concrete and practical grasp of our procedure, several numerical examples are illustrated. Then, medium-sized simulation experiment is performed to evaluate the precision of maximum likelihood estimates obtained in our procedure. The PND is also applied in the other univariate observation case. These are generally called mixed observation case since they involve both ungrouped (exactly specified) and grouped observations. Furthermore, it is applied in the observations subjected to upper constraint and in discrete observations.

In Chapter 3 and 4, we focus on two variables situation and consider two variables grouped in a correlation table, grouped bivariate regression, and simple linear regression when one or both variables are given in a grouped form. Thus, the observations are from a bivariate power-normal distribution (BPND) which is an extension of the PND to bivariate case. Correlation tables and bivariate regression are equivalent situations from our viewpoint. Thus, the solution provided for correlation tables is easily modified to include bivariate regression.

In Chapter 5, some knowledge obtained by these investigations and considerations are summarized as concluding remarks, and further developments are mentioned.

1.2 Grouped Observations

The problem of the statistical analysis of grouped observations has received a great deal of attention in the literature, so we will just briefly review some aspects of this general problem which are of interest to development of the present work. We first mention some situations in which we have to deal with grouped observations:

- The original observations are collected in grouped form (times of failure of objects subjected to inspection at regular intervals and census observations in general, where the respondents check the box corresponding to an appropriate level of age, income, etc),
- Individual observations are deemed confidential (this occurs very often with economic and sociological information),
- Original observations show a tendency toward heaping at selected digits (usually happens with age tabulations),
- The set of observation is too vast and limited research funds prohibit the use of the exact values.

Compared with many other kinds of observations, the frequency theory of grouped observations analysis is poorly developed. The main problem is that the sampling distributions of many potential estimates are complex and unattractive, so that the usual sort of decision-theoretic analysis has not been undertaken. Thus, for a long time the bulwark of the sampling theory of grouped observations was the Sheppard moment corrections, whose sampling properties are poor as Heitjan (1988) has indicated.

On the other hand, maximum likelihood estimates has been shown in the normal and exponential cases to be consistent and asymptotically efficient (Kulldorff, 1961). Expect for these distributions, the asymptotic theory has not been carefully examined. This has not deterred practice, however, where it has become more common to see examples of grouped observations maximum likelihood estimation. The main subjects related with this paper are reviewed as follows. See Gjeddebaek (1968), Haitovsky (1982) and Heitjan (1989) for detailed reviews.

Maximum Likelihood from Grouped Observations

Maximum likelihood estimation has widely used with grouped observations. For the grouped univariate normal observation, the work of Gjeddebaek (1949,1956, 1957, 1959a, 1959b, 1961, 1968) is foremost.

Other contributors were Stevens (1948) and Yoneda and Uchiyama (1956). Taillis and Young (1962) and Heitjan (1987) considered arbitrarily grouped bivariate normal data. Tallis (1967), extensions of Lindley's method are provided for multivariate distributions under equal grouping and univariate distributions under unequal grouping. Deken (1983) discussed higher dimensional multivariate date, but his algorithm is appropriate only for very small grouping rectangles, in which case corrections can be a viable alternative. Kulldorff (1961) presents a coherent recompilation of results on maximum likelihood estimations for grouped observations from exponential distributions, and in the 1-parameter-unknown case, for the normal distribution. Rao (1973) gives results for general multinomial model with cell probabilities depending on the unknown parameters. Aigner and Goldberger (1970) covered the Pareto's law and Flygare *et al.* (1985) have treated the Weibull. Boardman (1973) considered the compound exponential, a special case of bivariate grouping in which the grouping sets are unions of rectangles and triangles. Most recently Pettitt (1985) and Beckman and Johnson (1987) have fit the t distribution, including a degrees of freedom parameter, the latter by Newton's method, the former by EM.

Choice of Algorithms

Because the usual definition of grouped observations subsumes standard kinds of censoring as well, good algorithms for grouped observation likelihood calculations can be quite generally useful. Available methods are variants of either EM, Newton-Raphson or Fisher scoring. Newton-Raphson appears to be the swiftest, followed by scoring and EM (Schader and Schmid, 1984). Burrigde (1981) has suggested using concavity parameterizations to speed up Newton-Raphson algorithms. EM algorithms, although they have at best a linear rate of convergence, are guaranteed never to decrease the likelihood and so are quite robust, a virtue not to be taken lightly in the face of the vagaries of real observations. They are also easier to program and less costly per iteration than the other methods.

Computer programs in the literature are by Swan (Normal by Newton-Raphson, 1969), Benn and Sidebotton (Scale and location families by scoring, 1976) and Wolynetz (Normal linear model by quasi-EM, 1979a, 1979b). Stirling (1984) has recommended the use of iteratively reweighted least squares for the linear part (i.e., means and regression coefficients) in grouped observation models. This would permit fitting these models in generalized linear model (GLIM), but difficulties in estimation of the nuisance parameters may render this approach impractical.

Bivariate/Multivariate Grouped Observations and Regression

A bivariate frequency distribution is called a correlation table if the relationship between the two variables is of interest (Kitagawa and Inaba, 1979). Yule and Kendall (1968) state that " The difference between a correlation table and a contingency table lies in the fact that the latter term may be, and usually is, applied to tables classified according to unmeasured quantities or imperfectly defined intervals."

Pearson (1901) is apparently the first to attempt fitting a bivariate normal distribution to the observations given in a correlation table (even though he was only interested in estimating the correlation coefficient). He considered that that some 2×2 tables were obtained by dividing a bivariate normal population into for sections. Pearson then proposed a procedure for obtaining an estimate of ρ (the true correlation coefficient) and the resulting estimate was called the tetrachoric correlation, because it is based on the tetrachoric (four-entry) table. Among the known properties of tetrachoric correlation is its equivalence with the maximum likelihood estimations [see Hamdan (1970)].

It was Pearson again who considered the problem of estimation of the correlation coefficient when only one of the variables is dichotomized and the other is continuous, the same assumption of bivariate normality is made in order to find what is termed the biserial correlation. Tate (1955) reviews this problem form several viewpoints and provides may new results.

Generalization of the tetrachoric method to a $p \times q$ correlation table by Ritchie-Scott (1918) leads to the polychoric correlation.

The problem of regression when the observations are grouped has been reviewed by Haitovsky (1973), particularly for the case when both the cell means and counts are available (a very frequency encountered case in economic work). However, only one small chapter of that monograph is dedicated to the bivariate normal regression model when the cell means are not available. Some results regarding bias and loss of efficiency, due to the use of midpoints instead of cell means, are developed and connections with regression when both variables are subjected to error are established.

In Fryer and Pethybridge (1972), a generalization of Lindley's methods is proposed to obtain simplified estimates of the maximum likelihood estimations in the bivariate normal regression model. They considered the following cased; (i) both variables grouped, and (ii) one variables grouped and the other continuous.

Some departures of the assumptions made by Fryer and Pethybridge (1972) are studied numerically by Pethybridge (1975), these departures are; small samples, non-normality (the case studied is when non-normality is barely detectable, since gross non-normality can be detected and invalidates the use of the corrections) and unequal group widths "which might arise for instance when a transformation to normality is deemed necessary."

Effect of Grouping or Categorizing

Converting a continuous variable into a grouped or categorical one will result in loss of information. But it can be, and often is argued that with three or more interval or categories the loss is small and is offset by a gain in simplicity and the avoidance of assumptions. Corner (1972) quantified the loss of information when grouping a normal distributed variable which is linearly related to the outcome variables using the relative efficiency, which is based on the ratio of the expected variances of estimated regression coefficients under two models. For 2, 3, 4 and 5 groups, efficiency relative to an ungrouped analysis is 65%, 81%, 88% and 92% respectively. And it is almost identical for an exponentially distributed variable.

Given the decision to grouping, it is not at all obvious how many groups to create. In practice, the sample size should be one factor that influences the decision, as it is undesirable to have sparsely populated groups. The placing of the cutpoints may also not be obvious. To decide to the number of grouping or interval, Sturges' rule (Sturges, 1926) is commonly and widely used. Sakamoto *et al.* (1986) proposed the procedure using Akaike's Information Criterion (AIC)(Akaike, 1972). In stead of AIC Nagahata (1984) used the Fisher information and Mori (1975) employed the maximization criterion of integrated mean squares of error. Using optimally placed intervals, as derived by Corner (1972) following Cox (1957), is little different, in terms of efficiency, from using equally spaced intervals when the variables is normally distributed, but when the variables has an exponential distribution there is reduced efficiency with equiprobable intervals (Lagakos, 1988). This result is important, as it is common to grouping in such situations, and equiprobable intervals are the norm. Morgan and Elashoff (1986) examined the effect of grouping a continuous covariate when comparing survival times. Here too, unequal-grouping give increased efficiency.

1.3 Power Transformation and Power-normal Distribution

Data Investigation (DI) is particularly important in first step of process of statistical data analysis (Goto, 1986). In this step, it should be checked whether any assumptions of statistical methods to be applied were satisfied or not by looking at the data form various points of view. The standard statistical methods such as regression analysis or analysis of variance (ANOVA), which assume the linear model, are derived in the basis of (i) additivity or linearity of the model, (ii) constant variance of error term, and

(iii) normality of response observation. But, it is experientially known or seen in practice that all or some of these three assumptions would not be satisfied. Thus, it is not easy to discuss the methods of copying which reduce the discrepancy between the data and assumed model or build the bridge across the discrepancy. It is considerably difficult to diagnose what the cause of the discrepancy is, and to present consistently the methods of copying. It can think that the approach, which can avoid these assumptions or not be affected by these assumptions, is developed. Alternatively, it can think that observation is transformed to another scale in some methods in accordance with the premise and the nature of the statistical method to be applied.

In data analysis, we must proceed considering the five points of view, i.e., methodology, interpretation, theory, numerical operation (calculation), and practical application. It is needless to say "transformation" also need to be arranged with respect to each if the five points of view (Goto, 1986). Of transformations, "power-transformation" proposed by Box and Cox (1964) is comprehensive one, which includes log, square, reciprocal transformation, and so on.

Unfortunately, the power-transformed variable $Y^{(\lambda)}$ of the positive variable Y lies in lower or upper bounded region according to sign of transforming parameter λ, and the distribution of the power-transformed variables can not be full normal for $\lambda \neq 0$. Only if original variable Y has a log-normal distribution, the power-transformed variable $Y^{(\lambda)}$ will have a normal distribution (Hernandez and Johnson, 1980). However, it is possible to make observations after power-transformation be near-normal, and it is useful (Draper and Cox, 1969).

As Kruscal (1986) has suggested, in the order of (iii) normality of response observation, (ii) constant variance of error term, (i) additivity or linearity of the models, these three assumptions will become important in the practical data analysis. However, in univariate problems such as deciding of "normal" or reference range of laboratory data in clinical study, the evaluation of (iii) is most important, where linearity or non-linearity is not assumed in the structure of data.

If it is careful of these things, it is desirable that power-transformation is applied in the practice from two viewpoints.

First approach assumes that distribution shape dose not change by power-transformation, and achieve (ii) or (iii). In this approach, normal distribution is usually assumed, and the name "power-transformation". The inference procedure will be based on the maximum likelihood in which the power-transformed observations have exactly a normal distribution.

Second approach assumes that the original variable Y has "the power-normal distribution" and the power-transformed variable $Y^{(\lambda)}$ has near-normal distribution. It is clear that this approach intends to achieve normality and the name " power-normal transformation" stresses the fact. As above mentioned, the distribution of power-transformed variables become a right or left truncated normal distribution according to the sign of transforming parameter. In this case, the inference procedure is based on the maximum likelihood allowing for the magnitude of truncated probability in the distribution of the power-transformed variable. Because the evaluation of truncated probability become difficult in the practice, the inference on parameters such as mean or variance is usually performed in which the transforming parameter is assumed to be fixed. Then, it will be similar to the first procedure.

The applications of the PND are developed in the various models and analysis such as regression, survival, a multi-sample problem, bioassay, ANOVA and multivariate analysis, and it shows good performances. The genealogy of development and investigation on the power-transformation and the PND is shown in Fig.1.3.1. The main subjects related with this paper are reviewed as follows. See Goto *et al.* (1991) and Sakia (1992) for detailed reviews.

Data with Structure

Power-transformation is widely used in regression analysis. Power-transformation is performed to both or either of response and explanatory variable, where it is focused only on the improvements of the models. Ordinary power-transformation of observation with non-normal error in regression analysis dose not necessary imply the linearity of regression model. Nelder (1968) considered various transformations models satisfy both normality of error and linearity of regression model, and proposed a transformation methods of applying power-transformation to the former aim and log-transformation to the latter. The method has been followed by Wood (1974). An impressive example has been given in his paper. Goto *et al.* (1987) propose to apply the power-transformation to intend the achievement of two of (i) additivity or linearity of the model, (ii) constant variance of error term, and (iii) normality of response observation, and evaluate some performance of it. They refer to the transformation as double power-transformation.

Diagnostic and Robust Estimation

Maximum likelihood estimation is generally used to obtain parameter estimates. Several other methods are proposed as alternative estimation. Andrews (1971) recommends the estimators based on significant level of hypotheses testing for transforming parameter, and Hinkley (1975) gives the estimators based on the ordered statistics of sample (data). In addition, Hinkley (1977) and Carroll (1980) show the robust estimation for transforming parameter.

For diagnostic based on power-transformation, since Atkinson (1982) has proposed constructed variable and related plots, Cook and Wang (1983), Carroll and Ruppert (1985), Goto and Hatanaka (1985) and Atkinson (1986) improve the method or propose new type diagnostic tool.

Transformation for Counts Data and Discrete Distribution

We encounter responses with more than two ordered categories. Several explanatory variables or covariate are usually observed corresponding to the response, we are often interested in examining the effect due to the explanatory variables on the response. In these circumstances, models in which a function of the response probability is expressed by linear combination of the explanatory variables are often applied. Logistic and proportional odds models are well known among such models. However, the response function, which indicates transformation of the response probabilities in these models, is merely hypothetical. In general, there is no evidence that symmetry distribution is assumed as a distribution on the latent scale to ordered categorical response. Thus, it is necessary to diagnose and examine their appropriateness in any way. In order to answer to these requirements, the two models of power-transformation can be applied. One is asymmetric power-transformation (Guerrero and Johnson, 1982), and the other is symmetric power-transformation (Arandaz-Ordaz, 1981). These two transformations can also be applied to observation subjected to an upper constraint (Atkinson, 1985).

Bioassay

Finney (1978) has discussed various transformations of metameters of dose and response in bioassay with quantitative response, and describes in detail the application of power-transformation to examine the adequacy.

In general dichotomous dose-response relation, use of the PND as tolerance distribution is proposed as an extension view of probit analysis by Uesaka *et al.* (1981).

ANOVA

Applications of variable transformation to analysis of variance have been discussed for a long time by many researchers. Many of them have aimed to achieve additivity of model and homoscedasticity of error variance. Power-transformation using maximum likelihood method tries to satisfy normality of residuals in addition to above two aims. In selecting the best transformation, it is said to be appropriate to graph the changes on F-statistics of each factors, homoscedasticity test criteria, variance and skewness and/or kurtosis of residuals, caused by the changes of transforming parameter. Criteria of homoscedasticity and normality are considered to be appropriate for selection criteria when the observations are measured repeatedly [see Moore and Tukey (1954), Draper and Hunter (1969), Schlesselman (1971, 1973), Fuchs (1978,1979) and Goto *et al.* (1983) for detailed discussion about this topics].

Multivariate Analysis

Many of the theories of multivariate analysis are based on multivariate normal distribution. Thus, to assure high theoretical properties in application, it is often desired to apply transformation bridging the distribution of multivariate observations close to multivariate normal. In most of these cases applications of power-transformation seem to be useful. Then, we can estimate the transforming parameter variables-wise or jointly following some procedures of Andrews *et al.* (1971) and Gnanadesikan (1977).

To evaluate multivariate normality of power-transformed observations, we can utilize test on multivariate skewness and kurtosis (Mardia, 1970, 1974), the third and the fourth cumulants test (Uesaka and Goto, 1979) or directional normality test [see Andrews *et al.* (1971), Goto *et al.* (1978) for details].
Other transformation procedures based on directional projection have been proposed by Andrews *et al.* (1971). Hatanaka, Goto and Nagai (1980) and Hatanaka, Inoue and Goto (1981) have considered discriminant analysis on the PND and evaluated some relative performances of it to ordinary Fisher's linear discrimination. Furthermore, Goto *et al.* (1980, 1981a, 1981b, 1982), Kawai *et al.* (1997) and Jimura *et al.* (1997) have given some theoretical examination of definition and properties of the bivariate power-normal distribution as the first step of generalization of univariate to multivariate power-normal distribution

Analysis of Repeated Measurement Experiment and Growth Curve

In this case, a variate with repeated measurements has been observed on a common characteristic. And all the variables can be uniformly power-transformed; selecting transforming parameter affirms the adequacy of the analysis on the scale after transformation.

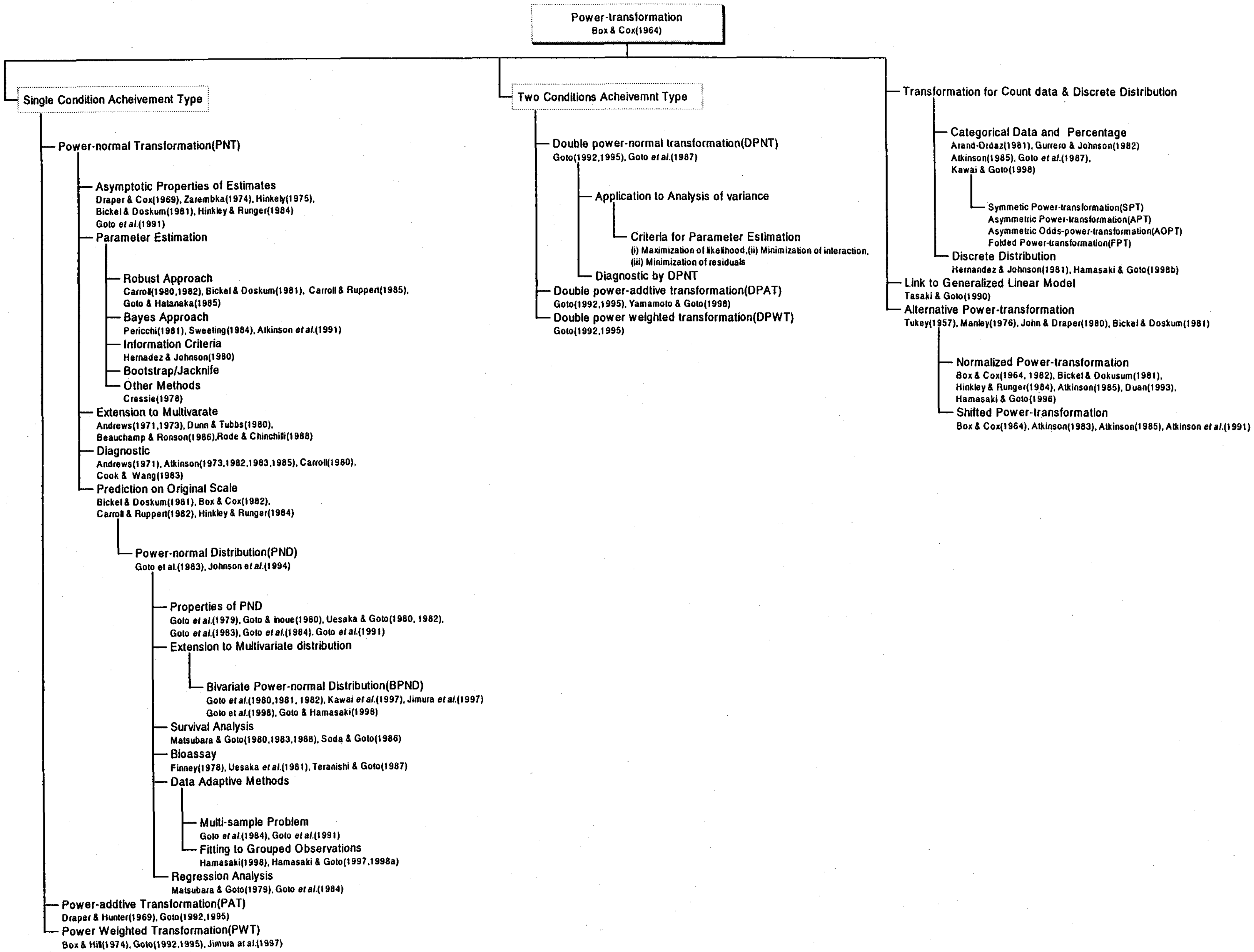

Figure 1.3.1 The genealogy of development and investigation on the power-transformation and the PND

2
Univariate Grouped Observations

2.1 Power-normal Distribution

For a positive random variable Y, the power-transformation is defined as

$$Y^{(\lambda)} = \begin{cases} \dfrac{Y^{\lambda}-1}{\lambda}, & \lambda \neq 0, \\ \log Y, & \lambda = 0 \end{cases} \tag{2.1.1}$$

where λ is called the transforming parameter (Box and Cox, 1964). Then, the distribution of power-transformed variable $Y^{(\lambda)}$ will become a near normal, which is the right or left-truncated distribution according to the sign of λ. Thus, Y has the power-normal distribution (PND), which probability density function is given by

$$g(y:\lambda,\mu,\sigma) = \frac{y^{\lambda-1}}{A(\kappa)\sqrt{2\pi}\sigma}\exp\left\{-\frac{(y^{(\lambda)}-\mu)^2}{2\sigma^2}\right\} \tag{2.1.2}$$

where μ and σ^2 are mean and variance in which the power-transformed variable $Y^{(\lambda)}$ of Y have a near-normal distribution, $A(\kappa)$ is the probability proportional constant term of the PND given by

$$A(\kappa) = \begin{cases} \Phi(\mathrm{sgn}(\lambda)|\kappa|), & \lambda \neq 0, \\ 1, & \lambda = 0 \end{cases} \tag{2.1.3}$$

where $\kappa = (\lambda\mu + 1)/(\lambda\sigma)$ is standardized truncation point of the near-normal distribution, and $\Phi(\cdot)$ is cumulative distribution function of standard normal distribution (Goto *et al.*, 1983: Johnson *et al.*, 1994).

Probability density function $g(y:\lambda,\mu,\sigma^2)$ have the 6 typical shapes; namely $\lambda>1$ (J-shaped distribution), $\lambda=1$ (truncated normal distribution), $c<\lambda<1$, $\lambda=c$, $0<\lambda<c$, and $\lambda<0$ (L-shaped distribution), where $c=4/(\kappa^2+4)$.

The m th moment about zero for the PND is given by

$$\mathrm{E}\left[Y^m\right]=\begin{cases}\dfrac{(\sqrt{2}\lambda\sigma)^{m/\lambda}}{\sqrt{2}}\dfrac{\phi(\kappa)}{A(\kappa)}\displaystyle\sum_{v=0}^{\infty}\frac{(\sqrt{2}\kappa)^v}{v!}\Gamma\left(\frac{m}{2\lambda}+\frac{v+1}{2}\right), & \lambda>0,\\ \exp\left(m\mu+\dfrac{1}{2}m^2\sigma^2\right), & \lambda=0\end{cases} \tag{2.1.4}$$

where $\phi(\cdot)$ is a probability density function of standard normal distribution, and $\Gamma(\cdot)$ is Gamma function. However, for $\lambda<0$, it will become

$$g(y:\lambda,\mu,\sigma)=o(y^{\lambda-1}). \tag{2.1.5}$$

Then, the moments over $|\lambda|$ do not exist. Therefore, only if $\lambda\geq 0$, mean, variance, skewness and kurtosis can be obtained (Uesaka and Goto, 1980). Also, s quintiles $\varsigma(s)$ is represented by

$$\varsigma(s)=(1+\lambda z_s)^{1/\lambda} \tag{2.1.6}$$

where

$$\Phi\left(\frac{z_s-\mu}{\sigma}\right)=\begin{cases}s\Phi(\kappa)+\Phi(-\kappa), & \lambda\neq 0,\\ s\Phi(-\kappa), & \lambda=0.\end{cases} \tag{2.1.7}$$

Therefore, midpoint of this distribution M_e is given by

$$M_e=\begin{cases}\lambda\sigma(\eta+\kappa)^{1/\lambda}, & \lambda\neq 0,\\ \exp(\mu), & \lambda=0\end{cases} \tag{2.1.8}$$

where η is a quintiles defined by

$$\int_{-\infty}^{\eta}\phi(\lambda)dt=\begin{cases}\dfrac{1}{2}A(\kappa)+\Phi(-\kappa), & \lambda>0,\\ \dfrac{1}{2}A(\kappa), & \lambda<0.\end{cases} \tag{2.1.9}$$

If $A(k)$ is nearly one, $(1+\lambda\mu)^{1/\mu}$ will be an approximation value of midpoint M_e. See Goto *et al.* (1979), Goto and Inoue (1980), Uesaka and Goto (1980), Uesaka and Goto (1982), Goto *et al.* (1983), Goto *et al.* (1984), Goto *et al.* (1991) and Johnson *et al.* (1994) for the detailed discussions about the properties and performances of the PND. Thus, setting $\boldsymbol{\theta}^{\mathrm{T}}=(\lambda,\mu,\sigma^2)$, the log-likelihood function for the sample of size n is given by

$$l_n(\boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i^{(\lambda)} - \mu)^2 + (\lambda - 1)\sum_{i=1}^{n}\log y_i - n\log A(\kappa). \tag{2.1.10}$$

The maximum likelihood estimates of λ, μ and σ^2 are obtained by maximizing this log-likelihood function over λ, μ and σ^2. Then, assuming $A(\kappa) = 1$ in this situation equals that $Y_i^{(\lambda)}$ have a normal distribution with mean μ and variance σ^2 in Box and Cox (1964). For fixed λ, the maximum likelihood estimates of μ and σ^2 are

$$\hat{\mu}_n(\lambda) = \sum_{i=1}^{n}\frac{y_i^{(\lambda)}}{n}, \tag{2.1.11}$$

$$\hat{\sigma}_n^2(\lambda) = \sum_{i=1}^{n}\frac{\{y_i^{(\lambda)} - \hat{\mu}(\lambda)\}^2}{n} \tag{2.1.12}$$

respectively. Substitution of the these maximum likelihood estimates into the log-likelihood function given by (2.1.4) yields, apart from constant

$$l_n(\hat{\boldsymbol{\theta}}_n(\lambda)) = -\frac{n}{2}\log\hat{\sigma}_n^2(\lambda) + (\lambda - 1)\sum_{i=1}^{n}\log y_i. \tag{2.1.13}$$

The maximum likelihood estimate $\hat{\lambda}$ is the value of the transforming parameter λ for which the maximized log-likelihood is a maximum. Furthermore, substitution of $\hat{\lambda}_n$ into $\hat{\mu}_n(\lambda)$ and $\hat{\sigma}_n^2(\lambda)$ yields the maximum likelihood estimates $\hat{\mu}_n(\hat{\lambda}_n)$ and $\hat{\sigma}_n^2(\hat{\lambda}_n)$ respectively.

2.2 Completely Grouped Observations

2.2.1 Fitting the PND to Grouped Observations

Let the observations of variable Y be grouped into $k(k \geq 3)$ intervals denoted by $I_1 = [y_0, y_1)$, $I_2 = [y_1, y_2), \cdots, I_k = [y_{k-1}, y_k)$, where $0 = y_0 < y_1 < \cdots < y_{k-1} < y_k = \infty$. The frequency of the observations lying within interval $I_i = [y_{i-1}, y_i)(i = 1, 2, \cdots, k)$ will be denoted by n_i, and then the sum of the frequency is $n = \sum_{i=1}^{k} n_i$. Let the probability that the observations lie within k exclusive interval $I_1, I_2, \cdots, I_k$ be denoted by $p_1, p_2, \cdots, p_k$ respectively, where $p = \sum_{i=1}^{k} p_i = 1$. Suppose that the frequency of the observations lying within the interval I_i is denoted by N_i, then its distribution will be a multinomial distribution.

$$\Pr(N_1 = n_1, N_2 = n_2, \cdots, N_k = n_k) = \frac{n!}{n_1!n_2!\cdots n_k!}\prod_{i=1}^{k} p_i^{n_i}. \tag{2.2.1}$$

Usually, the maximum likelihood estimate $\hat{p}_i$ of p_i is given by $\hat{p}_i = n_i/n$.

Table 2.2.1 Fitting the PND to grouped observations

Interval	Upper and lower limit of the interval		Frequency	Probability based on the PND $p_{\mathrm{PN}i}(\boldsymbol{\theta})$
	PND	Near-normal		
I_1	$[0, y_1)$	$[-\infty, y_1^{(\lambda)})$	n_1	$\Phi(z_1)/A(\kappa)$
I_2	$[y_1, y_2)$	$[y_1^{(\lambda)}, y_2^{(\lambda)})$	n_2	$\{\Phi(z_2) - \Phi(z_1)\}/A(\kappa)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
I_i	$[y_{i-1}, y_i)$	$[y_{i-1}^{(\lambda)}, y_i^{(\lambda)})$	n_i	$\{\Phi(z_i) - \Phi(z_{i-1})\}/A(\kappa)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
I_{k-1}	$[y_{k-2}, y_{k-1})$	$[y_{k-2}^{(\lambda)}, y_{k-1}^{(\lambda)})$	n_{k-1}	$\{\Phi(z_{k-1}) - \Phi(z_{k-2})\}/A(\kappa)$
I_k	$[y_{k-1}, \infty)$	$[y_{k-1}^{(\lambda)}, \infty)$	n_k	$\{1 - \Phi(z_{k-1})\}/A(\kappa)$

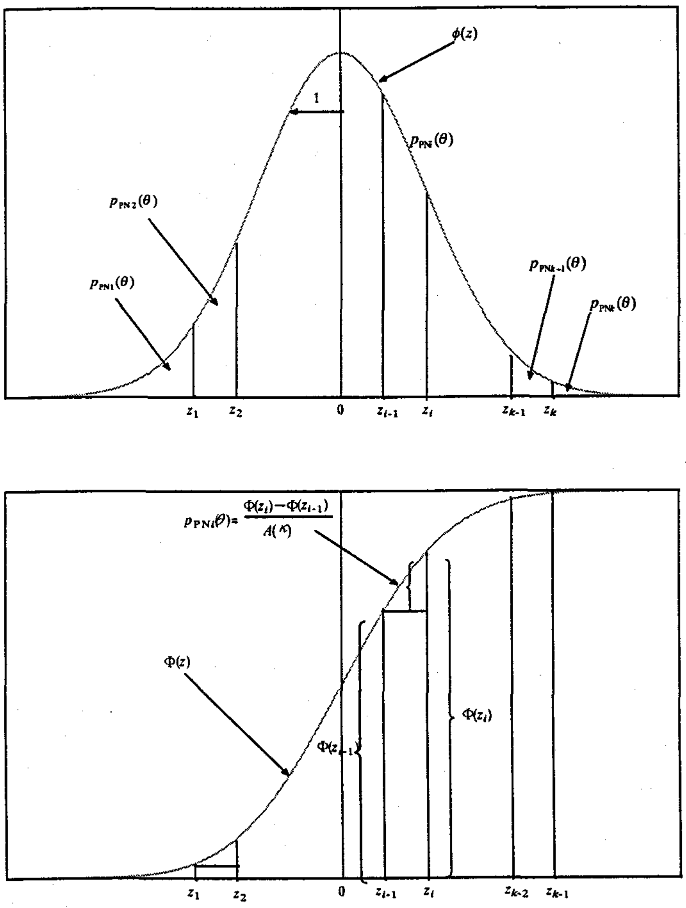

Figure 2.2.1 Fitting the PND to grouped observations

In order to consider the likelihood for grouped observations based on the PND, the probability p_i of the observations lying within the interval I_i is found by first power-transforming the interval $[y_{i-1}, y_i)$ and then applying the PND as a underlying distribution. Therefore, setting $\boldsymbol{\theta}^{\mathrm{T}} = (\lambda, \mu, \sigma)$, the probability $p_{\mathrm{PN}i}(\boldsymbol{\theta})$ based on the PND is given by

$$p_{\mathrm{PN}i}(\boldsymbol{\theta}) = \frac{\Phi(z_i) - \Phi(z_{i-1})}{A(\kappa)} \tag{2.2.2}$$

where $z_i = \{y_i^{(\lambda)} - \mu\}/\sigma$, $p_{\mathrm{PN}1}(\boldsymbol{\theta}) = \Phi(z_1)/A(\kappa)$ and $p_{\mathrm{PN}k}(\boldsymbol{\theta}) = \{1 - \Phi(z_{k-1})\}/A(\kappa)$. Table 2.2.1 and Figure 2.2.1 show these relationships.

Therefore, the likelihood function for the sample of size n is given by

$$L_n(\boldsymbol{\theta}) = \frac{n!}{n_1! n_2! \cdots n_k!} \prod_{i=1}^{k} p_{\mathrm{PN}i}^{n_i}(\boldsymbol{\theta}). \tag{2.2.3}$$

Then, the log-likelihood function becomes

$$l_n(\boldsymbol{\theta}) = \log L_n(\boldsymbol{\theta}) = \log n! - \sum_{i=1}^{k} \log n_i! + \sum_{i=1}^{k} n_i \log p_{\mathrm{PN}i}(\boldsymbol{\theta}). \tag{2.2.4}$$

This log-likelihood function (2.2.4) are non-linear with respect to parameter λ, so the maximum likelihood estimates $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\sigma}$ of λ, μ and σ can not be presented explicitly. Even when λ is given as λ_0 and $A(\kappa) = 1$ is assumed, the maximum likelihood estimates $\hat{\mu}_n(\lambda_0)$ and $\hat{\sigma}_n(\lambda_0)$ of μ and σ can not be presented explicitly as (2.1.4). Then, $\hat{\mu}_n(\lambda_0)$ and $\hat{\sigma}_n(\lambda_0)$ are given by the iterative formula

$$\hat{\mu}_{j+1,n}(\lambda_0) = \hat{\mu}_{j,n}(\lambda_0) + \frac{1}{G}\left(E \frac{\partial l_n(\boldsymbol{\theta}(\lambda_0))}{\partial \mu} - D \frac{\partial l_n(\boldsymbol{\theta}(\lambda_0))}{\partial \sigma} \right), \tag{2.2.5}$$

$$\hat{\sigma}_{j+1,n}(\lambda_0) = \hat{\sigma}_{j,n}(\lambda_0) + \frac{1}{G}\left(-D \frac{\partial l_n(\boldsymbol{\theta}(\lambda_0))}{\partial \mu} + C \frac{\partial l_n(\boldsymbol{\theta}(\lambda_0))}{\partial \sigma} \right) \tag{2.2.6}$$

where

$$C = -\frac{\partial^2 l_n(\boldsymbol{\theta}(\lambda_0))}{\partial \mu^2}, D = -\frac{\partial^2 l_n(\boldsymbol{\theta}(\lambda_0))}{\partial \mu \partial \sigma}, E = -\frac{\partial^2 l_n(\boldsymbol{\theta}(\lambda_0))}{\partial \sigma^2}, G = CE - D^2. \tag{2.2.7}$$

These derivatives associated with (2.2.5), (2.2.6) and (2.2.7) are given in Appendix 1. However, $\hat{\sigma}_n(\lambda_0)$ can be presented explicitly in some situations*. Thus, it is possible to gain some insight into

---

* Once λ and μ are given as λ_0 and μ_0, if $k = 3$ and if the interval limits Y_1 and Y_2 are symmetric with respect to μ_0, the MLE $\hat{\sigma}(\lambda_0, \mu_0)$ of σ exists if and only if $0 < n_2 < n$, and is then given by

$$\hat{\sigma}(\lambda_0, \mu_0) = (y_2 - \mu_0)/\Phi^{-1}(1/2 + n_2/2n)$$

where $\Phi^{-1}(\cdot)$ denote the inverse-function of $\Phi(\cdot)$.

our proposed procedure by investigating the asymptotic properties of $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\sigma}$. In deriving these properties, we will follow the approach to discrete distributions in Hernandez and Johnson (1981). These results will be shown in next section.

2.2.2 Properties of Parameter Estimates

The asymptotic properties of the maximum likelihood estimates $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\sigma}$ of λ, μ and σ will be investigated. In the deriving properties of them, we will follow the approach to discrete distributions in Hernandez and Johnson (1981).

Let $\boldsymbol{\theta}^{\mathrm{T}} = (\theta_1, \theta_2, \theta_3) = (\lambda, \mu, \sigma)$, Suppose that (i) the parameter space Θ is a compact subset $\Re^3$, which is given by $\Theta = \{\boldsymbol{\theta}^{\mathrm{T}} = (\lambda, \mu, \sigma)\}$, and (ii) $H(\boldsymbol{\theta}) = \sum_{i=1}^{k} p_i \log p_{\mathrm{PN}i}(\boldsymbol{\theta})/p_i$ which has a unique global maximum at $\boldsymbol{\theta}_0^{\mathrm{T}} = (\lambda_0, \mu_0, \sigma_0) \in \Theta$ is a continuos function, where p_i = $\int_{I_i} g^*(y)dy$ and g^* is the true probability density function. Also, suppose that (iii) $\boldsymbol{\theta}_0$ be an interior point of Θ, and (iv) the Hessian of $H(\boldsymbol{\theta})$, $\nabla^2 H(\boldsymbol{\theta}_0)$ be nonsingular at $\boldsymbol{\theta}_0$.

We know that $\hat{p}_{i,n} \to p_i$ on an almost sure set, where $\hat{p}_{i,n} = n_i/n$. On this set, Stirling's approximation yields

$$\frac{1}{n}\log n! - \frac{1}{n}\sum_{i=1}^{k}\log n_i! = \frac{1-k}{2n}\log 2\pi - \sum_{i=1}^{k}\hat{p}_{i,n}\log\hat{p}_{i,n} + \frac{1}{2n}\left\{(1-k)\log n - \sum_{i=1}^{k}\log\hat{p}_{i,n}\right\} + O(n^{-1}). \tag{2.3.1}$$

where $O(n^{-1})$ is uniform in $\boldsymbol{\theta}$. By (2.2.4), almost surely

$$\frac{1}{n}l_n(\boldsymbol{\theta}) = \sum_{i=1}^{k}\hat{p}_{i,n}\log p_{\mathrm{PN}i}(\boldsymbol{\theta}) - \sum_{i=1}^{k}\hat{p}_{i,n}\log\hat{p}_{i,n} + o(1). \tag{2.3.2}$$

Therefore, the inequality

$$\left|\frac{1}{n}l_n(\boldsymbol{\theta}) - \sum_{i=1}^{k}p_i\log\frac{p_{\mathrm{PN}i}(\boldsymbol{\theta})}{p_i}\right| \le \left|\sum_{i=1}^{k}\hat{p}_{i,n}\log p_{\mathrm{PN}i}(\boldsymbol{\theta}) - \sum_{i=1}^{k}p_i\log p_{\mathrm{PN}i}(\boldsymbol{\theta})\right| + \left|\sum_{i=1}^{k}\hat{p}_{i,n}\log\hat{p}_{i,n} - \sum_{i=1}^{k}p_i\log p_i\right| + o(1) \tag{2.3.3}$$

hold. Thus, because for $\boldsymbol{\theta} \in \Theta$

$$\left|\sum_{i=1}^{k}\hat{p}_{i,n}\log p_{\mathrm{PN}i}(\boldsymbol{\theta}) - \sum_{i=1}^{k}p_i\log p_{\mathrm{PN}i}(\boldsymbol{\theta})\right| \xrightarrow{a.s.} 0 \tag{2.3.4}$$

as $n \to \infty$ (see Appendix 2) and continuity of $x\log x$, and $\hat{p}_{i,n} \xrightarrow{a.s.} p_i$ as $n \to \infty$, the right hand side of inequality (2.3.3) goes to zero with probability one uniformly in $\boldsymbol{\theta} \in \Theta$. Hence, as $n \to \infty$, for $\boldsymbol{\theta} \in \Theta$

$$\frac{1}{n} l_n(\boldsymbol{\theta}) \xrightarrow{a.s.} H(\boldsymbol{\theta}). \tag{2.3.5}$$

Therefore, it follows from Appendix 3 that $\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}_0$ as $n \to \infty$ by identifying $H(\boldsymbol{\theta})$ with the function $f(\boldsymbol{\theta})$ of that Appendix.

In order to show the asymptotic normality of $\hat{\boldsymbol{\theta}}_n$, the gradient and the Hessian of the log-likelihood function are considered. The gradient of $l_n(\boldsymbol{\theta})$ is given by

$$\nabla l_n(\boldsymbol{\theta}) = \left(\frac{\partial l_n(\boldsymbol{\theta})}{\partial \theta_1}, \frac{\partial l_n(\boldsymbol{\theta})}{\partial \theta_2}, \frac{\partial l_n(\boldsymbol{\theta})}{\partial \theta_3} \right) \tag{2.3.6}$$

and the Hessian of $l_n(\boldsymbol{\theta})$, $\nabla^2 l_n(\boldsymbol{\theta}) = (h_{uv,n}(\boldsymbol{\theta}))$ is the 3×3 symmetric matrix with it elements $h_{uv,n}(\boldsymbol{\theta}) = \partial^2 l_n(\boldsymbol{\theta}) / \partial\theta_u \partial\theta_v$ $(u, v = 1,2,3)$, that is

$$\nabla^2 l_n(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial^2 l_n(\boldsymbol{\theta})}{\partial^2 \theta_1^2}, & \frac{\partial^2 l_n(\boldsymbol{\theta})}{\partial\theta_1 \partial\theta_2}, & \frac{\partial^2 l_n(\boldsymbol{\theta})}{\partial\theta_1 \partial\theta_3} \\ \frac{\partial^2 l_n(\boldsymbol{\theta})}{\partial\theta_2 \partial\theta_1}, & \frac{\partial^2 l_n(\boldsymbol{\theta})}{\partial^2 \theta_2^2}, & \frac{\partial^2 l_n(\boldsymbol{\theta})}{\partial\theta_2 \partial\theta_3} \\ \frac{\partial^2 l_n(\boldsymbol{\theta})}{\partial\theta_3 \partial\theta_1}, & \frac{\partial^2 l_n(\boldsymbol{\theta})}{\partial\theta_3 \partial\theta_2}, & \frac{\partial^2 l_n(\boldsymbol{\theta})}{\partial^2 \theta_3^2} \end{pmatrix}. \tag{2.3.7}$$

It is readily from Appendix 1 that second partial derivatives are continuous on Θ. Using Taylor's formula to expand $n^{-1/2} \nabla l_n(\hat{\boldsymbol{\theta}}_n)$ about $\boldsymbol{\theta}_0$

$$\frac{1}{\sqrt{n}} \nabla l_n(\hat{\boldsymbol{\theta}}_n) = \frac{1}{n} \nabla l_n(\boldsymbol{\theta}_0) + \frac{1}{n} \nabla^2 l_n(\boldsymbol{\theta}_n^*) \left\{ \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \right\} \tag{2.3.8}$$

is obtained, where, $\boldsymbol{\theta}_n^* = \gamma_n \boldsymbol{\theta}_0 + (1 - \gamma_n) \hat{\boldsymbol{\theta}}_n (0 < \gamma_n < 1)$.

Next, since $\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}_0$ as $n \to \infty$, with the assumption that $\boldsymbol{\theta}_0$ is an interior point of Θ, $\nabla l_n(\hat{\boldsymbol{\theta}}_n) = \mathbf{0}$ for all sufficiently large n, on an almost sure set. Therefore, as $n \to \infty$

$$\frac{1}{\sqrt{n}} \nabla l_n(\boldsymbol{\theta}_0) + \frac{1}{n} \nabla^2 l_n(\boldsymbol{\theta}_n^*) \left\{ \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \right\} \xrightarrow{a.s.} \mathbf{0} \tag{2.3.9}$$

so that $n^{-1/2} \nabla l_n(\boldsymbol{\theta}_0)$ and $-n^{-1} \nabla^2 l_n(\boldsymbol{\theta}_n^*) \left\{ \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \right\}$ have the same limiting distribution.

Let D_i be the indicator function of the set I_i, that is, $D_{I_i}(y_i) = 1$ if $y_i \in I_i$ and $D_{I_i}(y_i) = 0$ if $y_i \notin I_i$. And $\boldsymbol{\alpha}_i(\boldsymbol{\theta}_0) = (\alpha_{i1}(\boldsymbol{\theta}_0), \alpha_{i2}(\boldsymbol{\theta}_0), \alpha_{i3}(\boldsymbol{\theta}_0))^{\mathrm{T}}$ have the elements given by

$$\alpha_{ir}(\boldsymbol{\theta}_0) = \left. \frac{\partial \log p_{\mathrm{PN}i}(\boldsymbol{\theta})}{\partial \theta_r} \right|_{\boldsymbol{\theta} = \boldsymbol{\theta}_0}, \quad r = 1,2,3. \tag{2.3.10}$$

Also, suppose that the random vectors $X_1(\boldsymbol{\theta}_0), X_2(\boldsymbol{\theta}_0), \cdots, X_n(\boldsymbol{\theta}_0)$ are identically and independently distributed with $\mathrm{E}[X_1(\boldsymbol{\theta}_0)] = \nabla H(\boldsymbol{\theta}_0) = \mathbf{0}$ and $\mathrm{E}[X_1(\boldsymbol{\theta}_0) X^{\mathrm{T}}(\boldsymbol{\theta}_0)] = \boldsymbol{W}$, where $\boldsymbol{W} = (w_{uv})$ is given by

$$\begin{aligned} w_{uv} &= \sum_{i=1}^{k} p_i \alpha_{iu}(\boldsymbol{\theta}_0) \alpha_{iv}(\boldsymbol{\theta}_0) \\ &= \sum_{i=1}^{k} p_i \left(\left. \frac{\partial \log p_{\mathrm{PN}i}(\boldsymbol{\theta})}{\partial \theta_u} \right|_{\theta_0} \right) \left(\left. \frac{\partial \log p_{\mathrm{PN}i}(\boldsymbol{\theta})}{\partial \theta_v} \right|_{\theta_0} \right) \end{aligned} \tag{2.3.11}$$

Thus, $n^{-1} \nabla l_n(\boldsymbol{\theta}_0)$ can be written as

$$\frac{1}{\sqrt{n}} \nabla l_n(\boldsymbol{\theta}_0) = \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{k} D_{I_i}(y_j) \boldsymbol{\alpha}_i(\boldsymbol{\theta}_0) = \frac{1}{n} \sum_{j=1}^{n} X_j(\boldsymbol{\theta}_0). \tag{2.3.12}$$

Therefore, an application of the multivariate central limit theorem yields that, as $n \to \infty$

$$\frac{1}{\sqrt{n}} l_n(\boldsymbol{\theta}_0) \xrightarrow{d} \mathrm{N}_3(\mathbf{0}, \boldsymbol{W}). \tag{2.3.13}$$

So, the fact that $\hat{p}_{i,n} \xrightarrow{a.s.} p_i$ as $n \to \infty$ implies that

$$\frac{1}{n} h_{uv,n}(\boldsymbol{\theta}) \xrightarrow{a.s.} \frac{\partial^2}{\partial \theta_u \partial \theta_v} \left(\sum_{i=1}^{k} p_i \log \frac{p_{\mathrm{PN}i}(\boldsymbol{\theta})}{p_i} \right) = \frac{\partial^2 H(\boldsymbol{\theta})}{\partial \theta_u \partial \theta_v}. \tag{2.3.14}$$

Hence, as $n \to \infty$, with probability one uniformly in $\boldsymbol{\theta} \in \Theta$

$$\frac{1}{n} h_{lm,n}(\boldsymbol{\theta}_n^*) \xrightarrow{p} \left. \frac{\partial^2 H(\boldsymbol{\theta})}{\partial \theta_l \partial \theta_m} \right|_{\theta_0}.$$

Therefore, Setting $V = \{\nabla^2 H(\boldsymbol{\theta}_0)\}^{-1}$, and applying Slutsky's theorem, as $n \to \infty$

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{a.s.} \mathrm{N}_3(\mathbf{0}, \boldsymbol{VWV}^{\mathrm{T}}) \tag{2.3.15}$$

is obtained.

The importance of the true probability density function g^* is reflected in the asymptotic variance of $\hat{\boldsymbol{\theta}}_n$ derived in above. If one is faced with the task of transforming grouped observations to near normality, the true probabilities p_i can be estimated by the observed frequencies $\hat{p}_{i,n}$. Doing this one obtains a consistent estimate of asymptotic variance of $\hat{\boldsymbol{\theta}}_n$.

These results show that strong consistency and asymptotic normality of the maximum likelihood estimates of λ, μ and σ.

It should be noticed that the asymptotic results of this sections were obtain by pretending that the distribution of the power-transformed variables $Y^{(\lambda_0)}$ of Y was exactly normal for some parameter values μ_0 and σ_0^2. Since $Y^{(\lambda_0)}$ can not have a normal distribution, the likelihood functions were incorrectly derived. Here we determine the meaning of the limit of the maximum likelihood estimates $\hat{\boldsymbol{\theta}}_n$.

Hernandez and Johnson (1980) gave a new interpretation to the transforming parameter λ obtained by Box and Cox (1964) method. This new interpretation is based on the fact that the parameter value found by minimizing the Kullback-Leibler information number between the probability density function corresponding to $Y^{(\lambda)}$ and the normal probability density function is asymptotically equivalent to the maximum likelihood estimates by Box and Cox (1964) method. It is in this sense that we say that the distribution of the power-transformed variable $Y^{(\lambda_0)}$ is "close" to a normal distribution with parameters μ_0 and σ_0^2.

The generalization of the Kullback-Leibler information number interpretation to the both cases of grouped and grouped observations combined with ungrouped is fairly obvious given that the Kullback-Leibler information number is defined for probability density functions with respect to an arbitrary measure ν.

According to Kullback (1968), the Kullback-Leibler information number is defined as

$$\mathrm{I}[f_1; f_2] = \int f_1(y) \log\left[\frac{f_1(y)}{f_2(y)}\right] d\nu(y) \tag{2.3.16}$$

where $f_i(y)(i = 1,2)$ are probability density functions with respect to the measure ν. This number provides a measure of the mean amount of information per observation from f_1 for discriminating between f_1 and f_2.

The definition of mean information per observation is motivated by the following considerations: Let $\mathrm{H}_i(i = 1,2)$ be the hypothesis that Y is from the statistical population with probability density function f_i. Assume we have prior probabilities $\rho(\mathrm{H}_1) = \Pr[\mathrm{H}_1 \text{ is true}]$ and $\rho(\mathrm{H}_2) = 1 - \rho(\mathrm{H}_2)$, then Bayes' theorem gives the posterior probabilities

$$\Pr\left[\mathrm{H}_1 \middle| Y\right] = \frac{f_1(y)\rho(\mathrm{H}_1)}{f_1(y)\rho(\mathrm{H}_1) + f_2(y)\rho(\mathrm{H}_2)} \tag{2.3.17}$$

$$\Pr\left[\mathrm{H}_2 \middle| Y\right] = 1 - \Pr\left[\mathrm{H}_1 \middle| Y\right] \tag{2.3.18}$$

from which, taking the log-odds ratio, we obtain

$$\log\left\{\frac{\Pr\left[\mathrm{H}_1 \middle| Y\right]}{\Pr\left[\mathrm{H}_2 \middle| Y\right]}\right\} = \log\left\{\frac{f_1(y)}{f_2(y)}\right\} + \log\left\{\frac{\rho(\mathrm{H}_1)}{\rho(\mathrm{H}_2)}\right\} \tag{2.3.19}$$

or

$$\log\left\{\frac{f_1(y)}{f_2(y)}\right\} = \log\left\{\frac{\Pr[H_1|Y]}{\Pr[H_2|Y]}\right\} - \log\left\{\frac{\rho(H_1)}{\rho(H_2)}\right\}$$
$$= \text{posterior log} - \text{odds} - \text{prior log} - \text{odds} \qquad (2.3.20)$$
$$= \text{gain in information per observation.}$$

That is, $\log[f_1(y)/f_2(y)]$ is the information in $Y = y$ for discrimination in favor of H_1 against H_2. Its expected value, under H_1 is

$$I[f_1; f_2] = E_{f_1}\left\{\log \frac{f_1(y)}{f_2(y)}\right\} \qquad (2.3.21)$$

i.e., the mean information for discriminating in favor of H_1 against H_2 per observation*.

To make explicit the asymptotic equivalence of the minimizing values for the Kullback-Leibler information number and the maximum likelihood estimates in the completely-grouped observation case (the case of grouped observations combined with ungrouped ones is handled similarly) we consider the Kullback-Leibler information number between two discrete distributions $P_j = \{p_{ji} | p_{ji} > 0, i = 1, \cdots k,$ and $\sum_{i=1}^{k} p_{ji} = 1\} (j = 1,2)$ defined on the same support. According to (2.3.16), the Kullback-Leibler information numbers between P_1 and P_2 is

$$I[P_1; P_2] = \sum_{i=1}^{k} p_{1i} \log\left(\frac{p_{1i}}{p_{2i}}\right). \qquad (2.3.22)$$

Setting

$$Q = \{p_i | p_i = \int_{D_i} g^*(y)dy > 0, i = 1, \cdots, k\},$$

$$N = \{p_{PNi}(\theta) | p_{PNi}(\theta) = \Phi(z_i) - \Phi(z_{i-1}), i = 1, \cdots, k\}$$

it follows easily that, under the condition mentioned to show consistency and asymptotic normality of $\boldsymbol{\theta}_0$ is also that value of $\boldsymbol{\theta}$ which minimize $I[Q; N]$.

---

* (i) Let P_1 and P_2 be two probability measures, since $\nu = P_1 + P_2$ dominates both, then some probability density functions f_1 and f_2 (associated with P_1 and P_2 respectively) always exist.

(ii) $I[f_1; f_2] \geq 0$ with equality holding if and only if $f_1 = f_2$ on a set of f_1-probability one.

Hence, obtaining the maximum likelihood estimates of $\boldsymbol{\theta}$ using the likelihood function (2.2.4) is asymptotically equivalent to finding the minimum of the Kullback-Leibler information number between Q and N.

2.2.3 Algorithm for Parameter Estimation

In order to find the maximum likelihood estimate of $\boldsymbol{\theta}^{\mathrm{T}} = (\lambda, \mu, \sigma)$ in our power-normal distribution approach, we need to use an iterative procedure and this requires some initial estimates. Several methods have been devised for obtaining approximate maximum likelihood estimates for μ and σ [*c.f.* Lindley (1950), Swan (1969), Benn and Sidebottom (1976) and Wolynetz (1979a, 1979b)] which may be used for getting initial values for those parameters. But the problem still remains with respect to selecting an initial value for λ. Thus, we cam apply a specialization to our case, of a two-stage procedure. Namely,

- For fixed λ, maximize $l_n(\boldsymbol{\theta})$ with respect to μ and σ, thus obtaining a value $\hat{\boldsymbol{\theta}}_n^{\mathrm{T}}(\lambda) = (\lambda, \hat{\mu}_n(\lambda), \hat{\sigma}_n(\lambda))$
- maximize the already maximized log-likelihood $l_n(\hat{\boldsymbol{\theta}}_n(\lambda))$ with respect to λ to get $\hat{\boldsymbol{\theta}}_n(\hat{\lambda})$

In practice, the procedures of estimating parameters are based on Newton-Raphson iteration method. Defining the estimates of λ, μ and σ in the l-th iteration by $\hat{\lambda}_l$, $\hat{\mu}_l$ and $\hat{\sigma}_l$ respectively, the algorithm is as follows:

STEP1: Select $\hat{\lambda}_1$ for initial value of $\hat{\lambda}_i$.

STEP2: Select $\hat{\mu}_1(\hat{\lambda}_1)$ and $\hat{\sigma}_1(\hat{\lambda}_1)$ for initial values of $\hat{\mu}_l$ and $\hat{\sigma}_l$.

Obtain $\hat{\kappa}_l$ by solving the equation

$$\kappa_l = \frac{\lambda_i \mu_l + 1}{\lambda_i \sigma_l}.$$

And then calculate $A(\kappa) = A_l$.

Estimate $\hat{\mu}_{l+1}$ and $\hat{\sigma}_{l+1}$ by the equation

$$\begin{bmatrix} \hat{\mu}_{l+1} \\ \hat{\sigma}_{l+1} \end{bmatrix} = \begin{bmatrix} \hat{\mu}_l \\ \hat{\sigma}_l \end{bmatrix} + \begin{bmatrix} \dfrac{\partial^2 l(\boldsymbol{\theta})}{\partial \mu^2} & \dfrac{\partial^2 l(\boldsymbol{\theta})}{\partial \mu \partial \sigma} \\ \dfrac{\partial^2 l(\boldsymbol{\theta})}{\partial \mu \partial \sigma} & \dfrac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma^2} \end{bmatrix}^{-1} \begin{bmatrix} \dfrac{\partial l(\boldsymbol{\theta})}{\partial \mu} \\ \dfrac{\partial l(\boldsymbol{\theta})}{\partial \sigma} \end{bmatrix} \Bigg|_{\lambda=\hat{\lambda}_i, \mu=\hat{\mu}_l, \sigma=\hat{\sigma}_l, \kappa=\hat{\kappa}_l}$$

If $|\hat{\mu}_{l+1} - \hat{\mu}_l| + |\hat{\sigma}_{l+1} - \hat{\sigma}_l| > \delta$, replace l by $l+1$ and return to STEP3, where δ is a constant to judge convergence. In this paper, δ is set at 0.0001.

Estimate $\hat{\lambda}_{i+1}$ by using $\hat{\lambda}_i$, $\hat{\mu}_l$, $\hat{\sigma}_l$ and $\hat{\kappa}_l$ by the equation

$$\hat{\lambda}_{i+1} = \hat{\lambda}_i + \left[\frac{\partial l(\boldsymbol{\theta})}{\partial \lambda} \Big/ \left(-\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \lambda^2}\right)\right]\Bigg|_{\lambda=\hat{\lambda}_i, \mu=\hat{\mu}_l, \sigma=\hat{\sigma}_l, \kappa=\hat{\kappa}_l}$$

STEP3: if $|\hat{\lambda}_{i+1} - \hat{\lambda}_i| > \varepsilon$, replace i by $i+1$ and return to STEP 3, where ε is a constant to judge convergence. In this paper, ε is set at 0.0001.

STEP4: Calculate the log-likelihood form (2.2.4) by using $\hat{\lambda}_{i+1}$, $\hat{\mu}_{l+1}$, $\hat{\sigma}_{l+1}$ and $\hat{\kappa}_{l+1}$.

STEP5: Stop the iteration, and $\hat{\lambda}_{i+1}$, $\hat{\mu}_{i,l+1}$ and $\hat{\sigma}_{i,l+1}$ by $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\sigma}$, respectively.

In the above algorithm, it is important to select $\hat{\mu}_1(\hat{\lambda}_1)$ and $\hat{\sigma}_1(\hat{\lambda}_1)$ for initial values of $\hat{\mu}_l$ and $\hat{\sigma}_l$. Methods such as weighted least square or specifying mid-point in the interval can be used to select initial values. In this paper, weighted least square method was employed. Weighted least square estimates $\tilde{\mu}$ and $\tilde{\sigma}$ of μ and σ are found as following:

Letting

$$\bar{n}_i = \frac{1}{n}\sum_{j=1}^{i-1} n_j, \quad 2 \le i \le k, \tag{2.2.23}$$

$$u_i = \Phi^{-1}(\bar{n}_i) \tag{2.2.24}$$

we have

$$\mathrm{var}(u_i) \cong \frac{\mathrm{var}(\bar{n}_i)}{\{\phi(u_i)\}^2} \cong \frac{\bar{n}_i(1-\bar{n}_i)}{n\{\phi(u_i)\}^2}. \tag{2.2.25}$$

Therefore, the weights w_i is

$$w_i = \begin{cases} \dfrac{\{\phi(u_i)\}^2}{\bar{n}_i(1-\bar{n}_i)}, & 0 < \bar{n}_i < 1 \\ 0, & \bar{n}_i = 0,1 \end{cases} \tag{2.2.26}$$

Hence, weighted least square estimates $\tilde{\mu}$ and $\tilde{\sigma}$ of μ and σ used as initial estimates are given by

$$\tilde{\mu} = \frac{\sum_{i=2}^{k} w_i \{y_i^{(\lambda)}\}^2 - \tilde{\sigma}\sum_{i=2}^{k} w_i u_i}{\sum_{i=2}^{k} w_i}, \tag{2.2.27}$$

$$\tilde{\sigma} = \frac{\sum_{i=2}^{k} w_i \{y_i^{(\lambda)}\}^2 - \left\{\sum_{i=2}^{k} w_i y_i^{(\lambda)}\right\}^2 \Big/ \sum_{i=2}^{k} w_i}{\sum_{i=2}^{k} w_i u_i y_i^{(\lambda)} - \sum_{i=2}^{k} w_i u_i \sum_{i=2}^{k} w_i y_i^{(\lambda)} \Big/ \sum_{i=2}^{k} w_i}. \tag{2.2.28}$$

respectively.

To see that this maximum likelihood estimates $\hat{\boldsymbol{\theta}}_n^{\mathrm{T}} = (\hat{\lambda}, \hat{\mu}_n(\hat{\lambda}), \hat{\sigma}_n(\hat{\lambda}))$ indeed provides the absolute maximum of $l_n(\boldsymbol{\theta})$, we offer the following argument. Let $\hat{\boldsymbol{\theta}}^{\mathrm{T}} = (\hat{\lambda}, \hat{\mu}, \hat{\sigma})$ be the value of θ which maximizes $l_n(\boldsymbol{\theta})$, then

$$\begin{aligned} l_n(\boldsymbol{\theta}) &= \max_{\lambda,\mu,\sigma} l_n(\lambda, \mu, \sigma) \\ &\geq \max_{\lambda} l_n(\lambda, \hat{\mu}_n(\lambda), \hat{\sigma}_n(\lambda)) \\ &\geq l_n(\hat{\lambda}, \hat{\mu}_n(\hat{\lambda}), \hat{\sigma}_n(\hat{\lambda})) \geq l_n(\hat{\boldsymbol{\theta}}). \end{aligned} \tag{2.2.29}$$

The last inequality will be true in view of above step.

In practice, it is not possible to get explicit expressions for $\hat{\mu}_n(\lambda)$ and $\hat{\sigma}_n(\lambda)$, therefore the procedure relies heavily upon the use of the Implicit Function Theorem [*c.f.* Mangasarian (1969)] to assure the existence of such quantities. In the present situation, the conditions required for a correct application of this theorem become:

(i) The matrix $\nabla^2_{(\mu,\sigma)} l_n(\boldsymbol{\theta})$, which elements are the second order partial derivatives of $l_n(\boldsymbol{\theta})$ with respect to μ and σ, is continuous at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, and interior point of Θ,

(ii) $\nabla_{(\mu,\sigma)} l_n(\hat{\boldsymbol{\theta}}_n) = \mathbf{0}$, where $\nabla_{(\mu,\sigma)} l_n(\boldsymbol{\theta}) = (\partial l_n(\boldsymbol{\theta})/\partial\mu, \partial l_n(\boldsymbol{\theta})/\partial\sigma)^{\mathrm{T}}$,

(iii) $\nabla^2_{(\mu,\sigma)} l_n(\boldsymbol{\theta})$ is nonsingular.

Under the assumptions of Section 2.2.2, above conditions (i) and (ii) are clearly satisfied, at least for large samples, and we need only to further assume that (iii) holds in order to apply our procedure.

To find the values $\hat{\boldsymbol{\theta}}_n(\lambda)$, we will have solve, iteratively, the maximum likelihood equations for μ and σ given by Appendix 1. The estimated variance-covariance matrix of $\hat{\boldsymbol{\theta}}_n$ is obtained from

$$\frac{1}{n}\hat{V}\hat{W}\hat{V}^{\mathrm{T}} = n(h_{uv,n}(\boldsymbol{\theta}_0))^{-1}(\hat{w}_{uv})(h_{uv,n}(\boldsymbol{\theta}_0))^{-1} \tag{2.2.30}$$

where the h 's and w 's are given by Appendix 1 and (2.3.12) respectively.

2.2.4 Examples

To have a concrete and practical grasp of the aim in the previous section, and make the impression of them clear, we will fit the PND to some examples cited from published literature. Thus, we consider mainly the following two points;

(i) To consider the estimates and performances in fitting the PND to grouped observations
(ii) To consider the effects of grouping on estimates when observations are subsequently grouped into some intervals

(i) The estimates and performances in fitting the PND to grouped observations

We will take up seven examples cited from literatures for discussion to consider the estimates and performances in fitting the PND to grouped observations. These results of fitting and consideration are as follows.

Table 2.2.1 Exmaple1: The results of fitting of the PND

Maximum log-likelihood		Performance	
Maximum log-likelihood $l(\hat{\theta}(\hat{\lambda}))$	−68.0317	**Goodness of fit**	
Truncated probability $1-A(\kappa)$	1.0000	Chi-square statistics	25.254
Estimate of λ		p -value	0.1919
$\hat{\lambda}$	1.1107	**Shape of original observations**	
95% confidence interval for $\hat{\lambda}$	0.6804 1.5414	Skewness Kurtosis	−0.0125 3.2244
Estimate of μ		**Shape of power-transformed observations**	
$\hat{\mu}(\hat{\lambda})$	96.0289	Skewness	0.0014
Back to the original scale $\hat{\mu}^*(\hat{\lambda})$	67.5341	Kurtosis	3.2236
$\hat{\mu}(1)$	67.5209		
Estimate of σ			
$\hat{\sigma}(\hat{\lambda})$	4.0753		
$\hat{\sigma}(1)$	2.5563		

Example 1(Stuart and Ord, 1986): The data are grouped observations of height for 8,585 adult males in the United Kingdom cited from Stuart and Ord (1986). The results of fitting the PND to these grouped observations are shown in Table 2.4.1. The value of transforming parameter λ was estimated as 1.1107 with the approximate 95% confidence interval (0.6804, 1.5415). This optimized value suggests that these observations have a truncated normal distribution, and the likelihood ratio test provides a convenience value of λ as $\hat{\lambda}=1.0$ (p -value was 0.6312). For the optimized value, we obtained $\hat{\mu}(\hat{\lambda})=96.0289$ and $\hat{\sigma}(\hat{\lambda})=4.0753$. The value of back-transformed $\hat{\mu}(\hat{\lambda})$ to the original scale, $\hat{\mu}^{*}(\hat{\lambda})$ was 67.5341. The back-transformed value $\hat{\mu}^{*}(\hat{\lambda})$ was not far from the corresponding one for the observations on the original scale, $\hat{\mu}(1)$ =67.5209. The plot of the profile of maximized log-likelihood as a function of λ is shown in Figure 2.4.1, together with the value of chi-square statistics. Figure 2.4.1 shows that the values of chi-square statistics have a minimum at the neighborhood of $\hat{\lambda} = 1.1107$. From the estimated variance-covariance matrix of $\hat{\boldsymbol{\theta}}$, we had

$$\hat{\mathrm{var}}[\hat{\lambda}] = 0.000000017400, \quad \hat{\mathrm{var}}[\hat{\mu}(\hat{\lambda})] = 0.00000239513,$$

$$\hat{\mathrm{var}}[\hat{\sigma}(\hat{\lambda})] = 0.000134234351, \quad \hat{\mathrm{cov}}[\lambda, \hat{\mu}(\hat{\lambda})] = 0.000000010042,$$

$$\hat{\mathrm{cov}}[\lambda, \hat{\sigma}(\hat{\lambda})] = 0.000001527671, \quad \hat{\mathrm{cov}}[\hat{\mu}(\hat{\lambda}), \hat{\sigma}(\hat{\lambda})] = 0.000134234351.$$

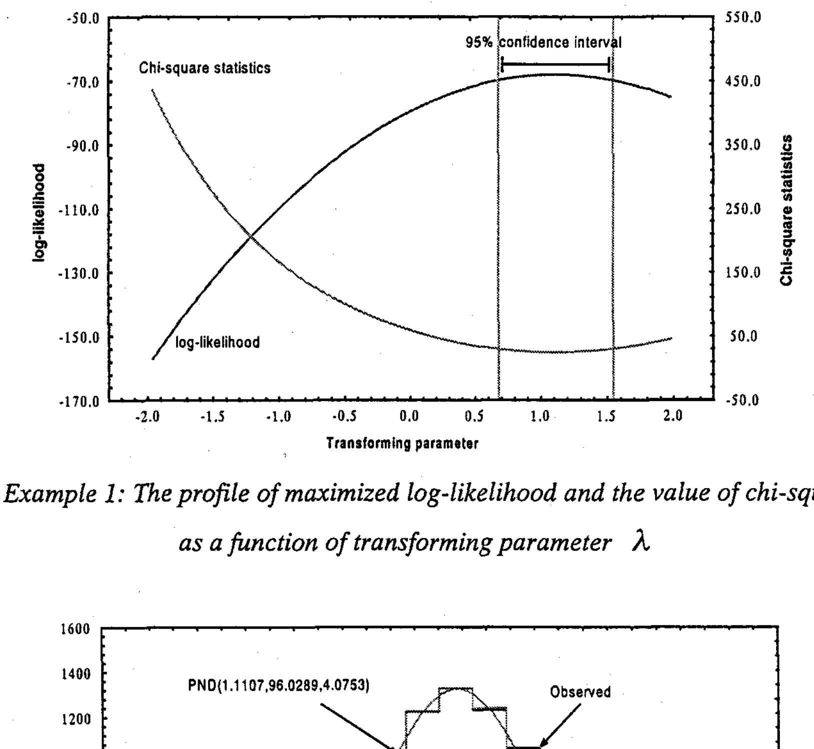

Figure 2.2.1 Example 1: The profile of maximized log-likelihood and the value of chi-square statistics as a function of transforming parameter λ

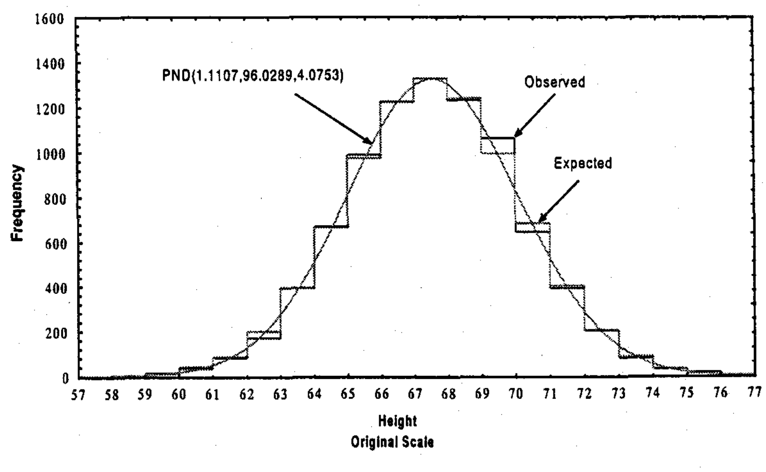

Figure 2.2.2 Example 1: The distribution of observations on the original scale

The distributions of observations on the original and the power-transformed scale are shown in Figure 2.4.2 and 2.4.3 respectively. The value of chi-square statistics for goodness of fit was given as 25.254, which was not quite significant at 5 % level. The values of the skewness and the kurtosis for the observations on the power-transformed scale were given as 0.0014 and 3.2236 respectively. There were little differences between those values on the original and power-transformed scale.

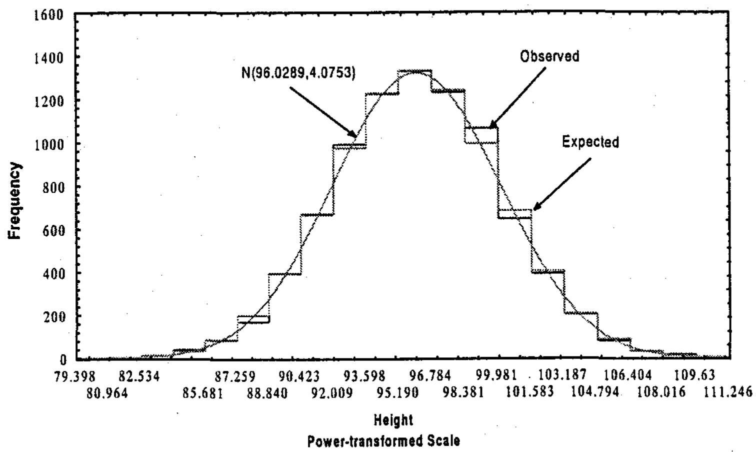

Figure 2.2.3 Example 1: The distribution of observations on the power-transformed scale

Table 2.2.2 Example 2: The results of fitting of the PND

Maximum log-likelihood		**Performance**	
Maximum log-likelihood $l(\hat{\theta}(\hat{\lambda}))$	-106.5696	**Goodness of fit**	
Truncated probability $1-A(\kappa)$	1.0000	Chi-square statistics	138.169
Estimate of λ		p -value	near 0
$\hat{\lambda}$	-0.4691	**Shape of original observations**	
95% confidence interval for $\hat{\lambda}$	-0.5241 -0.4590	Skewness Kurtosis	0.7523 4.5679
Estimate of μ		**Shape of power-transformed observations**	
$\hat{\mu}(\hat{\lambda})$	1.9318	Skewness	-0.0554
Back to the original scale $\hat{\mu}^*(\hat{\lambda})$	155.2640	Kurtosis	3.6820
$\hat{\mu}(1)$	157.2285		
Estimate of σ			
$\hat{\sigma}(\hat{\lambda})$	0.0123		
$\hat{\sigma}(1)$	21.1453		

Example 2 (Stuart and Ord, 1986): The data are grouped observations of weight for 7,749 adult males in the United Kingdom cited from Stuart and Ord (1986). The results of fitting the PND to these grouped observations are shown in Table 2.2.2. The value of transforming parameter λ was estimated as -0.4691 with the approximate 95% confidence interval (-0.5241, -0.4590). This optimized

value suggests that these observations have an L-shaped distribution, and the likelihood ratio test provides a convenience value of λ as $\hat{\lambda} = -0.5$ (p -value was 0.6477). For the optimized value, we obtained $\hat{\mu}(\hat{\lambda}) = 1.9318$ and $\hat{\sigma}(\hat{\lambda}) = 0.0123$. The value of back-transformed $\hat{\mu}(\hat{\lambda})$ to the original scale, $\hat{\mu}^*(\hat{\lambda})$ was 155.2640. The back-transformed value $\hat{\mu}^*(\hat{\lambda})$ was smaller than the corresponding one for the observations on the original scale, $\hat{\mu}(1)$ =157.2285. The plot of the profile of maximized log-likelihood as a function of λ is shown in Figure 2.2.4, together with the value of chi-square statistics. Figure 2.2.4 shows that the values of chi-square statistics have a minimum at the neighborhood of $\hat{\lambda}$ = -0.4691. from the estimated variance-covariance matrix of $\hat{\boldsymbol{\theta}}$, we had

$$\hat{\text{var}}[\hat{\lambda}] = 0.00000000002, \ \ \hat{\text{var}}[\hat{\mu}(\hat{\lambda})] = 0.000000000019,$$

$$\hat{\text{var}}[\hat{\sigma}(\hat{\lambda})] = 0.000000012688, \ \ \hat{\text{cov}}[\lambda, \hat{\mu}(\hat{\lambda})] = -0.000000000006,$$

$$\hat{\text{cov}}[\lambda, \hat{\sigma}(\hat{\lambda})] = 0.000000000000, \ \ \hat{\text{cov}}[\hat{\mu}(\hat{\lambda}), \hat{\sigma}(\hat{\lambda})] = -0.000000000005.$$

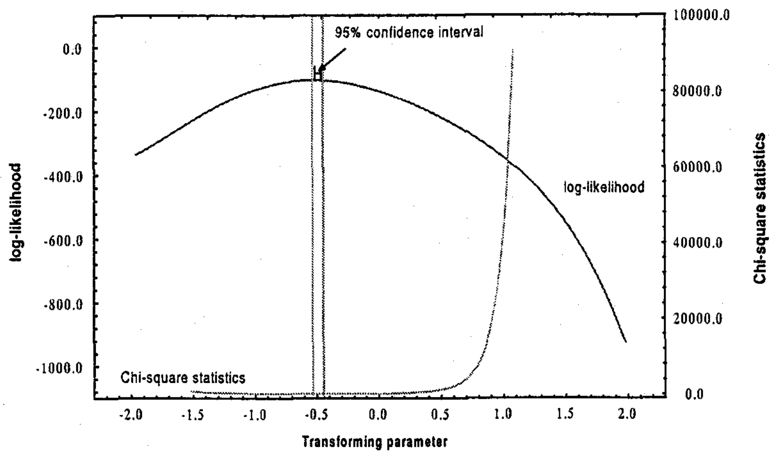

Figure 2.2.4 Example 2: The profile of maximized log-likelihood and the value of chi-square statistics as a function of transforming parameter λ

The distributions of observations on the original and the power-transformed scale are shown in Figure 2.2.5 and 2.2.6, respectively. The value of chi-square statistics for goodness of fit was given as 138.1689, which was very large and significant at 5% level. The value of skewness for the observations on the power-transformed scale was given as -0.0554, which was not farther from zero than the value of 0.7325 for the observations on the original scale. But the value for the kurtosis of the observations on the power-transformed scale was given as 3.6820, which was not farther from 3 than the value of 4.5679 for the observations on the original scale. This shows that the power-transformed observations achieve the near-normality in comparison with the original observations.

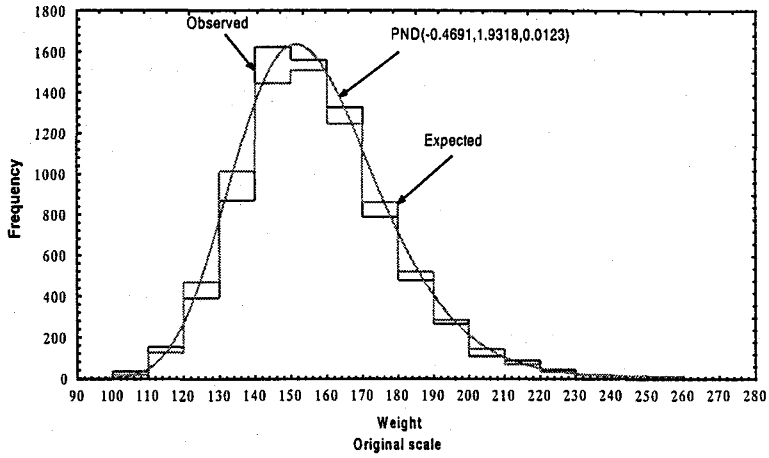

Figure 2.2.5 Example 2: The distribution of observations on the original scale

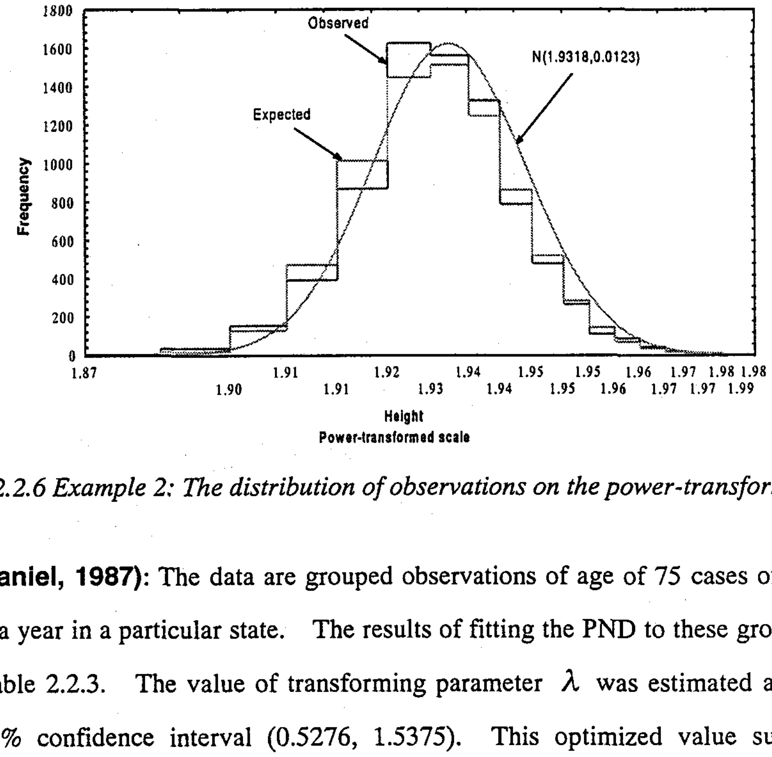

Figure 2.2.6 Example 2: The distribution of observations on the power-transformed scale

Example 3 (Daniel, 1987): The data are grouped observations of age of 75 cases of a certain disease reported during a year in a particular state. The results of fitting the PND to these grouped observations are shown in Table 2.2.3. The value of transforming parameter λ was estimated as 1.0270 with the approximate 95% confidence interval (0.5276, 1.5375). This optimized value suggests that these observations have a truncated normal distribution, and the likelihood ratio test provides a convenience value of λ as $\hat{\lambda} = 1.0$ (p-value was 0.9207). For the optimized value, we obtained $\hat{\mu}(\hat{\lambda}) = 37.4010$ and $\hat{\sigma}(\hat{\lambda}) = 13.7454$. The value of back-transformed $\hat{\mu}(\hat{\lambda})$ to the original scale, $\hat{\mu}^*(\hat{\lambda})$ was given as 35.7822. The back-transformed value $\hat{\mu}^*(\hat{\lambda})$ was not far from the corresponding one for the observations on the original scale, $\hat{\mu}(1)$ =35.7144. The plot of the profile of maximized log-likelihood as a function of λ is shown in Figure 2.2.7, together with the value of chi-square statistics. Figure 2.2.7 shows that the value of chi-square statistics have a minimum at the neighborhood of $\hat{\lambda} = 1.0270$. From the estimated variance-covariance matrix of $\hat{\boldsymbol{\theta}}$, we had

$$\widehat{\text{var}}[\hat{\lambda}] = 0.0000056198041,\quad \widehat{\text{var}}[\hat{\mu}(\hat{\lambda})] = 0.057487932881,$$

$$\widehat{\text{var}}[\hat{\sigma}(\hat{\lambda})] = 1.014278128289,\quad \widehat{\text{cov}}[\lambda, \hat{\mu}(\hat{\lambda})] = -0.005168787157,$$

$$\widehat{\text{cov}}[\lambda, \hat{\sigma}(\hat{\lambda})] = 0.000233051678,\quad \widehat{\text{cov}}[\hat{\mu}(\hat{\lambda}), \hat{\sigma}(\hat{\lambda})] = -0.046395362893.$$

Table 2.2.3 Example 3: The results of fitting of the PND

Maximum log-likelihood		Performance	
Maximum log-likelihood $l(\hat{\boldsymbol{\theta}}(\hat{\lambda}))$	−1.0919	**Goodness of fit**	
Truncated probability $1-A(\kappa)$	0.9974	Chi-square statistics	3.115
Estimate of λ		p -value	0.7943
$\hat{\lambda}$	1.0270	**Shape of original observations**	
95% confidence interval for $\hat{\lambda}$	0.5276 1.5375	Skewness Kurtosis	−0.1407 2.7070
Estimate of μ		**Shape of power-transformed observations**	
$\hat{\mu}(\hat{\lambda})$	37.4010	Skewness	−0.1144
Back to the original scale $\hat{\mu}^*(\hat{\lambda})$	35.7822	Kurtosis	2.6925
$\hat{\mu}(1)$	35.7144		
Estimate of σ			
$\hat{\sigma}(\hat{\lambda})$	13.7454		
$\hat{\sigma}(1)$	12.5062		

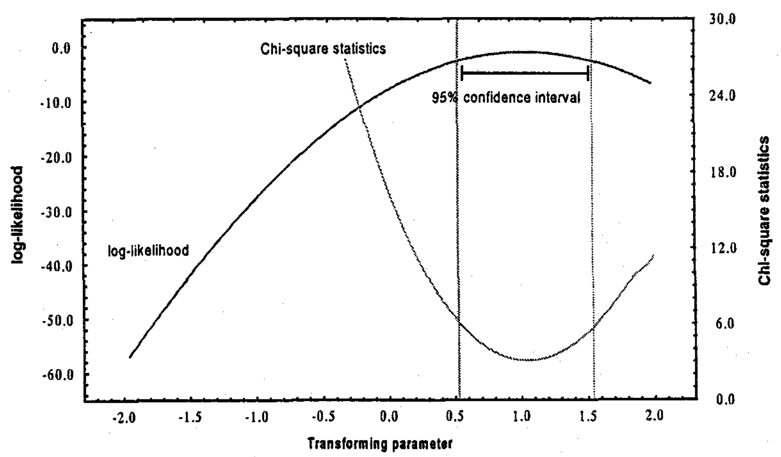

Figure 2.2.7 Example 3: The profile of maximized log-likelihood and the value of chi-square statistics as a function of transforming parameter λ

The distributions of observations on the original and the power-transformed scale are shown in Figure 2.2.7 and 2.2.8 respectively. The value of chi-square statistics for goodness of fit was given as 3.1149, which was not quite significant at 5% level. The values of skewness and the kurtosis for the observations on the power-transformed scale were given as −0.1144 and 2.6925 respectively. There were little difference between those values on the original and the power-transformed scale.

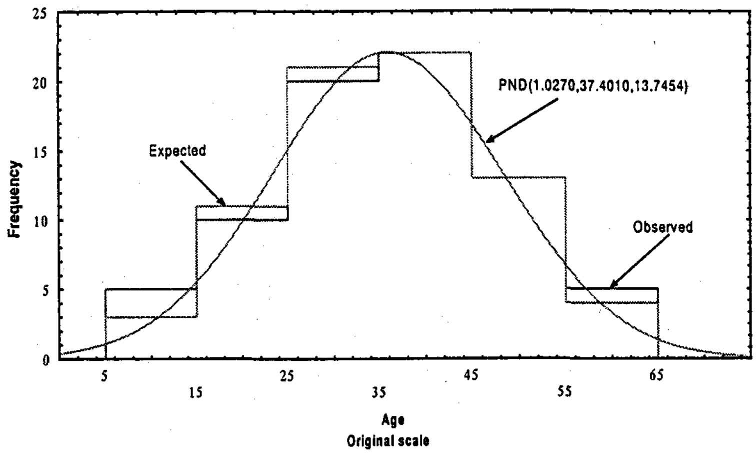

Figure 2.2.8 Example 3: The distribution of observations on the original scale

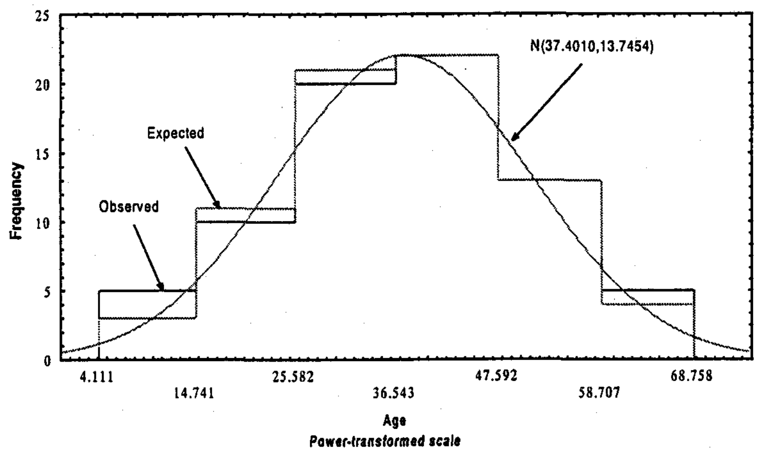

Figure 2.2.9 Example 3: The distribution of observations on the power-transformed scale

Example 4 (Pagano and Gauvreau, 1993): The data are grouped observations of birth weight for 3,751,275 infants cited from Pagano and Gauvreau (1993). The results of fitting the PND to these grouped observations are shown in Table 2.2.4. The value of transforming parameter λ was estimated as 1.9234 with the approximate 95% confidence interval (1.9133, 1.9335). This optimized value suggests that these observations have a J-shaped distribution. For the optimized value, we obtained $\hat{\mu}(\hat{\lambda}) = 3221897.1896$ and $\hat{\sigma}(\hat{\lambda}) = 1019537.7825$. The value of back-transformed $\hat{\mu}(\hat{\lambda})$ to the original scale, $\hat{\mu}^*(\hat{\lambda})$ was given as 3398.920. The back-transformed value $\hat{\mu}^*(\hat{\lambda})$ was smaller than the corresponding one for the observations on the original scale, $\hat{\mu}(1)$ =3348.4705. The plot of the profile of maximized log-likelihood as a function of λ is shown in Figure 2.2.10, together with the value of chi-square statistics. Figure 2.2.10 shows that the value of chi-square statistics have a minimum at the neighborhood of $\hat{\lambda} = 1.9234$. From the estimated variance-covariance matrix of $\hat{\boldsymbol{\theta}}$, we have

$$\widehat{\text{var}}[\hat{\lambda}] = 0.000000000000, \quad \widehat{\text{var}}[\hat{\mu}(\hat{\lambda})] = 40.139600672481,$$

$$\widehat{\text{var}}[\hat{\sigma}(\hat{\lambda})] = 219607.995501813100, \quad \widehat{\text{cov}}[\lambda, \hat{\mu}(\hat{\lambda})] = -0.00001676849,$$

$$\widehat{\text{cov}}[\lambda, \hat{\sigma}(\hat{\lambda})] = 0.000003002326, \quad \widehat{\text{cov}}[\hat{\mu}(\hat{\lambda}), \hat{\sigma}(\hat{\lambda})] = 167.995051562599.$$

Table 2.2.4 Example 4: The results of fitting of the PND

Maximum log-likelihood		**Performance**	
Maximum log-likelihood $l(\hat{\boldsymbol{\theta}}(\hat{\lambda}))$	−56159.2044	**Goodness of fit**	
Truncated probability $1-A(\kappa)$	0.9992	Chi-square statistics	168478.550
Estimate of λ		p -value	near 0
$\hat{\lambda}$	1.9234	**Shape of original observations**	
95% confidence interval for $\hat{\lambda}$	1.9133 1.9335	Skewness Kurtosis	−0.8077 6.0253
Estimate of μ		**Shape of power-transformed observations**	
$\hat{\mu}(\hat{\lambda})$	3221897.1896	Skewness	0.2232
Back to the original scale $\hat{\mu}^*(\hat{\lambda})$	3398.9208	Kurtosis	4.3399
$\hat{\mu}(1)$	3348.4705		
Estimate of σ			
$\hat{\sigma}(\hat{\lambda})$	1019537.7825		
$\hat{\sigma}(1)$	600.0442		

The distributions of observations on the original and the power-transformed scale are shown in Figure 2.2.11 and 2.2.12 respectively. The value of chi-square statistics for goodness of fit was given as 168478.550, which was very extensively large and was significant at 5% level. The value of the skewness for the observations on the power-transformed scale was given as 0.2232, which was not from zero than the value of −0.8077 for the observations on the original scale. But the value of kurtosis for the observations on the power-transformed scale was given as 4.3399, which was not farther from 3 than the value of 6.0253 of the observations on the original scale. This shows that the power-transformed observations achieve the near-normality in comparison with the original observations.

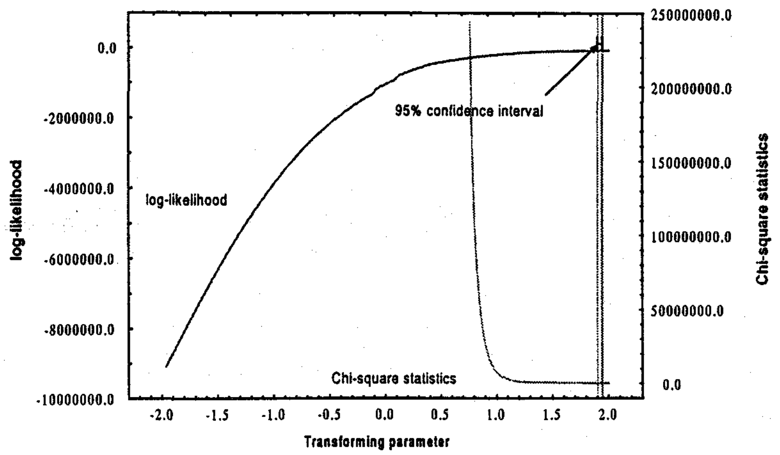

Figure 2.2.10 Example 4: The profile of maximized log-likelihood and the value of chi-square statistics as a function of transforming parameter λ

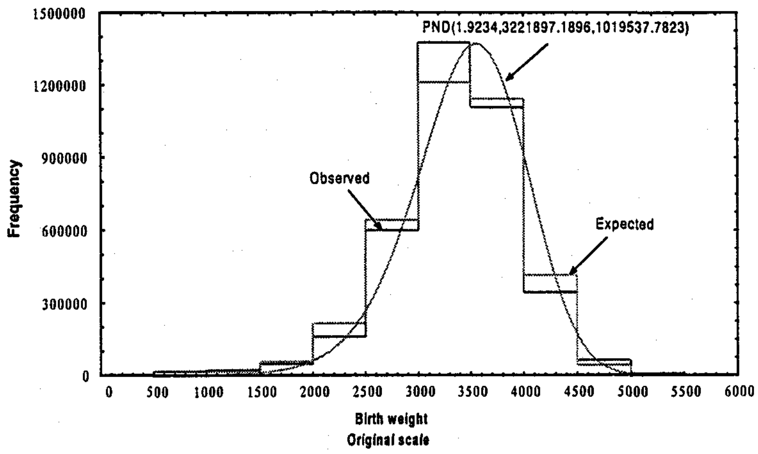

Figure 2.2.11 Example 4: The distribution of observations on the original scale

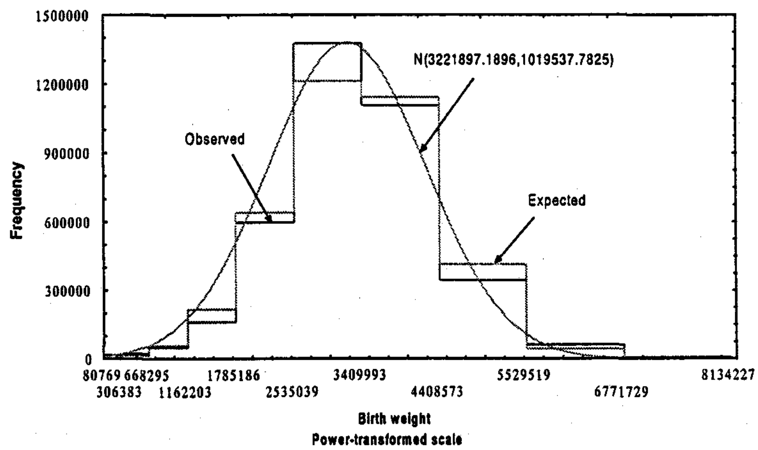

Figure 2.2.12 Example 4: The distribution of observations on the power-transformed scale

Table 2.2.5 Example 5: The results of fitting of the PND

Maximum log-likelihood		**Performance**	
Maximum log-likelihood $l(\hat{\theta}(\hat{\lambda}))$	-1.4422	**Goodness of fit**	
Truncated probability $1-A(\kappa)$	0.9986	Chi-square statistics	31.090
Estimate of λ		p -value	near 0
$\hat{\lambda}$	-1.3852	**Shape of original observations**	
95% confidence interval for $\hat{\lambda}$	-2.5724	Skewness	1.0766
	-0.3833	Kurtosis	3.0067
Estimate of μ		**Shape of power-transformed observations**	
$\hat{\mu}(\hat{\lambda})$	0.7098	Skewness	-0.3990
Back to the original scale $\hat{\mu}^*(\hat{\lambda})$	19.1187	Kurtosis	3.9907
$\hat{\mu}(1)$	20.8372		
Estimate of σ			
$\hat{\sigma}(\hat{\lambda})$	0.0040		
$\hat{\sigma}(1)$	6.0658		

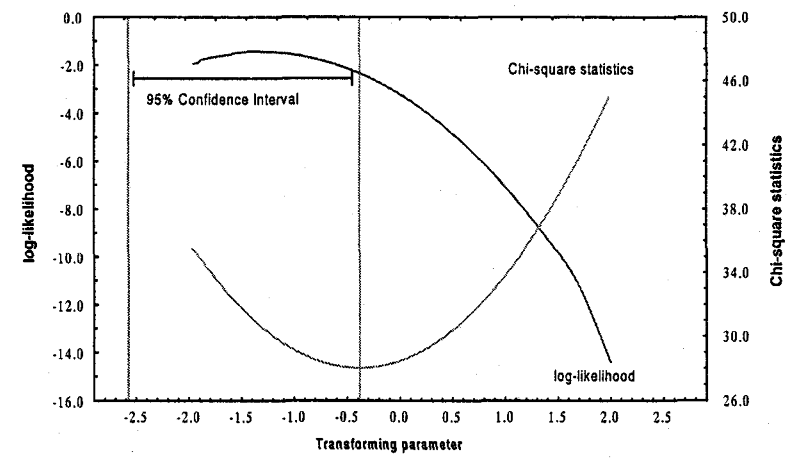

Figure 2.2.13 Example 5: The profile of maximized log-likelihood and the value of chi-square statistics as a function of transforming parameter λ

Example 5 (Siegel and Morgan, 1996): The data are grouped observations of women's professional golf tournaments in 1979 paid the prizes cited from Siegel and Morgan (1996). The results of fitting the PND to these grouped observations are shown in Table 2.2.5. The value of transforming parameter λ was estimated as -1.3852 with the approximate 95% confidence interval (-2.5724, -0.3833). This optimized value suggests that these observations have an L-shaped distribution, and the likelihood ratio test provides a convenience value of λ as $\hat{\lambda} = -1.5$ (p -value was 0.9359). For the optimized value, we obtained $\hat{\mu}(\hat{\lambda}) = 0.7098$ and $\hat{\sigma}(\hat{\lambda}) = 0.0040$. The value of back-transformed $\hat{\mu}(\hat{\lambda})$ to the original scale $\hat{\mu}^*(\hat{\lambda})$ was given as 19.1187. The back-transformed value $\hat{\mu}^*(\hat{\lambda})$ was smaller than the corresponding one for the observations on the original scale, $\hat{\mu}(1)$ =20.8372. The plot of the profile of maximized log-likelihood as a function of λ is shown in Figure 2.2.13, together with the value of chi-square statistics. Figure 2.2.13 shows that the value of chi-square statistics have a minimum at the neighborhood of $\hat{\lambda} = -1.3852$. from the estimated variance-covariance matrix of $\hat{\boldsymbol{\theta}}$, we had

$$\hat{\text{var}}[\hat{\lambda}] = 0.000000016361, \quad \hat{\text{var}}[\hat{\mu}(\hat{\lambda})] = 0.000000003486,$$

$$\hat{\text{var}}[\hat{\sigma}(\hat{\lambda})] = 0.000000202341, \quad \hat{\text{cov}}[\lambda, \hat{\mu}(\hat{\lambda})] = -0.000000007638,$$

$$\hat{\text{cov}}[\lambda, \hat{\sigma}(\hat{\lambda})] = 0.000000009041, \quad \hat{\text{cov}}[\hat{\mu}(\hat{\lambda}), \hat{\sigma}(\hat{\lambda})] = 0.000000003856.$$

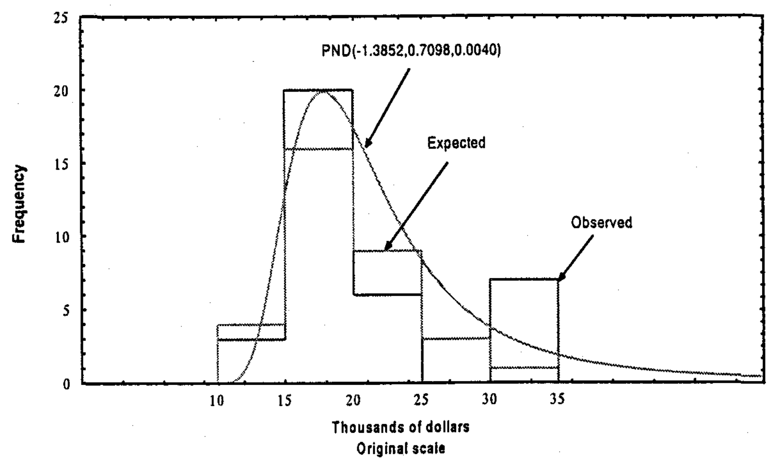

Figure 2.2.14 Example 5: The distribution of observations on the original scale

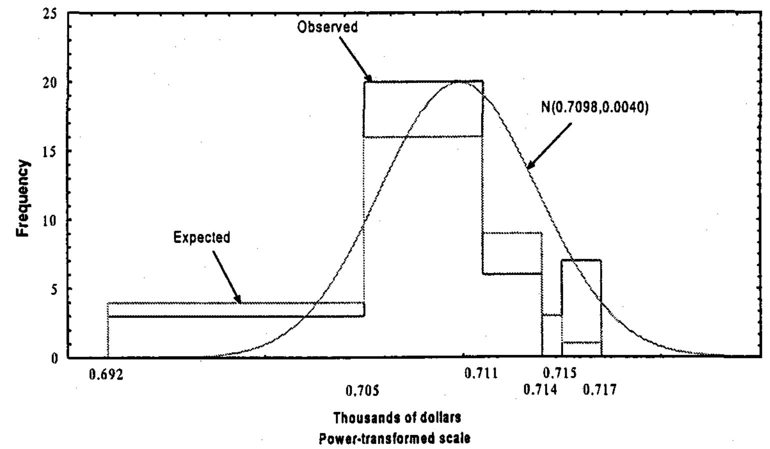

Figure 2.2.15 Example 5: The distribution of observations on the power-transformed scale

The distributions of observations on the original and the power-transformed scale are shown in Figure 2.2.14 and 2.2.15 respectively. The value of chi-square statistics for goodness of fit was given as 31.0904, which was significant at 5% level. The value of the skewness for the observations on the power-transformed scale was given as -0.3990, which was not farther from zero than the value of 1.0766 for the observations on the original scale. But the value of kurtosis for the observations on the power-transformed scale was given as 3.9907, which was farther from 3 than the value of 3.0067 for the observations on the original scale. This shows that the observations on the power-transformed scale achieve only the symmetry of distribution in comparison with the observations on the original scale.

Table 2.2.6 Example 6: The results of fitting of the PND

Maximum log-likelihood		**Performance**	
Maximum log-likelihood $l(\hat{\theta}(\hat{\lambda}))$	-16.7152	**Goodness of fit**	
Truncated probability $1-A(\kappa)$	0.9881	Chi-square statistics	30.3485
Estimate of λ		p -value	0.0001
$\hat{\lambda}$	-2.0010	**Shape of original observations**	
95% confidence interval for $\hat{\lambda}$	-2.7076 -1.7175	Skewness Kurtosis	15.1208 308.0621
Estimate of μ		**Shape of power-transformed observations**	
$\hat{\mu}(\hat{\lambda})$	0.4984	Skewness	-1.2273
Back to the original scale $\hat{\mu}^*(\hat{\lambda})$	19.2109	Kurtosis	5.9269
$\hat{\mu}(1)$	26.0651		
Estimate of σ			
$\hat{\sigma}(\hat{\lambda})$	0.0007		
$\hat{\sigma}(1)$	3.7477		

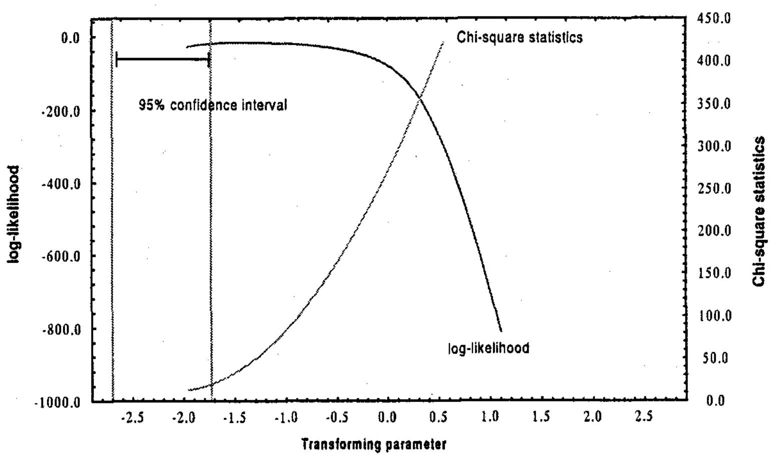

Figure 2.2.16 Example 6: The profile of maximized log-likelihood and the value of chi-square statistics as a function of transforming parameter λ

Example 6 (Bland, 1995): The data are grouped observations of the age of death associated with volatile substance abuse (VSA) mortality for Great Britain cited from Bland (1995). The results of fitting the PND to these grouped observations are shown in Table 2.2.6. The value of transforming parameter λ was estimated as -2.0010 with the approximate 95% confidence interval (-2.6985, -1.7283). This optimized value suggests that these observations had a J-shaped distribution, and the likelihood ratio test provides a convenience value of λ as $\hat{\lambda} = 2.0$ (p-value was 1.0000). For the optimized value, we obtained $\hat{\mu}(\hat{\lambda}) = 0.4984$ and $\hat{\sigma}(\hat{\lambda}) = 0.0007$. The value of back-transformed $\hat{\mu}(\hat{\lambda})$ to the original scale, $\hat{\mu}^*(\hat{\lambda})$ was given as 19.2109. The back-transformed value $\hat{\mu}^*(\hat{\lambda})$ was smaller than the corresponding one for the observations one the original scale, $\hat{\mu}(1)$ =26.0651. The plot of the profile of maximized log-likelihood as a function of λ is shown in Figure 2.2.16, together with the value of chi-square statistics. Figure 2.2.16 shows that the value of chi-square statistics have a minimum at the neighborhood of $\hat{\lambda} = -2.0010$. from the estimated variance-covariance matrix of $\hat{\boldsymbol{\theta}}$, we had

$$\hat{\text{var}}[\hat{\lambda}] = 0.000000000045, \quad \hat{\text{var}}[\hat{\mu}(\hat{\lambda})] = 0.000000000002,$$

$$\hat{\text{var}}[\hat{\sigma}(\hat{\lambda})] = 0.000000001340, \quad \hat{\text{cov}}[\lambda, \hat{\mu}(\hat{\lambda})] = -0.000000000011,$$

$$\hat{\text{cov}}[\lambda, \hat{\sigma}(\hat{\lambda})] = -0.000000000062, \quad \hat{\text{cov}}[\hat{\mu}(\hat{\lambda}), \hat{\sigma}(\hat{\lambda})] = 0.000000000003.$$

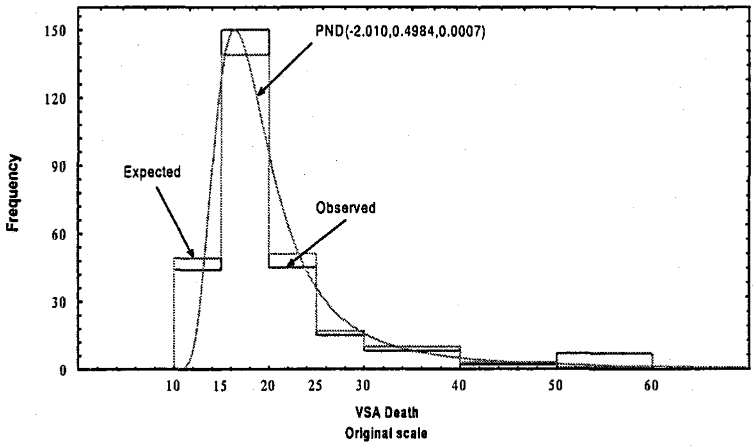

Figure 2.2.17 Example 6: The distribution of observations on the original scale

The distributions of observations on the original and the power-transformed scale are shown in Figure 2.2.17 and 2.2.18 respectively. The value of chi-square statistics for goodness of fit was given as 30.3485, which was significant at 5% level. The value of the skewness for the observations on the power-transformed scale was given as -1.2273, which was not farther from zero than the value of 15.1208 for the observations on the original scale. But the value of the kurtosis for the observations on the power-transformed scale was given 5.9269, which was not farther from 3 than the value of 308.0621 for the observations on the original scale. This shows that the observations on the power-transformed scale achieve the near-normality in comparison with the observations on the original scale.

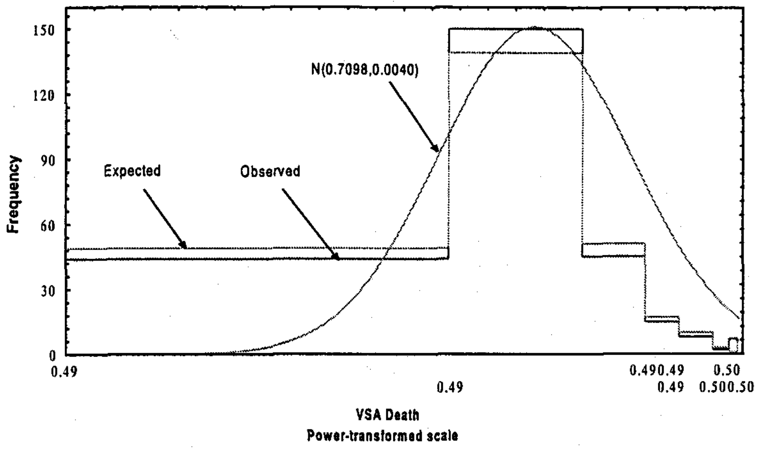

Figure 2.2.18 Example 6: The distribution of observations on the power-transformed scale

Table 2.2.7 Example 7: The results of fitting of the PND

Maximum log-likelihood		**Performance**	
Maximum log-likelihood $l(\hat{\theta}(\hat{\lambda}))$	-28.6022	**Goodness of fit**	
Truncated probability $1-A(\kappa)$	1.0000	Chi-square statistics	5.9625
Estimate of λ		p -value	0.9885
$\hat{\lambda}$	0.6735	**Shape of original observations**	
95% confidence interval for $\hat{\lambda}$	0.4221 0.9304	Skewness Kurtosis	0.3486 3.2930
Estimate of μ		**Shape of power-transformed observations**	
$\hat{\mu}(\hat{\lambda})$	63.2102	Skewness	0.0476
Back to the original scale $\hat{\mu}^*(\hat{\lambda})$	271.5596	Kurtosis	3.0939
$\hat{\mu}(1)$	275.1416		
Estimate of σ			
$\hat{\sigma}(\hat{\lambda})$	12.5686		
$\hat{\sigma}(1)$	77.8525		

Example 7 (Brown and Hollander, 1977): The data are grouped observations of the measurement of serum cholesterol on 500 adult males in a very small city cited from Brown and Hollander (1977). The results of fitting the PND to these grouped observations are shown in Table 2.2.7. The value of transforming parameter λ was estimated as 0.6735 with the approximate 95% confidence interval (0.4221, 0.9304). This optimized value suggests that these observations have a distribution similar to exponential one, and the likelihood ratio test provides a convenience value of λ as $\hat{\lambda} = 0.5$ (p -value was 0.1787). For the optimized value, we obtained $\hat{\mu}(\hat{\lambda}) = 63.2102$ and $\hat{\sigma}(\hat{\lambda}) = 12.5686$. The

value of back-transformed $\hat{\mu}(\hat{\lambda})$ to the original scale $\hat{\mu}^*(\hat{\lambda})$ was given as 271.5596. The back-transformed value $\hat{\mu}^*(\hat{\lambda})$ was smaller than the corresponding one for the observations on the original scale, $\hat{\mu}(1)$ =275.1416. The plot of the profile of maximized log-likelihood as a function of λ is shown in Figure 2.2.19, together with the value of chi-square statistics. Figure 2.2.19 shows that the value of chi-square statistics have a minimum at the neighborhood of $\hat{\lambda} = 0.6735$. From the estimated variance-covariance matrix of $\hat{\boldsymbol{\theta}}$, we had

$$\hat{\text{var}}[\hat{\lambda}] = 0.000000575718, \quad \hat{\text{var}}[\hat{\mu}(\hat{\lambda})] = 0.004200134132,$$

$$\hat{\text{var}}[\hat{\sigma}(\hat{\lambda})] = 0.149770774271, \quad \hat{\text{cov}}[\lambda, \hat{\mu}(\hat{\lambda})] = -0.000015447515,$$

$$\hat{\text{cov}}[\lambda, \hat{\sigma}(\hat{\lambda})] = -0.000003699686, \quad \hat{\text{cov}}[\hat{\mu}(\hat{\lambda}), \hat{\sigma}(\hat{\lambda})] = -0.001334047965.$$

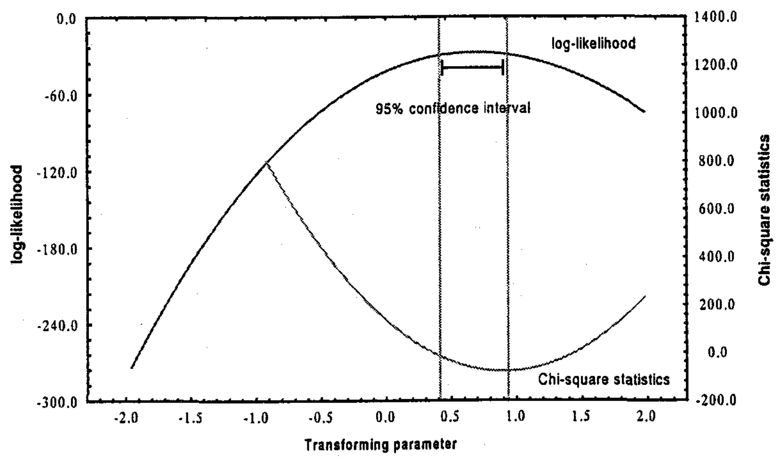

Figure 2.2.19 Example 7: The profile of maximized log-likelihood and the value of chi-square statistics as a function of transforming parameter λ

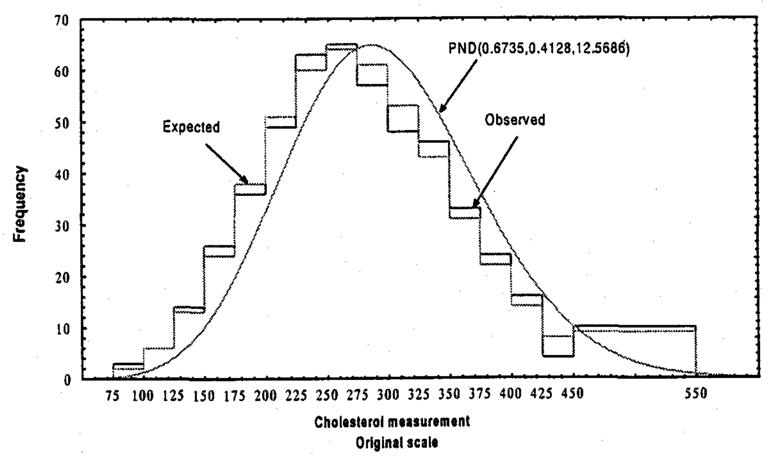

Figure 2.2.20 Example 7: The distribution of observations on the original scale

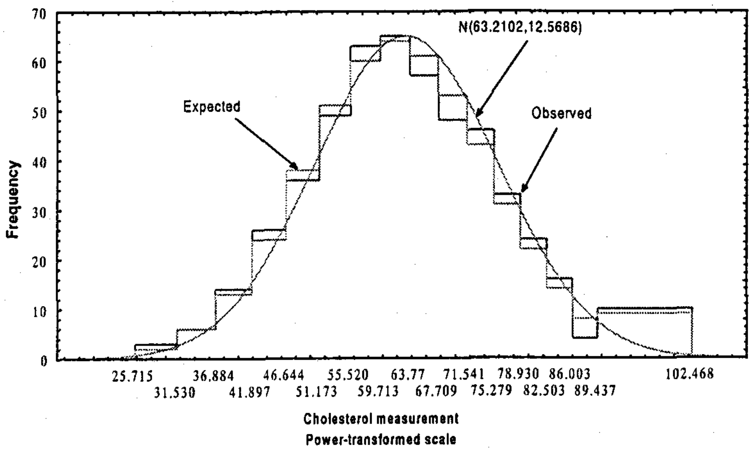

Figure 2.2.21 Example 7: The distribution of observations on the power-transformed scale

The distributions of observations on the original and the power-transformed scale are shown in Figure 2.2.20 and 2.2.21 respectively. The value of chi-square statistics for goodness of fit was given as 5.9625, which was small and not quite significant at 5% level. The value of the skewness for the observations on the power-transformed scale was given as 0.0476, which was nor farther from zero than the value of 0.3486 for the observations on the original scale. But the value of the kurtosis for the observations on the power-transformed scale was 3.0939, which was not farther from 3 than the value of 0.2930 for the observations on the original scale. This shows that the observations on the power-transformed scale achieve the near-normality in comparison with the observation on the original scale.

The results of fitting the PND to seven examples suggest the following knowledge:

- The back transformed mean $\hat{\mu}^*(\hat{\lambda})$ was larger than $\hat{\mu}(1)$ in the range of $\lambda > 1$, though $\hat{\mu}^*(\hat{\lambda})$ was smaller than $\hat{\mu}(1)$ in the range of $\lambda < 1$. This due that the power-transformed mean of original observations is smaller than the mean of power-transformed observations in the range of $\lambda > 1$, though in the range of $\lambda < 1$, it is larger than the mean of power-transformed observations.
- In the both range of $\lambda > 2$ and $\lambda < 0$, goodness of fit for the PND was not well. This is in line with the simulation results for ungrouped observation case by Goto *et al.* (1983) and Hamasaki and Goto (1996).
- Strict normality was not achieved, although the power-transformation yield a nearly symmetrical distribution. Draper and Cox (1969) notice the fact that in some cases of ungrouped observations, even when the transformation procedure dose not yield normality, it helps to "regularize" the data.

(ii) The effects of grouping on estimates when observations are subsequently grouped into some intervals

Next, we consider the effects of grouping on estimates when observations are subsequently grouped into some intervals. Then, we will employ the following four criteria to determine the length and interval.

- AIC minimization criterion (Sakamoto *et al.*, 1981): AIC
- Fisher Information maximization criterion (Nagahata, 1984): Fisher
- Maximization criterion of integrated mean squares of error (Mori, 1975): MSE
- Sturges' criterion (Sturges, 1926): Sturges

We take up three examples for discussion to the effects of grouping on estimates when observations are subsequently grouped into some intervals.

Example 8 (Silverman, 1986): The data are the lengths of 86 spells of psychiatric treatment undergone by patients used as controls in a study if suicide risks. Silverman (1986) divides the domain [0,800] into 20 intervals of length 40 to applying the density smoothing to the data set

Table 2.4.8 Example 8: The results of fitting of the PND

	Ungroup	AIC	Fisher	MSE	Sturges
Maximum log-likelihood					
Maximum log-likelihood $l(\hat{\theta}(\hat{\lambda}))$	−375.7840	−2.9821	−7.8102	1.4488	−5.7401
Truncated probability $1-A(\kappa)$	1.0000	0.6752	1.0000	1.0000	1.0000
Estimate of λ					
$\hat{\lambda}$	0.1930	0.5589	0.0564	0.1485	0.0954
95% confidence interval for $\hat{\lambda}$		−0.4312 1.4710	−0.4697 0.4415	−0.2016 0.3836	−0.2867 0.3575
Estimate of μ					
$\hat{\mu}(\hat{\lambda})$	6.6996	9.7234	4.9456	5.9080	5.2462
Back to the original scale $\hat{\mu}^*(\hat{\lambda})$	73.6887	27.9645	78.4255	69.5083	70.3557
$\hat{\mu}(1)$	122.3140	27.9681	64.9784	69.4934	70.3401
Estimate of σ					
$\hat{\sigma}(\hat{\lambda})$	2.8094	25.3451	1.3398	2.3125	1.7456
$\hat{\sigma}(1)$	146.7513	314.5602	208.8818	194.1371	197.6197
Performance					
Goodness of fit					
Chi-square statistics	-	8.1743	13.9479	29.3819	11.1405
p -value	-	0.0425	0.1242	0.0215	0.6750
Shape of original observations					
Skewness	2.2976	0.8589	1.1510	1.3642	1.3313
Kurtosis	8.4310	1.6204	2.8293	3.6910	3.5557
Shape of power-transformed observations					
Skewness	0.0094	0.8303	0.6442	0.5809	0.5714
Kurtosis	3.0828	1.4967	1.4877	1.5837	1.5667

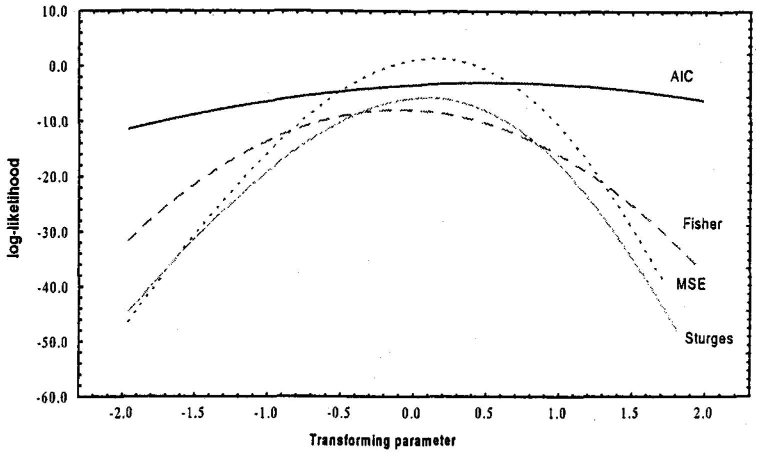

Figure 2.2.22 Example 8: The profile of maximized log-likelihood as a function of transforming parameter λ

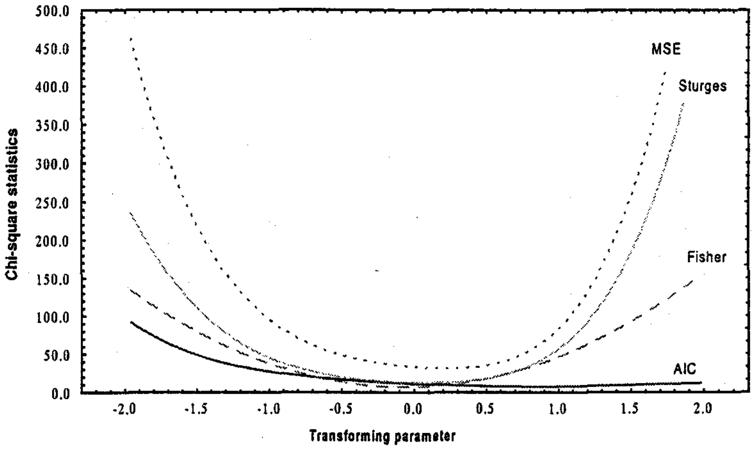

Figure 2.2.23 Example 8: The profile of value of chi-square as a function of transforming parameter λ

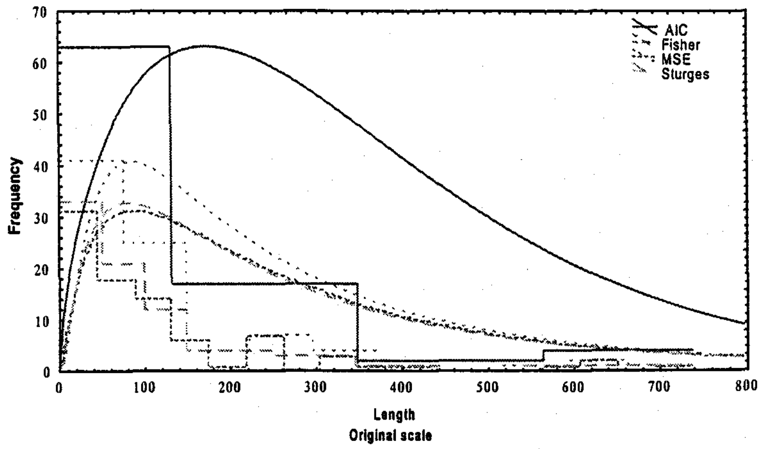

Figure 2.2.24 Example 8: The distribution of observations on the original scale

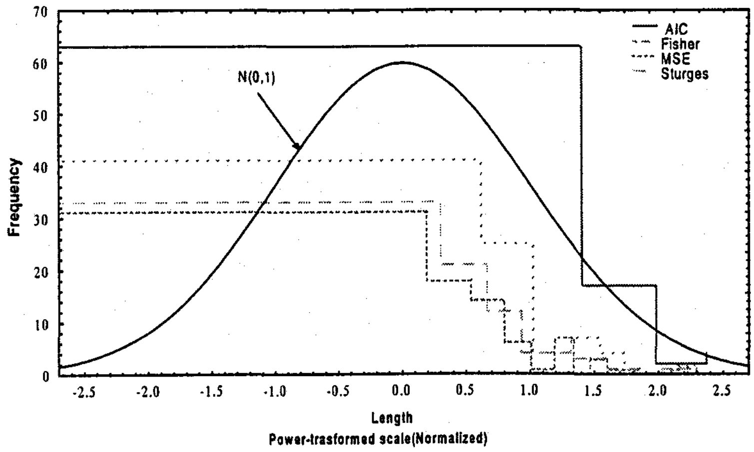

Figure 2.2.25 Example 8: The distribution of observations on the power-transformed scale

The ungrouped observations of this example were grouped into some interval by using the four methods of AIC, Fisher, MSE and Sturges, and so the four grouped observations were obtained. Thus, the numbers of intervals k were 4, 10, 17 and 15 for each of four grouping methods. Then, the PND was applied to each of these four grouped observations. The results of fitting the PND to these four grouped observations are shown in Table 2.4.8 together with that of fitting the PND to ungrouped observations. The optimized value $\hat{\lambda} = 0.1485$ for MSE was closest to $\hat{\lambda} = 0.1930$ for ungrouped observations. Thus, likelihood ratio test provides a convenience value of λ as $\hat{\lambda} = 0.5$ for Fisher, MSE and Sturges, but $\hat{\lambda} = 0.0$ for AIC. The both values of $\hat{\mu}(\hat{\lambda}) = 5.9080$ and $\hat{\sigma}(\hat{\lambda}) = 2.3125$ for MSE were closest to $\hat{\mu}(\hat{\lambda}) = 6.6996$ and $\hat{\sigma}(\hat{\lambda}) = 2.8094$ for ungrouped observations. The value of back-transformed $\hat{\mu}(\hat{\lambda})$ to the original scale for Sturges, $\hat{\mu}^*(\hat{\lambda}) = 70.3557$ was closest to the corresponding one for ungrouped observations on the original scale, $\hat{\mu}^*(\hat{\lambda}) = 73.6887$. The plot of the profiles of maximized log-likelihood for these four grouped observations as a function of λ is shown in Figure 2.2.22. Also, the related plot of profiles of value of chi-square as a function of transforming parameter λ is shown in Figure 2.2.23. These two plot show that the behavior of MSE and Sturges seem similar.

The distributions of observations on the original and the power-transformed scale are shown in Figure 2.2.24 and 2.2.25 respectively. The value of 8.1743 of chi-square statistics for AIC was smallest, which was significant at 5% level. The next smallest value was 11.1405 for Sturges, which was not significant at 5% level. All values of skewness for the observations on the power-transformed scale were not farther from zero than those values for the observations on the original scale. But all values of the kurtosis of power-transformed observations were farther from 3 than those values for the observations on the original scale. These results show that the observations on the power-transformed scale achieve only symmetry of distribution in comparison with the observations on the original scale. Thus, the values of the skewness and the kurtosis for the ungrouped observations on the power-transformed scale were given as 0.0094 and 3.0828 respectively. These two values suggest the satisfaction of normality of distribution for the observations on the power-transformed scale.

Example 9 (Daniel, 1987): The data are the weights in ounces of malignant tumors removed from the abdomens of 57 subjects. Daniel (1987) divides the data into the 7 interval of length 10 by using Sturges' rule.

Table 2.4.9 Example 9: The results of fitting of the PND

	Ungroup	AIC	Fisher	MSE	Sturges
	Maximum log-likelihood				
Maximum log-likelihood $l(\hat{\theta}(\hat{\lambda}))$	−154.3191	−11.2144	−14.1412	−19.1662	−21.4330
Truncated7 probability $1-A(\kappa)$		1.0000	1.0000	1.0000	1.0000
	Estimate of λ				
$\hat{\lambda}$	0.2907	0.1428	0.1210	0.3692	0.4793
95% confidence interval for $\hat{\lambda}$		−0.7043 0.9179	−0.6021 0.7621	−0.3019 0.9833	−0.1898 1.1174
	Estimate of μ				
$\hat{\mu}(\hat{\lambda})$	6.1547	4.5914	4.3662	7.2856	9.3542
Back to the original scale $\hat{\mu}^*(\hat{\lambda})$	34.0739	34.1505	33.2973	34.3380	34.8295
$\hat{\mu}(1)$	36.5263	34.1458	33.2933	34.3341	34.8251
	Estimate of σ				
$\hat{\sigma}(\hat{\lambda})$	1.2641	0.6851	0.6604	1.6775	2.4412
$\hat{\sigma}(1)$	16.1002	16.4320	18.2896	17.0816	16.5143
	Performance				
	Goodness of fit				
Chi-square statistics	-	12.6569	11.6311	15.7027	18.6947
p -value	-	0.0488	0.2349	0.2052	0.1329
	Shape of original observations				
Skewness	0.6852	0.7128	0.7400	0.6978	0.6497
Kurtosis	2.9153	2.1511	1.8267	2.2618	2.2437
	Shape of power-transformed observations				
Skewness	−0.0182	0.1277	0.1482	0.2352	0.2626
Kurtosis	2.5384	2.1142	2.1218	2.0103	2.0472

The ungrouped observations of this example were grouped into some intervals by using the four methods of AIC, Fisher, MSE and Sturges, and so the four grouped observations were obtained. Thus, the numbers of intervals k for each of grouping methods were 7, 10, 13 and 14. Then, the PND was applied to each of these four grouped observations. The results of fitting the PND to these four grouped observations are shown in Table 2.4.9 together with that of fitting the PND to ungrouped observations. $\hat{\lambda}=0.3692$ for MSE was closest to $\hat{\lambda}=0.2907$ for ungrouped observations. Thus, likelihood ratio test suggests a convenience value of λ as $\hat{\lambda} = 0.5$ for MSE and Sturges, but $\hat{\lambda} = 0.0$ for AIC and Fisher. The both values of $\hat{\mu}(\hat{\lambda}) = 7.2856$ and $\hat{\sigma}(\hat{\lambda}) = 1.6775$ for MSE were close to $\hat{\mu}(\hat{\lambda}) = 6.1547$ and $\hat{\sigma}(\hat{\lambda}) = 1.2641$ for ungrouped observations. The value of back-transformed $\hat{\mu}(\hat{\lambda})$ to the original scale for AIC, $\hat{\mu}^*(\hat{\lambda}) = 34.1505$ was closest to the corresponding one for ungrouped observations on the original scale, $\hat{\mu}^*(\hat{\lambda}) = 34.0739$. The plot of the profiles of maximized log-likelihood to these four grouped observations as a function of λ is shown in Figure 2.2.26. Also, the related plot of profiles of value of chi-square as a function of transforming parameter λ is shown in Figure 2.2.27. These two plot show that the behavior of MSE and Sturges seem similar, whereas that of AIC and Fisher seem similar.

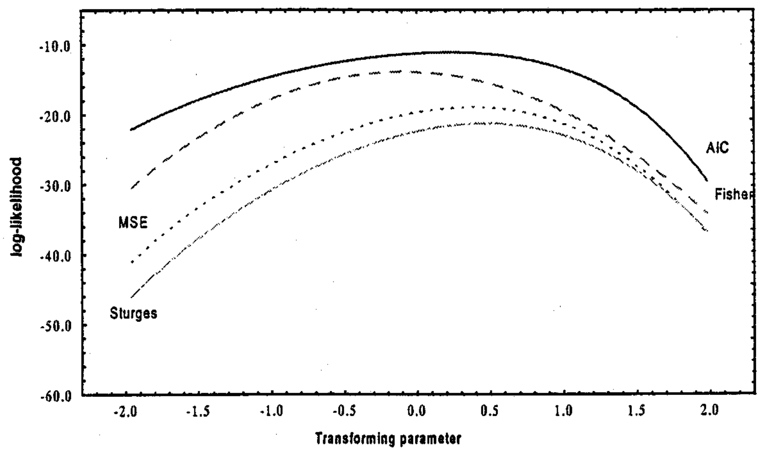

Figure 2.2.26 Example 9: The profile of maximized log-likelihood as a function of transforming parameter λ

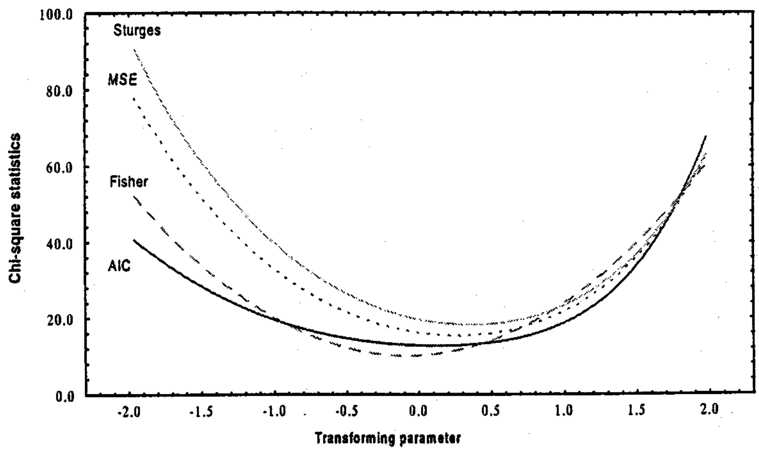

Figure 2.2.27 Example 9: The profile of value of chi-square statistics as a function of transforming parameter λ

The distributions of observations on the original and the power-transformed scale are shown in Figure 2.2.28 and 2.2.29 respectively. The value of 11.6311 of chi-square statistics for Fisher was smallest, which was not quite significant at 5% level. The next smallest value was 12.6569 for AIC, which was significant at 5% level. All values of the skewness for the observation on the power-transformed scale were not farther from zero than those values for the observation on the original scale. But all values of the kurtosis for the observations on the power-transformed scale were farther from 3 than those values for the observation on the original scale. These results show that the observations on the power-transformed scale achieve only symmetry of distribution than the observation on the original scale. Thus, the values of the skewness and the kurtosis for the ungrouped observations on the power-transformed scale were given as –0.0182 and 2.5384 respectively. These two values suggest the only satisfaction of symmetry of distribution for the observations on the power-transformed scale.

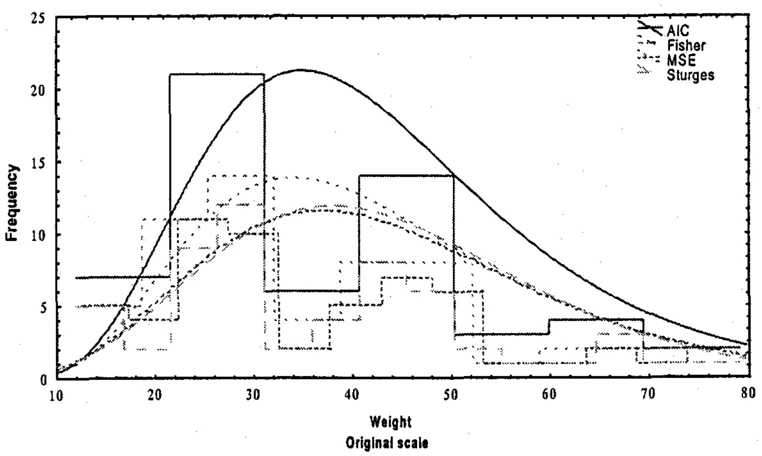

Figure 2.2.28 Example 9: The distribution of observations on the original scale

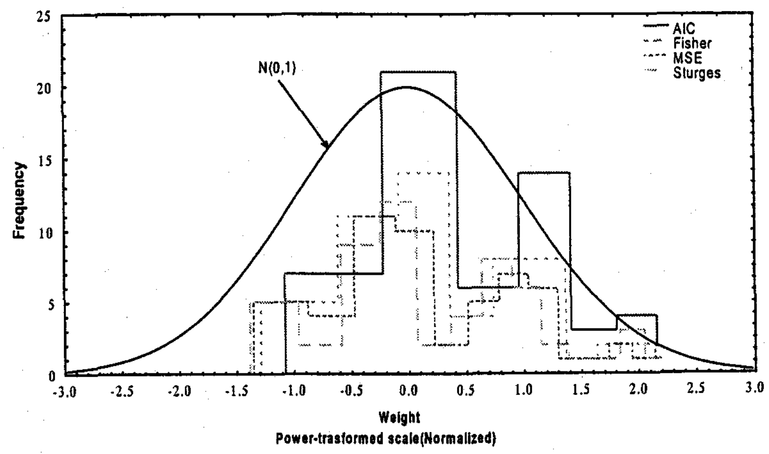

Figure 2.2.29 Example 9: The distribution of observations on the power-transformed scale

Example 10 (Chen, 1995): The data given below are the numbers of cycles to failure for a group of 60 electrical appliance in a life test.

Table 2.4.10 Example 10: The results of fitting of the PND

	Ungroup	AIC	Fisher	MSE	Sturges
Maximum log-likelihood					
Maximum log-likelihood $l(\hat{\theta}(\hat{\lambda}))$	−160.0081	0.7591	1.5721	6.6875	5.8117
Truncated probability $1-A(\kappa)$		0.8052	0.9654	0.9414	0.9477
Estimate of λ					
$\hat{\lambda}$	0.3799	0.8501	0.5519	0.5908	0.5774
95% confidence interval for $\hat{\lambda}$		−0.2010 1.8522	0.0368 0.9800	0.2287 0.8999	0.1643 0.9245
Estimate of μ					
$\hat{\mu}(\hat{\lambda})$	4.9169	11.0851	7.1356	7.4753	7.2828
Back to the original scale $\hat{\mu}^*(\hat{\lambda})$	16.0120	15.7587	18.0585	17.4540	17.4094
$\hat{\mu}(1)$	21.8257	15.7575	18.0569	17.4525	17.4080
Estimate of σ					
$\hat{\sigma}(\hat{\lambda})$	3.1682	14.2510	4.9243	5.8527	5.5547
$\hat{\sigma}(1)$	19.2461	23.8951	22.1464	21.9659	22.1003
Performance					
Goodness of fit					
Chi-square statistics	-	0.0756	6.9247	18.2273	12.8712
p -value	-	0.9946	0.6450	0.3746	0.4578
Shape of original observations					
Skewness	1.2554	0.8244	0.9871	1.0041	0.9888
Kurtosis	5.1255	2.0483	2.6925	2.9859	2.8448
Shape of power-transformed observations					
Skewness	−0.1667	0.6234	0.5585	0.5130	0.5305
Kurtosis	2.3160	1.9906	1.7018	1.7180	1.6661

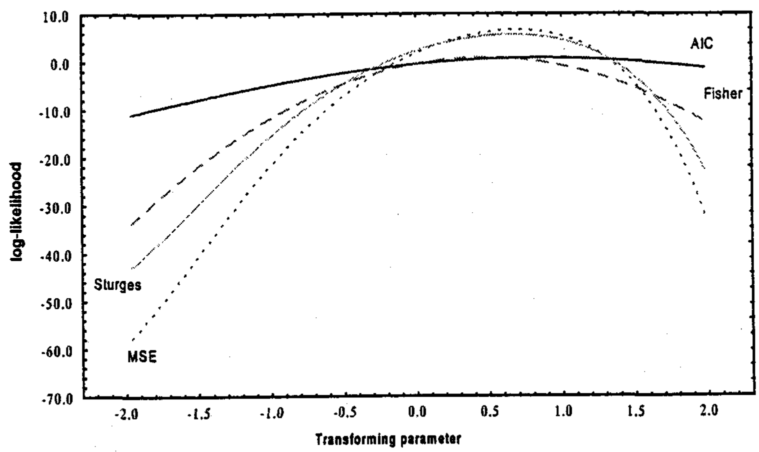

Figure 2.2.30 Example 10: The profile of maximized log-likelihood as a function of transforming parameter λ

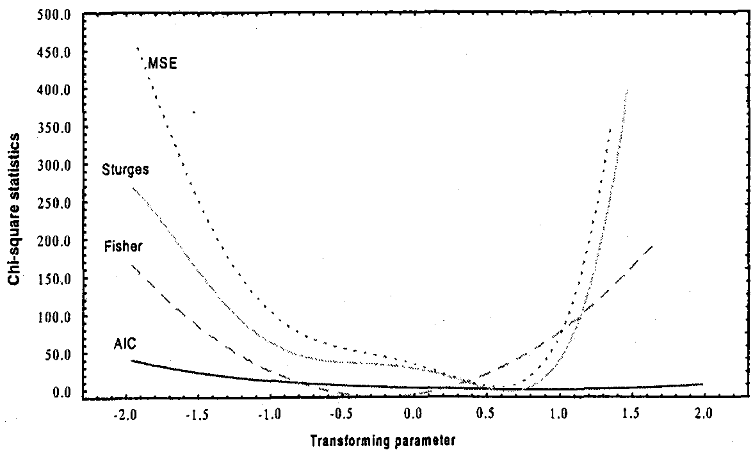

Figure 2.2.31 Example 10: The profile of value of chi-square statistics as a function of transforming parameter λ

The ungrouped observations of this example were grouped into some intervals by using the four methods of AIC, Fisher, MSE and Sturges, and so the four grouped observations were obtained. Thus, the numbers of intervals k were 4, 10, 18 and 14 for each of grouping methods. Then, the PND was applied to each of these four grouped observations. The results of fitting the PND to these four grouped observations are shown in Table 2.4.9 together with that of fitting the PND to ungrouped observations. The optimized value $\hat{\lambda}=0.5519$ for Fisher was closest to $\hat{\lambda}=0.3799$ for ungrouped observations. Thus, likelihood ratio test provides a convenience value of λ as $\hat{\lambda}=0.5$ for Fisher, MSE and Sturges, but $\hat{\lambda}=1.0$ for AIC. The both values of $\hat{\mu}(\hat{\lambda})=7.1356$ and $\hat{\sigma}(\hat{\lambda})=4.9243$ for MSE were close to $\hat{\mu}(\hat{\lambda})=4.9169$ and $\hat{\sigma}(\hat{\lambda})=3.1682$ for ungrouped observations. The value of back-transformed $\hat{\mu}(\hat{\lambda})$ to the original scale for AIC, $\hat{\mu}^*(\hat{\lambda})=15.7587$ was closest to the corresponding one on the original scale for ungrouped observations $\hat{\mu}^*(\hat{\lambda})=16.0120$. The plot of the profiles of maximized log-likelihood to these four grouped observations as a function of λ is shown in Figure 2.2.30. Also, the related plot of profiles of value of chi-square as a function of transforming parameter λ is shown in Figure 2.2.31. These two plot show that the behavior of MSE and Sturges seem similar, whereas that of AIC and Fisher seem similar at neighborhood of $\lambda=0.5$.

The distributions of observations on the original and the power-transformed scale are shown in Figure 2.2.32 and 2.2.33 respectively. The value of 0.0756 of chi-square statistics for AIC was smallest, which was not quite significant at 5% level. The next smallest value was 6.9247 for Fisher, which was not significant at 5% level. All values of the skewness for the observations on the power-transformed scale were not farther from zero than those values for the observations on the original scale. But all values of the kurtosis for the observations on the power-transformed scale were farther from 3 than those values for the observations on the original scale. These results show that the observations on the power-transformed scale achieve only symmetry of distribution in comparison with the observations on the original scale. Thus, the values of the skewness and the kurtosis of ungrouped observations on the power-transformed scale were given as −0.1667 and 2.3160. These two values suggest the satisfaction of normality for the observations on the power-transformed scale.

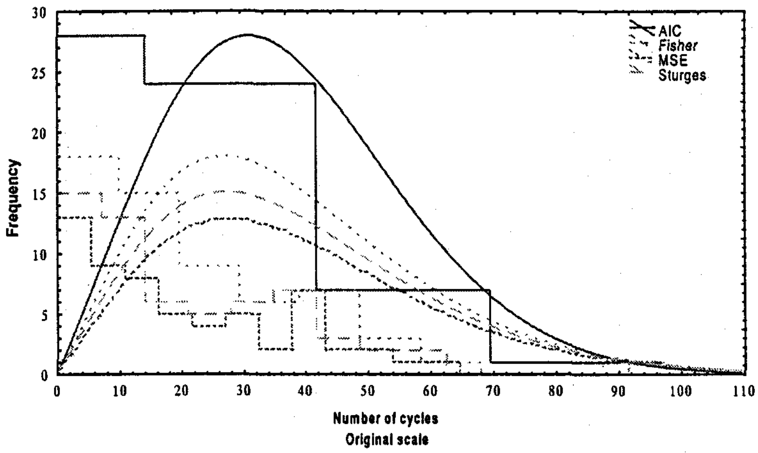

Figure 2.2.32 Example 10: The distribution of observations on the original scale

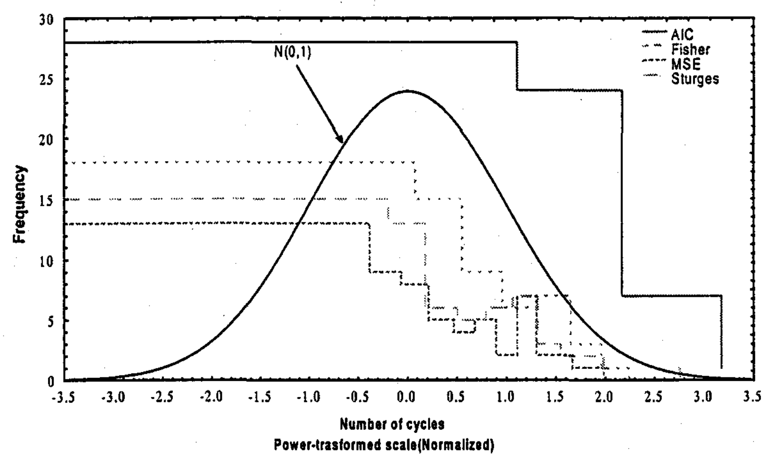

Figure 2.2.33 Example 10: The distribution of observations on the power-transformed scale

The results of fitting the PND to the above three examples suggest that it is preferable to transform the ungrouped observation because grouping results in loss of information and that the loss of information depends on the number of intervals. Only when the ungrouped observations are not available should grouped observations be transformed. Thus, we may want to transform to normality

- for better description or for model building purposes,
- to interpolate between the given observations,

- and to smooth count observations.

2.2.5 Simulation Experiment

2.2.5.1 Objective and Design

The log-likelihood function (2.2.4) are non-linear with respect to parameter λ, so the maximum likelihood estimates $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\sigma}$ of λ, μ and σ can not be presented explicitly. And the performances of the transformation can not be evaluated without observations. Therefore, the performance of fitting the PND to grouped observations will be evaluated in some simulation experiments in which main aims were

(i) to evaluate the effects of factors on estimates in fitting the PND to grouped observations,

(ii) to evaluate the difference in estimates between fitting to the PND to grouped and ungrouped observations, namely to evaluate the loss of information in subsequently grouping

The each of designs corresponding to above two aims is as follows

(i) Evaluation of effects of factors on estimates in fitting the PND to grouped observations

In order to evaluate the effects of factors on estimates $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\sigma}$ in fitting the PND to grouped observations, we employed the three indexes, namely mean square error of $\hat{\lambda}$

$$M_{\hat{\lambda}} = MSE[\hat{\lambda}] = E[\hat{\lambda} - \lambda]^2,$$

mean square error of $\hat{\mu}$

$$M_{\hat{\mu}(\hat{\lambda})} = MSE[\hat{\mu}(\hat{\lambda})] = E[\hat{\mu}(\hat{\lambda}) - \mu]^2$$

and mean square error of $\hat{\sigma}$

$$M_{\hat{\sigma}(\hat{\lambda})} = MSE[\hat{\sigma}(\hat{\lambda})] = E[\hat{\sigma}(\hat{\lambda}) - \sigma]^2.$$

Also, conditional mean square error of $\hat{\mu}(\hat{\lambda})$ and $\hat{\sigma}(\hat{\lambda})$ given λ

$$M_{\hat{\mu}(\hat{\lambda}|\lambda)} = MSE[\hat{\mu}(\hat{\lambda}|\lambda)] = E[\hat{\mu}(\hat{\lambda}) - \hat{\mu}(\lambda)]^2$$

and

$$M_{\hat{\sigma}(\hat{\lambda}|\lambda)} = MSE[\hat{\sigma}(\hat{\lambda}|\lambda)] = E[\hat{\sigma}(\hat{\lambda}) - \hat{\sigma}(\lambda)]^2$$

were used in order to evaluate the difference in $\hat{\mu}(\hat{\lambda})$ and $\hat{\mu}(\lambda)$, $\hat{\sigma}(\hat{\lambda})$ and $\hat{\sigma}(\lambda)$.

In the simulation, the PND was assumed as a distribution of the original observations, and sample seize n, shape of the PND λ^*, coefficient of variation after the transformation τ, and truncated point κ were taken up as a factor which has a effect on estimates of $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\sigma}$. Three levels of κ (κ = 1, 2 and 3) were defined corresponding to the value of $1 - A(\kappa)$ = 15.87, 2.88 and 0.14%. The value of λ were determined in 6 way corresponding to the 6 typical shaped of the PND, namely $\lambda^* > 1$, $\lambda^* = 1$ (truncated normal distribution), $\lambda_\kappa < \lambda^* < 1$, $\lambda^* = \lambda_\kappa$, $0 < \lambda^* < \lambda_\kappa$ and $\lambda < 0$ (L-shaped distribution), where $\lambda_\kappa = 4/(\kappa^2 + 4)$. And 3 values of τ (τ = 2, 4 and 16) were selected. The selected value of κ, λ^* and τ are shown in Table 2.5.1 and 2.52.

Incidentally, once the value of κ, λ^* and τ are given, μ and σ are given by

$$\mu = \frac{1}{\lambda^*(\kappa\tau - 1)}$$

and

$$\sigma = \frac{\tau}{\lambda^*(\kappa\tau - 1)}$$

respectively.

Then, for all combinations of κ, λ^*, τ, pseudorandom numbers were generated from a standard normal distribution using an acceptance/rejection technique by Kinderman and Ramage (1976). These normal random numbers were inverse-transformed by power-transformation (2.1.1), and then the 1000 sample for sample size n, which has the PND, were generated. Thus, the sample sizes were specified in 4 ways, namely n = 25, 50, 100 and 200. Furthermore, these samples were grouped by using the procedure by Sturges (1926), and then the PND were fitted to them.

Table 2.5.1 The truncation point κ and the truncated probability

Truncation point κ	Truncated probability $1 - A(\kappa)$	λ_κ	τ
1	0.1587	0.8	2
2	0.0288	0.5	4
3	0.0014	0.3077	16

Table 2.5.2 Selected truncation points κ and the transforming parameters λ^*

κ	$\lambda^* > 1$	$\lambda^* = 1$	$\lambda_\kappa < \lambda^* < 1$	$\lambda^* = \lambda_\kappa$	$0 < \lambda^* < \lambda_\kappa$	$\lambda < 0$
1	2.0	1.0	0.90	0.8	0.40	−1.0
2	2.0	1.0	0.75	0.5	0.25	−1.0
3	2.0	1.0	0.60	0.3077	0.20	−1.0

(ii) Evaluation of the loss of information in subsequently grouping

In order to evaluate the difference in estimates between fitting of the PND to grouped and ungrouped observations, namely to evaluate the loss of information in subsequently grouping, we employed the three indexes, namely

$$D_{\hat{\lambda}} = E[\hat{\lambda}_G - \hat{\lambda}_U],$$

$$D_{\hat{\mu}(\hat{\lambda})} = \mathrm{E}[\hat{\mu}_G(\hat{\lambda}_G) - \hat{\mu}_U(\hat{\lambda}_U)]^2$$

and

$$D_{\hat{\sigma}(\hat{\lambda})} = \mathrm{E}[\hat{\sigma}_G(\hat{\lambda}_G) - \hat{\sigma}_U(\hat{\lambda}_U)]^2,$$

where $\hat{\lambda}_G$, $\hat{\mu}_G(\hat{\lambda}_G)$ and $\hat{\sigma}_G(\hat{\lambda}_G)$ are estimates for grouped observations, and $\hat{\lambda}_U$, $\hat{\mu}_U(\hat{\lambda}_U)$ and $\hat{\sigma}_U(\hat{\lambda}_U)$ are for ungrouped observations.

In the simulation, the PND was assumed as a distribution of the original observations, and sample seize n, shape of the PND λ^*, coefficient of variation after the transformation τ, and truncated point κ were taken up as a factor which has a effect on estimates of $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\sigma}$ in the same way as (i). In addition to these three factors, the number of intervals k was taken up.

Then, for all combinations of κ, λ^*, τ, pseudorandom numbers were generated from a standard normal distribution using an acceptance/rejection technique by Kinderman and Ramage (1976). These normal random numbers were inverse-transformed by power-transformation (2.1.1), and then the 1000 sample for sample size n, which has the PND, were generated. Thus, the sample sizes were specified in 4 ways, namely n = 25, 50, 100 and 200. Furthermore, these samples were grouped by three level of number of interval k^* ($k^* = k/2$, k and $2k$) obtained by using Sturges' rule, and then the PND were fitted to them. Also, to evaluate the difference in estimates between fitting of the PND to grouped and ungrouped observations, the PND was fitted to ungrouped observations before grouping.

2.2.5.2 Results and Interpretations

(i) Evaluation of effects of factors on estimates in fitting the PND to grouped observations

To evaluate the effects of sample size n, shape of the PND λ^*, coefficient of variation τ, and truncation point κ on the maximum likelihood estimates $\hat{\lambda}$, $\hat{\mu}(\hat{\lambda})$ and $\hat{\sigma}(\hat{\lambda})$ of λ, μ and σ, mean square errors $\mathrm{M}_{\hat{\lambda}}$, $\mathrm{M}_{\hat{\mu}(\hat{\lambda})}$ and $\mathrm{M}_{\hat{\sigma}(\hat{\lambda})}$ of $\hat{\lambda}$, $\hat{\mu}(\hat{\lambda})$ and $\hat{\sigma}(\hat{\lambda})$ were calculated for generated 1000 samples for combinations of factors. The mean square errors $\mathrm{M}_{\hat{\lambda}}$, $\mathrm{M}_{\hat{\mu}(\hat{\lambda})}$ and

$M_{\hat{\sigma}(\hat{\lambda})}$ were then subjected to the 4-way analysis of variance (ANOVA) allowing for sample size n, shape of the PND λ^*, coefficient of variation τ, and truncation point κ. The results of ANOVA for $M_{\hat{\lambda}}$, $M_{\hat{\mu}(\hat{\lambda})}$ and $M_{\hat{\sigma}(\hat{\lambda})}$ are shown in Table 2.5.3, 2.5.4 and 2.5.5 respectively. In order to compare the results with those in ungrouped observation case, the results of ANOVA for $M_{\hat{\lambda}}$, $M_{\hat{\mu}(\hat{\lambda})}$ and $M_{\hat{\sigma}(\hat{\lambda})}$ in ungrouped observation case are shown in Table 2.5.6, 2.3.7 and 2.3.8. The both results of ANOVA were interpreted as follows by p-value associated with F-value for variation of factor and the proportion of variation explained (PVE) by factor.

Table 2.5.3 Table of ANOVA for $M_{\hat{\lambda}}$*: Grouped observation case*

Source		d.f.	M.S.	F-value	p-value	PVE
Sample size n		3	4.6833	3476.905	near 0	34.02%
Shape of the PND λ^*		5	8.7607	6503.932	near 0	63.64%
Coefficient of variation τ		2	0.0016	1.222	0.2974	0.01%
Truncation point κ		2	0.0042	3.111	0.0474	0.03%
Interaction	$n\times\lambda^*$	15	0.3054	226.739	near 0	2.22%
	$n\times\tau$	6	0.0010	0.752	0.6088	0.01%
	$\lambda^*\times\tau$	10	0.0003	0.216	0.9946	0.00%
	$n\times\kappa$	6	0.0016	1.214	0.3023	0.01%
	$\lambda^*\times\kappa$	10	0.0080	5.947	near 0	0.06%
	$\tau\times\kappa$	4	0.0003	0.230	0.9209	0.00%
Residuals		152	0.0013			
Total		215	13.768			

Table 2.5.4 Table of ANOVA for $M_{\hat{\mu}(\hat{\lambda})}$*: Grouped observation case*

Source		d.f.	M.S.	F-value	p-value	PVE
Sample size n		3	0.0865	0.499	0.6839	0.31%
Shape of the PND λ^*		5	5.4283	31.287	near 0	19.57%
Coefficient of variation τ		2	9.0261	52.024	near 0	32.54%
Truncation point κ		2	3.9051	22.508	near 0	14.08%
Interaction	$n\times\lambda^*$	15	0.0042	0.024	1.0000	0.02%
	$n\times\tau$	6	0.0070	0.040	0.9997	0.03%
	$\lambda^*\times\tau$	10	4.1509	23.925	near 0	14.96%
	$n\times\kappa$	6	0.0008	0.005	1.0000	0.00%
	$\lambda^*\times\kappa$	10	1.7355	10.003	near 0	6.26%
	$\tau\times\kappa$	4	3.3661	19.401	near 0	12.13%
Residuals		152	0.1735			
Total		215	27.8840			

For $M_{\hat{\lambda}}$ in grouped observation case, all of main effects except for τ were significant at 5 % level. The significant interactions at 5 % level were $n\times\lambda^*$ and $\lambda^*\times\kappa$. The value of 63.64% of PVD by λ^* was the largest, and the next largest PVE was the value of 34.02% by n. On the other hand, for $M_{\hat{\lambda}}$ in ungrouped observation case, as we have seen in grouped observation case, all of main effects except for τ, the two interactions $n \times \lambda^*$ and $\lambda^* \times \kappa$ were significant at 5 % level. But the largest PVE was the value of 98.62% by λ^*, and so the values of PVE's by remaining factors were very little. These results show that precison of maximum likelihood estimate of the transforming parameter λ in grouped observation case is not so much influenced by the shape of the PND λ^* as the precision of the estimate in ungrouped observation case. The estimate of main effect λ^* and the 95% confidence interval for $M_{\hat{\lambda}}$ in grouped and ungrouped observation case is shown in Figure 2.5.1. And, the estimate of main effect n and the 95% confidence interval for $M_{\hat{\lambda}}$ in grouped and ungrouped observation case is shown in Figure 2.5.2. Figure 2.5.1 shows that the values of $M_{\hat{\lambda}}$ in grouped and ungrouped case in the range of $\lambda^* < 0$ and $\lambda^* > 1$ are larger than those values in the remaining ranges, and the 95% confidence interval for $M_{\hat{\lambda}}$ in the range of $\lambda^* > 0$ is the largest. Figure 2.5.2 shows that the values of $M_{\hat{\lambda}}$ in grouped observation case decrease as n increase. But the values of $M_{\hat{\lambda}}$ in ungrouped observation case do not so much decrease with n increase as grouped observation case.

Table 2.5.5 Table of ANOVA for $M_{\hat{\sigma}(\hat{\lambda})}$ *: Grouped observation case*

Source		d.f.	M.S.	F-value	p-value	PVE
Sample size n		3	0.1218	0.103	0.9581	0.04%
Shape of the PND λ^*		5	96.4313	81.724	near 0	34.96%
Coefficient of variation τ		2	33.6486	28.517	near 0	12.20%
Truncation point κ		2	74.9241	63.497	near 0	27.16%
Interaction	$n \times \lambda^*$	15	0.0562	0.048	1.0000	0.02%
	$n \times \tau$	6	0.0168	0.014	1.0000	0.01%
	$\lambda^* \times \tau$	10	14.9725	12.689	near 0	5.43%
	$n \times \kappa$	6	0.0189	0.016	1.0000	0.01%
	$\lambda^* \times \kappa$	10	31.4560	26.659	near 0	11.40%
	$\tau \times \kappa$	4	22.7872	19.312	near 0	8.26%
Residuals		152	1.1800			
Total		215	275.6133			

For $M_{\hat{\mu}(\hat{\lambda})}$ in grouped observation case, all of main effects except for n were significant at 5 % level. The significant interactions at 5 % level were $\lambda^* \times \tau$, $\lambda^* \times \kappa$ and $\tau \times \kappa$. The value of 32.54% for PVE by τ was the largest and the next largest PVE's were the values of 19.57% by λ^*, 14.96% by $\lambda^* \times \tau$, 14.08% by κ, and 12.13% by $\tau \times \kappa$. On the other hand, for $M_{\hat{\mu}(\hat{\lambda})}$ in ungrouped observation case, all of main effects except for n, the three interactions $\lambda^* \times \tau$, $\lambda^* \times \kappa$ and $\tau \times \kappa$ were significant at 5% level. These results show that precison of maximum likelihood estimate of the mean μ in grouped observation case is as much influenced by the shape of the PND λ^* and the coefficient of variation τ as the precision of the estimate in ungrouped observation case. The estimate of interaction $\lambda^* \times \tau$ and the 95% confidence interval for $M_{\hat{\mu}(\hat{\lambda})}$ in grouped and ungrouped observation case is shown in Figure 2.5.3. Figure 2.5.3 shows that the value of $M_{\hat{\mu}(\hat{\lambda})}$ and the 95% confidence interval in grouped and ungrouped observation case in the range of $0 < \lambda^* < \lambda_\kappa$ is largest and the values of $M_{\hat{\mu}(\hat{\lambda})}$ decrease as τ increase.

Table 2.5.6 Table of ANOVA for $M_{\hat{\lambda}}$ *: Ungrouped observation case*

Source		d.f.	M.S.	F-value	p-value	PVE
Sample size n		3	0.0438	412.510	near 0	0.63%
Shape of the PND λ^*		5	6.9020	64964.630	near 0	98.62%
Coefficient of variation τ		2	0.0001	0.920	0.4020	0.00%
Truncation point κ		2	0.0267	251.460	near 0	0.38%
Interaction	$n \times \lambda^*$	15	0.0152	143.150	near 0	0.22%
	$n \times \tau$	6	0.0001	1.250	0.2841	0.00%
	$\lambda^* \times \tau$	10	0.0001	0.820	0.6095	0.00%
	$n \times \kappa$	6	0.0001	0.620	0.7144	0.00%
	$\lambda^* \times \kappa$	10	0.0106	99.370	near 0	0.15%
	$\tau \times \kappa$	4	0.0001	0.820	0.5160	0.00%
Residuals		152	0.0001			
Total		215	6.9988			

Table 2.5.7 Table of ANOVA for $M_{\hat{\mu}(\hat{\lambda})}$ *: Ungrouped observation case*

Source		d.f.	M.S.	F-value	p-value	PVE
Sample size n		3	0.0132	0.072	0.9747	0.04%
Shape of the PND λ^*		5	6.6269	36.384	near 0	19.64%
Coefficient of variation τ		2	11.8657	65.148	near 0	35.17%
Truncation point κ		2	4.8266	26.500	near 0	14.31%
Interaction	$n \times \lambda^*$	15	0.0006	0.003	1.0000	0.00%
	$n \times \tau$	6	0.0001	0.001	1.0000	0.00%
	$\lambda^* \times \tau$	10	4.6225	25.379	near 0	13.70%
	$n \times \kappa$	6	0.0000	0.000	1.0000	0.00%
	$\lambda^* \times \kappa$	10	1.8362	10.082	near 0	5.44%
	$\tau \times \kappa$	4	3.9145	21.492	near 0	11.60%
Residuals		152	0.1821			
Total		215	33.8884			

For $M_{\hat{\sigma}(\hat{\lambda})}$ in grouped observation case, all of main effects except for n were significant at 5 % level, and the significant interactions at 5 % were $\lambda^* \times \tau$, $\lambda^* \times \kappa$ and $\tau \times \kappa$. On the other hand, for $M_{\hat{\sigma}(\hat{\lambda})}$ in ungrouped observation case, all of main effects except for n, the three interactions $\lambda^* \times \tau$, $\lambda^* \times \kappa$ and $\tau \times \kappa$ were significant at 5% level. The value of 34.96% for PVE by λ^* was the largest and the next largest PVE's were the values of 27.16% by κ, 12.20% by τ, and 11.04% by $\lambda^* \times \kappa$. These results show that precison of maximum likelihood estimate of the mean μ in grouped observation case is as much influenced by the shape of the PND λ^* and the truncation point k as the precision of the estimate in ungrouped observation case. The estimate of interaction $\lambda^* \times \kappa$ and the 95%

confidence interval for $M_{\hat{\sigma}(\hat{\lambda})}$ is shown in Figure 2.5.4. Figure 2.5.4 shows that the value of $M_{\hat{\sigma}(\hat{\lambda})}$ and the 95% confidence interval in grouped and ungrouped observation case in the range of $0 < \lambda^* < \lambda_\kappa$ is largest and the values of $M_{\hat{\sigma}(\hat{\lambda})}$ decrease as κ increase.

Table 2.5.8 Table of ANOVA for $M_{\hat{\sigma}(\hat{\lambda})}$ *: Ungrouped observation case*

Source	d.f.	M.S.	F-value	p-value	PVE
Sample size n	3	0.0592	0.103	0.9581	0.04%
Shape of the PND λ^*	5	70.7229	81.724	near 0	34.96%
Coefficient of variation τ	2	33.6919	28.517	near 0	12.20%
Truncation point κ	2	80.6211	63.497	near 0	27.16%
Interaction $n\times\lambda^*$	15	0.0086	0.048	1.0000	0.02%
$n\times\tau$	6	0.0011	0.014	1.0000	0.01%
$\lambda^*\times\tau$	10	12.6688	12.689	near 0	5.43%
$n\times\kappa$	6	0.0025	0.016	1.0000	0.01%
$\lambda^*\times\kappa$	10	27.2458	26.659	near 0	11.40%
$\tau\times\kappa$	4	24.1596	19.312	near 0	8.26%
Residuals	152	1.0409			
Total	215	250.2222			

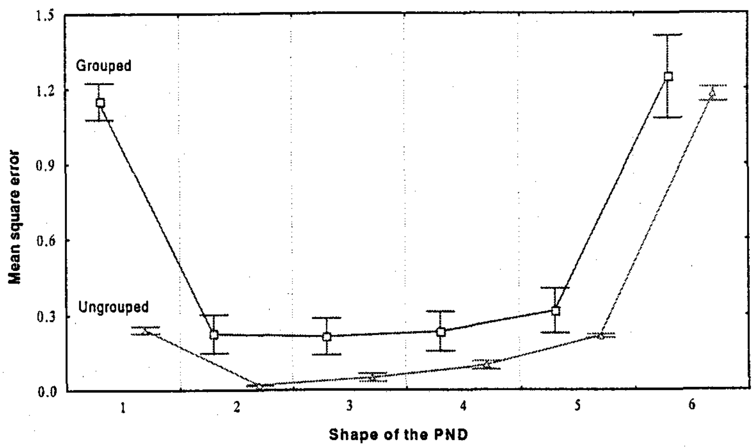

Figure 2.5.1 The estimate of main effect λ^* *and the 95% confidence interval for* $M_{\hat{\lambda}}$
1: $\lambda^* < 0$, *2:* $0 < \lambda^* < \lambda_\kappa$, *3:* $\lambda^* = \lambda_\kappa$, *4:* $\lambda_\kappa < \lambda^* < 1$, *5:* $\lambda^* = 1$, *6:* $\lambda^* > 1$

Next, to evaluate the effects of sample size n, shape of the PND λ^*, coefficient of variation τ, and truncation point κ on the difference in $\hat{\mu}(\hat{\lambda})$ and $\hat{\mu}(\lambda)$, $\hat{\sigma}(\hat{\lambda})$ and $\hat{\sigma}(\lambda)$, conditional mean square errors $M_{\hat{\mu}(\hat{\lambda}|\lambda)}$ and $M_{\hat{\sigma}(\hat{\lambda}|\lambda)}$ of $\hat{\mu}(\hat{\lambda})$ and $\hat{\sigma}(\hat{\lambda})$ given λ were calculated for generated 1000 samples for combinations of factors. The conditional mean square error $M_{\hat{\mu}(\hat{\lambda}|\lambda)}$ and $M_{\hat{\sigma}(\hat{\lambda}|\lambda)}$ were then subjected to the 4-way ANOVA allowing for sample size n, shape of the PND λ^*, coefficient of variation τ, and truncation point κ. The results of ANOVA for grouped observation case are shown in Table 2.5.9 and 2.5.10. The results of ANOVA were interpreted as follows by p-value associated with F-value for variation of factor and the PVE by factor.

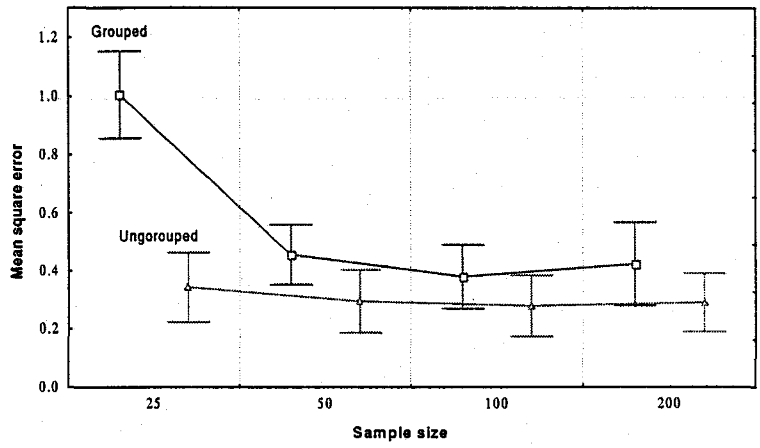

Figure 2.5.2 The estimate of main effect n *and the 95% confidence interval for* $M_{\hat{\lambda}}$

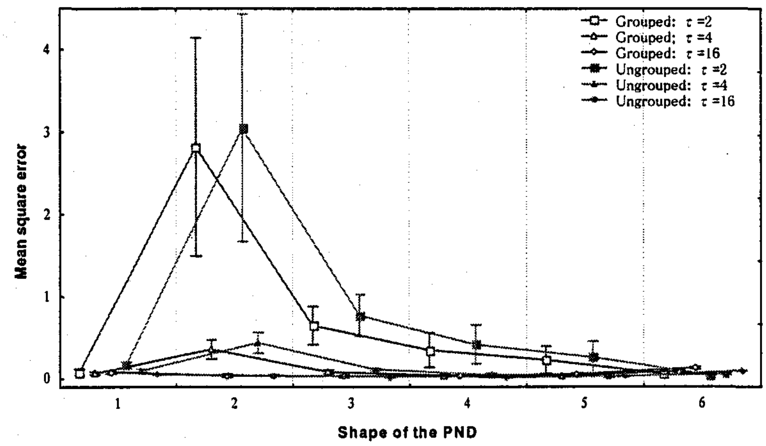

Figure 2.5.3 The estimate of interaction $\lambda^* \times \tau$ *and the 95% confidence interval for* $M_{\hat{\mu}(\hat{\lambda})}$

1: $\lambda^* < 0$, *2:* $0 < \lambda^* < \lambda_K$, *3:* $\lambda^* = \lambda_K$, *4:* $\lambda_K < \lambda^* < 1$, *5:* $\lambda^* = 1$, *6:* $\lambda^* > 1$

For $M_{\hat{\mu}(\hat{\lambda}|\lambda)}$, all of main effects except for λ^* were significant at 5 per cent level. The significant interactions at 5 % level were $n \times \tau$, $n \times \kappa$ and $\tau \times \kappa$. The value of 87.07% for PVE by n was the largest, and the values of PVE's by the remaining factor were very little. The estimate of main effect n and the 95% confidence interval for $M_{\hat{\mu}(\hat{\lambda}|\lambda)}$ is shown in Figure 2.5.5. Figure 2.5.5 shows that the values of $M_{\hat{\mu}(\hat{\lambda}|\lambda)}$ decrease as n increase. Namely, this means that the difference between $\hat{\mu}(\hat{\lambda})$ and $\hat{\mu}(\lambda)$ decrease as n increase.

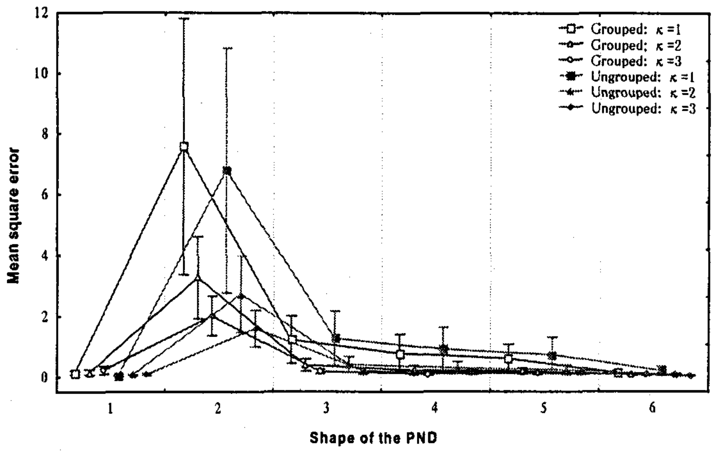

Figure 2.5.4 The estimate of interaction $\lambda^* \times \kappa$ *and the 95% confidence interval for* $M_{\hat{\sigma}(\hat{\lambda})}$
1: $\lambda^* < 0$, *2:* $0 < \lambda^* < \lambda_\kappa$, *3:* $\lambda^* = \lambda_\kappa$, *4:* $\lambda_\kappa < \lambda^* < 1$, *5:* $\lambda^* = 1$, *6:* $\lambda^* > 1$

Table 2.5.9 Table of ANOVA for $M_{\hat{\mu}(\hat{\lambda}|\lambda)}$

Source		d.f.	M.S.	F-value	p-value	PVE
Sample size n		3	0.0352	444.035	near 0	87.07%
Shape of the PND λ^*		5	0.0001	0.791	0.5580	0.16%
Coefficient of variation τ		2	0.0010	13.040	near 0	2.56%
Truncation point κ		2	0.0013	16.733	near 0	3.28%
Interaction	$n \times \lambda^*$	15	0.0000	0.486	0.9450	0.09%
	$n \times \tau$	6	0.0006	7.313	near 0	1.43%
	$\lambda^* \times \tau$	10	0.0001	0.805	0.6244	0.16%
	$n \times \kappa$	6	0.0009	11.956	near 0	2.34%
	$\lambda^* \times \kappa$	10	0.0001	0.791	0.6379	0.16%
	$\tau \times \kappa$	4	0.0010	13.040	near 0	2.56%
Residuals		152	0.0001			
Total		215	0.0404			

Table 2.5.10 Table of ANOVA for $M_{\hat{\sigma}(\hat{\lambda}|\lambda)}$

Source		d.f.	M.S.	F-value	p-value	PVE
Sample size n		3	0.0267	507.236	near 0	82.13%
Shape of the PND λ^*		5	0.0000	0.792	0.5568	0.13%
Coefficient of variation τ		2	0.0013	23.878	near 0	3.87%
Truncation point κ		2	0.0023	43.626	near 0	7.06%
Interaction	$n \times \lambda^*$	15	0.0000	0.342	0.9898	0.06%
	$n \times \tau$	6	0.0004	7.690	near 0	1.24%
	$\lambda^* \times \tau$	10	0.0000	0.574	0.8333	0.09%
	$n \times \kappa$	6	0.0004	7.749	near 0	1.25%
	$\lambda^* \times \kappa$	10	0.0000	0.792	0.6362	0.13%
	$\tau \times \kappa$	4	0.0013	23.878	near 0	3.87%
Residuals		152	0.0001			
Total		215	0.0325			

For $\mathrm{M}_{\hat{\sigma}(\hat{\lambda}|\lambda)}$, all of main effects except for λ^* were significant at 5 % level. The significant interactions at 5 % level were $n\times\tau$, $n\times\kappa$ and $\tau\times\kappa$. The value of 82.13% for PVE by n was the largest, and the PVE's by the remaining factors were very little. The estimate of main effect n and the 95% confidence interval for $\mathrm{M}_{\hat{\sigma}(\hat{\lambda}|\lambda)}$ is shown in Figure 2.5.6. Figure 2.5.6 shows that the values of $\mathrm{M}_{\hat{\sigma}(\hat{\lambda}|\lambda)}$ decrease as n increase. Namely, this means that the difference between $\hat{\sigma}(\hat{\lambda})$ and $\hat{\sigma}(\lambda)$ decrease as n increase.

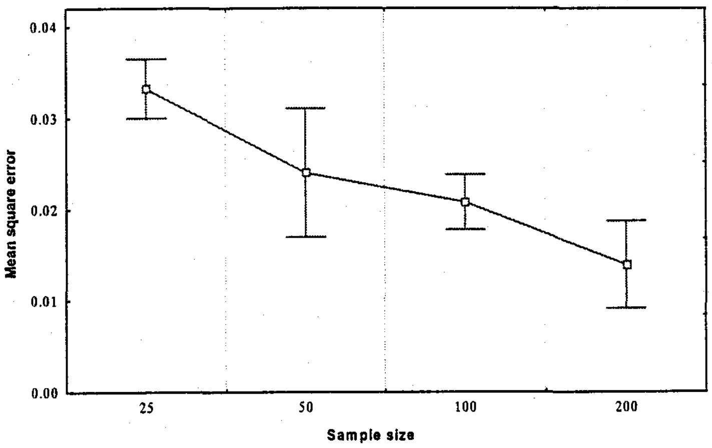

Figure 2.5.5 The estimate of main effect n and the 95% confidence interval for $\mathrm{M}_{\hat{\mu}(\hat{\lambda}|\lambda)}$

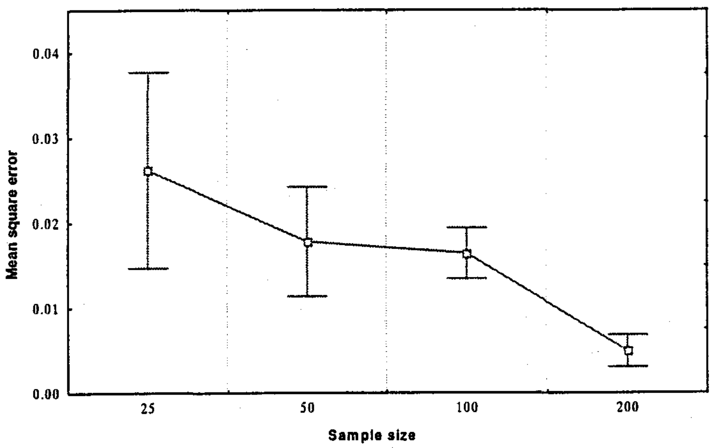

Figure 2.5.6 The estimate of main effect n and the 95% confidence interval for $\mathrm{M}_{\hat{\sigma}(\hat{\lambda}|\lambda)}$

(ii) Evaluation of the loss of information in subsequently grouping

To evaluate the effects of the number of intervals k, sample size n, shape of the PND λ^*, coefficient of variation τ, and truncation point κ on the difference in estimates between fitting to the PND to grouped and ungrouped observations, the mean square differences $D_{\hat{\lambda}}$, $D_{\hat{\mu}(\hat{\lambda})}$ and $D_{\hat{\sigma}(\hat{\lambda})}$ were calculated for generated 1000 samples for combinations of factors. $D_{\hat{\lambda}}$, $D_{\hat{\mu}(\hat{\lambda})}$ and $D_{\hat{\sigma}(\hat{\lambda})}$ were subjected to the 5-way ANOVA allowing for the number of intervals k, sample size n, shape of the PND λ^*, coefficient of variation τ, and truncation point κ. The results of ANOVA for $D_{\hat{\lambda}}$, $D_{\hat{\mu}(\hat{\lambda})}$ and $D_{\hat{\sigma}(\hat{\lambda})}$ are shown in Table 2.5.11, 2.5.12 and 2.5.13. The results of ANOVA were interpreted as follows by p-value associated with F-value for variation of factor and the PVE by factor.

For $D_{\hat{\mu}(\hat{\lambda})}$, k^*, n and λ^* were significant at 5 % level. The significant interactions at 5 % were $k^* \times n$, $k^* \times \lambda^*$, $n \times \lambda^*$, $k^* \times \kappa$, $n \times \kappa$ and $\lambda^* \times \kappa$. The value of 47.66% for PVE by k^* was the largest, and the next largest PVE's were the values of 22.55% by sample size n, 8.86% by $k^* \times n$, 7.64% by $n \times \lambda^*$ and 5.62% by $k^* \times \lambda^*$. The estimate of interaction $k^* \times n$ and the 95% confidence interval for $D_{\hat{\mu}(\hat{\lambda})}$ is shown in Figure 2.5.8. Figure 2.5.8 shows that the values of $D_{\hat{\mu}(\hat{\lambda})}$ decrease as n and k^* both increase. Namely, this means that the difference between $\hat{\mu}_G(\hat{\lambda}_G)$ and $\hat{\mu}_U(\hat{\lambda}_U)$ decrease as k^* and n both increase.

Table 2.5.11 Table of ANOVA for $D_{\hat{\lambda}}$

Source		d.f.	M.S.	F-value	p-value	PVE
The number of intervals k^*		2	732.3102	4946.349	near 0	59.44%
Sample size n		3	243.2190	1642.809	near 0	19.74%
Shape of the PND λ^*		5	35.5458	240.092	near 0	2.89%
Coefficient of variation τ		2	0.0165	0.111	0.8948	0.00%
Truncation point κ		2	0.0355	0.240	0.7869	0.00%
Interaction	$k^* \times n$	6	208.8811	1410.876	near 0	16.95%
	$k^* \times \lambda^*$	10	8.5664	57.861	near 0	0.70%
	$n \times \lambda^*$	15	2.6091	17.623	near 0	0.21%
	$k \times \tau$	4	0.0062	0.042	0.9966	0.00%
	$n \times \tau$	6	0.0075	0.051	0.9995	0.00%
	$\lambda^* \times \tau$	10	0.0072	0.049	1.0000	0.00%
	$k \times \kappa$	4	0.0198	0.134	0.9699	0.00%
	$n \times \kappa$	6	0.2796	1.889	0.0807	0.02%
	$\lambda^* \times \kappa$	10	0.3627	2.450	0.0072	0.03%
	$\tau \times \kappa$	4	0.0098	0.066	0.9920	0.00%
Residuals		558	0.1481			
Total		647	1232.024			

For $D_{\hat{\lambda}}$, all of main effects except for τ were significant at 5 % level. The significant interactions at 5 % level were $k^* \times n$, $k^* \times \lambda^*$, $n \times \lambda^*$ and $\lambda^* \times \kappa$. The value of 59.44% for PVD by k^* was the largest, and the next largest PVE's were the valued of 34.02% by sample size n and 16.95% by $k^* \times n$. The estimate of interaction $k^* \times n$ and the 95% confidence interval for $D_{\hat{\lambda}}$ is shown in Figure 2.5.7. Figure 2.5.7 shows that the values of $D_{\hat{\lambda}}$ and the 95% confidence interval decrease as

k^* and n both increase. Namely, this means that the difference between $\hat{\lambda}_G$ and $\hat{\lambda}_U$ decrease as k^* and n both increase.

Table 2.5.12 Table of ANOVA for $D_{\hat{\mu}(\hat{\lambda})}$

Source	d.f.	M.S.	F-value	p-value	PVE
The number of intervals k^*	2	0.9276	262.266	near 0	47.66%
Sample size n	3	0.1133	32.022	near 0	5.82%
Shape of the PND λ^*	5	0.4389	124.096	near 0	22.55%
Coefficient of variation τ	2	0.0005	0.127	0.8804	0.02%
Truncation point κ	2	0.0072	2.048	0.1300	0.37%
Interaction $k^* \times n$	6	0.1725	48.761	near 0	8.86%
$k^* \times \lambda^*$	10	0.1093	30.899	near 0	5.62%
$n \times \lambda^*$	15	0.1488	42.055	near 0	7.64%
$k \times \tau$	4	0.0002	0.062	0.9929	0.01%
$n \times \tau$	6	0.0003	0.089	0.9974	0.02%
$\lambda^* \times \tau$	10	0.0003	0.077	0.9999	0.01%
$k \times \kappa$	4	0.0099	2.792	0.0257	0.51%
$n \times \kappa$	6	0.0085	2.413	0.0260	0.44%
$\lambda^* \times \kappa$	10	0.0051	1.455	0.1530	0.26%
$\tau \times \kappa$	4	0.0005	0.132	0.9707	0.02%
Residuals	558	0.0035			
Total	647	1.9464			

Table 2.5.13 Table of ANOVA for $D_{\hat{\sigma}(\hat{\lambda})}$

Source	d.f.	M.S.	F-value	p-value	PVE
The number of intervals k^*	2	35.2033	134.156	near 0	39.03%
Sample size n	3	14.7641	56.264	near 0	16.37%
Shape of the PND λ^*	5	9.1583	34.901	near 0	10.15%
Coefficient of variation τ	2	0.0088	0.034	0.9668	0.01%
Truncation point κ	2	1.4264	5.436	0.0046	1.58%
Interaction $k^* \times n$	6	15.7201	59.907	near 0	17.43%
$k^* \times \lambda^*$	10	7.7863	29.673	near 0	8.63%
$n \times \lambda^*$	15	3.2564	12.410	near 0	3.61%
$k \times \tau$	4	0.0059	0.022	0.9990	0.01%
$n \times \tau$	6	0.0044	0.017	1.0000	0.00%
$\lambda^* \times \tau$	10	0.0086	0.033	1.0000	0.01%
$k \times \kappa$	4	1.3709	5.224	0.0004	1.52%
$n \times \kappa$	6	0.6919	2.637	0.0157	0.77%
$\lambda^* \times \kappa$	10	0.5078	1.935	0.0383	0.56%
$\tau \times \kappa$	4	0.0142	0.054	0.9945	0.02%
Residuals	558	0.2624			
Total	647	90.1898			

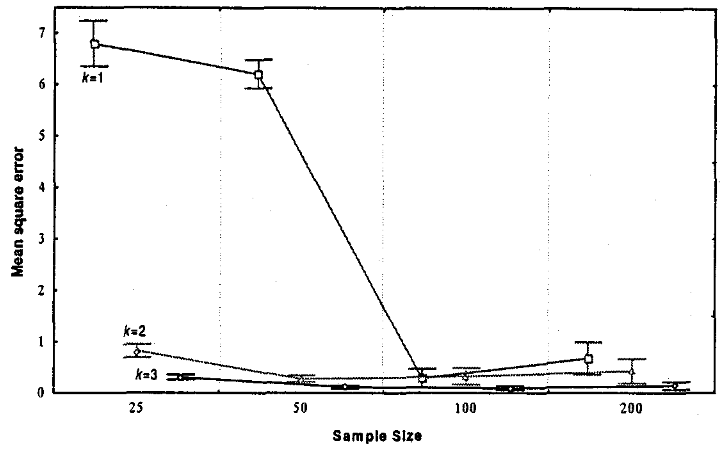

Figure 2.5.7 The estimate of interaction $k^* \times n$ *and the 95% confidence interval for* $D_{\hat{\lambda}}$

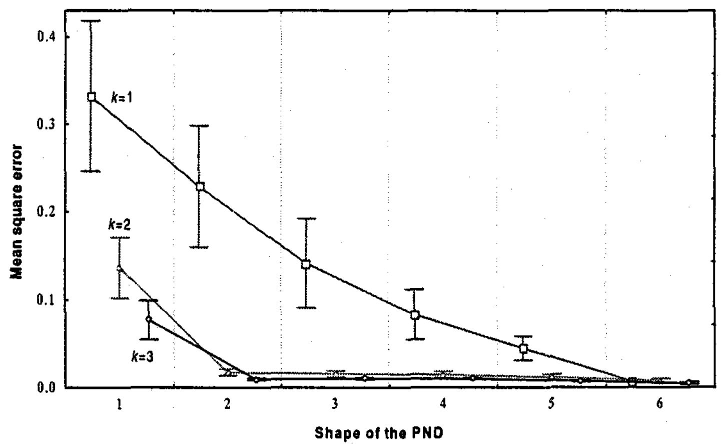

Figure 2.5.8 The estimate of interaction $k^* \times n$ *and the 95% confidence interval for* $D_{\hat{\mu}(\hat{\lambda})}$

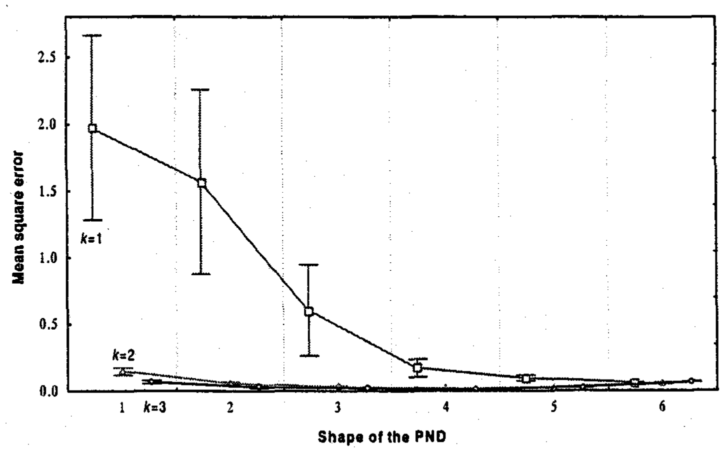

Figure 2.5.9 The estimate of interaction $k^* \times n$ *and the 95% confidence interval for* $D_{\hat{\sigma}(\hat{\lambda})}$

For $D_{\hat{\sigma}(\hat{\lambda})}$, all of main effects except for τ were significant at 5 % level. The significant interactions at 5 % level were $k \times n$, $k \times \lambda^*$, $n \times \lambda^*$, $k \times \kappa$, $n \times \kappa$, and $\lambda^* \times \kappa$. The value of 39.03% of PVD by k^* was the largest, and the next largest PVE's were the values of 16.37% by sample size n, 10.15% by λ^*, 17.43% by $k^* \times n$ and 8.63% by $k \times \lambda^*$. The estimate of interaction $k^* \times n$ and the 95% confidence interval for $D_{\hat{\sigma}(\hat{\lambda})}$ is shown in Figure 2.5.9. Figure 2.5.9 shows that the values of $D_{\hat{\sigma}(\hat{\lambda})}$ decrease as k^* and n both increase. Namely, this means that the difference between $\hat{\sigma}_G(\hat{\lambda}_G)$ and $\hat{\sigma}_U(\hat{\lambda}_U)$ decrease as k^* and n both increase.

2.3 Grouped Observations Combined with Ungrouped Observations

We now assume that it is possible to obtain some exactly specified values besides the counts of observations in other specified intervals. The following cases of mixed (grouped and ungrouped) observations are considered.

- Ungrouped observations available in the tail
- Right and/or left censored observations
- Individual observations arbitrarily chosen
- Ungrouped observations at fixed proportions of the observation

We now proceed to study each one of these cases separately.

2.3.1 Ungrouped Observations Available in the Tail

Here, we should considered a censoring scheme where exact values are recorded only for observations either in the lower or the upper tail, or both extremes. The rest of the observations are grouped. More precisely, the values of all observations below the (specified) fixed value y_1 or above the fixed value y_{k-1} are recorded exactly, while those in between y_1 and y_{k-1} are grouped into the intervals $[y_1, y_2), \cdots, [y_{k-2}, y_{k-1})$, where $0 < y_1 < \cdots < y_{k-1} < \infty$. This is a form of fixed time censoring which produces mixed observations of exact and grouped.

The motivation behind this formulation is that it is good statistical practice to report extreme values exactly for further scrutiny. This may be done either to check for outliers or to delineate unique features of the phenomenon under study, particularly those related to the tail behavior of the distributions.

Pretending that Y_j have the power-normal distribution, namely $\mathrm{PND}(\lambda_0, \mu_0, \sigma_0^2)$ for some $\boldsymbol{\theta}_0^{\mathrm{T}} = (\lambda_0, \mu_0, \sigma_0)$, the log-likelihood $l_n(\boldsymbol{\theta}|y)$ of the mixed observations for the sample of size $n = \sum_{i=1}^{k} n_i$, becomes

$$l_n(\boldsymbol{\theta}|y) = \log n! - \sum_{i=1}^{k} \log n_i! + \sum_{i=2}^{k-1} n_i \log p_{\mathrm{PN}i}(\boldsymbol{\theta})$$
$$-\frac{1}{2}(n_1 + n_k)\log 2\pi - \frac{1}{2}(n_1 + n_k)\log\sigma^2 - \frac{1}{2}\sum_{j}^{n_1,n_k}\left(\frac{y_j^{(\lambda)} - \mu}{\sigma}\right)^2 \tag{2.3.1}$$
$$+ (\lambda - 1)\sum_{j}^{n_1,n_k}\log y_j - (n_1 + n_k)A(\kappa)$$

where $p_{\mathrm{PN}i}(\boldsymbol{\theta})$ is defined in (2.2.2) and $\sum_{j}^{n_1,n_k}$ stands for the sum over the smallest n_1 and the largest n_k order statistics. Notice that (2.3.1) can also be written as

$$l_n(\boldsymbol{\theta}|y) = l_{n-n_1-n_k}^{(g)}(\boldsymbol{\theta}) + l_{n_1+n_k}^{(u)}(\boldsymbol{\theta}|y) \tag{2.3.2}$$

where

$$l_{n-n_1-n_k}^{(g)}(\boldsymbol{\theta}) = \log n! - \sum_{i=1}^{k} \log n_i! + \sum_{i=2}^{k-1} n_i \log p_{\mathrm{PN}i}(\boldsymbol{\theta}) \tag{2.3.3}$$

is the part of the log-likelihood corresponding to the grouped observations, and

$$l_{n_1+n_k}^{(u)}(\boldsymbol{\theta}|y) = -\frac{1}{2}(n_1 + n_k)\log 2\pi - \frac{1}{2}(n_1 + n_k)\log\sigma^2 - \frac{1}{2}\sum_{j}^{n_1,n_k}\left(\frac{y_j^{(\lambda)} - \mu}{\sigma}\right)^2$$
$$+ (\lambda - 1)\sum_{j}^{n_1,n_k}\log y_j - (n_1 + n_k)A(\kappa) \tag{2.3.4}$$

is the part corresponding to the ungrouped observations.

Now, we consider the properties of estimates of λ, μ and σ based on the behavior of $n^{-1}l_n(\boldsymbol{\theta}|y)$.

Let p_i and $p_{\mathrm{PN}i}(\boldsymbol{\theta})$ be defined as in Section 2.2.2, and suppose that (i) the parameter space Θ is the compact subset of $\Re^3$ defined by $\Theta = \{\boldsymbol{\theta}^{\mathrm{T}} = (\lambda, \mu, \sigma) \| |\mu| \le M, s_1 \le \sigma \le s_2, a \le \lambda \le b$ for some $0 < M, s_1, s_2, b < \infty$ and $-\infty < a < 0\}$, (ii) the moments $\mathrm{E}_{g*}[Y^{2a}]$ and $\mathrm{E}_{g*}[Y^{2b}]$ are finite, and (iii) $H(\boldsymbol{\theta}) + \mathrm{E}_{g*}[D_{I_1 \cup I_k}(y) l_1^{(u)}(\boldsymbol{\theta}|y)]$ has a unique global maximum at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, where $H(\boldsymbol{\theta}) = \sum_{i=2}^{k-1} p_i \log(p_{\mathrm{PND}i}(\boldsymbol{\theta})/p_i)$. Furthermore, suppose that (iv) $\boldsymbol{\theta}_0$ be an interior point of Θ, and furthermore suppose that (v) $\mathrm{E}_{g*}[Y^a \log Y]^2$ and $\mathrm{E}_{g*}[Y^b \log Y]^2$ both are finite, (vi)

$$\nabla H(\boldsymbol{\theta}_0) + \mathrm{E}_{g*}[D_{I_1 \cup I_k}(y) \nabla l_1^{(u)}(\boldsymbol{\theta}_0|y)] = \mathbf{0}$$

and (vii) $V = \{\nabla^2 H(\boldsymbol{\theta}_0) + \mathrm{E}_{g^*}[D_{I_1 \cup I_k}(y)\nabla^2 l_1^{(u)}(\boldsymbol{\theta}_0|y)]\}^{-1}$ exists.

By strong low of large number, $n_i/n \xrightarrow{a.s.} p_i$ as $n \to \infty$, and Stirling's formula and (2.3.3) give

$$l_{n-n_1-n_k}^{(g)}(\boldsymbol{\theta}) = \frac{1-k}{2}\log 2\pi - \sum_{i=1}^{k} n_i \log\frac{n_i}{n} + \frac{1}{2}\left[(1-k)\log n - \sum_{i=1}^{k}\log\frac{n_i}{n}\right] + \sum_{i=2}^{k-1} n_i \log p_{\mathrm{PN}i}(\boldsymbol{\theta}) + O(n^{-1}) \tag{2.3.5}$$

where $O(n^{-1})$ is uniform in $\boldsymbol{\theta}$ for an almost sure set. If $p_i = 0$ for some i, then $n_i = 0$ and the term involving n_i is excluded from the log-likelihood. Thus, writing $\hat{p}_{i,n} = n_i/n\,(i = 1,\cdots,k)$, we have with probability one, as $n \to \infty$

$$\frac{1}{n} l_{n-n_1-n_k}^{(g)}(\boldsymbol{\theta}) = \sum_{i=2}^{k-1} \hat{p}_{i,n} \log p_{\mathrm{PN}i}(\boldsymbol{\theta}) - \sum_{i=1}^{k} \hat{p}_{i,n} \log \hat{p}_{i,n} + o(1). \tag{2.3.6}$$

A procedure similar to that of Section 2.2.2 establishes that as $n \to \infty$, uniformly on Θ

$$\left|\sum_{i=2}^{k-1} \hat{p}_{i,n} \log p_{\mathrm{PN}i}(\boldsymbol{\theta}) - \sum_{i=2}^{k-1} p_i \log p_{\mathrm{PN}i}(\boldsymbol{\theta})\right| \xrightarrow{a.s.} 0. \tag{2.3.7}$$

It also follows from the continuity of $x\log x$ that as $n \to \infty$

$$\left|\sum_{i=1}^{k} \hat{p}_{i,n} \log \hat{p}_{i,n} - \sum_{i=1}^{k} p_i \log p_i\right| \xrightarrow{a.s.} 0.$$

This last results in conjunction with (2.3.6) and (2.3.7) yields almost surely

$$\lim_{n\to\infty} l_{n-n_1-n_k}^{(g)}(\boldsymbol{\theta}) = \sum_{i=2}^{k-1} p_i \log\frac{p_{\mathrm{PN}i}(\boldsymbol{\theta})}{p_i} + (p_1 \log p_1 + p_k \log p_k) \tag{2.3.8}$$

uniformly in $\boldsymbol{\theta} \in \Theta$.

Next we consider the approximation

$$l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y) = D_{(0,y_1-\varepsilon)}(y) l_1^{(u)}(\boldsymbol{\theta}|y) + D_{[y_1-\varepsilon,y_1)}\left(\frac{y_1 - y}{\varepsilon}\right) l_1^{(u)}(\boldsymbol{\theta}|y_1 - \varepsilon) + D_{[y_{k-1},y_{k-1}+\varepsilon)}\left(\frac{y - y_1}{\varepsilon}\right) l_1^{(u)}(\boldsymbol{\theta}|y_{k-1} + \varepsilon) + D_{[y_{k-1}+\varepsilon,\infty)}(y) l_1^{(u)}(\boldsymbol{\theta}|y_{k-1} + \varepsilon) \tag{2.3.9}$$

to $I_{D_1 \cup D_k}(y) l_1^{(u)}(\boldsymbol{\theta}|y)$ which is continuous in Y for each $\varepsilon > 0$. It follows from Appendix 7 that as $n \to \infty$, uniformly on Θ

$$\frac{1}{n}\sum_{j=1}^{n} l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y_j) \xrightarrow{a.s.} \mathrm{E}_{g^*}[l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y)] \tag{2.3.10}$$

with $\mathrm{E}_{g^*}[l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y)]$ is continuous.

Here, let us consider $l_{n_1+n_k}^{(u)}(\boldsymbol{\theta}|y)$ given by (2.3.4) which we observe can be expressed as

$$\sum_{j=1}^{n} D_{I_1 \cup I_k}(y_j) l_1^{(u)}(\boldsymbol{\theta}|y_j).$$

By analogy of the above approximation, we define

$$l_{n_1+n_k,\varepsilon}^{(u)}(\boldsymbol{\theta}|y) = \sum_{j=1}^{n} l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y_j). \tag{2.3.11}$$

Also, to simplify the notation, let $I_{1,\varepsilon} = [y_1 - \varepsilon, y_1)$ and $I_{k,\varepsilon} = [y_{k-1}, y_{k-1} + \varepsilon)$. Then

$$\frac{1}{n} l_{n_1+n_k}^{(u)}(\boldsymbol{\theta}|y) - \frac{1}{n} l_{n_1+n_k,\varepsilon}^{(u)}(\boldsymbol{\theta}|y) = \frac{1}{n}\sum_{j=1}^{n}\left\{D_{I_{1,\varepsilon}}(y_j)\left[l_1^{(u)}(\boldsymbol{\theta}|y_j) - \left(\frac{y_1 - y_j}{\varepsilon}\right)l_1^{(u)}(\boldsymbol{\theta}|y_1 - \varepsilon)\right]\right.$$
$$\left. + D_{I_{k,\varepsilon}}(y_j)\left[l_1^{(u)}(\boldsymbol{\theta}|y_j) - \left(\frac{y_j - y_{k-1}}{\varepsilon}\right)l_1^{(u)}(\boldsymbol{\theta}|y_{k-1} + \varepsilon)\right]\right\} \tag{2.3.12}$$

so that

$$\left|\frac{1}{n} l_{n_1+n_k}^{(u)}(\boldsymbol{\theta}|y) - \frac{1}{n} l_{n_1+n_k,\varepsilon}^{(u)}(\boldsymbol{\theta}|y)\right| \le \frac{1}{n}\sum_{j=1}^{n} D_{I_{1,\varepsilon}\cup I_{k,\varepsilon}}(y_j)\left|l_1^{(u)}(\boldsymbol{\theta}|y)\right|$$
$$\le \frac{1}{n}\sum_{j=1}^{n} D_{I_{1,\varepsilon}\cup I_{k,\varepsilon}}(y_j) r(y_j) \tag{2.3.13}$$

where $r(\cdot)$ is defined in Appendix 7. Thus, applying the strong low of large number, for $\varepsilon < \delta(\eta)$

$$\frac{1}{n}\sum_{j=1}^{n} D_{I_{1,\varepsilon}\cup I_{k,\varepsilon}}(y_j) r(y_j) \xrightarrow{a.s.} \mathrm{E}_{g*}[D_{I_{1,\varepsilon}\cup I_{k,\varepsilon}}(y) r(y)] < \frac{\eta}{2} \tag{2.3.14}$$

say for ε sufficiently small, since $r(\cdot)$ is intergerable. We also have

$$\left|\mathrm{E}_{g*}[l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y)] - \mathrm{E}_{g*}[D_{I_1\cup I_k}(y) l_1^{(u)}(\boldsymbol{\theta}|y)]\right| \le \mathrm{E}_{g*}\left|l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y) - D_{I_1\cup I_k}(y) l_1^{(u)}(\boldsymbol{\theta}|y)\right|$$
$$\le \mathrm{E}_{g*}\left|D_{I_{1,\varepsilon}\cup I_{k,\varepsilon}}(y) r(y)\right| < \frac{n}{2}. \tag{2.3.15}$$

Adding and subtracting $n^{-1} l_{n_1+n_k,\varepsilon}^{(u)}(\boldsymbol{\theta}|y) - \mathrm{E}_{g*}[l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y)]$ to $n^{-1} l_{n_1+n_k,\varepsilon}^{(u)}(\boldsymbol{\theta}|y) - \mathrm{E}_{g*}[D_{I_1\cup I_k}(y) l_1^{(u)}(\boldsymbol{\theta}|y)]$, we get

$$\left|\frac{1}{n} l_{n_1+n_k}^{(u)}(\boldsymbol{\theta}|y) - \mathrm{E}_{g*}[D_{I_1\cup I_k}(y) l_1^{(u)}(\boldsymbol{\theta}|y)]\right| \le \left|\frac{1}{n} l_{n_1+n_k}^{(u)}(\boldsymbol{\theta}|y) - \frac{1}{n} l_{n_1+n_k,\varepsilon}^{(u)}(\boldsymbol{\theta}|y)\right|$$
$$+ \left|\frac{1}{n} l_{n_1+n_k,\varepsilon}^{(u)}(\boldsymbol{\theta}|y) - \mathrm{E}_{g*}[l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y)]\right| \tag{2.3.16}$$
$$+ \left|\mathrm{E}_{g*}[l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y)] - \mathrm{E}_{g*}[D_{I_1\cup I_k}(y) l_1^{(u)}(\boldsymbol{\theta}|y)]\right|.$$

Results (2.2.10), (2.2.13) and (2.2.15) applied to the right hand side of (2.2.16) yield the almost sure limit

$$\lim_{n\to\infty}\left|\frac{1}{n}l^{(u)}_{n_1+n_k}(\boldsymbol{\theta}|y)-\mathrm{E}_{g^*}[D_{I_1\cup I_k}(Y)l^{(u)}_1(\boldsymbol{\theta}|y)]\right|$$

$$\leq \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}I_{I_{1,\varepsilon}\cup I_{k,\varepsilon}}(y_j)r(y_j)+\mathrm{E}_{g^*}[D_{I_{1,\varepsilon}\cup I_{k,\varepsilon}}(y)r(y)]+0$$

$$<\frac{\eta}{2}+\frac{\eta}{2}$$

uniformly in $\boldsymbol{\theta}\in\Theta$, for given $\eta>0$, whenever $0<\varepsilon<\delta(\eta)$. Since η is arbitrary, it follows that, with probability one, uniformly in $\boldsymbol{\theta}\in\Theta$

$$\lim_{n\to\infty}\frac{1}{n}l^{(u)}_{n_1+n_k}=\mathrm{E}_{g^*}[D_{I_1\cup I_k}(y)l^{(u)}_1(\boldsymbol{\theta}|y)]. \tag{2.3.17}$$

It follows from expression (2.3.2) and results (2.3.8) and (2.3.17) that as $n\to\infty$, uniformly in $\boldsymbol{\theta}\in\Theta$

$$\begin{aligned}\frac{1}{n}l_n(\boldsymbol{\theta}|y)\xrightarrow{a.s.}&\sum_{i=2}^{k-1}p_i\log\frac{p_{\mathrm{PN}i}(\boldsymbol{\theta})}{p_i}-(p_1\log p_1+p_k\log p_k)\\&+\mathrm{E}_{g^*}\left[D_{I_1\cup I_k}(y)l^{(u)}_1(\boldsymbol{\theta}|y)\right]\end{aligned} \tag{2.3.18}$$

Result (2.3.15) implies that $\mathrm{E}_{g^*}[D_{I_1\cup I_k}(y)l^{(u)}_1(\boldsymbol{\theta}|y)]$ is also continuous in $\boldsymbol{\theta}$ since $\mathrm{E}_{g^*}[l^{(u)}_{1,\varepsilon}(\boldsymbol{\theta}|y)]$ is continuous. Thus, from (2.3.18) and Appendix 3, we obtain the strong consistency of $\hat{\boldsymbol{\theta}}_n$.

To establish the asymptotic normality of $\hat{\boldsymbol{\theta}}_n$, we consider $n^{-1/2}$ times the gradient of the log-likelihood evaluated at $\hat{\boldsymbol{\theta}}_n$. After an expansion in Taylor's series about $\boldsymbol{\theta}_0$, the fact that $\nabla l_n(\hat{\boldsymbol{\theta}}_n|y)=\mathbf{0}$ almost surely for n large enough, implies that $n^{-1/2}\nabla l_n(\boldsymbol{\theta}_n|y)$ and $-n^{-1}\nabla^2 l_n(\boldsymbol{\theta}^*_n|y)[\sqrt{n}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)]$ have the same limiting distribution, where $\boldsymbol{\theta}^*_n=\gamma_n\boldsymbol{\theta}_0+(1-\gamma_n)\hat{\boldsymbol{\theta}}_n\,(0<\gamma_n<1)$.

Next, we can write

$$\begin{aligned}\frac{1}{n}\nabla l_n(\boldsymbol{\theta}_0|y)&=\frac{1}{n}\left\{\nabla l^{(g)}_{n-n_1-n_k}(\boldsymbol{\theta}_0)+\nabla l^{(u)}_{n_1+n_k}(\boldsymbol{\theta}_0 y)\right\}\\&=\frac{1}{n}\sum_{j=1}^{n}X_j(\boldsymbol{\theta}_0)\end{aligned} \tag{2.3.19}$$

where the random vectors $X_1(\boldsymbol{\theta}_0),\cdots,X_n(\boldsymbol{\theta}_0)$ are identically and independently distributed with mean vector

$$\mathrm{E}_{g^*}[X(\boldsymbol{\theta}_0)]=\nabla H(\boldsymbol{\theta}_0)+\mathrm{E}_{g^*}[D_{I_1\cup I_k}(y)\nabla l^{(u)}_1(\boldsymbol{\theta}_0|y)]=\mathbf{0}$$

and covariance matrix $W=(w_{uv})(u,v=1,2,3)$ with elements given by

$$w_{uv} = \sum_{i=2}^{k-1} p_i \left(\frac{\partial \log p_{\mathrm{PN}i}(\boldsymbol{\theta})}{\partial \theta_u}\bigg|_{\theta_0} \right)\left(\frac{\partial \log p_{\mathrm{PN}i}(\boldsymbol{\theta})}{\partial \theta_v}\bigg|_{\theta_0} \right)$$
$$+ \mathrm{E}_{g^*}\left[I_{I_1 \cup I_k}(y) \left(\frac{\partial l_1^{(u)}(\boldsymbol{\theta}|y)}{\partial \theta_u}\bigg|_{\theta_0} \right)\left(\frac{\partial l_1^{(u)}(\boldsymbol{\theta}|y)}{\partial \theta_v}\bigg|_{\theta_0} \right) \right]$$

Applying the multivariate central limit theorem gives that, as $n \to \infty$, the asymptotic normality of $n^{-1/2}\nabla l_n(\boldsymbol{\theta}_0|y)$ and consequently

$$\frac{1}{n}\nabla^2 l_n(\boldsymbol{\theta}_n^*|y)\left[\sqrt{n}(\hat{\boldsymbol{\theta}}_n^* - \boldsymbol{\theta}_0)\right] \xrightarrow{d} \mathrm{N}_3(\mathbf{0},\boldsymbol{W}). \tag{2.3.20}$$

Premultiplying the left hand side of (2.3.20) by the matrix $\boldsymbol{V}$, defined by $\boldsymbol{V} = \{\nabla^2 H(\boldsymbol{\theta}_0) + \mathrm{E}_{g^*}[D_{I_1\cup I_k}(y)\nabla^2 l_1^{(u)}(\boldsymbol{\theta}_0|y)]\}^{-1}$ gives

$$\boldsymbol{V}\left\{\frac{1}{n}\nabla^2 l_n(\boldsymbol{\theta}_n^*|y)\left[\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right]\right\} \xrightarrow{d} \mathrm{N}_3(\mathbf{0},\boldsymbol{W}\boldsymbol{V}\boldsymbol{W}^{\mathrm{T}}). \tag{2.3.21}$$

Next, we will use the representation

$$\frac{1}{n}\nabla^2 l_n(\boldsymbol{\theta}|y) = \frac{1}{n}\nabla^2 l_{n-n_1-n_k}^{(g)}(\boldsymbol{\theta}) + \frac{1}{n}\nabla^2 l_{n_1+n_k}^{(u)}(\boldsymbol{\theta}|y) \tag{2.3.22}$$

to consider the grouped and ungrouped Hessians separately. By the same argument in the Section 2.2.2, we concluded that, as $n \to \infty$

$$\frac{1}{n}\nabla^2 l_{n-n_1-n_k}^{(g)}(\boldsymbol{\theta}_n^*) \xrightarrow{a.s.} \nabla^2 H(\boldsymbol{\theta}_0). \tag{2.3.23}$$

In the Appendix, we have mentioned, for easier reference, some properties of the second order partial derivatives of $l_1^{(u)}(\boldsymbol{\theta}|y)$ established by Hernandez (1978). By Appendix 5, we know that these derivatives are continuous in $(\boldsymbol{\theta}, Y)$ and Appendix 6, they are also g-integrable. These facts are used in what follows.

For u and v arbitrary but fixed, define

$$d_1(\boldsymbol{\theta}|y) = \frac{\partial^2 l_1^{(u)}(\boldsymbol{\theta}|y)}{\partial\theta_u \partial\theta_v}, \quad d_{1,\eta}(\boldsymbol{\theta}|y) = \frac{\partial^2 l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y)}{\partial\theta_u \partial\theta_v},$$
$$d_{n_1+n_k}(\boldsymbol{\theta}|y) = \sum_{j=1}^{n} D_{I_1\cup I_k}(y_j) d_1(\boldsymbol{\theta}|y), \quad d_{n_1+n_k,\varepsilon}(\boldsymbol{\theta}|y) = \sum_{j=1}^{n} d_{1,\varepsilon}(\boldsymbol{\theta}|y_j) \tag{2.3.24}$$

with $l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y)$ as defined in (2.2.8). Then, $|d_{1,\varepsilon}(\boldsymbol{\theta}|y)| \le |d_1(\boldsymbol{\theta}|y)|$ implies that $|d_{1,\varepsilon}(\boldsymbol{\theta}|y)|$ is also dominated by g-integrable function $H_{uv}(y)$ that dominates $|d_1(\boldsymbol{\theta}|y)|$. By considering the sequence of sets $\{S_i\}_{i=1}^{\infty}$ used in Appendix 7, we deduce that $|d_{1,\varepsilon}(\boldsymbol{\theta}|y)|$ is equicontinuous in $\boldsymbol{\theta}$ for $Y \in S_i$. Thus, applying Appendix 4, as $n \to \infty$, uniformly in $\boldsymbol{\theta} \in \Theta$

$$\frac{1}{n} d_{n_1+n_k,\varepsilon}(\boldsymbol{\theta}|y) \xrightarrow{a.s.} \mathrm{E}_{g^*}[d_{1,\varepsilon}(\boldsymbol{\theta}|y)] \tag{2.3.25}$$

and the limit function is continuous.

Next, we have

$$\begin{aligned}\left|\frac{1}{n} d_{n_1+n_k,\varepsilon}(\boldsymbol{\theta}_n^*|y) - \mathrm{E}_{g^*}[d_1(\boldsymbol{\theta}_0|y]\right| &\le \left|\frac{1}{n} d_{n_1+n_k,\varepsilon}(\boldsymbol{\theta}_n^*|y) - \mathrm{E}_{g^*}[d_1(\boldsymbol{\theta}_n^*|y]\right| \\ &+ \left|\mathrm{E}_{g^*}[d_{1,\varepsilon}(\boldsymbol{\theta}_0|y)] - \mathrm{E}_{g^*}[d_{1,\varepsilon}(\boldsymbol{\theta}_0|y)]\right|.\end{aligned} \tag{2.3.26}$$

The first term on the right hand side of this expression tends almost surely to zero because of (2.3.25), while the second term also converges to zero with probability one because of the continuity of $\mathrm{E}_{g^*}[d_{1,\varepsilon}(\boldsymbol{\theta}_0|y)]$ and the fact $\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}_0$. Hence, with probability one

$$\lim_{n\to\infty}\left|\frac{1}{n} d_{n_1+n_k,\varepsilon}(\boldsymbol{\theta}_n^*|y) - \mathrm{E}_{g^*}[d_{1,\varepsilon}(\boldsymbol{\theta}_0|y)]\right| = 0. \tag{2.3.27}$$

Now, adding and subtracting $n^{-1} d_{n_1+n_k,\varepsilon}(\boldsymbol{\theta}_n^*|y) - \mathrm{E}_{g^*}[d_{1,\varepsilon}(\boldsymbol{\theta}_0|y)]$ to $n^{-1} d_{n_1+n_k}(\boldsymbol{\theta}_n^*|y) - \mathrm{E}_{g^*}[D_{I_1 \cup I_k}(y) d_1(\boldsymbol{\theta}_0|y)]$, we obtain

$$\begin{aligned}&\left|\frac{1}{n} d_{n_1+n_k}(\boldsymbol{\theta}_n^*|y) - \mathrm{E}_{g^*}[D_{I_1 \cup I_k}(y) d_1(\boldsymbol{\theta}_0|y)]\right| \\ &\le \left|\frac{1}{n} d_{n_1+n_k}(\boldsymbol{\theta}_n^*|y) - \frac{1}{n} d_{n_1+n_k,\varepsilon}(\boldsymbol{\theta}_n^*|y)\right| + \left|\frac{1}{n} d_{n_1+n_k,\varepsilon}(\boldsymbol{\theta}_n^*|y) - \mathrm{E}_{g^*}[d_{1,\varepsilon}(\boldsymbol{\theta}_0|y)]\right| \\ &+ \left|\mathrm{E}_{g^*}[d_{1,\varepsilon}(\boldsymbol{\theta}_0|y)] - \mathrm{E}_{g^*}[D_{I_1 \cup I_k}(y) d_1(\boldsymbol{\theta}_0|y)]\right|.\end{aligned} \tag{2.3.28}$$

Therefore, following the same reasoning as in steps (2.3.13) through (2.3.17)

$$\begin{aligned}\left|\frac{1}{n} d_{n_1+n_k}(\boldsymbol{\theta}_n^*|y) - \frac{1}{n} d_{n_1+n_k,\varepsilon}(\boldsymbol{\theta}_n^*|y)\right| &\le \frac{1}{n}\sum_{j=1}^{n} D_{I_{1,\varepsilon} \cup I_{k,\varepsilon}}(y_j) H_{uv}(y_j) \\ &\xrightarrow{\text{a.s.}} \mathrm{E}_{g^*}[D_{I_{1,\varepsilon} \cup I_{k,\varepsilon}}(y) H_{uv}(y_j)]\end{aligned} \tag{2.3.29}$$

as $n \to \infty$, and for $\eta > 0$

$$\begin{aligned}\left|\mathrm{E}_{g^*}[d_{1,\varepsilon}(\boldsymbol{\theta}_0|y) - \mathrm{E}_{g^*}[D_{I_1 \cup I_k}](y) d_1(\boldsymbol{\theta}_0|y)\right| &\le \mathrm{E}_{g^*}[D_{I_{1,\varepsilon} \cup I_{k,\varepsilon}}(y) H_{uv}(y)] \\ &< \frac{\eta}{2}.\end{aligned} \tag{2.3.30}$$

if ε is sufficiently small. Thus, by taking limit as $n \to \infty$ in (2.3.28) and using (2.3.27), (2.3.29) and (2.3.30), we get the almost sure condition

$$\lim_{n\to\infty}\left|\frac{1}{n}d_{n_1+n_k}(\boldsymbol{\theta}_n^*|y) - \mathrm{E}_{g^*}[D_{I_1\cup I_k}(y)d_1(\boldsymbol{\theta}_0|y)]\right| < \eta$$

Since η is arbitrary, it follows that, as $n \to \infty$

$$\frac{1}{n}d_{n_1+n_k}(\boldsymbol{\theta}_n^*|y) \xrightarrow{a.s.} \mathrm{E}_{g^*}[D_{I_1\cup I_k}(y)d_1(\boldsymbol{\theta}_0|y)] \tag{2.3.31}$$

As u and v were arbitrarily chosen in (2.3.24), results (2.3.31) implies that, as $n \to \infty$

$$\frac{1}{n}\nabla^2 l_{n_1+n_k}^{(u)}(\boldsymbol{\theta}_n^*|y) \xrightarrow{a.s.} \mathrm{E}_{g^*}[D_{I_1\cup I_k}(y)\nabla^2 l_1^{(u)}(\boldsymbol{\theta}_0|y)]. \tag{2.3.32}$$

Using results (2.3.23) and (2.3.31), we finally conclude that, as $n \to \infty$

$$\frac{1}{n}\nabla^2 l_n(\boldsymbol{\theta}_n^*|y) \xrightarrow{a.s.} \nabla^2 H(\boldsymbol{\theta}_0) + \mathrm{E}_{g^*}[D_{I_1\cup I_k}(y)\nabla^2 l_1^{(u)}(\boldsymbol{\theta}_0|y)] = V^{-1}. \tag{2.3.33}$$

Consequently, $n[\nabla^2 l_n(\boldsymbol{\theta}_n^*|y)]^{-1}V^{-1} \xrightarrow{p} \mathbf{I}_3$ as $n \to \infty$, which together with (2.3.21), gives the asymptotic normality of the maximum likelihood estimates $\hat{\boldsymbol{\theta}}_n$, where $\mathbf{I}_3$ denotes the 3×3 identity matrix.

2.3.2 Right and/or Left Censored Observations

The following type of mixed observations occurs when a censoring scheme operates in such a way that observations falling between the specified limits y_1 and y_2 are recorded exactly, while only the counts of values lower y_1 and upper y_2 $(0 < y_1 < y_2 < \infty)$ are available. This is the usual double Type I censoring scheme [*c.f.* Kendall and Stuart (1973)], which for our case, divides the positive real line into three intervals, namely $I_1 = [0, y_1), I_2 = [y_1, y_2)$ and $I_3 = [y_2, \infty)$, and yields observations of grouped and ungrouped.

Then the log-likelihood function of the mixed observations of size n is given by

$$\begin{aligned} l_n(\boldsymbol{\theta}|y) &= \log n! - \sum_{i=1}^{3}\log n_i! + n_1 \log p_{\mathrm{PN1}}(\boldsymbol{\theta}) + n_3 \log p_{\mathrm{PN3}}(\boldsymbol{\theta}) \\ &\quad - \frac{1}{2}n_2 \log 2\pi - \frac{1}{2}n_2 \log\sigma^2 - \frac{1}{2}\sum_{j=n_1+1}^{n-n_3}\left(\frac{y_j^{(\lambda)} - \mu}{\sigma}\right)^2 \\ &\quad + (\lambda - 1)\sum_{j=n_1+1}^{n-n_3}\log y_j - n_2 A(\kappa) \end{aligned} \tag{2.3.34}$$

where $n = n_1 + n_2 + n_3$, and

$$p_{\mathrm{PN1}}(\boldsymbol{\theta}) = \frac{1}{A(\kappa)}\Phi\left(\frac{y_1^{(\lambda)} - \mu}{\sigma}\right) \text{ and } p_{\mathrm{PN3}}(\boldsymbol{\theta}) = \frac{1}{A(\kappa)}\left[1 - \Phi\left(\frac{y_2^{(\lambda)} - \mu}{\sigma}\right)\right]. \tag{2.3.35}$$

Here, we notice again that (2.3.34) can be rewritten as

$$l_n(\boldsymbol{\theta}|y) = l^{(g)}_{n_1+n_3}(\boldsymbol{\theta}) + l^{(u)}_{n_2}(\boldsymbol{\theta}|y) \tag{2.3.36}$$

where

$$l^{(g)}_{n_1+n_3}(\boldsymbol{\theta}) = \log n! - \sum_{i=1}^{3} \log n_i! + n_1 \log p_{\text{PN1}}(\boldsymbol{\theta}) + n_3 \log p_{\text{PN3}}(\boldsymbol{\theta}) \tag{2.3.37}$$

is the part of the log-likelihood linked to the grouped observations, and

$$\begin{aligned} l^{(u)}_{n_2}(\boldsymbol{\theta}|y) = &-\frac{1}{2}n_2 \log 2\pi - \frac{1}{2}n_2 \log\sigma^2 - \frac{1}{2}\sum_{j=n_1+1}^{n-n_3}\left(\frac{y_j^{(\lambda)} - \mu}{\sigma}\right)^2 \\ &+ (\lambda - 1)\sum_{j=n_1+1}^{n-n_3} \log y_j - n_2 A(\kappa) \end{aligned} \tag{2.3.38}$$

is the part linked to the ungrouped observations.

Now, we consider the properties of estimates of λ, μ and σ based on the behavior of $n^{-1}l_n(\boldsymbol{\theta}|y)$. This procedure goes essentially as that of previous section, so we just present a few main intermediate steps.

Let $p_i = \int_{I_i} g^*(y)dy$ for $i = 1,2,3$ and let $p_{\text{PN1}}(\boldsymbol{\theta})$ and $p_{\text{PN3}}(\boldsymbol{\theta})$ be as in (2.3.35). Suppose that (i) the parameter space Θ is the compact subset of $\Re^3$ defined in previous section, (ii) the moments $\mathrm{E}_{g^*}[Y^{2a}]$ and $\mathrm{E}_{g^*}[Y^{2b}]$ are finite, and (iii) $H(\boldsymbol{\theta}) + \mathrm{E}_{g^*}[D_{I_2}(y) l_1^{(u)}(\boldsymbol{\theta}|y)]$ has a unique global maximum at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, where

$$H(\boldsymbol{\theta}) = p_1 \log\frac{p_{\text{PN1}}(\boldsymbol{\theta})}{p_1} + p_3 \log\frac{p_{\text{PN3}}(\boldsymbol{\theta})}{p_3}.$$

Furthermore, suppose that (iv) $\boldsymbol{\theta}_0$ be an interior point of Θ, (v) $\mathrm{E}_{g^*}[Y^a \log Y]^2$ and $\mathrm{E}_{g^*}[Y^b \log Y]^2$ both are finite, (vi)

$$\nabla H(\boldsymbol{\theta}_0) + \mathrm{E}_{g^*}[D_{I_2}(y)\nabla l_1^{(u)}(\boldsymbol{\theta}_0|y)] = \mathbf{0} \tag{2.3.39}$$

and (vii) $V = \{\nabla^2 H(\boldsymbol{\theta}_0) + \mathrm{E}_{g^*}[D_{I_2}(y)\nabla^2 l_1^{(u)}(\boldsymbol{\theta}_0|y)]\}^{-1}$ exists. We immediately have that, uniformly on Θ

$$\lim_{n\to\infty} l^{(g)}_{n_1+n_3}(\boldsymbol{\theta}) = p_1 \log\frac{p_{\text{PN1}}(\boldsymbol{\theta})}{p_1} + p_2 \log p_2 + p_3 \log\frac{p_{\text{PN3}}(\boldsymbol{\theta})}{p_3}. \tag{2.3.40}$$

Here, we define an approximation to $D_{I_2}(y) l_1^{(u)}(\boldsymbol{\theta}|y)$ which is continuous in Y $\forall\varepsilon > 0$

$$l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y) = D_{[y_1,y_1+\varepsilon)}(y)\left(\frac{y-y_1}{\varepsilon}\right)l_1^{(u)}(\boldsymbol{\theta}|y) + D_{[y_1+\varepsilon,y_2+\varepsilon)}(y)l_1^{(u)}(\boldsymbol{\theta}|y) + D_{[y_2-\varepsilon,y_2)}(y)\left(\frac{y_2-y}{\varepsilon}\right)l_1^{(u)}(\boldsymbol{\theta}|y) \tag{2.3.41}$$

By the uniform strong low of large number described in Appendix 4, we obtain that, as $n \to \infty$, uniformly in $\boldsymbol{\theta} \in \Theta$, with $\mathrm{E}_{g^*}[l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y)]$ continuous

$$\frac{1}{n}\sum_{j=1}^{n} l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y_j) \xrightarrow{a.s.} \mathrm{E}_{g^*}[l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y)]. \tag{2.3.42}$$

Therefore, (2.3.41) provides the almost sure limit

$$\lim_{n\to\infty}\frac{1}{n} l_{n_2}^{(u)}(\boldsymbol{\theta}|y) = \mathrm{E}_{g^*}[D_{I_2}(y)l_1^{(u)}(\boldsymbol{\theta}|y)] \tag{2.3.43}$$

uniformly in $\boldsymbol{\theta} \in \Theta$. Thus, combining (2.4.38) and (2.3.42), we finally obtain the probability one limit

$$\lim_{n\to\infty}\frac{1}{n} l_n(\boldsymbol{\theta}|y) = H(\boldsymbol{\theta}) + p_2 \log p_2 + \mathrm{E}_{g^*}[D_{I_2}(y)l_1^{(u)}(\boldsymbol{\theta}|y)] \tag{2.3.44}$$

uniformly in $\boldsymbol{\theta} \in \Theta$. Hence, the strong consistency of $\hat{\boldsymbol{\theta}}_n$ is deduced from Appendix 3.

A Taylor's series expansion of $n^{-1/2}\nabla l_n(\boldsymbol{\theta}|y)$ about $\boldsymbol{\theta}_0$ then allows for establishing that, as $n \to \infty$

$$\frac{1}{n}\nabla^2 l_n(\boldsymbol{\theta}_n^*|y)\left[\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right] \xrightarrow{d} \mathrm{N}_3(\mathbf{0}, W) \tag{2.3.45}$$

for $\boldsymbol{\theta}_n^*$ lying on the segment joining $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$, where the 3×3 matrix $W = (w_{uv})$ is defined by

$$\begin{aligned} w_{uv} &= p_1\left(\left.\frac{\partial \log p_{\mathrm{PN1}}(\boldsymbol{\theta})}{\partial \theta_u}\right|_{\theta_0}\right)\left(\left.\frac{\partial \log p_{\mathrm{PN1}}(\boldsymbol{\theta})}{\partial \theta_v}\right|_{\theta_0}\right) \\ &\quad + p_3\left(\left.\frac{\partial \log p_{\mathrm{PN3}}(\boldsymbol{\theta})}{\partial \theta_u}\right|_{\theta_0}\right)\left(\left.\frac{\partial \log p_{\mathrm{PN3}}(\boldsymbol{\theta})}{\partial \theta_v}\right|_{\theta_0}\right) \\ &\quad + \mathrm{E}_{g^*}\left[D_{I_2}(y)p_1\left(\left.\frac{\partial l_1^{(u)}(\boldsymbol{\theta}|y)}{\partial \theta_u}\right|_{\theta_0}\right)\left(\left.\frac{\partial l_1^{(u)}(\boldsymbol{\theta}|y)}{\partial \theta_v}\right|_{\theta_0}\right)\right]. \end{aligned}$$

Since the second order partial derivatives

$$\frac{\partial^2 \log p_{\mathrm{PN}i}(\boldsymbol{\theta})}{\partial \theta_u \partial \theta_v}, \quad i = 1,3$$

are uniformly continuous in $\boldsymbol{\theta} \in \Theta$, it easily follows that, as $n \to \infty$

$$\frac{1}{n}\nabla^2 l_{n_1+n_3}^{(g)}(\boldsymbol{\theta}_n^*) \xrightarrow{a.s.} \nabla^2 H(\boldsymbol{\theta}_0). \tag{2.3.46}$$

Now, defining

$$d_{1,\varepsilon}(\boldsymbol{\theta}|y) = \frac{\partial^2 l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y)}{\partial\theta_u \partial\theta_v}$$

for fixed u and v, we can use the uniform strong low of large number to concluded that, as $n \to \infty$

$$\frac{1}{n}\sum_{j=n_1+1}^{n-n_3} d_{1,\varepsilon}(\boldsymbol{\theta}|y_j) \xrightarrow{a.s.} \mathrm{E}_{g^*}[d_{1,\varepsilon}(\boldsymbol{\theta}|y)] \tag{2.3.47}$$

uniformly on Θ, where the limit function is continuous. Then, using the fact that the second order partial derivatives of $l_1^{(u)}(\boldsymbol{\theta}|y)$ are g-integralbe, we can also obtain that, as $n \to \infty$

$$\frac{1}{n}\nabla^2 l_{n_2}^{(u)}(\boldsymbol{\theta}|y) \xrightarrow{a.s.} \mathrm{E}_{g^*}[D_{I_2}(y)\nabla^2 l_1^{(u)}(\boldsymbol{\theta}_0|y)]. \tag{2.3.48}$$

Finally, it follows from (2.3.45) and (2.3.47) that

$$\lim_{n\to\infty}\frac{1}{n}\nabla^2 l_n(\boldsymbol{\theta}_n^*|y) = \nabla^2 H(\boldsymbol{\theta}_0) + \mathrm{E}_{g^*}[D_{I_2}(y)\nabla^2 l_1^{(u)}(\boldsymbol{\theta}|y)] = V^{-1}. \tag{2.3.49}$$

Applying Stutsky's theorem to (2.3.44) under the (2.3.48), we have the asymptotic normality of maximum likelihood estimate $\hat{\boldsymbol{\theta}}_n$.

In the case just discussed, the censoring arose because some values fell outside an observable range; the censoring tool place at certain fixed points and is called Type I censoring. Type II censoring is said to occur when a fixed proportion of the observations is censored at the lower and/or upper ends of the range of Y. In practice, Type II censoring often occurs when Y, the variate under study, is a time-period (e.g., the period to failure of the piece of equipment undergoing testing) and the experimental time available is limited.

From the theoretical point of view, the prime distribution between Type I and Type II censoring is that the former case, the number of censoring observations is a random variable, while in the latter case it is fixed in advance. The distribution theory of Type II censoring is correspondingly simpler.

If the censoring is of Type II, with n_1 and n_3 fixed, the log-likelihood function is

$$\begin{aligned} l_n(\boldsymbol{\theta}|y) &= \log n! - \sum_{i=1}^{3}\log n_i! + n_1\log p_{\mathrm{PN1}}(\boldsymbol{\theta}) + n_3\log p_{\mathrm{PN3}}(\boldsymbol{\theta}) \\ &\quad -\frac{1}{2}n_2\log 2\pi - \frac{1}{2}n_2\log\sigma^2 - \frac{1}{2}\sum_{j=n_1+1}^{n-n_3}\left(\frac{y_j^{(\lambda)}-\mu}{\sigma}\right)^2 \\ &\quad +(\lambda-1)\sum_{j=n_1+1}^{n-n_3}\log y_j - n_2 A(\kappa) \end{aligned} \tag{2.3.50}$$

with

$$p_{\text{PN1}}(\boldsymbol{\theta}) = \frac{1}{A(\kappa)}\Phi\left(\frac{y_{n_1+1}^{(\lambda)} - \mu}{\sigma}\right) \text{ and } p_{\text{PN3}}(\boldsymbol{\theta}) = \frac{1}{A(\kappa)}\left[1 - \Phi\left(\frac{y_{n-n_3}^{(\lambda)} - \mu}{\sigma}\right)\right] \tag{2.3.51}$$

where y_{n_1+1} and y_{n_3+1} denote the n_1+1 st and $n-n_3$ th order statistics respectively. Expression (2.3.34), (2.3.35), (2.3.50) and (2.3.51) are of exactly the same form. They differ in that the quantities involved in (2.3.51) are random variables, but not in (2.3.35) and that n_1 and n_3 are random variables in (2.3.34) but not in (2.3.50). Given a set of observations, however, the formal similarity permits the same methods of iteration to be used in obtaining the maximum likelihood estimators. Moreover, asymptotically, the behavior of the two cases is very similar.

2.3.3 Individual Observations Arbitrarily Chosen

In the present case, we assume that it is possible to record h arbitrarily chosen exact values besides the counts, with $\lim_{n\to\infty} h/n = p(0 \le p \le 1)$. A situation like this may arise for instance, when a large completely grouped observations are available and then it is decided to draw another small observations (independent of the previous observations) of exact values, to improve the accuracy of the results (in this situation p would be close to zero). It will be assumed that the proportion p is fixed and specified in advance.

Let the observations be grouped into $k(k \ge 3)$ intervals denoted by $I_1 = [y_0, y_1), I_2 = [y_1, y_2), \cdots, I_k = [y_{k-1}, y_k)$ as was assumed in Section 2.2. But now let the total grouped observations be of size $n-h=\sum_{i=1}^{k} n_i$, where the frequency of the i th interval, as before. The independent observations of h ungrouped observations make the total sample size of the mixed observations equal to n.

The log-likelihood of the mixed observations, based on the tentative assumption that $Y_i \sim \text{PND}(\lambda,_0 \mu_0, \sigma_0^2)$, is obtained as

$$\begin{aligned} l_n(\boldsymbol{\theta}|y) &= \log(n-h)! - \sum_{i=1}^{k} n_i! + \sum_{i=1}^{k} n_i p_{\text{PN}i}(\boldsymbol{\theta}) - \frac{h}{2}\log 2\pi \\ &- \frac{h}{2}\log\sigma^2 - \frac{1}{2}\sum_{j=1}^{h}\left(\frac{y_j^{(\lambda)} - \mu}{\sigma}\right)^2 + (\lambda - 1)\sum_{j=1}^{h}\log y_j - hA(\kappa) \end{aligned} \tag{2.3.52}$$

where $p_{\text{PN}i}(\boldsymbol{\theta})$ is given by (2.2.2).

The corresponding likelihood is $l_n(\boldsymbol{\theta}|y) = l_{n-h}^{(g)}(\boldsymbol{\theta}) + l_h^{(u)}(\boldsymbol{\theta}|y)$ with

$$l_{n-h}^{(g)}(\boldsymbol{\theta}) = \log(n-h)! - \sum_{i=1}^{k} n_i! + \sum_{i=1}^{k} n_i p_{\text{PN}i}(\boldsymbol{\theta}) \tag{2.3.53}$$

and

$$\begin{aligned} l_h^{(u)}(\boldsymbol{\theta}|y) = &-\frac{h}{2}\log 2\pi - \frac{h}{2}\log\sigma^2 - \frac{1}{2}\sum_{j=1}^{h}\left(\frac{y_j^{(\lambda)}-\mu}{\sigma}\right)^2 \\ &+ (\lambda-1)\sum_{j=1}^{h}\log y_j - hA(\kappa). \end{aligned} \tag{2.3.54}$$

Now, we consider the properties of estimates of λ, μ and σ based on the behavior of $n^{-1}l_n(\boldsymbol{\theta}|y)$. This procedure is based mainly on results obtained in Section 2.3.1 and 2.3.2

Let p_i and $p_{\text{PND}i}(\boldsymbol{\theta})$ be defined as in Section 2.2.2, and suppose that (i) the parameter space Θ is the compact subset of $\Re^3$ defined by $\Theta = \{\boldsymbol{\theta}^{\mathrm{T}} = (\lambda, \mu, \sigma) \| \mu| \le M, s_1 \le \sigma \le s_2, a \le \lambda \le b$ for some $0 < M, s_1, s_2, b < \infty$ and $-\infty < a < 0\}$, (ii) the moments $\mathrm{E}_{g^*}[Y^{2a}]$ and $\mathrm{E}_{g^*}[Y^{2b}]$ are finite, and (iii) $(1-p)H(\boldsymbol{\theta}) + p\mathrm{E}_{g^*}[l_1^{(u)}(\boldsymbol{\theta}|y)]$ has a unique global maximum at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, where $H(\boldsymbol{\theta}) = \sum_{i=1}^{k} p_i \log(p_{\text{PN}i}(\boldsymbol{\theta})/p_i)$. Also, suppose that (iv) $\boldsymbol{\theta}_0$ be an interior point of Θ, (v) $\mathrm{E}_{g^*}[Y^a \log Y]^2$ and $\mathrm{E}_{g^*}[Y^b \log Y]^2$ both are finite, (vi)

$$\sqrt{1-p}\nabla H(\boldsymbol{\theta}_0) + \sqrt{p}\mathrm{E}_{g^*}[\nabla l_1^{(u)}(\boldsymbol{\theta}_0|y)] = \mathbf{0}$$

and (vii) $V = \{(1-p)\nabla^2 H(\boldsymbol{\theta}_0) + p\mathrm{E}_{g^*}[\nabla^2 l_1^{(u)}(\boldsymbol{\theta}_0|y)]\}^{-1}$ exists.

Since $h/n \to p$ as $n \to \infty$, it follows that both $h \to \infty$ and $n-h \to \infty$ as $n \to \infty$. Using Stirling's approximation and Appendix 2, it is easy to show that, as $n \to \infty$, uniformly on Θ

$$\frac{1}{n} l_{n-h}^{(g)}(\boldsymbol{\theta}) = \frac{n-h}{n}\frac{1}{n-h} l_{n-h}^{(g)}(\boldsymbol{\theta}) \xrightarrow{a.s.} (1-p)H(\boldsymbol{\theta}) \tag{2.3.55}$$

and, by the uniform strong low of large number, uniformly on Θ

$$\frac{1}{n} l_h^{(u)}(\boldsymbol{\theta}|y) = \frac{h}{n}\frac{1}{h} l_h^{(u)}(\boldsymbol{\theta}|y) \xrightarrow{a.s.} p\mathrm{E}_{g^*}[l_1^{(u)}(\boldsymbol{\theta}|y)] \tag{2.3.56}$$

where $\mathrm{E}_{g^*}[l_1^{(u)}(\boldsymbol{\theta}|y)]$ is continuous. Together (2.3.55) and (2.3.56) imply that, uniformly on Θ

$$\lim_{n\to\infty}\frac{1}{n} l_n(\boldsymbol{\theta}|y) = (1-p)H(\boldsymbol{\theta}) + p\mathrm{E}_{g^*}[l_1^{(u)}(\boldsymbol{\theta}|y)]. \tag{2.3.57}$$

Then, by Appendix 3 with $f(\boldsymbol{\theta}) = (1-p)H(\boldsymbol{\theta}) + p\mathrm{E}_{g^*}[l_1^{(u)}(\boldsymbol{\theta}|y)]$, $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ as $n \to \infty$, with probability one.

Next, by the multivariate central limit theorem, as $n-h \to \infty$

$$\frac{1}{\sqrt{n-h}} \nabla l_{n-h}^{(g)}(\boldsymbol{\theta}_0) \xrightarrow{d} \mathrm{N}_3(U_1, W_1) \tag{2.3.58}$$

and

$$\frac{1}{\sqrt{h}}\nabla l_h^{(u)}(\boldsymbol{\theta}_0|y) \xrightarrow{d} \mathrm{N}_3(U_2, W_2) \tag{2.3.59}$$

where

$$W_1 = \left(\sum_{i=1}^{k} p_i(1-p_i)\left(\frac{\partial \log p_{\mathrm{PN}i}(\boldsymbol{\theta})}{\partial \theta_u}\bigg|_{\theta_0}\right)\left(\frac{\partial \log p_{\mathrm{PN}i}(\boldsymbol{\theta})}{\partial \theta_v}\bigg|_{\theta_0}\right)\right)_{3\times 3},$$

$$W_2 = \left(\mathrm{var}_{g^*}\left(\frac{\partial l_1^{(u)}(\boldsymbol{\theta}|y)}{\partial \theta_u}\bigg|_{\theta_0}\right)\left(\frac{\partial l_1^{(u)}(\boldsymbol{\theta}|y)}{\partial \theta_v}\bigg|_{\theta_0}\right)\right)_{3\times 3}$$

$U_1 = \nabla H(\boldsymbol{\theta})$, and $U_2 = \mathrm{E}_{g^*}[\nabla l_1^{(u)}(\boldsymbol{\theta}_0|y)]$. Therefore, as $n \to \infty$

$$\frac{1}{\sqrt{n}}\nabla l_n(\boldsymbol{\theta}_0|y) \xrightarrow{d} \mathrm{N}_3(\mathbf{0}, W) \tag{2.3.60}$$

where $W = (1-p)W_1 + pW_2$.

Now, expanding by Taylor's formula, it can be shown that $n^{-1/2}\nabla l_n(\boldsymbol{\theta}_0|y)$ and $n^{-1}\nabla^2 l_n(\boldsymbol{\theta}_n^*|y)[\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)]$ have the same limit distribution, where $\boldsymbol{\theta}_n^* = r_n\hat{\boldsymbol{\theta}}_n + (1-r_n)\boldsymbol{\theta}_0$ for some $0 < r_n < 1$. Also, as $n \to \infty$

$$\frac{1}{n}\nabla^2 l_{n-h}^{(g)}(\boldsymbol{\theta}_n^*) \xrightarrow{a.s.} (1-p)\nabla^2 H(\boldsymbol{\theta}_0) \tag{2.3.61}$$

and

$$\frac{1}{n}\nabla^2 l_h^{(u)}(\boldsymbol{\theta}_n^*|y) \xrightarrow{a.s.} p\mathrm{E}_{g^*}[\nabla^2 l_1^{(u)}(\boldsymbol{\theta}_0|y)] \tag{2.3.62}$$

so that

$$\frac{1}{n}\nabla^2 l_n(\boldsymbol{\theta}_n^*|y) \xrightarrow{a.s.} V^{-1} \tag{2.3.63}$$

as $n \to \infty$. Thus, Slutsky's theorem enables to conclude $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{a.s.} \mathrm{N}_3(\mathbf{0}, VWV)$ as $n \to \infty$.

2.3.4 Ungrouped Observations at Fixed Proportions of Observations

The last case of mixed observations to be studied arises when a fixed number of proportions $[h_1, \cdots, h_k$ with $0 < h_i < 1 (i = 1, \cdots, k)$ and $\sum_{i=1}^{k} h_i = 1]$ is determined at the outset. Suppose that it is permitted to make an exact observation after each proportion $h_i\ (i=1,\cdots,k-1)$ of order statistics has occurred. If $n_i = h_i n + 1$ for $i=1,\cdots,k-1$ and $n_k = n - \sum_{i=1}^{k-1} n_i$, where n is the total sample size, then the sampling scheme can be expressed as follows:

Obtain $k-1$ ordered observations $y_{m_1},\cdots,y_{m_{k-1}}$, and $n-k+1$ grouped values such that

n_1 observations fall in $[0, y_{m_1})$

n_2 observations fall in $[y_{m_1}, y_{m_2})$

$\vdots$

n_k observations fall in $[y_{m_{k-1}}, \infty)$

where y_{m_i} denotes the $\sum_{j=1}^{i} n_j + 1$st order statistic, so that $0 < y_{m_1} \le y_{m_2} \le \cdots \le y_{m_{k-1}} < \infty$. Notice that y_{m_i} is counted as one of the n_{i+1} observations falling in $[y_{m_i}, y_{m_{i+1}})$.

Therefore, probability $p_{\mathrm{PN}i}$ is defied as

$$p_{\mathrm{PN}i}(\boldsymbol{\theta}) = \frac{1}{A(\kappa)}\left[\Phi\left(\frac{y_{m_i}^{(\lambda)}-\mu}{\sigma}\right) - \Phi\left(\frac{y_{m_{i+1}}^{(\lambda)}-\mu}{\sigma}\right)\right] \tag{2.3.64}$$

where

$$p_{\mathrm{PN}1}(\boldsymbol{\theta}) = \frac{1}{A(\kappa)}\Phi\left(\frac{y_{m_1}^{(\lambda)}-\mu}{\sigma}\right).$$

Then, for $Y \in S$, the log-likelihood function obtained under the tentative assumption that $Y_j \sim \mathrm{PND}(\lambda_0, \mu_0, \sigma_0^2)$ becomes

$$\begin{aligned} l_n(\boldsymbol{\theta}|y) &= \log n! - \log n_1! - \sum_{i=2}^{k}\log(n_i-1)! + n_1 \log p_{\mathrm{PN}1}(\boldsymbol{\theta}) \\ &+ \sum_{i=2}^{k}(n_i-1)\log p_{\mathrm{PN}i}(\boldsymbol{\theta}) - \frac{k-1}{2}\log 2\pi - \frac{k-1}{2}\log\sigma^2 \\ &- \frac{1}{2}\sum_{i=1}^{k-1}\left(\frac{y_{m_i}^{(\lambda)}-\mu}{\sigma}\right) + (\lambda-1)\sum_{i=1}^{k-1}\log y_{m_i} - (k-1)A(\kappa) \end{aligned} \tag{2.3.65}$$

where S is the set in which the requirements of the sampling scheme are fulfilled.

To establish the asymptotic properties of the maximum likelihood estimates $\hat{\theta}_n$ in this situation, we will suse the concept of quantile. The p is the quantile $(0 < p < 1)$ of the distribution $F(y)$ is formally defined as real number ξ_p such that $F(\xi_p) \le p$ and $F(\xi_p + 0) \ge p$. The sample p the quantile then $F(\xi_p) = p$ and $F(\xi_p + \varepsilon) > p\ \forall \varepsilon > 0$. The sample p th quantile is correspondingly $y_{[np]+1}$. Assume that ξ_p is unique, then the following is well-known fact which follows from the Glivenko-Cantelli lemma. With probability one, as $n \to \infty$

$$y_{[np]+1} \to \xi_p. \tag{2.3.66}$$

Setting

$$y_0 = 0,\ y_i = \xi_{\sum_{j=1}^{i} h_j},\ \text{and}\ y_k = \infty$$

so that

$$\int_{y_{i-1}}^{y_i} g^*(y)dy = h_i \tag{2.3.67}$$

we can apply (2.3.66), assuming uniqueness of the Y_i 's, to get

$$y_{m_i} \xrightarrow{a.s.} y_i \tag{2.3.68}$$

as $n \to \infty$. Hence, by the continuity of $\Phi(y)$, letting $n \to \infty$, we get

$$\begin{aligned} p_{\text{PN1}}(\boldsymbol{\theta}|y) &\xrightarrow{a.s.} p_{\text{PN1}}(\boldsymbol{\theta}) = \frac{1}{A(\kappa)}\Phi\left(\frac{y_1^{(\lambda)} - \mu}{\sigma}\right), \\ p_{\text{PN}i}(\boldsymbol{\theta}|y) &\xrightarrow{a.s.} p_{\text{PN}i}(\boldsymbol{\theta}) = \frac{1}{A(\kappa)}\left[\Phi\left(\frac{y_i^{(\lambda)} - \mu}{\sigma}\right) - \Phi\left(\frac{y_{i-1}^{(\lambda)} - \mu}{\sigma}\right)\right]. \end{aligned} \tag{2.3.69}$$

Using Stirling's approximation for factorials, as we did in the procedure of Section 2.2.2, it follows that, for $Y \in S$

$$\begin{aligned} \frac{1}{n} l_n(\boldsymbol{\theta}|y) = &\left\{\sum_{i=1}^{k} h_i \log p_{\text{PN}i}(\boldsymbol{\theta}|y) - \sum_{i=1}^{k} h_i \log h_i + o(1)\right\} \\ &+ \left\{-\frac{k-1}{2n}\log 2\pi - \frac{k-1}{n}\log\sigma - \frac{1}{2n}\sum_{i=1}^{k-1}\left(\frac{y_{m_i}^{(\lambda)} - \mu}{\sigma}\right)^2 \right. \\ &\left. + \frac{\lambda - 1}{n}\sum_{i=1}^{k-1}\log y_{m_i} - \frac{k-1}{n}A(\kappa)\right\} \end{aligned} \tag{2.3.70}$$

Next, the first order partial derivatives of $l_n(\boldsymbol{\theta}|y)$ are

$$\begin{aligned} \frac{\partial l_n(\boldsymbol{\theta}|y)}{\partial \theta_1} = &\frac{n}{\sigma}\sum_{i=1}^{k} h_i \frac{\phi(y_{i-1}) - \phi(y_i)}{p_{\text{PND}i}(\boldsymbol{\theta}|y)} - \frac{1}{\sigma}\sum_{i=2}^{k}\frac{\phi(y_{i-1}) - \phi(y_i)}{p_{\text{PND}i}(\boldsymbol{\theta}|y)} \\ &- \frac{1}{\sigma^2}\left[\sum_{i=1}^{k-1} y_{m_i}^{(\lambda)} - (k-1)\mu\right] - \frac{n}{A(\kappa)}\frac{\partial A(\kappa)}{\partial \lambda} \end{aligned},$$

$$\frac{\partial l_n(\boldsymbol{\theta}|y)}{\partial \theta_2} = \frac{n}{\sigma}\sum_{i=1}^{k} h_i \frac{Z_{i-1}\phi(y_{i-1}) - Z_i\phi(y_i)}{p_{\mathrm{PND}i}(\boldsymbol{\theta}|y)} - \frac{1}{\sigma}\sum_{i=2}^{k}\frac{z_{i-1}\phi(y_{i-1}) - z_i\phi(y_i)}{p_{\mathrm{PND}i}(\boldsymbol{\theta}|y)} - \frac{k-1}{\sigma} + \frac{1}{\sigma^3}\left[\sum_{i=1}^{k-1} y_{m_i}^{(\lambda)} - \mu\right]^2 - \frac{n}{A(\kappa)}\frac{\partial A(\kappa)}{\partial \mu}, \quad (2.3.71)$$

$$\frac{\partial l_n(\boldsymbol{\theta}|y)}{\partial \theta_1} = \frac{n}{\sigma}\sum_{i=1}^{k} h_i \frac{\phi(y_{i-1})y_{m_i}^{\prime(\lambda)} - \phi(y_i)y_{m_{i-1}}^{\prime(\lambda)}}{p_{\mathrm{PN}i}(\boldsymbol{\theta}|y)} - \frac{1}{\sigma}\sum_{i=2}^{k}\frac{\phi(y_{i-1})y_{m_i}^{\prime(\lambda)} - \phi(y_i)y_{m_{-1i}}^{\prime(\lambda)}}{p_{\mathrm{PN}i}(\boldsymbol{\theta}|y)} - \frac{1}{\sigma}\sum_{i=1}^{k-1}\left(\frac{y_{m_i}^{(\lambda)} - \mu}{\sigma}\right)y_{m_i}^{\prime(\lambda)} + \sum_{i=1}^{k-1}\log y_{m_i} - \frac{n}{A(\kappa)}\frac{\partial A(\kappa)}{\partial \sigma}$$

where $z_i = (y_{m_i}^{(\lambda)} - \mu)/\sigma$ and $y_{m_i}^{\prime(\lambda)} = dy_{m_i}^{(\lambda)}/d\lambda$.

Thus, by comparing $n^{-1}l_n(\boldsymbol{\theta}|y)$ and $n^{-1}\nabla l_n(\boldsymbol{\theta}|y)$ with their analogues in the completely grouped observations case $n^{-1}l_n(\boldsymbol{\theta})$ and $n^{-1}\nabla l_n(\boldsymbol{\theta})$, we can readily observe the asymptotic equivalence, as $n \to \infty$. The same thing occurs in the Hessians. Thus, provided that $\mathrm{E}_{g^*}[Y^{2a}]$, $\mathrm{E}_{g^*}[Y^{2b}]$, $\mathrm{E}_{g^*}[Y^a \log Y]^2$ and $\mathrm{E}_{g^*}[Y^b \log Y]^2$ are all finite, the only thing we have to do is change p_i by h_i in Section 2.2.2 and that theorem holds true also for current case.

2.3.5 Examples

In this section, the two examples will consider the two situations of Section 2.3.1 and 2.3.2. One is the example of the weekly earnings of secretaries cited from Industry Wage Surveys (1976), which will be used to consider the case of the ungrouped observations available in the tail. The other is the aflatoxin levels of raw peanut kernels cited from Fisher and Belle (1993), which will be used to consider the case of the right and left censored observations.

Example 11 (Industry Wage Surveys, 1976): The data are the weekly earnings of secretaries. The observations outside the range [130,250) are given ungrouped. So five observations falling in [120,130) are 121,123,125,127,129 respectively, and the last three observations are 255, 285 and 325 respectively.

The results of fitting the PND to these grouped observations are shown in Table 2.3.1. The value of transforming parameter λ was estimated as -1.3145 with the approximate 95% confidence interval (-1.3354, -1.2936). This optimized value suggests that these observations have an L-shaped distribution. For the optimized value, we obtained $\hat{\mu}(\hat{\lambda}) = 0.7598$ and $\hat{\sigma}(\hat{\lambda}) = 0.0002$. The value of back-transformed $\hat{\mu}(\hat{\lambda})$ to the original scale, $\hat{\mu}^*(\hat{\lambda})$ was given as 162.3285. The back-transformed value $\hat{\mu}^*(\hat{\lambda})$ was smaller than the corresponding one on the original scale, $\hat{\mu}(1)$ =172.6924. The plot of the profile of maximized log-likelihood as a function of λ is shown in Figure 2.3.1, together with the value of chi-square statistics. Figure 2.3.1 shows that the value of chi-square statistics have a minimum at the neighborhood of $\hat{\lambda}$ = 0.7598.

Table 2.3.1 Example 11: The results of fitting of the PND

Maximum log-likelihood		**Performance**	
Maximum log-likelihood $l(\hat{\theta}(\hat{\lambda}))$	−61.7163	**Goodness of fit**	
Truncated probability $1-A(\kappa)$	1.0000	Chi-square statistics	60.0566
Estimate of λ		p -value	near 0
$\hat{\lambda}$	−1.3145	**Shape of original observations**	
95% confidence interval for $\hat{\lambda}$	−1.3354 −1.2936	Skewness Kurtosis	1.0742 5.4620
Estimate of μ		**Shape of power-transformed observations**	
$\hat{\mu}(\hat{\lambda})$	0.7598	Skewness	0.0337
Back to the original scale $\hat{\mu}^*(\hat{\lambda})$	162.3285	Kurtosis	2.5476
$\hat{\mu}(1)$	172.6924		
Estimate of σ			
$\hat{\sigma}(\hat{\lambda})$	0.0002		
$\hat{\sigma}(1)$	26.7211		

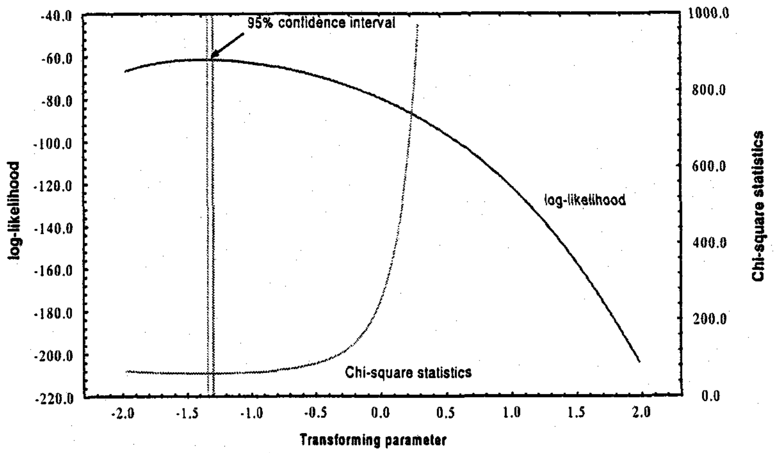

Figure 2.3.1 Example 11: The profile of maximized log-likelihood and the value of chi-square statistics as a function of transforming parameter λ

The distributions of observations on the original and the power-transformed scale are shown in Figure 2.3.2 and 2.3.3 respectively. The value of chi-square statistics for goodness of fit was given as 60.0566, which was significant at 5% level. The value of skewness for the observations on the power-transformed scale was given as 0.0337, which was not farther from zero than the value of 1.0742 for the observations on the original scale. But the value of the kurtosis for the observations on the power-transformed scale was 2.5476, which was not farther from 3 than the value of 5.4620 for the observations on the original scale. This shows that the observations on the power-transformed scale achieve the near-normality in comparison with the observation on the original scale.

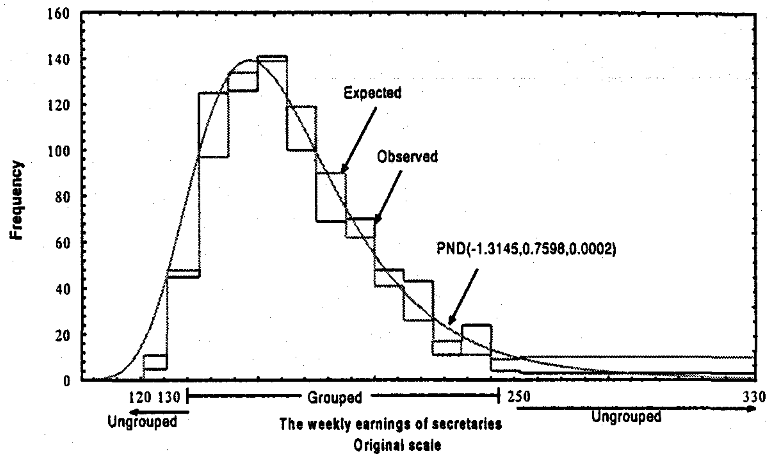

Figure 2.3.2 Example 11: The distribution of observations on the original scale

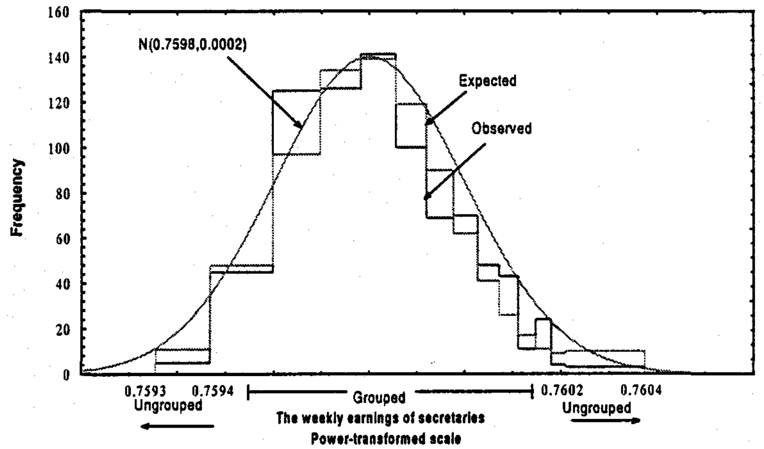

Figure 2.3.3 2 Example 11: The distribution of observations on the power-transformed scale

Example 12 (Fisher and Belle, 1993): The data are the aflatoxin levels of raw peanut kernels. The observations outside [20,50) are given grouped. So, the twelve observations falling in [20,50) are 26,26,22,27,23,28,30,36,31,35,37 and 48 respectively.

The results of fitting the PND to these grouped observations are shown in Table 2.3.2. The value of transforming parameter λ was estimated as -0.9460 with the approximate 95% confidence interval $(-3.1041,\ 1.2031)$. This optimized value suggests that these observations have an L-shaped distribution. For the optimized value, we obtained $\hat{\mu}(\hat{\lambda}) = 1.0165$ and $\hat{\sigma}(\hat{\lambda}) = 0.0130$. The value of back-transformed $\hat{\mu}(\hat{\lambda})$ to the original scale, $\hat{\mu}^*(\hat{\lambda})$ was given as 31.3752. The back-transformed value $\hat{\mu}^*(\hat{\lambda})$ was smaller than the corresponding one on the original scale, $\hat{\mu}(1)$ =3.8682. The plot of the profile of maximized log-likelihood as a function of λ is shown in Figure 2.3.4, together with the value of chi-square statistics. Figure 2.3.4 shows that the value of chi-square statistics have a minimum at the neighborhood of $\hat{\lambda} = -0.9460$.

Table 2.3.2 Example 12: The results of fitting of the PND

Maximum log-likelihood		Performance	
Maximum log-likelihood $l(\hat{\theta}(\hat{\lambda}))$	-9.2062	**Goodness of fit**	
Truncated probability $1-A(\kappa)$	0.9991	Chi-square statistics	10.2138
Estimate of λ		p -value	0.5972
$\hat{\lambda}$	-0.9460	**Shape of original observations**	
	-3.1041	Skewness	0.2963
95% confidence interval for $\hat{\lambda}$	1.2031	Kurtosis	1.0541
Estimate of μ		**Shape of power-transformed observations**	
$\hat{\mu}(\hat{\lambda})$	1.0165	Skewness	-0.0305
Back to the original scale $\hat{\mu}^*(\hat{\lambda})$	31.3752	Kurtosis	1.3964
$\hat{\mu}(1)$	33.8682		
Estimate of σ			
$\hat{\sigma}(\hat{\lambda})$	0.0130		
$\hat{\sigma}(1)$	11.8505		

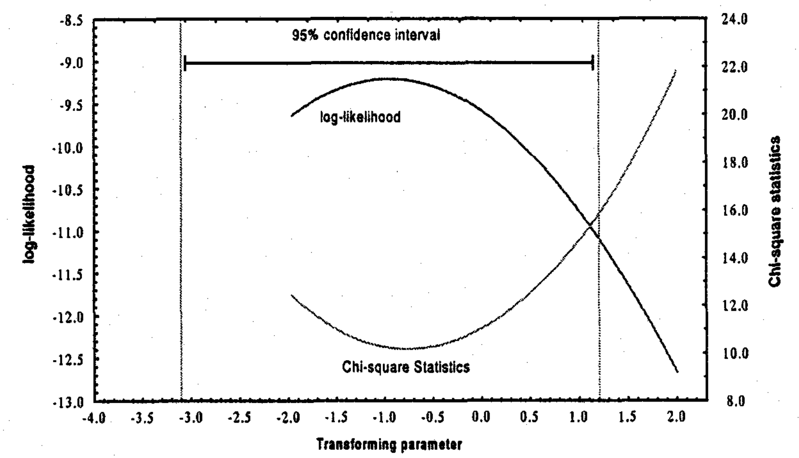

Figure 2.3.4 Example 12: The profile of maximized log-likelihood and the value of chi-square statistics as a function of transforming parameter λ

The distributions of observations on the original and the power-transformed scale are shown in Figure 2.3.5 and 2.3.6 respectively. The value of chi-square statistics for goodness of fit was given as 10.2138, which was not quite significant at 5% level. The value of skewness for the observations on the power-transformed scale was given as –0.3049, which was not farther from zero than the value of 0.2963 for the observations on the original scale. But the value of the kurtosis for theobservations on the power-transformed scale was given as 1.3964, which was far from 3 as same as the value of 1.0541 for the observations on the original scale. This shows that the observations on the power-transformed scale achieve the near-normality in comparison with the observations on the original scale.

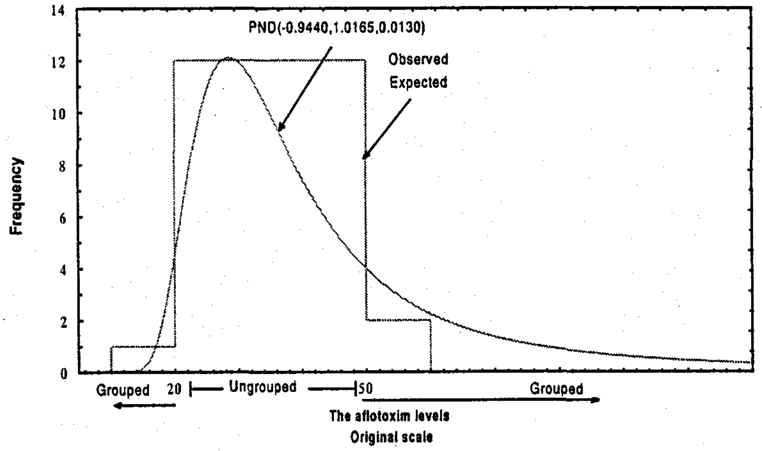

Figure 2 3.5 Example 12: The distribution of observations on the original scale

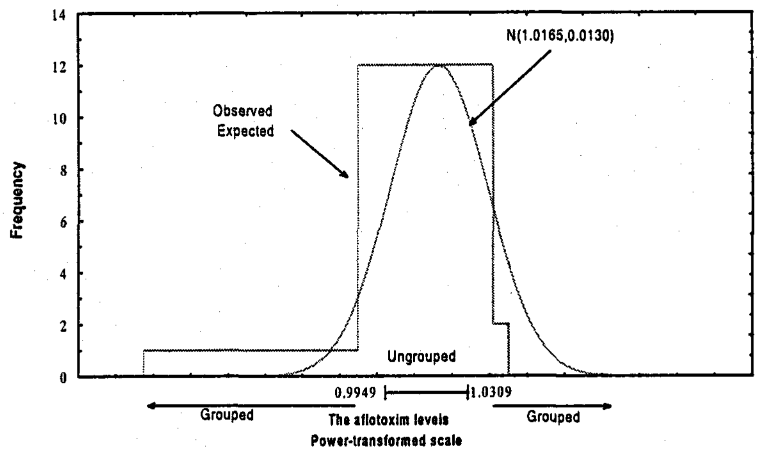

Figure 2.3.6 Example 12: The distribution of observations on the power-transformed scale

2.4 Further Problems, Other Analysis

2.4.1 Normalizing Power-transformation

The power-transformation defined by (2.2.1) dose not generally has scale invariance with the transformation of observations. For this reason, it is meaningless to perform subsequent statistical processing by quantities or statistics which depend on scale of observations in before and after of the transformation (Atkinson, 1982, 1985: Box and Cox, 1982: Hinkley and Runger, 1984: Goto *et al.*, 1991). Then, in order to take away scaling effect which depends on the transforming parameter, it may be recommended to apply "the normalized power-transformation (NPT)" which is adjusted ordinary power-transformation (OPT) by n th root of Jacobian of transformation (Box and Cox, 1964: Hinkley and Runger, 1984: Goto *et al.*, 1991: Duan, 1993). By working with NPT rather than OPT, quantities or statistics can be directly comparable in before and after of transformation, and the effect of the transforming parameter on the parameter estimates in the model will become much small (Hinkley and Runger, 1984: Duan, 1993: Hamasaki and Goto, 1996). Furthermore the asymptotic variance of the model parameters in which NPT is used will smaller than that of OPT (Bickel and Doksum, 1981: Carroll and Ruppert, 1981: Hinkley and Runger, 1984).

NPT corresponding to OPT defined by (2.1.1) is

$$Y_{\mathrm{NPT}}^{(\lambda)} = \begin{cases} \dfrac{Y^{\lambda}-1}{\lambda \dot{Y}^{\lambda-1}}, & \lambda \neq 0, \\ \dot{Y}\log Y, & \lambda = 0 \end{cases} \tag{2.4.1}$$

where $\dot{Y}$ denotes the geometric mean $\left(\prod_{i=1}^{k-1} y_i^{\,n_i}\right)^{1/n}$ of the observations $y_1, \cdots, y_{k-1}$. Thus, the Jacobian of transformation corresponding to (2.1.1) and (2.4.1) are

$$J_{\mathrm{OPT}} = \begin{cases} \dot{Y}^{n(\lambda-1)}, & \lambda \neq 0, \\ \dot{Y}^{n}, & \lambda = 0, \end{cases} \tag{2.4.2}$$

$$J_{\mathrm{NPT}} = \begin{cases} 1 + \dfrac{1-\lambda}{n\lambda} \displaystyle\sum_{i=1}^{k-1} (1 - y_i^{\lambda}), & \lambda \neq 0, \\ 1 + \log \dot{Y}, & \lambda = 0 \end{cases} \tag{2.4.3}$$

respectively. Thought the Jacobian (2.4.2) corresponding to (2.4.1) do not exactly equal to one so that the effect of λ can not entirely be removed on the scale of normalized power-transformed scale, $J_{\mathrm{NPT}}^{1/n} \xrightarrow{a.s.} 1$ for all λ as $n \to \infty$.

To illustrate above arguments, let us consider the seven examples in Section 2.2.4. The estimates of refitting the PND to seven examples in Section 2.2.4 by using NPT are shown in Table 2.4.1 together with that of OPT. Table 2.4.1 shows that there are little differences in the estimate of transforming parameter $\hat{\lambda}$ between OPT and NPT for all seven examples, though there are large difference in $\hat{\mu}(\hat{\lambda})$ and $\hat{\sigma}(\hat{\lambda})$.

Table 2.4.1 Comparison of estimates between OPT and NPT

Example		OPT	NPT	Example		OPT	NPT
	$\hat{\lambda}$	1.1107	1.1108		$\hat{\lambda}$	−1.3852	−1.4363
Example 1	$\hat{\mu}(\hat{\lambda})$	96.0289	60.2330	**Example 5**	$\hat{\mu}(\hat{\lambda})$	0.7098	1016.0138
	$\hat{\sigma}(\hat{\lambda})$	4.0753	2.5563		$\hat{\sigma}(\hat{\lambda})$	0.0040	4.9643
	$\hat{\lambda}$	−0.4691	−0.4691		$\hat{\lambda}$	−2.0010	−2.0010
Example 2	$\hat{\mu}(\hat{\lambda})$	1.9318	3215.1296	**Example 6**	$\hat{\mu}(\hat{\lambda})$	0.4984	3470.0952
	$\hat{\sigma}(\hat{\lambda})$	0.0123	20.4423		$\hat{\sigma}(\hat{\lambda})$	0.0007	4.7330
	$\hat{\lambda}$	1.0270	1.0270		$\hat{\lambda}$	0.6735	0.6736
Example 3	$\hat{\mu}(\hat{\lambda})$	37.4010	34.0309	**Example 7**	$\hat{\mu}(\hat{\lambda})$	63.2102	390.0697
	$\hat{\sigma}(\hat{\lambda})$	13.7454	12.5068		$\hat{\sigma}(\hat{\lambda})$	12.5686	77.5640
	$\hat{\lambda}$	1.9234	1.9234				
Example 4	$\hat{\mu}(\hat{\lambda})$	3221897.1896	1828.0513				
	$\hat{\sigma}(\hat{\lambda})$	1019537.7825	578.4517				

Next, to evaluate how stable working with normalized transformation can make the scale of the power-transformed observations, namely how it can removes the effect of the transforming parameter on the scale of the power-transformed observation, the n th root of the Jacobian of power-transformation were calculated for seven example. Also, to evaluate the effect of λ on μ and σ, we calculated two relative change rates defined by

$$R_{\hat{\mu}}(\lambda) = \lim_{\Delta \to 0} \frac{\hat{\mu}(\lambda + \Delta) - \hat{\mu}(\lambda)/\Delta}{\hat{\mu}(\lambda)} = \frac{\hat{\mu}'(\lambda)}{\hat{\mu}(\lambda)}, \tag{2.4.4}$$

$$R_{\hat{\sigma}}(\lambda) = \lim_{\Delta \to 0} \frac{\hat{\sigma}(\lambda + \Delta) - \hat{\sigma}(\lambda)/\Delta}{\hat{\sigma}(\lambda)} = \frac{\hat{\sigma}'(\lambda)}{\hat{\sigma}(\lambda)}. \tag{2.4.5}$$

The values of the n th root of the Jacobian of normalized power-transformation, and two relative change rates are shown in Table 2.4.2 together with those of OPT. Table 2.6.1 shows that the values of the n th root of the Jacobian of NPT are more close to one than that of OPT, for all seven examples. Also, two relative change rates which NPT was used were more smaller than that of OPT, in particular, Example 2, 5 and 6, where $\hat{\lambda}$ were estimated as a negative value. These results suggest that the scale of normalized power-transformed observations is more stable than that of power-transformed observations. However, the values of the n th root of the Jacobian of power-transformation in any examples do not exactly equal to one, so that the effect of λ can not entirely be removed on the scale of normalized power-transformed observations. However, note that, assuming that $J_{\mathrm{NPT}}^{1/n} \approx 1$ over the practically important range of λ, the absolute location of normalized power-transformed observation $Y_{\mathrm{NPT}}^{(\lambda)}$ will not be the same as that of Y and will depend on λ (Hinkley and Runger, 1984: Hamasaki and Goto, 1996).

Table 2.4.2 Comparison of Jacobian, $R_{\hat{\mu}}$ and $R_{\hat{\sigma}}$ between OPT and NPT

Example		OPT	NPT
	n th root of Jacobian	1.5945	1.0004
Example 1	$R_{\hat{\mu}}(\hat{\lambda})$	0.0000000466	0.0000001184
	$R_{\hat{\sigma}}(\hat{\lambda})$	−0.0000050754	−0.0000128975
	n th root of Jacobian	0.0006	1.0004
Example 2	$R_{\hat{\mu}}(\hat{\lambda})$	0.3009197352	0.0000000010
	$R_{\hat{\sigma}}(\hat{\lambda})$	−273.6778455808	−0.0000009247
	n th root of Jacobian	1.0990	1.0265
Example 3	$R_{\hat{\mu}}(\hat{\lambda})$	0.0000000039	0.0000000047
	$R_{\hat{\sigma}}(\hat{\lambda})$	−0.0000000249	−0.0000000301
	n th root of Jacobian	1761.5789	1.0000
Example 4	$R_{\hat{\mu}}(\hat{\lambda})$	0.0000000000	0.0000000127
	$R_{\hat{\sigma}}(\hat{\lambda})$	0.0000000000	−0.0000001566
	n th root of Jacobian	0.0007	1.0496
Example 5	$R_{\hat{\mu}}(\hat{\lambda})$	132.8646662489	0.0000007384
	$R_{\hat{\sigma}}(\hat{\lambda})$	−101852.0826853260	−0.0007223971
	n th root of Jacobian	0.0001	1.0076
Example 6	$R_{\hat{\mu}}(\hat{\lambda})$	7859.8394818762	0.0000000985
	$R_{\hat{\sigma}}(\hat{\lambda})$	−19552473.7652356000	−0.0003918026
	n th root of Jacobian	0.1621	1.0056
Example 7	$R_{\hat{\mu}}(\hat{\lambda})$	−0.0000000278	0.0000000000
	$R_{\hat{\sigma}}(\hat{\lambda})$	0.0000004982	−0.0000000004

2.4.2 Observations Subject to an Upper Constraint

The power transformation defined by (2.1.1) is appropriate for observations constrained to be non-negative. But in some cases the observations may also be subject to an upper constraint. For example, the observations might be the score of an examination paper out of 100, the true value of which must lie between 0 and 100. Unless the value lies near zero, when the upper constraint is of little importance, a power-transformation of the observations will not, in general, provide a simple model. If the values all lie near one hundred, for example in the range $90-100$, then a transformation from the family

$(100-Y)^{\lambda}$ will sometimes be useful. But for the complete range of values of Y, a more general class of transformations needs to be studied.

In this section, we introduce four analogues of the power-transformation for the observation subjected to an upper constraint c, so that the true values are constrained to lie between zero and c. These are respectively folded power-transformation (FPT), symmetric power-transformation (SPT), asymmetric power-transformation (APT) and asymmetric odds-ratio power-transformation (AOPT) (Arandaz-Ordaz, 1981: Guerrero and Johnson, 1982: Goto *et al.*, 1986: Atkinson, 1985).

Folded power-transformation (FPT): This transformation is defined by

$$Y_{\text{FPT}}^{(\lambda)} = \begin{cases} \dfrac{Y^{\lambda}-(c-Y^{\lambda})}{\lambda}, & \lambda \neq 0, \\ \log\left(\dfrac{Y}{c-Y}\right), & \lambda = 0 \end{cases} \tag{2.4.4}$$

where c is the upper constraint of variable Y. The FPT coincides with the logistic transformation for $\lambda = 0$ if $c = 1$. For values of Y near zero, the transformation behaves like the power-transformation Y^{λ}, whereas for values of Y near one it behaves like $(c-Y)^{\lambda}$. Suppose that, once the model has been fitted, expected value on the power-transformed scale is $\hat{\mu}$. Then, $\hat{Y}$, the expected value on the original scale, satisfies the relationship

$$\hat{Y}-(c-\hat{Y}) = \lambda\hat{\mu}. \tag{2.4.5}$$

Except for a few special values of λ, such as zero and one, analytical inversion of this transformation is not possible. Thus to obtain expected values on the original scale of the observations the iterative solution of relationships like (2.4.5) is required.

Asymmetric power-transformation (APT): This is a direct application of the power-transformation defined by (2.1.1) not to Y, but to the ratio $c/(c-Y)$, and defined by

$$Y_{\text{APT}}^{(\lambda)} = \begin{cases} \dfrac{1}{\lambda}\left[\left(\dfrac{c}{c-Y}\right)^{\lambda}-1\right], & \lambda \neq 0, \\ \log\left(\dfrac{c}{c-Y}\right), & \lambda = 0. \end{cases} \tag{2.4.6}$$

However this transformation does not yield the original observations for $\lambda = 1$.

The advantage of APT is that it is readily inverted. For a given λ, we can write

$$\frac{1}{\lambda}\left[\left(\frac{c}{c-Y}\right)^{\lambda}-1\right]=\mu \tag{2.4.7}$$

which can be inverted straightforwardly to yield

$$Y=\frac{c(\lambda\mu+1)^{1/\lambda}+c}{(\lambda\mu+1)^{1/\lambda}}. \tag{2.4.8}$$

Asymmetric odds-ratio power-transformation (AOPT): This is a direct application of the power-transformation defined by (2.1.1) not to Y, but to the odds-ratio $Y/(c-Y)$, and defined by

$$Y_{\text{AOPT}}^{(\lambda)}=\begin{cases}\dfrac{1}{\lambda}\left[\left(\dfrac{Y}{c-Y}\right)^{\lambda}-1\right], & \lambda\neq 0,\\ \log\left(\dfrac{Y}{c-Y}\right), & \lambda=0.\end{cases} \tag{2.4.9}$$

The AOPT coincides with the logistic transformation for $\lambda=0$ if $c=1$. However this transformation does not yield the original observations for $\lambda=1$.

The advantage of AOPT is that it is readily inverted. For a given λ, we can write

$$\frac{1}{\lambda}\left[\left(\frac{Y}{c-Y}\right)^{\lambda}-1\right]=\mu \tag{2.4.10}$$

which can be inverted straightforwardly to yield

$$Y=\frac{c(\lambda\mu+1)^{1/\lambda}}{1+(\lambda\mu+1)^{1/\lambda}}. \tag{2.4.11}$$

Thus the fitted model can readily be used to provide predictions and confidence intervals on the original scale of the observations one these have been calculated on the power-transformed scale.

Symmetric power-transformation (SPT): This transformation is more complicated than APT and AOPT. But it has the advantages of being invertible and of yielding the undefined observations for $\lambda=1$ as well as reducing to the logistic transformation for $\lambda=0$ if $c=1$. STP is defined by

$$Y_{\text{SPT}}^{(\lambda)}=\begin{cases}\dfrac{2}{\lambda}\left\{\dfrac{Y^{\lambda}-(c-Y)^{\lambda}}{Y^{\lambda}+(c-Y)^{\lambda}}\right\}, & \lambda\neq 0,\\ \log\left(\dfrac{Y}{c-Y}\right), & \lambda=0.\end{cases} \tag{2.4.12}$$

In term of odds $Y/(c-Y)$, rather than Y, (2.4.9) is

$$Y_{\mathrm{SPT}}^{(\lambda)} = \frac{2}{\lambda}\frac{[Y/(c-Y)]^{\lambda}-1}{[Y/(c-Y)]^{\lambda}+1}. \tag{2.4.13}$$

These two forms emphasize the connection between SPT and those of two preceding transformations. The number of (2.4.13) is the AOPT, while the number of (2.4.12) yields the FTP. The inverse of the transformation is found, by

$$\left(\frac{c}{c-Y}\right)^{\lambda} = \left(\frac{1+\lambda\mu/2}{1-\lambda\mu/2}\right)^{1/\lambda} \tag{2.4.14}$$

which can readily be further re-arranged to give a slightly cumbersome expression for Y as a function of μ.

A strange feature of the transformation given by (2.4.12) is that

$$Y_{\mathrm{SPT}}^{(\lambda)} = -Y_{\mathrm{SPT}}^{(-\lambda)} \tag{2.4.15}$$

which is easily shown since

$$\left(\frac{c}{c-Y}\right)^{-\lambda} = 1\Big/\left(\frac{c}{c-Y}\right)^{\lambda}. \tag{2.4.16}$$

The maximized log-likelihood is therefore symmetrical about $\lambda = 0$.

As one comparison of above four transformations, let us consider the total score of the common first-stage examination for university entrance cited from Sugiura (1980,1981).

Example 13 (Sugiura, 1980, 1981): This data are the grouped observations of total score of the common first-stage examination for university entrance conducted in fiscal 1979(13A), 1980(13B) and 1981(13C) cited from Sugiura (1980, 1981). Because these observations were the score of an examination paper out of 1000, the true value of which should lie between 0 and 1000.

The results of fitting the PND to these three grouped observations 13A, 13B and 13C by using FPT, APT, AOPT and SPT are shown in Table 2.4.3, 2.4.4, and 2.4.5 respectively together with those of the ordinary power-transformation (OPT). The plots of the profile of maximized log-likelihood $l(\hat{\theta}(\lambda))$ as a function of λ for 13A, 13B and 13C are shown in Figure 2.4.1, 2.4.5 and 2.4.9 respectively. The related plots of value of chi-square statistics are shown in Figure 2.4.2, 2.4.6 and 2.4.10 respectively. FPT, AOPT and SPT yield the logistic transformation for $\lambda = 0$ and so coincide at that point. OPT, FPT and SPT also coincide at $\lambda = 1$. For $\lambda = 0.5$, the behavior of FTP and SPT seem similar, though they diverge as λ increases toward 2. For SPT, because of the symmetry of $l(\hat{\theta}(\lambda))$ and chi-square statistics, the slope of the two functions is zero at $\lambda = 0$.

Table 2.4.3 Example 13A(1979): The results of fitting of the PND

	OPT	FPT	APT	AOPT	SPT
Maximum log-likelihood					
Maximum log-likelihood $l(\hat{\theta}(\hat{\lambda}))$	−2651.9942	−185.5613	−1155.7049	−150.1864	−196.8037
Truncated probability $1-A(\kappa)$	0.9994	1.0000	1.0000	1.0000	1.0000
Estimate of λ					
$\hat{\lambda}$	1.5066	0.1478	−0.6366	−0.0689	0.4269
95% confidence interval for $\hat{\lambda}$	1.4857 1.5275	0.1269 0.1687	−0.6575 −0.6157	−0.0898 −0.0480	0.4060 0.4478
Estimate of μ					
$\hat{\mu}(\hat{\lambda})$	11315.1434	1.5154	0.7598	0.5867	0.5981
Back to the original scale $\hat{\mu}^*(\hat{\lambda})$	643.8190	–	645.9787	645.4067	645.9738
$\hat{\mu}(1)$			636.7576		
Estimate of σ					
$\hat{\sigma}(\hat{\lambda})$	3489.4185	1.5446	0.1977	0.6044	0.6102
$\hat{\sigma}(1)$			134.9674		
Performance					
Goodness of fit					
Chi-square statistics	4755.0929	273.9908	2035.6975	155.8879	245.5072
p -value	near 0	near 0	near 0	near 0	near 0
Shape of original observations					
Skewness			−0.2855		
Kurtosis			2.5904		
Shape of power-transformed observations					
Skewness	−0.0281	0.0597	0.0900	0.0725	0.0926
Kurtosis	2.4163	2.8727	2.6133	2.9766	2.8746

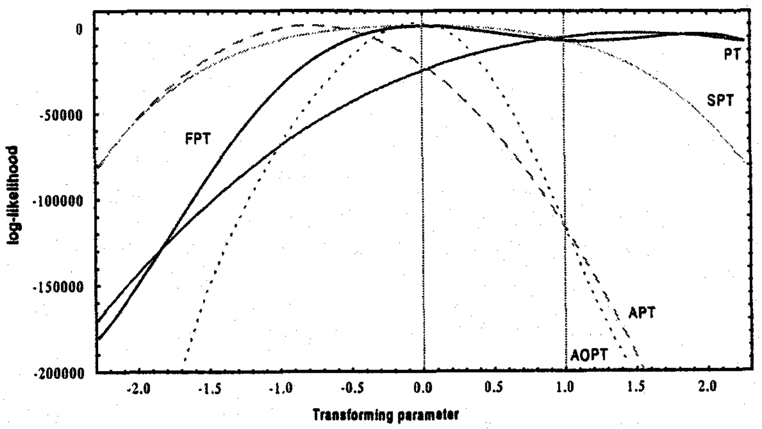

Figure 2.4.1 Example 13A(1979): The profile of maximized log-likelihood as a function of transforming parameter λ

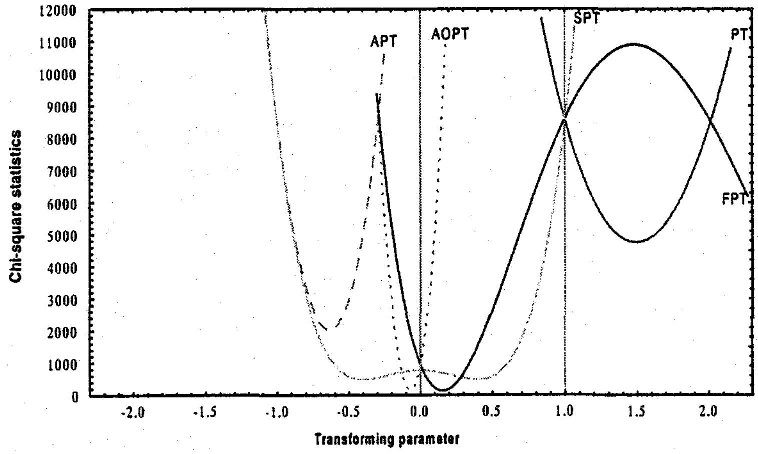

Figure 2.4.2 Example 13A(1979): The profile of the value of chi-square statistics as a function of transforming parameter λ

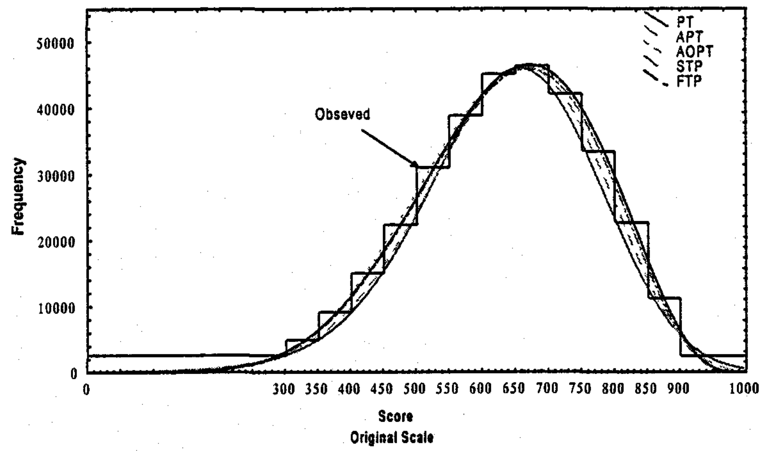

Figure 2.4.3 Example 13A(1979): The distributions of observations on the original scale

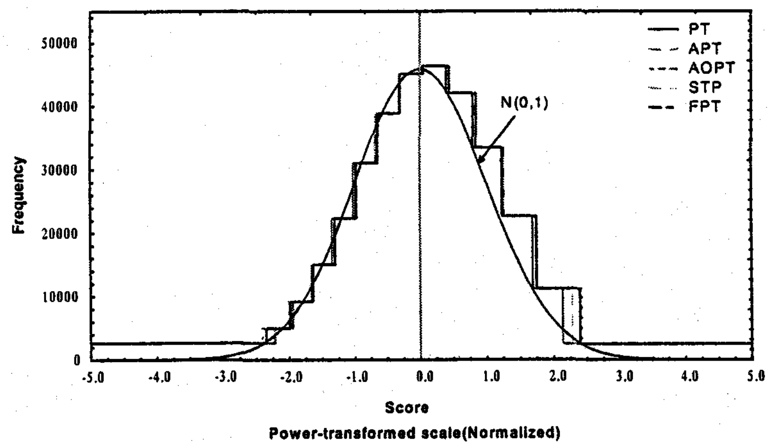

Figure 2.4.4 Example 13A(1979): The distributions of observations on the power-transformed scale

Table 2.4.4 Example 13B(1980): The results of fitting of the PND

	OPT	FPT	APT	AOPT	SPT
Maximum log-likelihood					
Maximum log-likelihood $l(\hat{\theta}(\hat{\lambda}))$	−2459.6568	−265.1823	−1301.8472	−255.3088	−265.1776
Truncated probability $1-A(\kappa)$	0.9996	1.0000	1.0000	1.0000	1.0000
Estimate of λ					
$\hat{\lambda}$	1.4760	0.0026	−0.6210	−0.0125	0.0307
95% confidence interval for $\hat{\lambda}$	1.4551	−0.0183	−0.6419	−0.0334	0.0000
	1.4969	0.0235	−0.6001	0.0084	0.1388
Estimate of μ					
$\hat{\mu}(\hat{\lambda})$	9057.2813	0.5177	0.7365	0.5141	0.5177
Back to the original scale $\hat{\mu}^*(\hat{\lambda})$	624.4131	–	626.3418	626.1623	626.6112
$\hat{\mu}(1)$			618.3322		
Estimate of σ					
$\hat{\sigma}(\hat{\lambda})$	2691.3160	0.5773	0.1867	0.5737	0.5772
$\hat{\sigma}(1)$			128.6309		
Performance					
Goodness of fit					
Chi-square statistics	4608.9699	388.5771	2338.4420	374.5717	388.6457
p -value	near 0	near 0	near 0	near 0	near 0
Shape of original observations					
Skewness			−0.2943		
Kurtosis			2.5295		
Shape of power-transformed observations					
Skewness	−0.0627	0.0688	0.0257	0.0130	0.0320
Kurtosis	2.3772	2.9796	2.4812	2.7439	2.7447

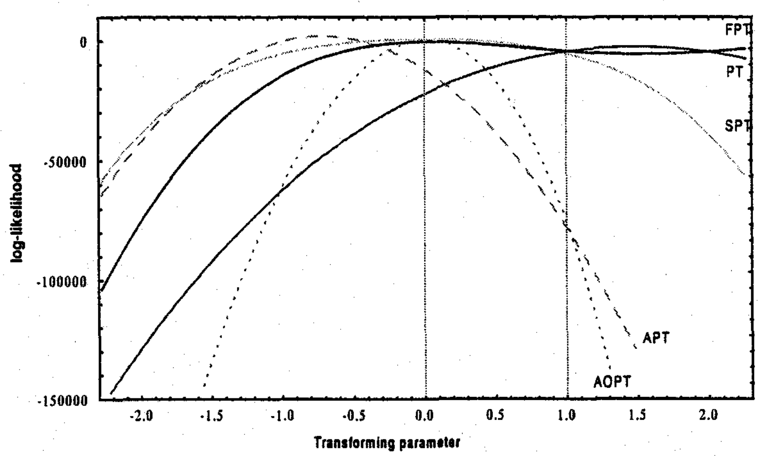

Figure 2.4.5 Example 13B(1980): The profile of maximized log-likelihood as a function of transforming parameter λ

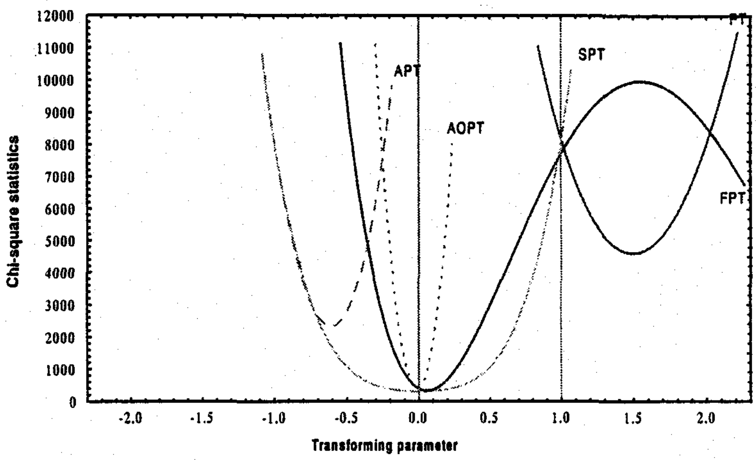

Figure 2.4.6 Example 13B(1980): The profile of the value of chi-square statistics as a function of transforming parameter λ

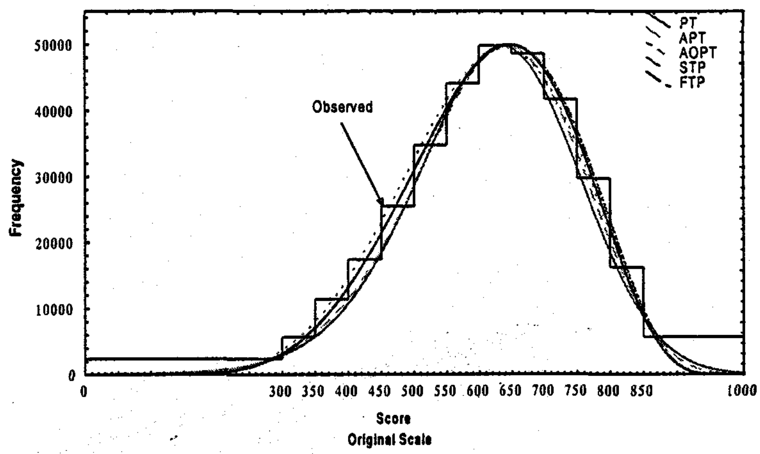

Figure 2.4.7 Example 13B(1980): The distributions of observations on the original scale

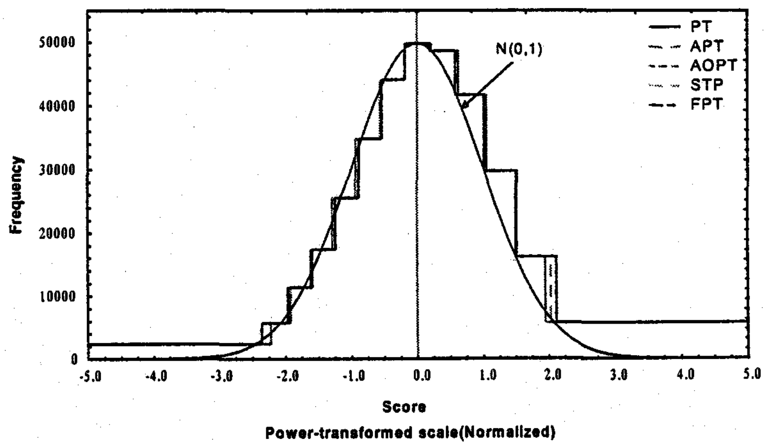

Figure 2.4.8 Example 13B(1980): The distributions of observations on the power-transformed scale

Table 2.4.5 Example 13C(1981): The results of fitting of the PND

	OPT	FPT	APT	AOPT	SPT
Maximum log-likelihood					
Maximum log-likelihood $l(\hat{\theta}(\hat{\lambda}))$	−2581.7629	−380.1049	−1567.3292	−110.5031	−401.4169
Truncated probability $1-A(\kappa)$	0.9998	0.9900	0.9999	1.0000	0.9999
Estimate of λ					
$\hat{\lambda}$	1.2713	0.2390	−0.7483	−0.1103	0.5427
95% confidence interval for $\hat{\lambda}$	1.2504 1.2922	0.2181 0.2599	−0.7692 −0.7274	−0.1312 −0.0894	0.5218 0.5636
Estimate of μ					
$\hat{\mu}(\hat{\lambda})$	2746.1563	2.0490	0.6814	0.4488	0.4649
Back to the original scale $\hat{\mu}^*(\hat{\lambda})$	612.3333	–	614.3975	613.0833	614.7633
$\hat{\mu}(1)$			608.0712		
Estimate of σ					
$\hat{\sigma}(\hat{\lambda})$	785.6198	2.6781	0.1784	0.6004	0.6077
$\hat{\sigma}(1)$			139.0661		
Performance					
Goodness of fit					
Chi-square statistics	4616.3647	597.8618	2836.8749	74.6738	640.2104
p -value	near 0	near 0	near 0	near 0	near 0
Shape of original observations					
Skewness			−0.1375		
Kurtosis			2.4327		
Shape of power-transformed observations					
Skewness	0.0041	0.1051	0.1146	0.0829	0.1385
Kurtosis	2.3736	2.7285	2.5250	2.9005	2.7591

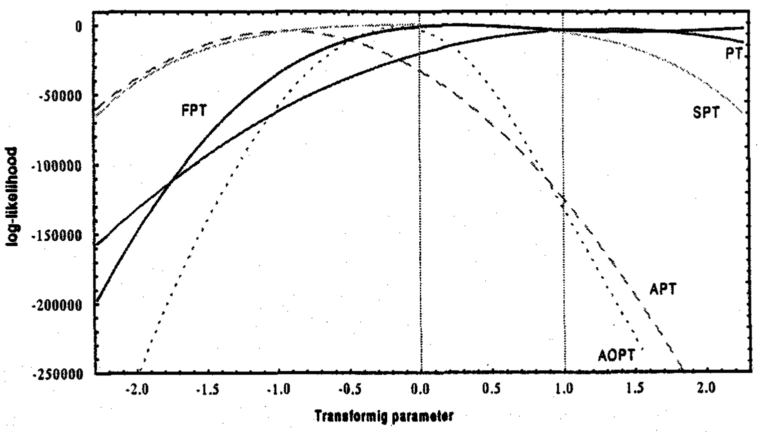

Figure 2.4.9 Example 13C(1981): The profile of maximized log-likelihood as a function of transforming parameter λ

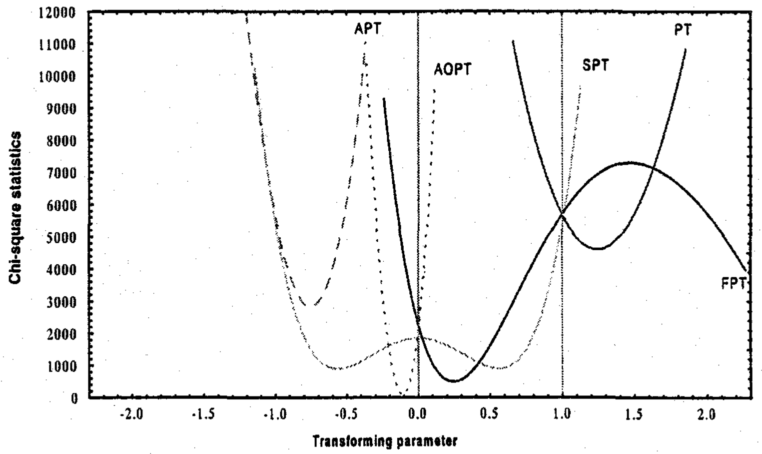

Figure 2.4.10 Example 13C(1981): The profile of the value of chi-square statistics as a function of transforming parameter λ

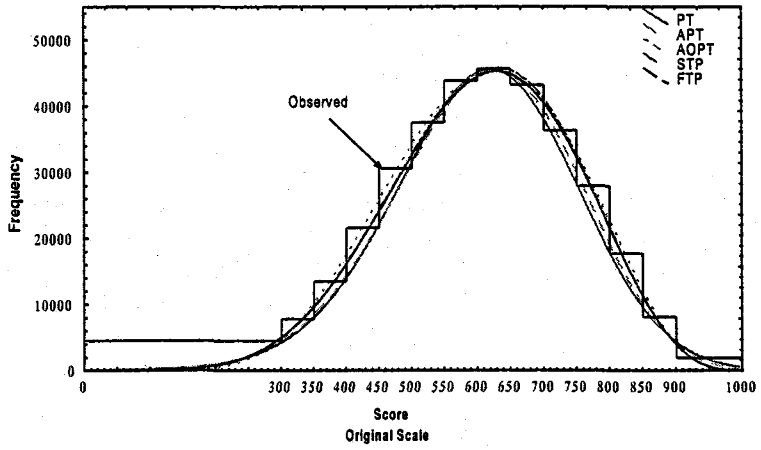

Figure 2.4.11 Example 13C(1981): The distributions of observations on the original scale

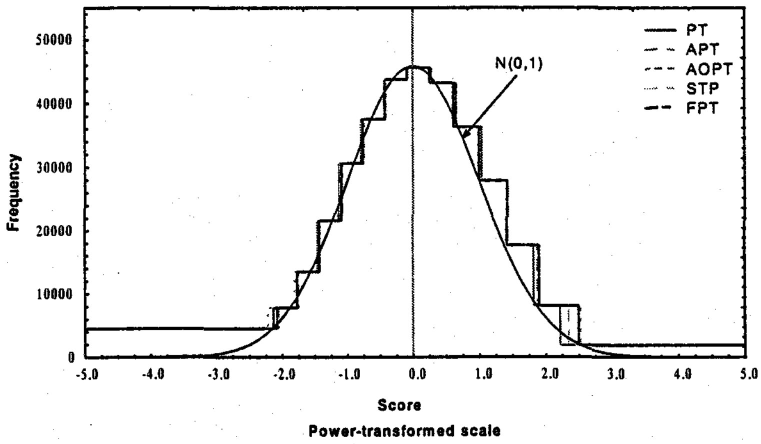

Figure 2.4.12 Example 13C(1981): The distributions of observations on the power-transformed scale

The distributions of observations on the original scale for 13A, 13B and 13C are shown in Figure 2.4.3, 2.4.7 and 2.4.11 respectively. The distributions of observations on the power-transformed scale are shown in Figure 2.4.4, 2.4.8 and 2.4.12 respectively. The values of chi-square statistics for AOPT in all three observations were smallest among five transformations. The values of the skewness and the kurtosis for the observations on the power-transformed scale show that the observations transformed by AOPT satisfy more normality than the remaining four transformations.

2.4.3 Application to Discrete Observations

The contents developed in previous sections can be applied directly to discrete observations such as count or the number of events. Let Y be a discrete random variable with the probability $p_i = \Pr\{Y = i\}$ at i for $0 < i < N$. Thus, the probability p_i can be approximate to the probability based on the PND, $p_{\mathrm{PND}i}$ which is given by

$$p_{\mathrm{PN}i}(\boldsymbol{\theta}) = \frac{1}{A(\kappa)}\left\{\Phi\left(\frac{(i+\Delta)^{(\lambda)}-\mu}{\sigma}\right) - \Phi\left(\frac{(i-\Delta)^{(\lambda)}-\mu}{\sigma}\right)\right\} \tag{2.4.17}$$

where Δ is the constant of correction for continuity with range $0 < \Delta < 1$. As the value of Δ, $\Delta = 0.5$ is often used. And then, for n trials, if $p_i = 0$ as $i > n$

$$p_{\mathrm{PN}0}(\boldsymbol{\theta}) = \frac{1}{A(\kappa)}\Phi\left(\frac{(1-\Delta)^{(\lambda)}-\mu}{\sigma}\right),$$

$$p_{\mathrm{PN}n}(\boldsymbol{\theta}) = \frac{1}{A(\kappa)}\left\{1 - \Phi\left(\frac{(n-1+\Delta)^{(\lambda)}-\mu}{\sigma}\right)\right\}. \tag{2.4.18}$$

Therefore, maximum likelihood estimates of λ, μ and σ can be obtained by maximizing the log-likelihood

$$l_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} n_i \log p_{\mathrm{PN}i}(\boldsymbol{\theta}) - n\log 2\Delta. \tag{2.4.19}$$

As an example of the above procedure, let us consider the number of major labour strikes in the U.K., 1948 − 1959, commencing in each week cited from Stuart and Ord (1986).

Example 14 (Stuart and Ord, 1986): The data are the number of major labour strikes in the U.K., 1948 − 1959, commencing in each week, which Stuart and Ord (1986) give as an example to describe the Poisson distribution. The results of fitting the PND to these observations are shown in Table 2.4.6.

The value of transforming parameter λ was estimated as 0.7307 with the approximate 95% confidence interval (0.5036, 0.9508). This optimized value was very close to $1/2$ which is theoretical value to Poisson distribution to achieve constant variance, but the 95% conficence interval did not include $1/2$. There was then no disagreement between the results obtained in this section and the theoretical considerations in Hamasaki and Goto (1998b), though the likelihood ratio test does not provide a convenience value of λ as $\hat{\lambda} = 0.5$ (p-value was 0.0499). For the optimized value, we obtained $\hat{\mu}(\hat{\lambda}) = -0.2893$ and $\hat{\sigma}(\hat{\lambda}) = 1.0290$. The value of back-transformed $\hat{\mu}(\hat{\lambda})$ to the original scale, $\hat{\mu}^*(\hat{\lambda})$ was given as 0.7225. The back-transformed value $\hat{\mu}^*(\hat{\lambda})$ was smaller than the corresponding one on the original scale, $\hat{\mu}(1)$=0.7509. The plot of the profile of maximized log-likelihood as a function of λ is shown in Figure 2.4.14, together with the value of chi-square statistics. Figure 2.4.14 shows that the value of chi-square statistics have a minimum at the neighborhood of $\hat{\lambda} = 0.7307$.

Table 2.4.6 Example 14: The results of fitting of the PND

Maximum log-likelihood		**Performance**	
Maximum log-likelihood $l(\hat{\theta}(\hat{\lambda}))$	-10.2939	**Goodness of fit**	
Truncated probability $1-A(\kappa)$	0.8529	Chi-square statistics	0.4435
Estimate of λ		p-value	0.9311
$\hat{\lambda}$	0.7307	**Shape of original observations**	
95% confidence interval for $\hat{\lambda}$	0.5036	Skewness	0.8254
	0.9508	Kurtosis	2.0737
Estimate of μ		**Shape of power-transformed observations**	
$\hat{\mu}(\hat{\lambda})$	-0.2893	Skewness	0.3587
Back to the original scale $\hat{\mu}^*(\hat{\lambda})$	0.7225	Kurtosis	2.0059
$\hat{\mu}(1)$	0.7509		
Estimate of σ			
$\hat{\sigma}(\hat{\lambda})$	1.0290		
$\hat{\sigma}(1)$	1.1077		

The distributions of observations on the original and the power-transformed scale are shown in Figure 2.4.14 and 2.4.15 respectively. The value of chi-square statistics for goodness of fit was given as 0.4435, which was quite significant at 5% level. The value of skewness for the observations on the power-transformed scale was given as 0.3587, which was not so much far from zero as the value of 0.8254 for the observations on the original scale. But the value of the kurtosis for the observations on the power-transformed scale was given as 2.0059, which was far from 3 as same as the value of 2.0737 for the observations on the original scale. This shows that the observations on the power-transformed scale achieve only symmetry of distribution in comparison with the observations on the original scale.

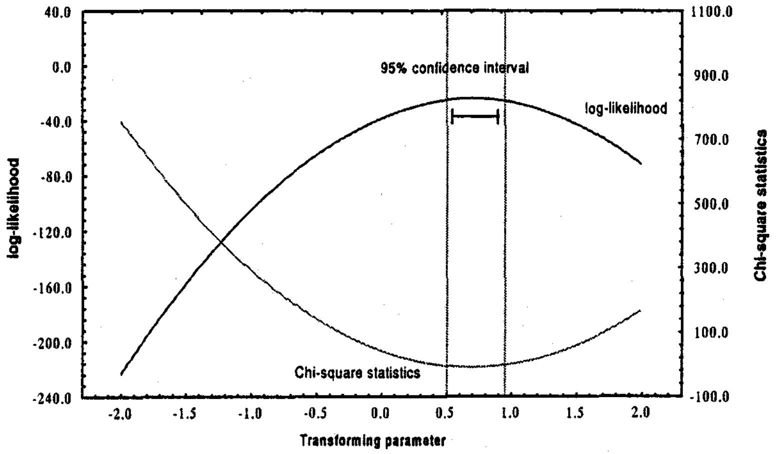

Figure 2.4.14 Example 14: The profile of maximized log-likelihood and the value of chi-square statistics as a function of transforming parameter λ

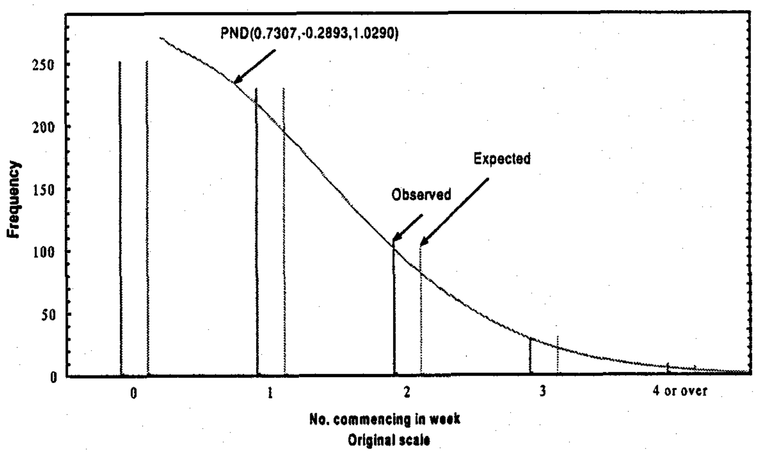

Figure 2.4.15 Example 14: The distribution of observations on the original scale

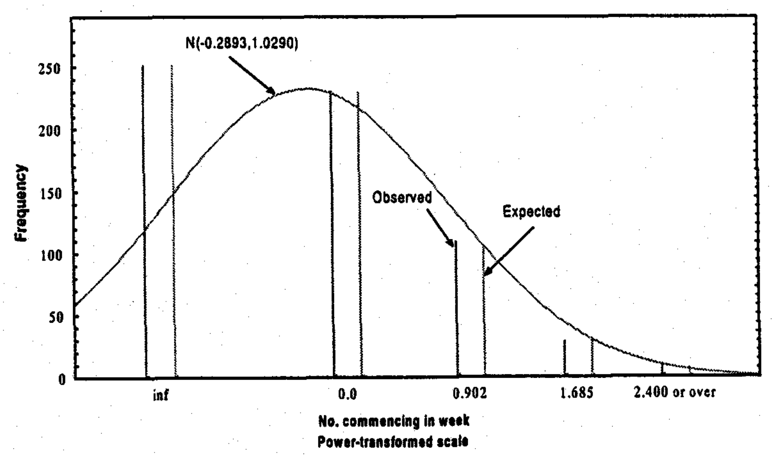

Figure 2.4.16 Example 14: The distribution of observations on the power-transformed scale

2.4.4 A Multi-sample Problem

In this section, we consider a multi-sample problem on the PND in which observations are grouped. Thus, the procedure developed by Goto *et al.* (1984) can be applied directly to this case.

Suppose we have p random samples of size $n_j(j=1,\cdots,p)$. Let us denote them by $\{Y_{ij}, i=1,\cdots,n_j; j=1,\cdots,p\}$. We consider a sequence of hypotheses

$$\mathrm{H}_{\mathrm{PN}}: Y_{ij} \sim g(y;\lambda_{,j},\mu_j,\sigma_j^2),$$

$$\mathrm{H}_U: \mathrm{H}_{\mathrm{PN}} \text{ and } \lambda_1=\lambda_2=\cdots=\lambda_p=\lambda,$$

$$\mathrm{H}_S: \mathrm{H}_U \text{ and } \sigma_1^2=\sigma_2^2=\cdots=\sigma_p^2=\sigma,$$

$$\mathrm{H}_{Li}: \mathrm{H}_S \text{ and } C_{i\mu}=0, i=1,\cdots,J$$

where $\boldsymbol{\mu}=(\mu_1,\mu_2,\cdots,\mu_p)^{\mathrm{T}}$ and C_i is $r_i \times p$ matrix which consists of the first r_i row of C_J, and C_J is a $r_J \times p$ full rank matrix. We assume $r_1 < r_2 < \cdots < r_J$.

Thus we have a sequence of hypotheses $\mathrm{H}_{\mathrm{PN}} \supset \mathrm{H}_U \supset \mathrm{H}_S \supset \mathrm{H}_{L_1} \supset \mathrm{H}_{L_1} \supset \cdots \supset \mathrm{H}_{L_J}$. Let us denote the maximum value of log-likelihood function under H_* by $l(\mathrm{H}_*)$ and the number of parameters to be estimated by $p(\mathrm{H}_*)$. AIC for model selection

$$AIC = -2l(\mathrm{H}_*)+2p(\mathrm{H}_*) \tag{2.4.19}$$

can be applied to choose the best model in the sequence of hypotheses.

Other hypotheses for the transforming parameter of the scale parameters can be incorporated into sequence.

When we fit the PND to the observations, we expect H_U to hold, and this implies H_{PN} holds. Description of observations by a model is to re-express the variation among the observations by signals through filter of the model.

As an example of our procedure, let us consider the body weight of male and female students who entered Saitama University in fiscal 1982 cited from Domae and Miyahara (1984).

Example 15 (Domae and Miyahara, 1984): The data are the body weight of 1,084 students who entered Saitama University in fiscal 1982 cited from Domae and Miyahara (1984). They consider whether the cube root of observations of body weight of 753 male and 355 female students who entered Saitama University in fiscal 1982 have a normal distribution or not. The results of fitting the PND to these grouped observations are shown in Table 2.4.7. The value of transforming parameter λ for male was estimated as -0.8013 with the approximate 95% confidence interval (-1.3345, -0.2790), suggesting that these observations have an L-shaped distribution. On the other hand, The value of transforming parameter λ for female was estimated as -0.8528 with the approximate 95% confidence interval (-1.6149, -0.1234), suggesting that these observations have an L-shaped distribution. The likelihood ratio tests provides a convenience value of λ as $\hat{\lambda} = -1.0$ for both observations (p-value was 0.4639 and 0.7208 respectively). These results show that both observations of male and female have the same shaped distribution.

Table 2.4.7 Example 15: The results of fitting of the PND

	Male	Female
Maximum log-likelihood		
Maximum log-likelihood $l(\hat{\theta}(\hat{\lambda}))$	-28.8719	-20.3891
Truncated probability $1-A(\kappa)$	1.0000	1.0000
Estimate of λ		
$\hat{\lambda}$	-0.8013	-0.8528
95% confidence interval for $\hat{\lambda}$	-1.3345 -0.2790	-1.6149 -0.1234
Estimate of μ		
$\hat{\mu}(\hat{\lambda})$	1.2015	1.1316
Back to the original scale $\hat{\mu}^*(\hat{\lambda})$	60.6604	50.9519
$\hat{\mu}(1)$	61.3929	51.6225
Estimate of σ		
$\hat{\sigma}(\hat{\lambda})$	0.0043	0.0042
$\hat{\sigma}(1)$	7.2563	6.2896
Performance		
Goodness of fit		
Chi-square statistics	10.7618	6.3100
p-value	0.7693	0.9741
Shape of original observations		
Skewness	0.5610	0.6346
Kurtosis	3.4996	3.8246
Shape of power-transformed observations		
Skewness	-0.0878	-0.1101
Kurtosis	2.9810	3.1465

Furthermore, in order to look at whether they have the same distribution, we consider them on the sequence of hypotheses. The results of fitting of the sequences of hypotheses are shown in Table 2.4.8. These results show that the value of AIC for hypotheses H_U was minimum, and then this hypotheses in which each of male and female has a power-normal distribution was adopted on AIC. The estimate of the transforming parameter for this hypotheses was -0.8181, suggesting that these observations have an

L-shaped distribution. Thus, the same result was also obtained by likelihood rate test. The distributions of observations for male and female on the original and the power-transformed scale are shown in Figure 2.4.17 and 2.4.18.

Table 2.4.8 Example 15: The results of fitting of the sequences of hypotheses

Hypotheses	$\hat{\lambda}$	Maximum log-likelihood	AIC	likelihood rate test	
				Chi-square	p -value
H_{PN}		−49.2160	110.5520		
H_U	−0.8181	−49.2672	108.5344	0.0124	0.9113
H_S	−0.4294	−53.1050	112.2100	7.6756	0.0056
H_L	−0.4019	−307.7293	619.4586	509.2486	near 0

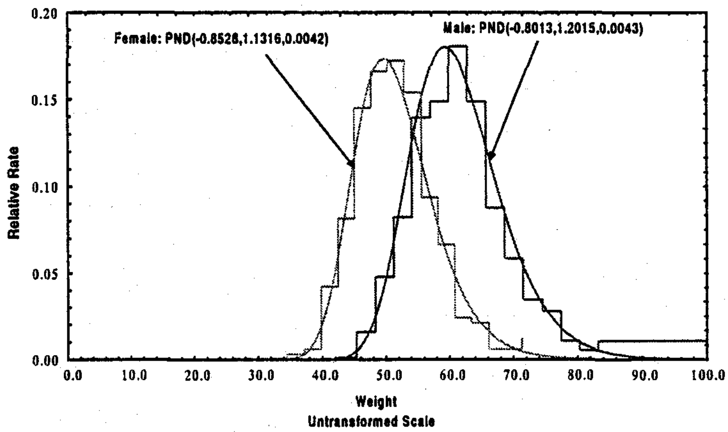

Figure 2.4.17 Example 15: The distribution of observations on the original scale

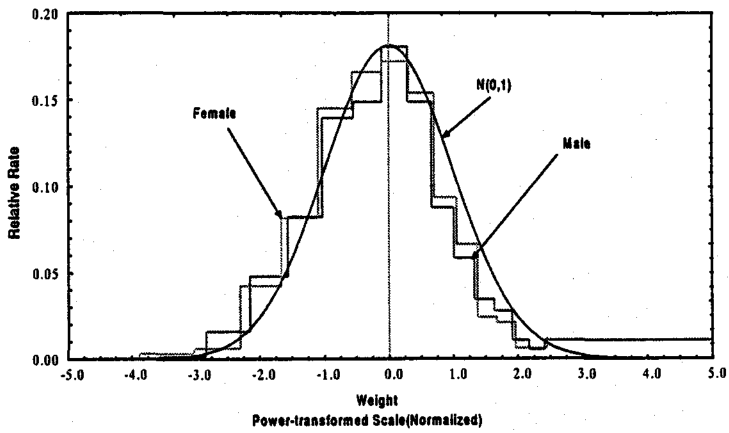

Figure 2.4.18 Example 15: The distribution of observations on the power-transformed scale

3
Bivariate Grouped Observations

3.1 Bivariate Power-normal Distribution

In this section, the relationship between a set of two positive variables (Y_1, Y_2) is considered. Suppose that (Y_1, Y_2) have the bivariate power-normal distribution (BPND), and then, through transforming parameters $\lambda = (\lambda_1, \lambda_2)$, the power-transformed variables $(Y_1^{(\lambda_1)}, Y_2^{(\lambda_2)})$ of (Y_1, Y_2) have nearly a bivariate normal distribution. Here, the BPND is an extension of the PND to two-dimensional case, which probability density function is given by

$$g(y_1, y_2) = \frac{y_1^{\lambda_1 - 1} y_2^{\lambda_2 - 1}}{A(\lambda, \mu, \Sigma)} f(y_1^{(\lambda_1)}, y_2^{(\lambda_2)}) \tag{3.1.1}$$

where

$$f(y_1^{(\lambda_1)}, y_2^{(\lambda_2)}) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2} Q(y_1^{(\lambda_1)}, y_2^{(\lambda_2)})\right\}, \tag{3.1.2}$$

$$Q(y_1^{(\lambda_1)}, y_2^{(\lambda_2)}) = \frac{1}{1-\rho^2} \times \left\{\left(\frac{y_1^{(\lambda_1)} - \mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{y_1^{(\lambda_1)} - \mu_1}{\sigma_1}\right)\left(\frac{y_2^{(\lambda_2)} - \mu_2}{\sigma_2}\right) + \left(\frac{y_2^{(\lambda_2)} - \mu_2}{\sigma_2}\right)^2\right\} \tag{3.1.3}$$

(Goto *et al.*, 1980). $\boldsymbol{\mu}$ and Σ are vector of mean parameter and variance-covariance matrix, in which $(Y_1^{(\lambda_1)}, Y_2^{(\lambda_2)})$ have nearly a bivariate normal distribution, given by

$$\boldsymbol{\mu} = (\mu_1, \mu_2)^{\mathrm{T}}, \tag{3.1.4}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \tag{3.1.5}$$

respectively. And, the probability proportional constant term of the BPND, $A(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is presented by

$$A(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int_{a_2}^{b_2}\int_{a_1}^{b_1} \phi(u_1, u_2 : \rho) du_1 du_2 \tag{3.1.6}$$

in term of joint probability density function of a bivariate standard normal distribution*

$$\phi_2(u_1, u_2 : \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{u_1^2 - 2\rho u_1 u_2 + u_2^2}{2(1-\rho^2)}\right\}. \tag{3.1.7}$$

Table 3.1.1 The values of a_1, a_2, b_1 and b_2

λ_1	λ_2	a_1	b_1	a_2	b_2
	$\lambda_2 < 0$			$-\infty$	$-\kappa_2$
$\lambda_1 < 0$	$\lambda_2 = 0$	$-\infty$	$-\kappa_1$	$-\infty$	∞
	$\lambda_2 > 0$			$-\kappa_2$	∞
	$\lambda_2 < 0$			$-\infty$	$-\kappa_2$
$\lambda_1 = 0$	$\lambda_2 = 0$	$-\infty$	∞	$-\infty$	∞
	$\lambda_2 > 0$			$-\kappa_2$	∞
	$\lambda_2 < 0$			$-\infty$	$-\kappa_2$
$\lambda_1 > 0$	$\lambda_2 = 0$	$-\kappa_1$	∞	$-\infty$	∞
	$\lambda_2 > 0$			$-\kappa_2$	∞

Then, if truncation point of Y_1 and Y_2 are $\kappa_1 = (\lambda_1\mu_1 + 1)/\lambda_1\sigma_1$ and $\kappa_2 = (\lambda_2\mu_2 + 1)/\lambda_2\sigma_2$ respectively, a_1, a_2, b_1 and b_2 are given as Table 3.1.1 respectively. Actual magnitude of $A(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ can be evaluated in term of the both distribution function $\Phi_2(\cdot,\cdot : \rho)$ and $\Phi(\cdot)$ of a bivariate standard normal distribution and univariate standard normal distribution. Namely, for $\lambda_1 < 0$, $A(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is denoted by

$$A(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \begin{cases} \Phi_2(-\kappa_1, -\kappa_2 : \rho), & \lambda_2 < 0, \\ \Phi(-\kappa_1), & \lambda_2 = 0, \\ \Phi(-\kappa_1) - \Phi_2(-\kappa_1, -\kappa_2 : \rho), & \lambda_2 > 0, \end{cases} \tag{3.1.8}$$

*2 Though an ordinary bivariate standard normal distribution is denoted by $\mathrm{N}_2(\mathbf{0}, \mathbf{I}_2)$, in this paper denoted by distribution $\mathrm{N}_2(\mathbf{0}, \mathbf{I}_2^*)$ with variance-covariance matrix $\mathbf{I}_2^* = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$.

for $\lambda_1 = 0$

$$A(\boldsymbol{\lambda}, \boldsymbol{\mu}, \Sigma) = \begin{cases} 1-\Phi_2(\kappa_2), & \lambda_2 < 0, \\ 1, & \lambda_2 = 0, \\ \Phi_2(\kappa_2), & \lambda_2 > 0, \end{cases} \tag{3.1.9}$$

and for $\lambda_1 > 0$

$$A(\boldsymbol{\lambda}, \boldsymbol{\mu}, \Sigma) = \begin{cases} \Phi(\kappa_1) - \Phi_2(\kappa_1, \kappa_2 : \rho), \lambda_2 < 0, \\ \Phi(\kappa_1),\ \lambda_2 = 0, \\ \Phi_2(\kappa_1, \kappa_2 : \rho),\ \lambda_2 > 0. \end{cases} \tag{3.1.10}$$

See Goto *et al.* (1980, 1981a, 1981b, 1982), Kawai *et al.* (1996) and Jimura *et al.* (1996) for detailed discussions about the properties of the BPND.

Therefore, setting $\boldsymbol{\theta}^{\mathrm{T}} = (\lambda_1, \lambda_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, the log-likelihood function for the sample of size n is given by

$$\begin{aligned} l_n(\boldsymbol{\theta}) &= -n\log 2\pi - \frac{n}{2}\left\{\log\sigma_1^2 + \log\sigma_2^2 + \log(1-\rho^2)\right\} \\ &- \frac{1}{2(1-\rho^2)}\sum_{l=1}^{n}\left\{\left(\frac{y_{1l}^{(\lambda_1)}-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{y_{1l}^{(\lambda_1)}-\mu_1}{\sigma_1}\right)\left(\frac{y_{2l}^{(\lambda_2)}-\mu_2}{\sigma_2}\right) + \left(\frac{y_{2l}^{(\lambda_2)}-\mu_2}{\sigma_2}\right)^2\right\} \\ &+ (\lambda_1 - 1)\sum_{l}^{n}\log y_{1l} + (\lambda_2 - 1)\sum_{l=1}^{n}\log y_{2l} - n\log A(\boldsymbol{\lambda}, \boldsymbol{\mu}, \Sigma). \end{aligned} \tag{3.1.11}$$

The maximum likelihood estimates of λ_1, λ_2, μ_1, μ_2, σ_1^2, σ_2^2 and ρ are obtained by maximizing this log-likelihood function (3.1.11) over each of these parameters. Thus, assuming $A(\lambda, \mu, \Sigma) = 1$, for fixed λ_1 and λ_2, the maximum likelihood estimates of μ_1, μ_2, σ_1^2, σ_2^2 and ρ are given by

$$\begin{aligned} &\hat{\mu}_{1n}(\lambda_1) = \sum_{l=1}^{n}\frac{y_{1l}^{(\lambda_1)}}{n}, \quad \hat{\mu}_{2n}(\lambda_2) = \sum_{l=1}^{n}\frac{y_{2l}^{(\lambda_2)}}{n}, \\ &\hat{\sigma}_{1n}^2(\lambda_1) = \sum_{l=1}^{n}\frac{\left\{y_{1l}^{(\lambda_1)} - \hat{\mu}_{1n}(\lambda)\right\}^2}{n}, \quad \hat{\sigma}_{2n}^2(\lambda_2) = \sum_{l=1}^{n}\frac{\left\{y_{2l}^{(\lambda_2)} - \hat{\mu}_{2n}(\lambda)\right\}^2}{n}, \\ &\hat{\rho}_n(\lambda_1, \lambda_2) = \sum_{l=1}^{n}\frac{1}{n}\left(\frac{y_{1l}^{(\lambda_1)} - \hat{\mu}_{1n}(\lambda_1)}{\hat{\sigma}_{1n}(\lambda_1)}\right)\left(\frac{y_{2l}^{(\lambda_2)} - \hat{\mu}_{2n}(\lambda_2)}{\hat{\sigma}_{2n}(\lambda_2)}\right) \end{aligned} \tag{3.1.12}$$

respectively. Substitution of the these maximum likelihood estimates into the log-likelihood function given by (3.1.11) yields, apart from constant

$$l_n(\boldsymbol{\theta}) = -\frac{n}{2}\left\{\log\hat{\sigma}^2_{1n}(\lambda_1) + \log\hat{\sigma}^2_{2n}(\lambda_2) + \log(1-\rho^2_n(\lambda))\right\}$$
$$+(\lambda_1 - 1)\sum_{l}^{n}\log y_{1l} + (\lambda_2 - 1)\sum_{l=1}^{n}\log y_{2l}. \tag{3.1.13}$$

The maximum likelihood estimates $\hat{\lambda}_{1n}$ and $\hat{\lambda}_{2n}$ are the values of the transforming parameter λ_1 and λ_2 for which the maximized log-likelihood is a maximum. Furthermore, substitution of $\hat{\lambda}_{1n}$ and $\hat{\lambda}_{2n}$ into $\hat{\mu}_{1n}(\lambda)$, $\hat{\mu}_{2n}(\lambda)$, $\hat{\sigma}^2_{1n}(\lambda)$, $\hat{\sigma}^2_{2n}(\lambda)$ and $\hat{\rho}_n(\lambda_1, \lambda_2)$ yields the maximum likelihood estimates $\hat{\mu}_{1n}(\hat{\lambda}_{1n})$, $\hat{\mu}_{2n}(\hat{\lambda}_{2n})$, $\hat{\sigma}^2_{1n}(\hat{\lambda}_{1n})$, $\hat{\sigma}^2_{2n}(\hat{\lambda}_{2n})$ and $\hat{\rho}_n(\hat{\lambda}_{1n}, \hat{\lambda}_{2n})$ respectively.

3.2 Completely Grouped Observations

3.2.1 Fitting the BPND to Grouped Observations

In this section, the procedure of fitting the BPND to two variables which observations are grouped into several intervals is developed.

Let the observations be grouped into $k \times l$ cells, specified beforehand and determined by $k(k \geq 3)$ interval with $I_{11} = [y_{10}, y_{11})$, $I_{12} = [y_{11}, y_{12}), \cdots, I_{1k} = [y_{1k-1}, y_k)$ on Y_1, and by $l(l \geq 3)$ interval with $I_{21} = [y_{20}, y_{21}), I_{22} = [y_{21}, y_{22}), \cdots, I_{2l} = [y_{2l-1}, y_{2l})$ on Y_2, where $0 = y_{10} < y_{11} < \cdots < y_{1k-1} < y_{1k} = \infty$ and $0 = y_{20} < y_{21} < \cdots < y_{2l-1} < y_{2l} = \infty$. Then, the frequency of the observations lying within the ij th cell is denoted by n_{ij}, and is given by the two-way frequency table shown as Table 3.2.1. In general, it is called a correlation table (Kendall and Buckland, 1982: Kitagawa and Inaba, 1979).

Table 3.2.1 Correlation table for two set of observations Y_1 and Y_2

Y_2 \ Y_1		I_{11}	I_{12}	$\cdots$	I_{1i}	$\cdots$	I_{1k}	Total
		$[y_{10}, y_{11})$	$[y_{11}, y_{12})$	$\cdots$	$[y_{1i-1}, y_{1i})$	$\cdots$	$[y_{1k-1}, y_{1k})$	
I_{21}	$[y_{20}, y_{21})$	n_{11}	n_{21}	$\cdots$	n_{i1}	$\cdots$	n_{k1}	$n_{\cdot 1}$
I_{22}	$[y_{21}, y_{22})$	n_{12}	n_{22}	$\cdots$	n_{i2}	$\cdots$	n_{k2}	$n_{\cdot 2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
I_{2j}	$[y_{2j-1}, y_{2j})$	n_{1j}	n_{2j}	$\cdots$	n_{ij}	$\cdots$	n_{kj}	$n_{\cdot j}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
I_{2l}	$[y_{2l-1}, y_{2l})$	n_{1l}	n_{2l}	$\cdots$	n_{il}	$\cdots$	n_{kl}	$n_{\cdot l}$
Total		$n_{1\cdot}$	$n_{2\cdot}$	$\cdots$	$n_{i\cdot}$	$\cdots$	$n_{k\cdot}$	n

Therefore, setting $\boldsymbol{\theta}^{\mathrm{T}} = (\lambda_1, \lambda_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, for grouped observations shown as the correlation table in Table 3.2.1, the likelihood function for the sample of size $n = \sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij}$ can be written as

$$L_n(\boldsymbol{\theta}) = \frac{n!}{\prod_{i=1}^{k} \prod_{j=1}^{l} n_{ij}!} \prod_{i=1}^{k} \prod_{j=1}^{l} p_{\mathrm{BPN}ij}^{n_{ij}}(\boldsymbol{\theta}) \tag{3.2.1}$$

where $p_{\mathrm{BPN}ij}(\boldsymbol{\theta})$ is the probability based on the BPND given by

$$p_{\mathrm{BPN}ij}(\boldsymbol{\theta}) = \frac{1}{A(\boldsymbol{\lambda}, \boldsymbol{\mu}, \Sigma)} \times \left\{ \Phi_2(z_{1,i}, z_{2,j}) + \Phi_2(z_{1,i-1}, z_{2,j-1}) - \Phi_2(z_{1,i}, z_{2,j-1}) - \Phi_2(z_{1,i-1}, z_{2,j}) \right\} \tag{3.2.2}$$

where $z_{1,i} = (z_{1,i}^{(\lambda_1)} - \mu_1)/\sigma_1$ and $z_{2,j} = (z_{2,j}^{(\lambda_2)} - \mu_2)/\sigma_2$. Then, the log-likelihood function become

$$l_n(\boldsymbol{\theta}) = \log L_n(\boldsymbol{\theta}) = \log n! - \sum_{i=1}^{k} \sum_{j=1}^{l} \log n_{ij} + \sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij} \log p_{\mathrm{BPN}ij}(\boldsymbol{\theta}) \tag{3.2.3}$$

The maximum likelihood estimates $\hat{\lambda}_1$, $\hat{\lambda}_2$, $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\sigma}_1$, $\hat{\sigma}_2$ and $\hat{\rho}$ of λ_1, λ_2, μ_1, μ_2, σ_1, σ_2 and ρ can be obtained by maximize the log-likelihood function (3.2.1). The asymptotic properties of the maximum likelihood estimates $\hat{\lambda}_1$, $\hat{\lambda}_2$, $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\sigma}_1$, $\hat{\sigma}_2$ and $\hat{\rho}$ in this situation are given in Section 3.3.

3.2.2 Properties of Parameter Estimates

In this section, the asymptotic properties of the maximum likelihood estimates $\hat{\lambda}_1$, $\hat{\lambda}_2$, $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\sigma}_1$, $\hat{\sigma}_2$ and $\hat{\rho}$ of λ_1, λ_2, μ_1, μ_2, σ_1, σ_2 and ρ are considered by using the analogy of the procedure in the Section 2.2.2.

Let $\boldsymbol{\theta}^{\mathrm{T}} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7) = (\lambda_1, \lambda_2, \mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$, $\hat{p}_{ij,n} = n_{ij}/n$, and

$$p_{ij} = \int_{y_{1i-1}}^{y_{1i}} \int_{y_{2j-1}}^{y_{2j}} g^*(y_1, y_2) dy_1 dy_2$$

where g^* is the true probability density function. Suppose that (i) the parameter space Θ is a compact set given by $\Theta = \{\boldsymbol{\theta}^{\mathrm{T}} = (\lambda_1, \lambda_2, \mu_1, \mu_2, \sigma_1, \sigma_2, \rho)\}$, and (ii)

$$H(\boldsymbol{\theta}) = \sum_{i=1}^{k} p_{ij} \log p_{\mathrm{BPN}ij}(\boldsymbol{\theta})/p_{ij}$$

which has a unique global maximum at $\boldsymbol{\theta}_0 \in \Theta$ is a continuos function. Also, suppose that (iii) $\boldsymbol{\theta}_0$ be an interior point of Θ, and (iv) the Hessian $\nabla^2 H(\boldsymbol{\theta}_0)$ of $H(\boldsymbol{\theta})$ be nonsingular at $\boldsymbol{\theta}_0$.

Stirling's approximation yields

$$\frac{1}{n}\log n! - \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{l}\log n_{ij}! = \frac{1-kl}{2n}\log 2\pi - \sum_{i=1}^{k}\sum_{j=1}^{l}\hat{p}_{ij,n}\log\hat{p}_{ij,n} + \frac{1}{2n}\left\{(1-kl)\log n - \sum_{i=1}^{k}\sum_{j=1}^{l}\log\hat{p}_{ij,n}\right\} + O(n^{-1}). \tag{3.2.4}$$

By (2.2.1)

$$\frac{1}{n}l_n(\boldsymbol{\theta}) = \sum_{i=1}^{k}\sum_{j=1}^{l}\hat{p}_{ij,n}\log p_{\mathrm{BP}ij}(\boldsymbol{\theta}) - \sum_{i=1}^{k}\sum_{j=1}^{l}\hat{p}_{ij,n}\log\hat{p}_{ij,n} + o(1). \tag{3.2.5}$$

Therefore, the inequality

$$\left|\frac{1}{n}l_n(\theta) - \sum_{i=1}^{k}\sum_{j=1}^{l}p_{ij}\log\frac{p_{\mathrm{BPN}ij}(\boldsymbol{\theta})}{p_{ij}}\right| \le \left|\sum_{i=1}^{k}\sum_{j=1}^{l}\hat{p}_{ij,n}\log p_{\mathrm{BPN}ij}(\boldsymbol{\theta}) - \sum_{i=1}^{k}\sum_{j=1}^{l}p_{ij}\log p_{\mathrm{BPN}ij}(\boldsymbol{\theta})\right| + \left|\sum_{i=1}^{k}\sum_{j=1}^{l}\hat{p}_{ij,n}\log\hat{p}_{ij,n} - \sum_{i=1}^{k}\sum_{j=1}^{l}p_{ij}\log p_{ij}\right| + o(1) \tag{3.2.6}$$

hold. Thus, because it follows from the analogy of Appendix 1 that as $n \to \infty$, for $\boldsymbol{\theta} \in \Theta$

$$\left|\sum_{i=1}^{k}\sum_{j=1}^{l}\hat{p}_{ij,n}\log p_{\mathrm{BPN}ij}(\boldsymbol{\theta}) - \sum_{i=1}^{k}\sum_{j=1}^{l}p_{ij}\log p_{\mathrm{BPN}ij}(\boldsymbol{\theta})\right| \xrightarrow{a.s.} 0 \tag{3.2.7}$$

and continuity of $x\log x$, and $\hat{p}_{ij,n} \xrightarrow{a.s.} p_{ij}$ as $n \to \infty$, the right hand side of inequality (3.2.6) goes to zero with probability one uniformly in $\boldsymbol{\theta} \in \Theta$. Hence, as $n \to \infty$, for $\boldsymbol{\theta} \in \Theta$

$$\frac{1}{n}l_n(\boldsymbol{\theta}) \xrightarrow{a.s.} H(\boldsymbol{\theta}) \tag{3.2.8}$$

Therefore, it follows form the analogy of Appendix 2 that $\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}_0$ as $n \to \infty$ by identifying $H(\boldsymbol{\theta})$ with the function $f(\boldsymbol{\theta})$ of the Appendix.

In order to show the asymptotic normality of $\hat{\boldsymbol{\theta}}_n$, the gradient and the Hessian of log-likelihood function are considered. The gradient of $l_n(\boldsymbol{\theta})$ is given by

$$\nabla l_n(\boldsymbol{\theta}) = \left(\frac{\partial l_n(\boldsymbol{\theta})}{\partial\theta_u}\right), \quad u = 1,\cdots,7 \tag{3.2.9}$$

and the Hessian of $l_n(\boldsymbol{\theta})$ is the 7×7 symmetric matrix $\nabla^2 l_n(\boldsymbol{\theta}) = (h_{uv,n}(\boldsymbol{\theta}))$ with its elements

$$h_{uv,n}(\boldsymbol{\theta}) = \frac{\partial^2 l_n(\boldsymbol{\theta})}{\partial\theta_u\partial\theta_v}, \quad u,v = 1,\cdots,7 \tag{3.2.10}$$

It is readily seen that the second partial derivatives are continuous on Θ. Using Taylor's formula to expand $n^{-1/2}\nabla l_n(\hat{\boldsymbol{\theta}}_n)$ about $\boldsymbol{\theta}_0$

$$\frac{1}{\sqrt{n}}\nabla l_n(\hat{\boldsymbol{\theta}}_n) = \frac{1}{n}\nabla l_n(\boldsymbol{\theta}_0) + \frac{1}{n}\nabla^2 l_n(\boldsymbol{\theta}_n^*)\left\{\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right\} \tag{3.2.11}$$

is obtained, where $\boldsymbol{\theta}_n^* = \gamma_n\boldsymbol{\theta}_0 + (1-\gamma_n)\hat{\boldsymbol{\theta}}_n \, (0 < \gamma_n < 1)$.

Next, since $\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}_0$ as $n \to \infty$, with the assumption that $\boldsymbol{\theta}_0$ is an interior point of Θ, $\nabla l_n(\hat{\boldsymbol{\theta}}_n) = \mathbf{0}$ for all sufficiently large n, on an almost sure set. Therefore, as $n \to \infty$

$$\frac{1}{\sqrt{n}}\nabla l_n(\boldsymbol{\theta}_0) + \frac{1}{n}\nabla^2 l_n(\boldsymbol{\theta}_n^*)\left\{\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right\} \xrightarrow{a.s.} \mathbf{0} \tag{3.2.12}$$

so that $n^{1/2}\nabla l_n(\boldsymbol{\theta}_0)$ and $-n^{1/2}\nabla^2 l_n(\boldsymbol{\theta}_n^*)\left\{\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right\}$ have the same limiting distribution.

Since we can write

$$\frac{1}{n}\nabla l_n(\boldsymbol{\theta}_0) = \frac{1}{n}\sum_{m=1}^{n}\left[\sum_{i=1}^{k}\sum_{j=1}^{l} I_{D_{ij}}(y_m)\alpha_{ij}(\boldsymbol{\theta}_0)\right] = \frac{1}{n}\sum_{m=1}^{n} X_m(\boldsymbol{\theta}_0) \tag{3.2.13}$$

where D_{ij} is the indicator function of the set I_{ij}, that is, $D_{I_{ij}}(y_m) = 1$ if $y_m \in I_{ij}$ and $D_{I_{ij}}(y_m) =$ 0 if $y_m \notin I_{ij}$, $\boldsymbol{\alpha}_{ij}(\boldsymbol{\theta}_0) = (\alpha_{ijr}(\boldsymbol{\theta}_0))$ and

$$\alpha_{ijr}(\boldsymbol{\theta}_0) = (\alpha_{ijr}(\boldsymbol{\theta}_0)) = \left(\left.\frac{\partial \log p_{\mathrm{BP}ij}(\boldsymbol{\theta})}{\partial \theta_r}\right|_{\theta=\theta_0}\right), \quad r = 1,\cdots,7 \tag{3.2.14}$$

then the random vectors $X_1(\boldsymbol{\theta}_0),\cdots,X_n(\boldsymbol{\theta}_0)$ are identically and independently distributed with $\mathrm{E}[X_1(\boldsymbol{\theta}_0)] = \nabla H(\boldsymbol{\theta}_0) = \mathbf{0}$ and $\mathrm{E}[X_1(\boldsymbol{\theta}_0)X^{\mathrm{T}}(\boldsymbol{\theta}_0)] = \boldsymbol{W}$, where $\boldsymbol{W} = (w_{uv})_{7\times 7}$ is given by

$$\begin{aligned} w_{uv} &= \sum_{i=1}^{k}\sum_{j=1}^{l} p_{ij}\alpha_{iju}(\boldsymbol{\theta}_0)\alpha_{ijv}(\boldsymbol{\theta}_0) \\ &= \sum_{i=1}^{k}\sum_{j=1}^{l} p_{ij}\left(\left.\frac{\partial \log p_{\mathrm{BP}ij}(\boldsymbol{\theta})}{\partial \theta_u}\right|_{\theta_0}\right)\left(\left.\frac{\partial \log p_{\mathrm{BP}ij}(\boldsymbol{\theta})}{\partial \theta_v}\right|_{\theta_0}\right). \end{aligned} \tag{3.2.15}$$

Thus, an application of the multivariate central limit theorem yields that, as $n \to \infty$

$$\frac{1}{\sqrt{n}} l_n(\boldsymbol{\theta}_0) \xrightarrow{d} \mathrm{N}_7(\mathbf{0},\boldsymbol{W}). \tag{3.2.16}$$

Furthermore, from Hessian of $l_n(\boldsymbol{\theta})$, we can write

$$\frac{1}{n}h_{uv,n}(\boldsymbol{\theta}) = \sum_{i=1}^{k}\sum_{j=1}^{l}\hat{p}_{ij,n}\delta_{ij,uv}(\boldsymbol{\theta}) \tag{3.2.17}$$

where $\delta_{ij,uv}(\boldsymbol{\theta})$ is uniformly continuous for $\boldsymbol{\theta} \in \Theta$ for each u and v. So, the fact that $\hat{p}_{ij,n} \xrightarrow{a.s.} p_{ij}$ as $n \to \infty$ implies that uniformly on Θ

$$\frac{1}{n}h_{uv,n}(\boldsymbol{\theta}) \xrightarrow{a.s.} \frac{\partial^2}{\partial\theta_u\partial\theta_v}\left(\sum_{i=1}^{k}\sum_{j=1}^{l} p_{ij}\log\frac{p_{\mathrm{BPN}ij}(\boldsymbol{\theta})}{p_{ij}}\right) = \frac{\partial^2 H(\boldsymbol{\theta})}{\partial\theta_u\partial\theta_v}. \tag{3.2.18}$$

Hence, as $\boldsymbol{\theta}_n^*$ lies on the segment jointing $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$, we have as $n \to \infty$

$$\frac{1}{n} h_{uv,n}(\boldsymbol{\theta}_n^*) \xrightarrow{p} \left. \frac{\partial^2 H(\boldsymbol{\theta})}{\partial\theta_u \partial\theta_v} \right|_{\theta_0} . \tag{3.2.19}$$

Therefore, Setting $V = \{\nabla^2 H(\boldsymbol{\theta}_0)\}^{-1}$, we can apply Slutsky's theorem to finally conclude that as $n \to \infty$

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{a.s.} \mathrm{N}_7(\mathbf{0}, VWV^{\mathrm{T}}). \tag{3.2.20}$$

These results show that strong consistency and asymptotic normality of the maximum likelihood estimates $\hat{\lambda}_1$, $\hat{\lambda}_2$, $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\sigma}_1$, $\hat{\sigma}_2$ and $\hat{\rho}$ of λ_1, λ_2, μ_1, μ_2, σ_1, σ_2, and ρ.

3.2.3 Examples

To have a concrete and practical grasp of the aim in this paper, and make the impression of them clear, the performance of fitting the BPND to grouped observations are numerically examined through some examples cited from published literature. Thus, to evaluate the bivariate normality of the simultaneous power-transformed observations, we adopted the following three tests.

Test 1: test whether third and fourth cumulants equal to zero or not (Dahiya and Gurland, 1973: Takeuchi, 1974)

Test 2: assess multi-dimensional skewness and assess multi-dimensional kurtosis (Mardia, 1970)

Test 3: assess the size of sample which falls into the each of four quadrants of two-dimensional coordinates (Takeuchi, 1974: Shibata, 1981)

Example 16: Cramér (1974) considers the relationship between the age of father and mother of 475,322 boys of live born children in Norway during nineteen period 1871 − 1900, which observations are the frequencies in several groups of these two variables. The results of fitting the BPND to these observations setting age of father as Y_1 and age of mother as Y_2 are shown in Table 3.2.1. The value of transforming parameter for father was estimated as −0.0630, which was close to zero, whereas for mother, it was estimated as 0.4330. These optimized values of transforming parameters suggest that the observations of age of father and mother have a log-normal and exponential respectively. In the fitting of PND to each of the two variables, the value of transforming parameter for father was given as 0.1204, and for mother it was given as 0.3613. There was thus no disagreement between the results on the transforming parameter drawn from the fitting of the BPND and the PND. The plot of the profile of maximized log-likelihood as a function of λ_1 and λ_2 is shown in Figure 3.2.1. For the optimized values of these two transforming parameters, we obtained $\hat{\mu}_1(\hat{\lambda}_1) = 34.5489$, $\hat{\mu}_2(\hat{\lambda}_2) = 8.0022$, $\hat{\sigma}_1(\hat{\lambda}_1) = 0.1904$ and

$\hat{\sigma}_2(\hat{\lambda}_2) = 0.9445$. The values of back-transformed $\hat{\mu}_1(\hat{\lambda}_1)$ and $\hat{\mu}_2(\hat{\lambda}_2)$ to the original scale, $\hat{\mu}_1^*(\hat{\lambda}_1)$ and $\hat{\mu}_2^*(\hat{\lambda}_2)$ were given as 34.5474 and 31.6761 respectively. Both of them were smaller than the corresponding values on the original scale, $\hat{\mu}_1(1)$ =35.6991 and $\hat{\mu}_2(1)$ =32.1275. Also, the value of $\hat{\rho}(\hat{\lambda}_1, \hat{\lambda}_2)$ was given as 0.6264, which was slightly larger than the value of 0.6183 for $\hat{\rho}(1,1)$ on the original scale.

Table 3.2.1 Example 16: The results of fitting of the BPND

	Original (Untransformed)	BPND (Transformed)	PND Age of father	PND Age of mother
	Maximum likelihood and estimate			
Maximum log-likelihood	–	−1608753.9991	−2282.3949	−6816.4912
$\hat{\lambda}_1$	–	−0.0630	0.1204	–
$\hat{\lambda}_2$	–	0.4330	–	0.3613
$\hat{\mu}_1$	35.6991	3.1749	4.3941	–
$\hat{\mu}_1^*$		34.5474	34.0205	
$\hat{\mu}_2$	32.1275	8.0022	–	6.8835
$\hat{\mu}_2^*$		31.6761		31.7243
$\hat{\sigma}_1$	8.5334	0.1904	0.3310	–
$\hat{\sigma}_2$	6.7001	0.9445	–	0.7266
$\hat{\rho}$	0.6183	0.6264	–	–

	Performance			
Test 1	$\hat{\gamma}_{12}$	0.1610	$\hat{\gamma}_{12}$	0.0421
	p-value	near 0	p-value	near 0
	$\hat{\gamma}_{21}$	0.1971	$\hat{\gamma}_{21}$	0.0077
	p-value	near 0	p-value	0.0024
Test 2	Skewness	0.0002	Skewness	0.0001
	p-value	near 0	p-value	0.0504
	Kurtosis	9.3914	Kurtosis	8.1434
	p-value	near 0	p-value	near 0
Test 3	χ^2	48140.7147	χ^2	47618.3182
	p-value	near 0	p-value	near 0

The distributions of observations on the original and the power-transformed scale are shown in Figure 3.2.2 and 3.2.3 respectively. The results of these three tests in Table 3.2.1 show that the observations on the power-transformed scale satisfied the bivariate normality in comparison with the observations on the original scale.

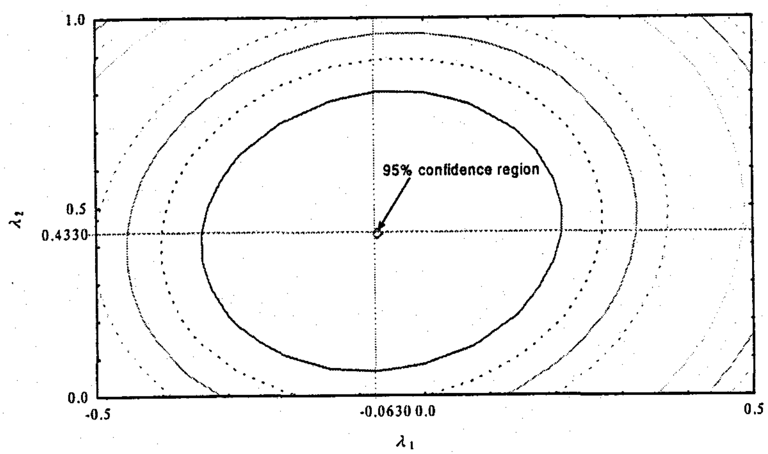

Figure 3.2.1 Example 16: The profile of maximized log-likelihood as a function of transforming parameters λ_1 *and* λ_2

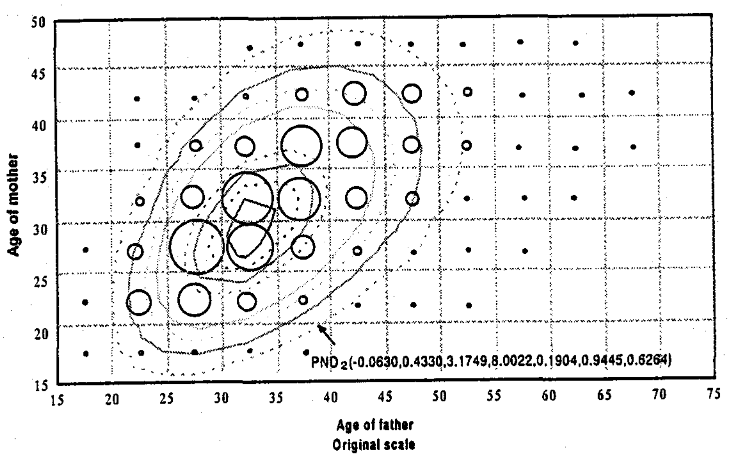

Figure7.2.2 Example 16: The distribution of observations on the original scale

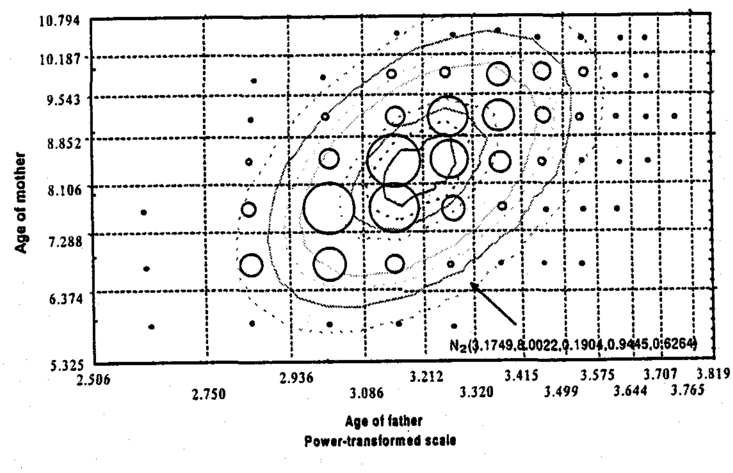

Figure 3.2.3 Example 16: The distribution of observations on the power-transformed scale

Table 3.2.2 Example 17: The results of fitting of the BPND

	Original (Untransformed)	BPND (Transformed)	PND Family income	PND Miles driven
	Maximum likelihood and estimate			
Maximum log-likelihood	–	−16621.4776	−45.1614	−51.0135
$\hat{\lambda}_1$	–	0.3600	0.3504	–
$\hat{\lambda}_2$	–	0.3550	–	0.4453
$\hat{\mu}_1$	12259.5713	76.3190	71.4135	–
$\hat{\mu}_1^*$		10969.0813	10946.5108	
$\hat{\mu}_2$	15.2430	4.2016	–	4.6688
$\hat{\mu}_2^*$		13.0865		12.4973
$\hat{\sigma}_1$	6802.4333	17.3144	17.0563	–
$\hat{\sigma}_2$	10.1955	1.7868	–	2.4593
$\hat{\rho}$	0.4009	0.4381	–	–

	Performance			
Test 1	$\hat{\gamma}_{12}$	0.1188	$\hat{\gamma}_{12}$	0.0450
	p-value	near 0	p-value	0.0430
	$\hat{\gamma}_{21}$	0.1261	$\hat{\gamma}_{21}$	0.0456
	p-value	near 0	p-value	0.0407
Test 2	Skewness	0.0097	Skewness	0.0075
	p-value	0.1640	p-value	0.2831
	Kurtosis	7.5353	Kurtosis	6.7365
	p-value	0.9999	p-value	1.0000
Test 3	χ^2	705.9977	χ^2	329.8819
	p-value	near 0	p-value	near 0

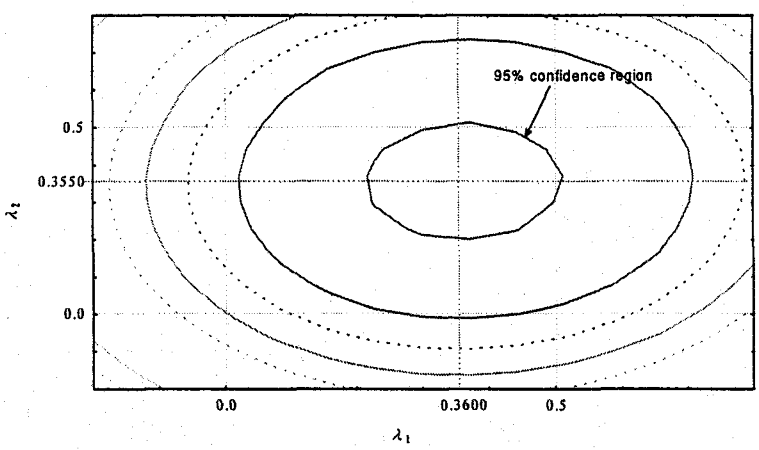

Figure 3.2.3 Example 17: The profile of maximized log-likelihood as a function of transforming parameters λ_1 and λ_2

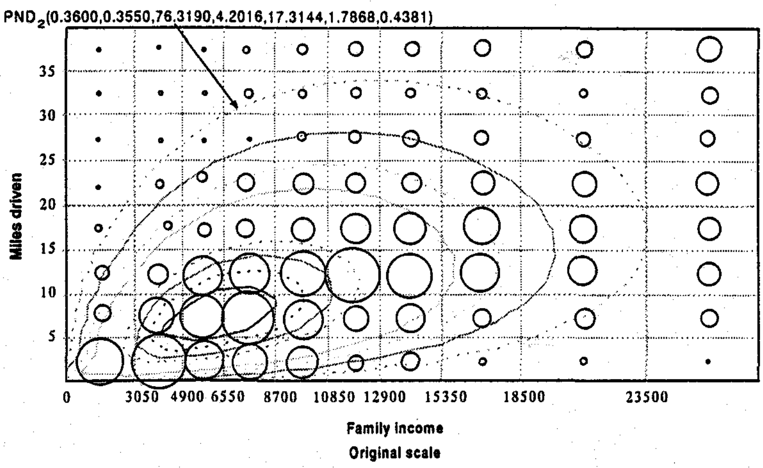

Figure 3.2.4 Example 17: The distribution of observations on the original scale

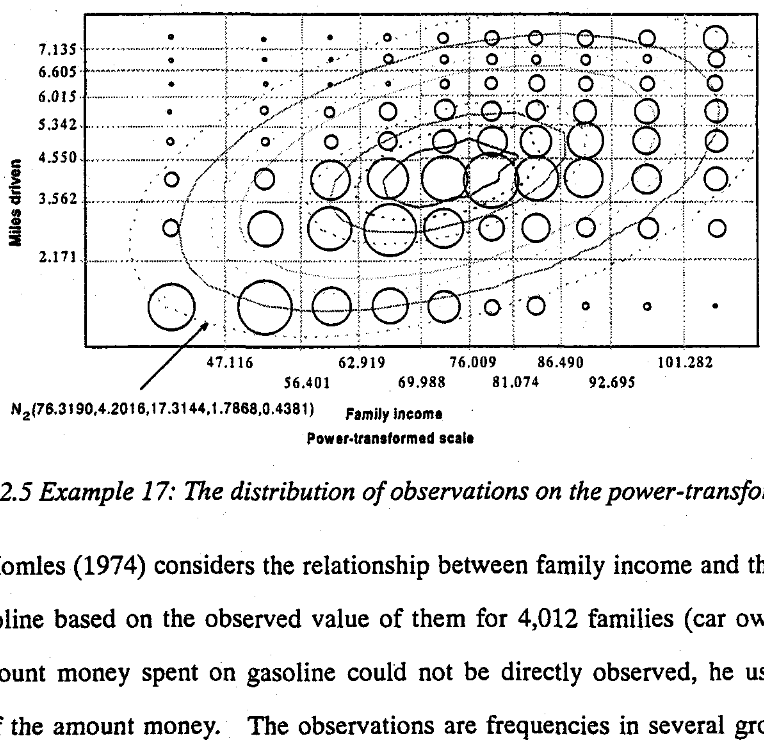

Figure 3.2.5 Example 17: The distribution of observations on the power-transformed scale

Example 17: Homles (1974) considers the relationship between family income and the amount of money spent on gasoline based on the observed value of them for 4,012 families (car owner) during 1973. Thus, as the amount money spent on gasoline could not be directly observed, he uses the total miles driven instead of the amount money. The observations are frequencies in several groups of total miles driven and family income. The results of fitting the BPND to these observations setting family income as Y_1 and total miles driven as Y_2 are shown in Table 3.2.2. The value of transforming parameter for family income was estimated as 0.3600, whereas for total miles driven, it was estimated as 0.3550. These optimized values suggest that the both observations of family income and total miles driven have an exponential distribution. In the fitting of PND to each of the two variables, the value of transforming parameter for family was given as 0.3504, and for total miles driven it was given as 0.4453. There was thus no disagreement between the results on the transforming parameter drawn from the fitting of the BPND and the PND. The plot of the profile of maximized log-likelihood as a function of λ_1 and λ_2

is shown in Figure 3.2.4. For the optimized values of these two transforming parameters, we obtained $\hat{\mu}_1(\hat{\lambda}_1) = 76.3190$, $\hat{\mu}_2(\hat{\lambda}_2) = 4.2016$, $\hat{\sigma}_1(\hat{\lambda}_1) = 17.3144$, and $\hat{\sigma}_2(\hat{\lambda}_2) = 1.7868$. The values of back-transformed $\hat{\mu}_1(\hat{\lambda}_1)$ and $\hat{\mu}_2(\hat{\lambda}_2)$ to the original scale, $\hat{\mu}_1^*(\hat{\lambda}_1)$ and $\hat{\mu}_2^*(\hat{\lambda}_2)$ were 10969.0813 and 13.0865. Both of them were smaller than the corresponding values on the original scale, $\hat{\mu}_1(1)$ =12259.5713 and $\hat{\mu}_2(1)$ =15.2430. Also, the value of $\hat{\rho}(\hat{\lambda}_1, \hat{\lambda}_2)$ was given as 0.4381, which was slightly larger than the value of 0.4009 for $\hat{\rho}(1,1)$ on the original scale.

The distributions of observations on the original and the power-transformed scale are shown in Figure 3.2.5 and 3.2.6 respectively. The results of these three tests in Table 3.2.2 show that the observations on the power-transformed scale satisfied the bivariate normality in comparison with the observations on the original scale.

Example 18: Winkelmann (1994) considers the relationship between the number of employers and the numbers of unemployment spells during the ten year period 1974−1984 for 1962 individuals, provided by German Socio-economic panel (SOEP). And using the information on the number of employers and the number of unemployment spells, Winkelmann (1994) investigates direct job change. The results of fitting the BPND to these observations setting the number of employers as Y_1 and the number of unemployment spells as Y_2 are shown in Table 3.2.3. The value of transforming parameter for the number of employers was estimated as −0.1270, whereas for the number of employment spells, it was estimated as −1.6470. These estimates suggest that the observations of employers and unemployment spells both have an L-shaped distribution. However, in the in fitting the PND to each of the two variables, the values of the transforming parameter for the number of employers and for the number of unemployment spells were estimated as 0.3011 and 0.1204 respectively. Both of them were near zero, suggesting a log-normal distribution. The plot of the profile of maximized log-likelihood as a function of λ_1 and λ_2 is shown in Figure 3.2.7. For the optimized values of these two transforming parameters, we obtained $\hat{\mu}_1(\hat{\lambda}_1) = -0.8054$, $\hat{\mu}_2(\hat{\lambda}_2) = -1.7013$, $\hat{\sigma}_1(\hat{\lambda}_1) = 0.7623$, and $\hat{\sigma}_2(\hat{\lambda}_2) = 0.7963$. The values of back-transformed $\hat{\mu}_1(\hat{\lambda}_1)$ and $\hat{\mu}_2(\hat{\lambda}_2)$ to the original scale, $\hat{\mu}_1^*(\hat{\lambda}_1)$ and $\hat{\mu}_2^*(\hat{\lambda}_2)$ were given as 0.4645 and 0.4445 respectively. The value of $\hat{\mu}_1^*(\hat{\lambda}_1)$ was smaller than the corresponding one on the original scale, $\hat{\mu}_1(1)$ =0.5395, whereas the value of $\hat{\mu}_2(\hat{\lambda}_2)$ was larger than the corresponding one on the original scale $\hat{\mu}_2(1)$ =0.3736. Also, the value of $\hat{\rho}(\hat{\lambda}_1, \hat{\lambda}_2)$ was given as 0.1096, which was slightly larger than the value of 0.0501 for $\hat{\rho}(1,1)$ on the original scale

The distributions of observations on the original and the power-transformed scale are shown in Figure 3.2.8 and 3.2.9 respectively. The results of these three tests in Table 3.2.3 show that observations on the power-transformed scale satisfied the bivariate normality in comparison with the observations on the original scale.

Table 3.2.3 Example 18: The results of fitting of the BPND

	Original (Untransformed)	BPND (Transformed)	PND Employers	PND Unemployment
	Maximum likelihood and estimate			
Maximum log-likelihood	–	−3642.3400	−24.1869	−31.2127
$\hat{\lambda}_1$	–	−0.1270	0.3011	–
$\hat{\lambda}_2$	–	−1.6470	–	0.1204
$\hat{\mu}_1$	0.5395	−0.8054	−1.2943	–
$\hat{\mu}_1^*$		0.4645	0.1940	
$\hat{\mu}_2$	0.3736	−1.7013	–	−2.0413
$\hat{\mu}_2^*$		0.4445		0.0961
$\hat{\sigma}_1$	1.0828	0.7623	1.4323	–
$\hat{\sigma}_2$	1.1049	0.7963	–	1.6545
$\hat{\rho}$	0.0501	0.1096	–	–

	Performance			
Test 1	$\hat{\gamma}_{12}$	0.1341	$\hat{\gamma}_{12}$	0.1603
	p-value	near 0	p-value	near 0
	$\hat{\gamma}_{21}$	0.3262	$\hat{\gamma}_{21}$	0.2013
	p-value	near 0	p-value	near 0
Test 2	Skewness	2.5966	Skewness	0.0516
	p-value	near 0	p-value	0.0021
	Kurtosis	70.4191	Kurtosis	11.3122
	p-value	near 0	p-value	near 0
Test 3	χ^2	2827.9281	χ^2	2765.4151
	p-value	near 0	p-value	near 0

Example 19: Mardia (1970) and Kendall and Stuart (1977) consider the relationship between length and breadth of 9,440 beans, which observations are the frequencies in several groups of these two variables. The results of fitting the BPND to these observations setting length as Y_1 and breadth as Y_2 are shown in Table 3.2.4. The value of transforming parameter for length was estimated as 4.0560, whereas for breadth, it was estimated as 2.6860. These optimized values suggest that the both observations of length and breadth have a J-shaped distribution. In the fitting of PND to each of the two variables, the value of transforming parameter for length was given as 3.0319, and for breadth, it was given as 3.1300. There was thus no disagreement between the results on the transforming parameter drawn from the fitting of the BPND and the PND. The plot of the profile of maximized log-likelihood as a function of λ_1 and λ_2

is shown in Figure 3.2.10. For the optimized values of these two transforming parameters, we obtained $\hat{\mu}_1(\hat{\lambda}_1) = 12642.3733$, $\hat{\mu}_2(\hat{\lambda}_2) = 14.4953$, $\hat{\sigma}_1(\hat{\lambda}_1) = 3014.5518$ and $\hat{\sigma}_2(\hat{\lambda}_2) = 11.4446$. The values of back-transformed $\hat{\mu}_1(\hat{\lambda}_1)$ and $\hat{\mu}_2(\hat{\lambda}_2)$ to the original scale $\hat{\mu}_1^*(\hat{\lambda}_1)$ and $\hat{\mu}_2^*(\hat{\lambda}_2)$ were given as 14.4953 and 7.9869 respectively. Both of them were slightly larger than the corresponding values on the original scale, $\hat{\mu}_1(1)$ =14.4035 and $\hat{\mu}_2(1)$ =7.9804. Also, the value $\hat{\rho}(\hat{\lambda}_1, \hat{\lambda}_2)$ was given as 0.7562, which was slightly smaller than the value of 0.7590 for $\hat{\rho}(1,1)$ on the original scale.

The distributions of observations on the original and the power-transformed scale are shown in Figure 3.2.11 and 3.2.12 respectively. The results of these three tests in Table 3.2.4 show that the observations on the power-transformed scale satisfied the bivariate normality in comparison with the observations on the original scale

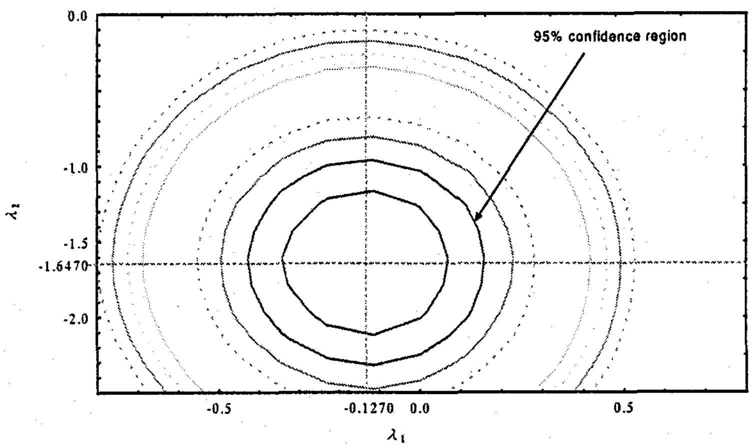

Figure 3.2.7 Example 18: The profile of maximized log-likelihood as a function of transforming parameters λ_1 and λ_2

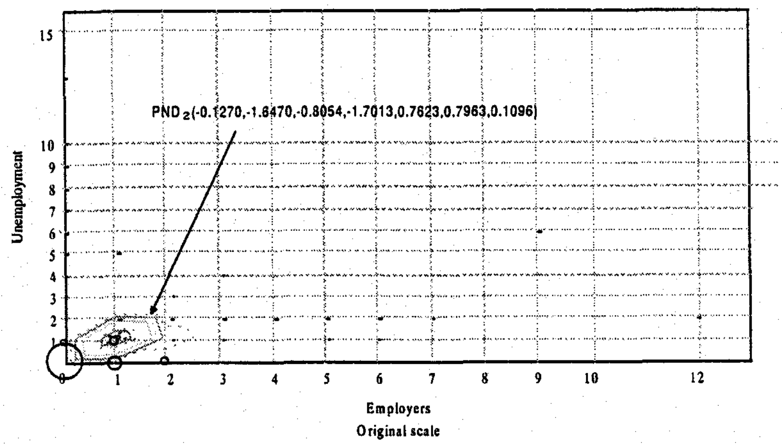

Figure 3.2.8 Example 18: The distribution of observations on the original scale

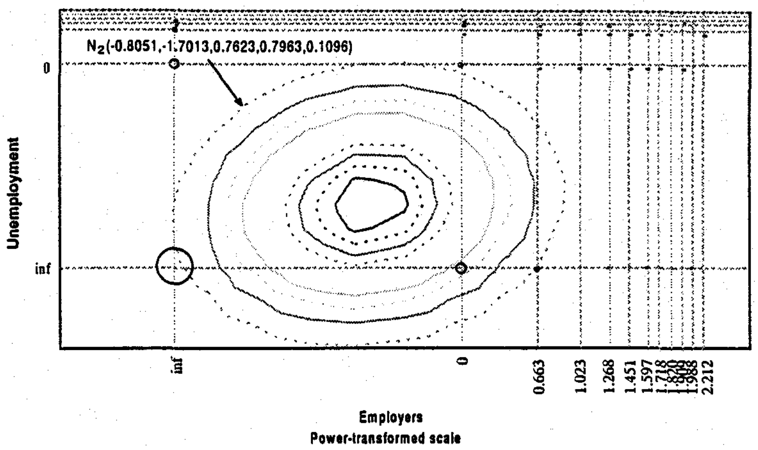

Figure 3.2.9 Example 18: The distribution of observations on the power-transformed scale

Table 3.2.4 Example 19: The results of fitting of the BPND

	Original (Untransformed)	BPND (Transformed)	PND Lengths	PND Breadth
	Maximum likelihood and estimate			
Maximum log-likelihood	–	−30806.3758	−185.7441	−73.1354
$\hat{\lambda}_1$	–	4.0560	3.0319	–
$\hat{\lambda}_2$	–	2.6860	–	3.1300
$\hat{\mu}_1$	14.4035	12642.3733	1033.7271	–
$\hat{\mu}_1^*$		14.4953	14.2285	
$\hat{\mu}_2$	7.9724	98.4120	–	172.8152
$\hat{\mu}_2^*$		7.9869		7.4724
$\hat{\sigma}_1$	0.9166	3014.5518	184.0468	–
$\hat{\sigma}_2$	0.3493	11.4446	–	22.0982
$\hat{\rho}$	0.7590	0.7508	–	–

Performance

		Original		BPND
Test 1	$\hat{\gamma}_{12}$	0.4605	$\hat{\gamma}_{12}$	0.1595
	p-value	near 0	p-value	near 0
	$\hat{\gamma}_{21}$	0.6177	$\hat{\gamma}_{21}$	0.1598
	p-value	near 0	p-value	near 0
Test 2	Skewness	0.0105	Skewness	0.0066
	p-value	0.0026	p-value	0.0351
	Kurtosis	10.2394	Kurtosis	8.7884
	p-value	near 0	p-value	near 0
Test 3	χ^2	1380.0420	χ^2	1378.7311
	p-value	near 0	p-value	near 0

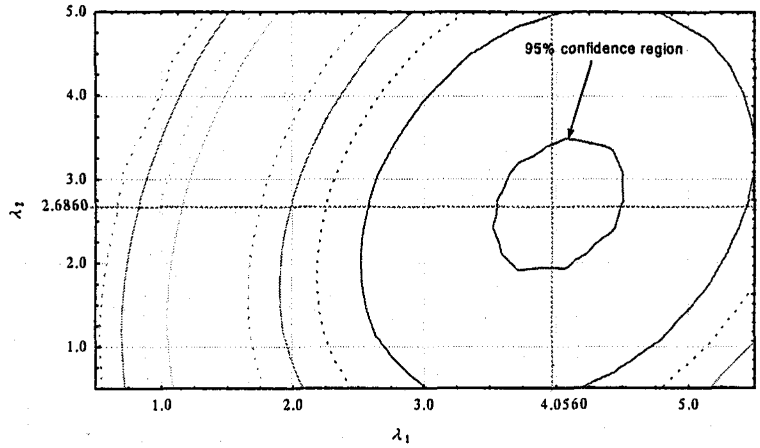

Figure 3.2.10 Example 19: The profile of maximized log-likelihood as a function of transforming parameters λ_1 and λ_2

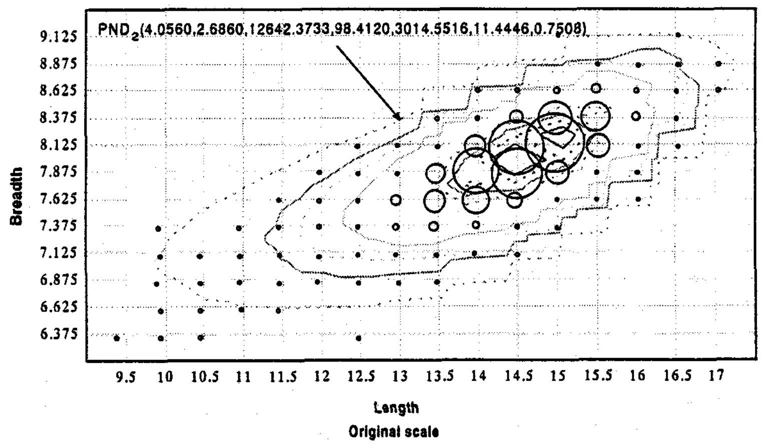

Figure 3.2.11 Example 19: The distribution of observations on the original scale

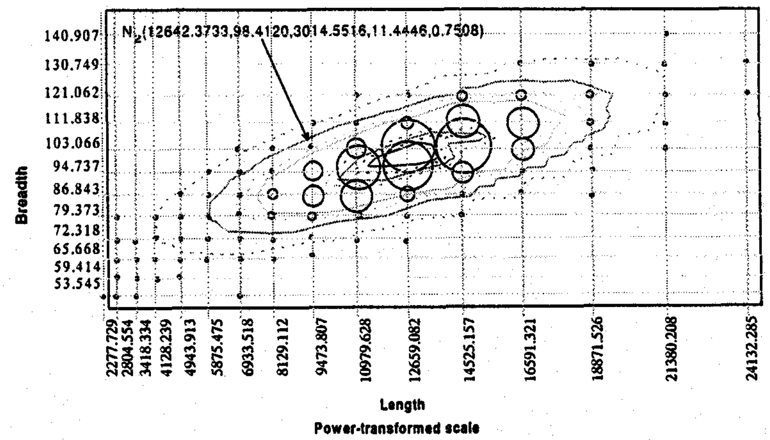

Figure 3.2.12 Example 19: The distribution of observations on the power-transformed scale

Table 3.2.5 Example 20: The results of fitting of the BPND

	Original (Untransformed)	BPND (Transformed)	PND Read	PND Looked at only
		Maximum likelihood and estimate		
Maximum log-likelihood	–	−2264.2991	−15.0338	−13.2485
$\hat{\lambda}_1$	–	0.4720	0.4818	–
$\hat{\lambda}_2$	–	0.7080	–	0.6022
$\hat{\mu}_1$	0.9777	0.1961	0.0779	–
$\hat{\mu}_1^*$		1.2063	1.0795	
$\hat{\mu}_2$	1.2327	0.1369	–	−0.2971
$\hat{\mu}_2^*$		1.1396		0.7208
$\hat{\sigma}_1$	1.0707	0.9087	0.8009	–
$\hat{\sigma}_2$	0.9298	0.8377	–	1.1460
$\hat{\rho}$	0.0936	0.0935	–	–

	Performance			
Test 1	$\hat{\gamma}_{12}$	0.0823	$\hat{\gamma}_{12}$	0.0679
	p-value	0.0419	p-value	0.0769
	$\hat{\gamma}_{21}$	0.0888	$\hat{\gamma}_{21}$	0.0828
	p-value	0.0311	p-value	0.0411
Test 2	Skewness	0.1187	Skewness	0.0657
	p-value	0.0014	p-value	0.0434
	Kurtosis	10.6200	Kurtosis	8.2923
	p-value	near 0	p-value	0.1368
Test 3	χ^2	377.3460	χ^2	377.4072
	p-value	near 0	p-value	near 0

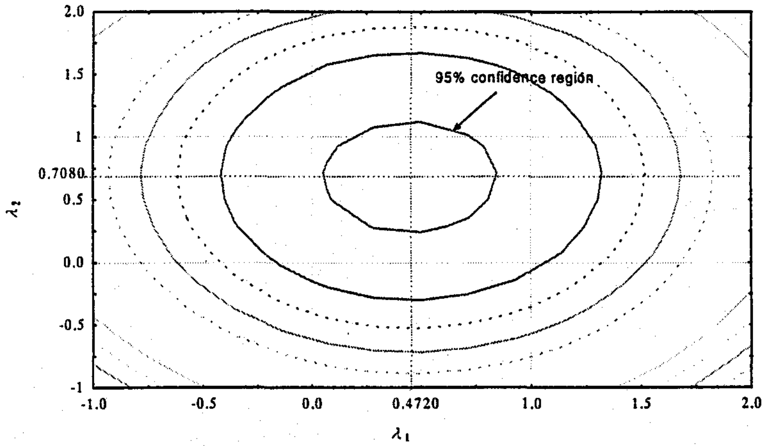

Figure 3.2.13 Example 19: The profile of maximized log-likelihood as a function of transforming parameters λ_1 and λ_2

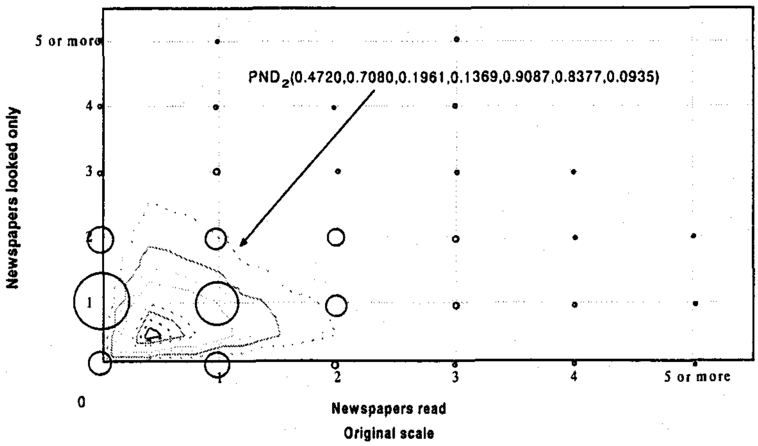

Figure 3.2.14 Example 20: The distribution of observations on the original scale

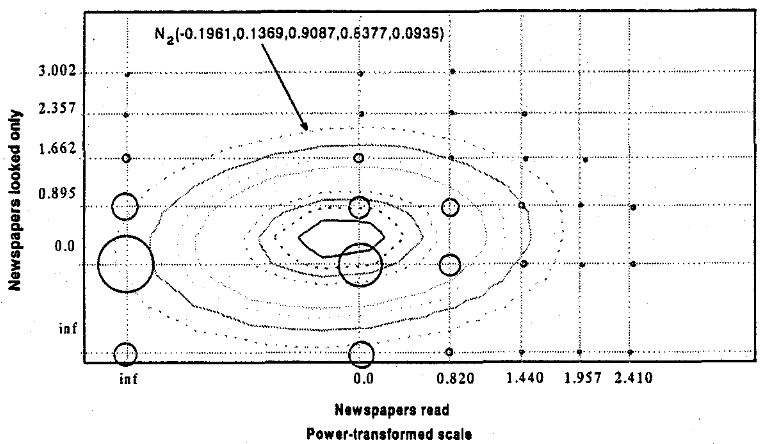

Figure 3.2.15 Example 20: The distribution of observations on the power-transformed scale

Example 20: Stuart and Ord (1986) considers the relationship between newspapers read and the number of newspapers looked at only for students in the University of London, which observations are the frequencies of the number of students classified by these two variables. The results of fitting the BPND to these observations setting length as Y_1 and breadth as Y_2 are shown in Table 3.2.5. The value of transforming parameter for newspapers read was estimated as 0.4720, whereas for newspapers looked at only, it was estimated as 0.7080. These optimized values suggest that the observations of newspaper read and looked at only both have an exponential distribution. In the fitting of PND to each of the two variables, the value of transforming parameter for newspaper read was given as 0.4818, and for look at only, it was given as 0.6022 There was thus no disagreement between the results on the transforming parameter drawn from the fitting of the BPND and the PND. The plot of the profile of maximized log-likelihood as a function of λ_1 and λ_2 is shown in Figure 3.2.13. For the optimized values of these two transforming parameters, we obtained $\hat{\mu}_1(\hat{\lambda}_1) = 0.1961$, $\hat{\mu}_2(\hat{\lambda}_2) = 0.1369$ $\hat{\sigma}_1(\hat{\lambda}_1) = 0.9087$ and $\hat{\sigma}_2(\hat{\lambda}_2) = 0.8377$. The values of back-transformed $\hat{\mu}_1(\hat{\lambda}_1)$ and $\hat{\mu}_2(\hat{\lambda}_2)$ to the original scale, $\hat{\mu}_1^*(\hat{\lambda}_1)$ and $\hat{\mu}_2^*(\hat{\lambda}_2)$ were given as 0.12063 and 1.1396 respectively. Both of them were smaller than the corresponding values on the original scale $\hat{\mu}_1(1)$ =0.9777 and $\hat{\mu}_2(1)$ =1.2327. Also, the value of $\hat{\rho}(\hat{\lambda}_1, \hat{\lambda}_2)$ was given as 0.0935, which was close to the value of 0.0936 for $\hat{\rho}(1,1)$ original scale.

The distributions of observations on the original and the power-transformed scale are shown in Figure 3.2.14 and 3.2.15 respectively. The results of these three tests in Table 3.2.5 show that the observations on the power-transformed scale satisfied the bivariate normality in comparison with the observations on the original scale.

These results of five examples show bivariate power-normal distribution would be helpful to "regularize" the observations even when strict bivariate normality was not achieved. Also, the correlation coefficient for power-transformed observations increases than the original observations. For ungrouped observations, if (X,Y) are jointly distributed in the bivariate normal distribution with correlation ρ, and transformation is made to $X^* = X^*(X)$ and $Y^* = Y^*(Y)$ with $E_X[X^{*2}]$ and $E_Y[Y^{*2}]$ both finite, then the correlation of the new variables X^* and Y^* is less in absolute value than ρ. Namely, for ungrouped observations, transforming so that bivariate normality is more nearly satisfied should increase the correlation coefficient in absolute value. This suggests that the same may be true in grouped observations.

4 Extensions to Regression

The bivariate normal regression problems are considered within the framework developed in the previous section. Now the major interest is to obtain a "good" prediction equation on the basis of observations presented in a correlation table. Also, simple regression problems are briefly mentioned.

4.1 Bivariate Regression

4.1.1 Bivariate Regression in Grouped Form

Regression of Y_2 on Y_1 is considered. The other case follows by symmetry. If the simultaneous power-transformed observation of (Y_1, Y_2), $(Y_1^{(\lambda_1)}, Y_2^{(\lambda_2)})$ have nearly a bivariate normal distribution, for $\lambda_2 \neq 0$, conditional probability density function of $Y_2^{(\lambda_2)}$ given $Y_1^{(\lambda_1)} = y_1^{(\lambda_1)}$ is given by

$$h(y_2^{(\lambda_2)} | Y_1^{(\lambda_1)} = y_1^{(\lambda_1)}) = \frac{f(y_2^{(\lambda_2)} | y_1^{(\lambda_1)})}{\Phi\left\{\frac{\operatorname{sgn}(\lambda_2)}{\sqrt{1-\rho^2}}\left(\kappa_2 + \rho\frac{y_1^{(\lambda_1)} - \mu_1}{\sigma_1}\right)\right\}} \tag{4.1.1}$$

where

$$f(y_2^{(\lambda_2)} | y_1^{(\lambda_1)}) = \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} \times \exp\left[-\frac{1}{2\sigma_2^2(1-\rho^2)}\left\{y_2^{(\lambda_2)} - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(y_1^{(\lambda_1)} - \mu_1)\right\}^2\right]. \tag{4.1.2}$$

Therefore, the conditional expected value of $Y_2^{(\lambda_2)}$ given $Y_1^{(\lambda_1)} = y_1^{(\lambda_1)}$, which for $\lambda_2 \neq 0$ is given by

$$\mathrm{E}\left[Y_2^{(\lambda_2)} \middle| Y_1^{(\lambda_2)} = y_1^{(\lambda_1)}\right] = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(y_1^{(\lambda_2)} - \mu_1)$$
$$+ \frac{\sigma_2\sqrt{1-\rho^2}\,\phi\left\{\dfrac{\mathrm{sgn}(\lambda_2)}{\sqrt{1-\rho^2}}\left(\kappa_2 + \rho\dfrac{y_1^{(\lambda_1)} - \mu_1}{\sigma_1}\right)\right\}}{\Phi\left\{\dfrac{\mathrm{sgn}(\lambda_2)}{\sqrt{1-\rho^2}}\left(\kappa_2 + \rho\dfrac{y_1^{(\lambda_1)} - \mu_1}{\sigma_1}\right)\right\}} \tag{4.1.3}$$

yields the regression function of $Y_2^{(\lambda_2)}$ on $Y_1^{(\lambda_1)}$ on the scale of transformed observations. Here is, the regression function of Y_2 on Y_1 on the BPND given by

$$\mathrm{E}\left[Y_2 \middle| Y_1 = y_1\right] = C_0 \sum_{\nu=0}^{\infty} \frac{(\sqrt{2})^{p+\nu-1}}{\nu!}(1-\rho^2)^{(p-\nu)/2}\left(\kappa_2 + \rho\frac{y_1^{(\lambda_1)} - \mu_1}{\sigma_1}\right)^{\nu} \Gamma\left(\frac{p+\nu+1}{2}\right) \tag{4.1.4}$$

where $p = 1/\lambda_2$, and

$$C_0 = \frac{(\lambda_2\sigma_2)^p}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\kappa_2 + \rho\frac{y_1^{(\lambda_1)} - \mu_1}{\sigma_1}\right)^2\right\}$$
$$\times\left[\Phi\left\{\frac{1}{\sqrt{1-\rho^2}}\left(\kappa_2 + \frac{y_1^{(\lambda_1)} - \mu_1}{\sigma_1}\right)\right\}\right]^{-1} \tag{4.1.5}$$

where $\Gamma(\cdot)$ denotes gamma function.

Our power-transformed normal approach requires the maximum likelihood estimates of the parameter λ_1, λ_2, μ_1, μ_2, σ_1, σ_2 and ρ. By invariance principle of maximum likelihood estimates [*c.f.* Zehna (1966)], we obtain the maximum likelihood estimate of $\boldsymbol{\theta}_n^* = (\lambda_1, \lambda_2, \beta_{21}, \mu_1, \mu_2, \sigma_1, \sigma_2)^{\mathrm{T}}$, for a given sample of size n, as $\boldsymbol{\theta}_n^* = (\hat{\lambda}_{1n}, \hat{\lambda}_{2n}, \hat{\rho}_n\hat{\sigma}_{2n}/\hat{\sigma}_{1n}, \hat{\mu}_{1n}, \hat{\mu}_{2n}, \hat{\sigma}_{1n}, \hat{\sigma}_{2n})^{\mathrm{T}}$, where $\beta_{12} = \rho\sigma_2/\sigma_1$, $\boldsymbol{\theta}_n = (\hat{\lambda}_{1n}, \hat{\lambda}_{2n}, \hat{\rho}_n, \hat{\mu}_{1n}, \hat{\mu}_{2n}, \hat{\sigma}_{1n}, \hat{\sigma}_{2n})^{\mathrm{T}}$ can be obtained as indicated in Section 3.2.

The asymptotic properties of $\hat{\boldsymbol{\theta}}_n^*$ can also easily obtained from those of $\hat{\boldsymbol{\theta}}_n$. Since $\hat{\boldsymbol{\theta}}_n^*$ is a continuous function of $\hat{\boldsymbol{\theta}}_n$, strong consistency of $\hat{\boldsymbol{\theta}}_n$ implies that $\lim_{n\to\infty} \hat{\boldsymbol{\theta}}_n^* = \hat{\boldsymbol{\theta}}_0^*$ with probability one. Furthermore [*c.f.* Rao (1973)], as $n \to \infty$

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n^* - \boldsymbol{\theta}_0^*) \xrightarrow{d} \mathrm{N}_7(\mathbf{0}, \boldsymbol{RVWV}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}})$$

where $\boldsymbol{R} = (\partial\theta_i^*/\partial\theta_j)_{7\times7}$ and $\boldsymbol{VWV}^{\mathrm{T}}$ is the matrix appearing in Section 3.2.2.

4.1.2 Example

The following numerical illustration uses Example 16 considered in Section 3.2.3. The parameter estimates for the regression Y_1 (age of father) on Y_2 (age of mother) became $\hat{\lambda}_1 = -0.0630$ $\hat{\lambda}_2 = 0.4330$, $\hat{\beta}_{12} = \hat{\rho}\hat{\sigma}_1/\hat{\sigma}_2 = 0.1263$, $\hat{\mu}_1 = 3.1749$, $\hat{\mu}_2 = 8.0022$, $\hat{\sigma}_1 = 0.1904$ and $\hat{\sigma}_2 = 0.9445$. The normal regression lines on the power-transformed scale and the corresponding regression curves after transforming back to the original scale shown in Figure 4.1.1 and 4.1.2 respectively. We can appreciate from these two plots how lines follow the modal points.

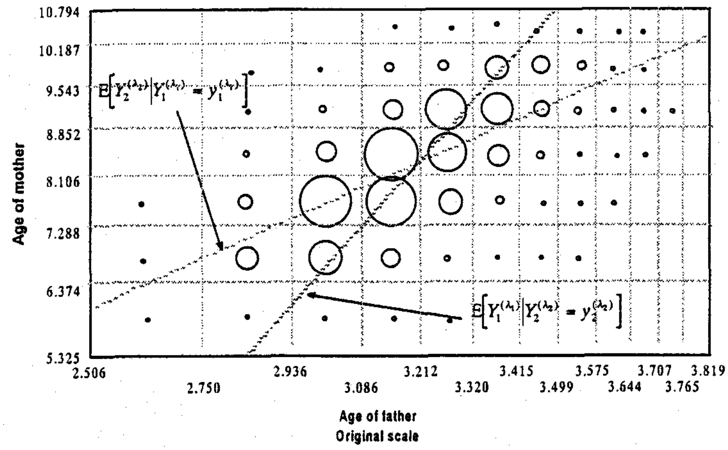

Figure 4.1.1 Example 16: Regression curve obtained on power-transformed scale

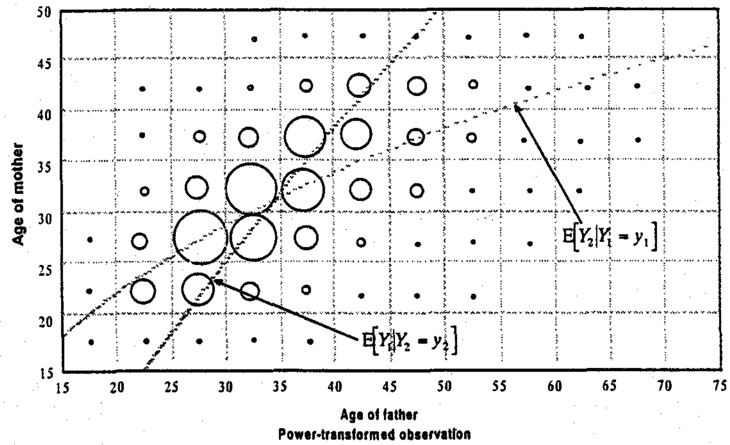

Figure 4.1.2 Example 16: Regression curve after transforming back to the original scale

4.2 Simple Regression

4.2.1 Both Variables in Grouped From

In the ordinary regression based on power-transformation, it is assumed that the power-transformed observations $Y^{(\lambda_2)}$ have the linear model. Namely, in the case of simple regression

$$Y_u^{(\lambda_2)} = \beta_0 + \beta_1 x_u^{(\lambda_1)} + \varepsilon_u, \quad (u = 1, \cdots, n) \tag{4.2.1}$$

where β_0 and β_1 are unknown parameters to be estimated, $x_u^{(\lambda_1)}$ are known power-transformed independent variables, ε_u are independently and identically distributed $\mathrm{N}(0, \sigma^2)$. The situations mentioned in Section 4.1 are applied to this case.

Suppose that the observations are given in a frequency table like Table 4.2.1, where $x_0 = y_0 = 0$ and $x_k = y_l = \infty$. For a given sample size n, the quantities $n_{i\cdot}$ are supposed to be fixed. In order to carry out calculations with grouped data, it is customary to assume that the observations are placed at the midpoints of the intervals. Namely, for the independent variables, if $x_u \in [x_{i-1}, x_i)$, we use the midpoint $x_{Mi}^{(\lambda_1)}$ given by

$$x_{Mi}^{(\lambda_1)} = \frac{x_{i-1}^{(\lambda_1)} + x_i^{(\lambda_1)}}{2} \tag{4.2.2}$$

in stead of $x_u^{(\lambda_1)}$. Therefore, model (4.2.1) will become

$$Y_u^{(\lambda_2)} = \beta_0 + \beta_1 x_{Mi}^{(\lambda_1)} + \varepsilon_i \tag{4.2.3}$$

Table 4.2.1 Frequency Table

Y \ x	$[x_0, x_1)$	$[x_1, x_2)$	$\cdots$	$[x_{i-1}, x_i)$	$\cdots$	$[x_{k-1}, x_k)$	Total
$[y_0, y_1)$	n_{11}	n_{21}	$\cdots$	n_{i1}	$\cdots$	n_{k1}	$n_{\cdot 1}$
$[y_1, y_2)$	n_{12}	n_{22}	$\cdots$	n_{i2}	$\cdots$	n_{k2}	$n_{\cdot 2}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
$[y_{j-1}, y_j)$	n_{1j}	n_{2j}	$\cdots$	n_{ij}	$\cdots$	n_{kj}	$n_{\cdot j}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
$[y_{l-1}, y_l)$	n_{1l}	n_{2l}	$\cdots$	n_{il}	$\cdots$	n_{kl}	$n_{\cdot l}$
Total	$n_{1\cdot}$	$n_{2\cdot}$	$\cdots$	$n_{i\cdot}$	$\cdots$	$n_{k\cdot}$	n

Since Y_u is independent of Y_v for $u \neq v$, setting $\boldsymbol{\theta}^{\mathrm{T}} = (\lambda_1, \lambda_2, \beta_0, \beta_1, \sigma)$, for given sample of size n, the log-likelihood function is given by

$$l_n(\boldsymbol{\theta}) = \sum_{i=1}^{k}\left\{\log n_{i\cdot}! - \sum_{j=1}^{l}\log n_{ij}!\right\} + \sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}\log p_{\mathrm{BPN}ij}(\boldsymbol{\theta}) \tag{4.2.4}$$

where

$$p_{\mathrm{BPN}ij}(\boldsymbol{\theta}) = \frac{1}{A(\kappa)}\left\{\Phi\left(\frac{y_j^{(\lambda_2)} - \beta_0 - \beta_1 x_i^{(\lambda_1)}}{\sigma}\right) - \Phi\left(\frac{y_{j-1}^{(\lambda_2)} - \beta_0 - \beta_1 x_i^{(\lambda_1)}}{\sigma}\right)\right\}. \tag{4.2.5}$$

and

$$A(\kappa) = \begin{cases} \Phi\left(\dfrac{\lambda_2(\beta_0 + \beta_1 x_i^{(\lambda_1)}) + 1}{\lambda_2 \sigma}\right), & \lambda_2 > 0, \\ 1, & \lambda_2 = 0, \\ \Phi\left(\dfrac{\lambda_2(\beta_0 + \beta_1 x_i^{(\lambda_1)}) + 1}{\lambda_2 \sigma}\right), & \lambda_2 < 0. \end{cases} \tag{4.2.6}$$

In order to obtain the adequately model in a regression situation, we assume that the true underlying distribution of Y varies according to level of x. That is, we assume Y_u has probability density function g_i^* concentrated on $(0,\infty)$, whenever x_u lies in $[x_{i-1}, x_i)$. Then, the true probabilities for the interval $[y_{j-1}, y_j)$ become

$$p_{ij} = \int_{y_{j-1}}^{y_j} g_i^*(y)dy, \quad (i = 1,\cdots k, j = 1,\cdots l). \tag{4.2.7}$$

Now, we will consider the asymptotic properties of estimates of λ_1, λ_2, β_0, β_1 and σ.

Let $p_{\mathrm{BPN}ij}(\boldsymbol{\theta})$ and p_{ij} be as in (4.2.5) and (4.2.7). Let $r_1,\cdots,r_k$ be some fixed numbers such that $0 < r_i < 1\, \forall i$ and $\sum_{i=1}^{k} r_i = 1$, and suppose that (i) the parameter space $\Theta \subset \Re^5$ is compact, (ii) $\lim_{n\to\infty} n_{\cdot i}/n = r_i$, and (iii) $H(\boldsymbol{\theta}) = \sum_{i=1}^{k}\sum_{j=1}^{l} r_i p_{ij}\log(p_{\mathrm{BPN}ij}(\boldsymbol{\theta})/p_{ij})$ attains a unique global maximum at $\boldsymbol{\theta}_0 = (\lambda_{10}, \lambda_{20}, \beta_{00}, \beta_{10}, \sigma_0)^{\mathrm{T}}$. Furthermore, suppose that (iv) $\boldsymbol{\theta}_0$ is an interior point of Θ and (v) the Hessian of $H(\boldsymbol{\theta})$ is nonsingular at $\boldsymbol{\theta}_0$.

For a fixed but arbitrary i, consider the sequences of independently and identically distributed indicator functions $\{D_{I_{ij}}(y_u), u = 1,\cdots,n_{\cdot i}\}$ with $\Pr_{g^*_i}[D_{I_{ij}}(y_1) = 1] = p_{ij}(j = 1,\cdots,l)$. The strong law of large number implies that $\hat{p}_{ij} = n_{ij}/n_{\cdot i} \xrightarrow{a.s.} p_{ij}$ as $n_{\cdot i} \to \infty$ for each j. Consequently, a procedure similar to that of Appendix 2 yields that, as $n_{\cdot i} \to \infty$

$$\left|\sum_{j=1}^{l}\hat{p}_{ij}\log p_{\mathrm{BPN}ij}(\boldsymbol{\theta}) - \sum_{j=1}^{l}p_{ij}\log p_{\mathrm{BPN}ij}(\boldsymbol{\theta})\right| \xrightarrow{a.s.} 0. \tag{4.2.8}$$

Now, since $n^{-1}l_n(\boldsymbol{\theta})$ can be approximated almost surely by

$$\sum_{i=1}^{k} \frac{n_{.i}}{n} \left[\sum_{j=1}^{l} \hat{p}_{ij} \log p_{\mathrm{BPN}ij}(\boldsymbol{\theta}) - \sum_{j=1}^{l} \hat{p}_{ij} \log \hat{p}_{ij} \right] + o(1) \tag{4.2.9}$$

we readily obtain the almost sure limit

$$\lim_{n \to \infty} \frac{1}{n} l_n(\boldsymbol{\theta}) = \sum_{i=1}^{k} \sum_{j=1}^{l} r_i p_{ij} \log \frac{p_{\mathrm{BPN}ij}(\boldsymbol{\theta})}{p_{ij}} = H(\boldsymbol{\theta}) \tag{4.2.10}$$

uniformly in $\boldsymbol{\theta}$. Then, Appendix 3 gives the desired strong consistency of $\hat{\boldsymbol{\theta}}_n$. The asymptotic normality of $\hat{\boldsymbol{\theta}}_n$ is established along same lines as the procedure for Section 2.2.2.

4.2.2 Only One Variable Grouped

Here, we consider first the case in which only the dependent variable is grouped. Suppose that $n_{i.}$ observations of the variable Y are made at the fixed value x_i, where $i = 1, \cdots, k$ and $\sum_{i=1}^{k} n_i = n$ is the total sample size. We can deal with this situation exactly as in the previous case. The only change needed is to substitute the approximate quantities $x_{Mi}^{(\lambda_1)}$ used previously, by the exact transformed observations $x_i^{(\lambda_1)}$.

The other case is when the outcomes of the random variable Y are exactly specified, but the value of x are grouped into the intervals $[0, x_1), [x_1, x_2), \cdots, [x_{k-1}, \infty)$. In the latter case, we again consider the model (4.2.1) which gives rise to the log-likelihood for given sample of size n

$$\begin{aligned} l_n(\boldsymbol{\theta}|y) = &-\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 \\ &- \frac{1}{2\sigma^2} \sum_{u=1}^{n} \sum_{i=1}^{k} \delta_{iu} \left(y_u^{(\lambda_2)} - \beta_0 - \beta_1 x_i^{(\lambda_1)} \right)^2 + n(\lambda_1 - 1)\log y_u - nA(\kappa) \end{aligned} \tag{4.2.11}$$

where $\delta_{iu} = 1$ if $x_u \in [x_{i-1}, x_i)$ and 0 otherwise.

Now, if let $n_i = \sum_{u=1}^{n} \delta_{iu} \, (i = 1, \cdots, k)$ and assume that Y_u has the probability density function g_i^* whenever x_u belongs to $[x_{i-1}, x_i)$, we are able to consider the asymptotic properties of λ_1, λ_2, β_0, β_1 and σ.

Let $r_1, \cdots, r_k$ be some fixed numbers such that $0 < r_i < 1 \ \forall i$ and $\sum_{i=1}^{k} r_i = 1$. Suppose that (i) the parameter space Θ is the compact set given by $\Theta = \{\boldsymbol{\theta} = (\lambda_1, \lambda_2, \beta_0, \beta_1, \sigma)^{\mathrm{T}} \| \alpha| \le M_1$, $|\beta| \le M_2, s_1 \le \sigma \le s_2, a \le \lambda_1 \le b, c \le \lambda_2 \le d$, with $0 < M_1, M_2, s_1, s_2, b, d < \infty$ and $-\infty < a$, $c < 0\}$, (ii) $\mathrm{E}_{g^*_i}[Y^{2a}]$ and $\mathrm{E}_{g^*_i}[Y^{2b}]$ are both finite $\forall i$, (iii) $\lim_{n \to \infty} n_i / n = r_i$, and (iv) $\sum_{i=1}^{k} r_i \mathrm{E}_{g^*_i}[l_1(\boldsymbol{\theta}|y)]$ has a unique global maximum at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Furthermore, suppose that (v) $\boldsymbol{\theta}_0$

is an interior point of Θ, (vi) $\mathrm{E}_{g^*_i}[Y^a \log Y]^2$ and $\mathrm{E}_{g^*_i}[Y^b \log Y]^2$ are both finite $\forall i$, (vii) $\sum_{i=1}^{k}\sqrt{r_i}\mathrm{E}_{g^*_i}[\nabla l_1(\boldsymbol{\theta}_0|y)] = \mathbf{0}$, and (viii) $V = \{\sum_{i=1}^{k}\sqrt{r_i}\mathrm{E}_{g^*_i}[\nabla^2 l_1(\boldsymbol{\theta}_0|y)]\}^{-1}$ exists.

Then, the log-likelihood for one observation of Y is

$$l_1(\boldsymbol{\theta}|y) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\left(y^{(\lambda_2)} - \beta_0 - \beta_1 x_i^{(\lambda_1)}\right)^2 + n(\lambda_1 - 1)\log y - A(\kappa). \tag{4.2.12}$$

To show that $\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}_0$ as $n \to \infty$, we will apply Appendix 4 to this log-likelihood function. First, since the power-transformation $x^{(\lambda_1)}$ is increasing in both x and λ_1, we see that, for all $\boldsymbol{\theta} \in \Theta$ and $Y > 0$, $l_1(\boldsymbol{\theta}|y) \le h(y)$, where

$$h(y) = \frac{1}{2}\log 2\pi + |\log s_1| + |\log s_2| + s_1^{-2}\left[\left\{y^{(a)}\right\}^2 + \left\{y^{(b)}\right\}^2 + M_1^2 + M_2^2 Z + (e+1)\right]|\log y| \tag{4.2.13}$$

with $Z = \max_i\{Z_i^2(c) + Z_i^2(d)\}$ and $e = \max\{|a|, b\}$. Since $h(y)$ is g_i^* - integrable, the first assumption of the strong law of large number is satisfied. The second assumption is also satisfied by considering the set $S_j = [j^{-1}, j](j = 1,2,\cdots)$ so that $\Pr_{g^*_i}[(0,\infty) - \bigcup_{j=1}^{\infty} S_j] = 0$. The last assumption of Appendix 4 is seen to hold since $l_1(\boldsymbol{\theta}|y)$ is continuous on $\Theta \times S_j$, a compact set, and this implies that l_1 is equicontinuous in $\boldsymbol{\theta}$ for $Y \in S_j$. Hence, as $n \to \infty$, with probability one

$$\frac{1}{n}\sum_{m=1}^{n_i} l_1(\boldsymbol{\theta}|y_m) \to \mathrm{E}_{g^*_i}[l_1(\boldsymbol{\theta}|y)] \tag{4.2.14}$$

uniformly for $\boldsymbol{\theta} \in \Theta$, and the limit expectation is continuous in $\boldsymbol{\theta}$.

Now, let us partition the n - dimensional vector $\boldsymbol{x}$ as $\boldsymbol{x}^{\mathrm{T}} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k]$ where $\boldsymbol{x}_i$ is the n_i - dimensional vector whose elements lie in $[x_{i-1}, x_i)$. Corresponding to this partition in $\boldsymbol{x}$, we will consider $\boldsymbol{Y} = [\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_k]$. Then, we observe that $\sum_{m=1}^{n_i} l_1(\boldsymbol{\theta}|y_m) = l_{n_i}(\boldsymbol{\theta}|\boldsymbol{Y}_i)$, so that (4.2.14) yields, that as $n \to \infty$

$$\frac{1}{n}l_n(\boldsymbol{\theta}|\boldsymbol{Y}) = \frac{1}{n}\sum_{i=1}^{k} l_{n_i}(\boldsymbol{\theta}|\boldsymbol{Y}_i) = \sum_{i=1}^{k}\frac{n_i}{n}\frac{1}{n_i}l_{n_i}(\boldsymbol{\theta}|\boldsymbol{Y}_i) \xrightarrow{a.s.} \sum_{i=1}^{k} r_i \mathrm{E}_{g^*_i}[l_1(\boldsymbol{\theta}|\boldsymbol{Y})] \tag{4.2.15}$$

uniformly in $\theta \in \Theta$, with the limit function continuous. Thus, we can apply Appendix 3 to get the desired strong consistency of $\hat{\boldsymbol{\theta}}_n$.

The asymptotic normality of $\hat{\boldsymbol{\theta}}_n$ is obtained by mean of an expansion of the gradient $\nabla l_n(\hat{\boldsymbol{\theta}}_n|\boldsymbol{Y})$ about $\boldsymbol{\theta}_0$, that is

$$\frac{1}{\sqrt{n}}\nabla l_n(\hat{\boldsymbol{\theta}}_n|Y)=\frac{1}{\sqrt{n}}\nabla l_n(\boldsymbol{\theta}_0|Y)+\frac{1}{n}\nabla^2 l_n(\boldsymbol{\theta}_n^*|Y)\sqrt{n}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0) \tag{4.2.16}$$

for some $\boldsymbol{\theta}_n^*$ lying on the segment joining $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$. Application of the multivariate central limit theorem, for each i

$$\frac{1}{\sqrt{n_i}}\nabla l_{n_i}(\boldsymbol{\theta}_0|Y_i)\xrightarrow{d}\mathrm{N}_5(\boldsymbol{U}_i,\boldsymbol{W}_i) \text{ as } n_i\to\infty$$

where $\boldsymbol{U}_i=\mathrm{E}_{g^*_i}[\nabla l_1(\boldsymbol{\theta}_0|Y)]$ and $\boldsymbol{W}_i=\mathrm{var}_{g^*_i}[\nabla l_1(\boldsymbol{\theta}_0|Y)]$. Consequently

$$\frac{1}{\sqrt{n_i}}\nabla l_n(\boldsymbol{\theta}_0|Y)=\sum_{i=1}^{k}\sqrt{\frac{n_i}{n}}\sqrt{\frac{1}{n_i}}l_{n_i}(\boldsymbol{\theta}_0|Y_i)\xrightarrow{d}\mathrm{N}_5(\mathbf{0},\boldsymbol{W}) \tag{4.2.17}$$

where $\boldsymbol{W}=\sum_{i=1}^{k}r_i\boldsymbol{W}_i$. Now, on an almost sure set, $\nabla l_n(\hat{\boldsymbol{\theta}}_n|Y)=\mathbf{0}$ for n sufficiently large. Therefore, (4.2.16) and (4.2.17) imply that

$$\frac{1}{n}\nabla^2 l_n(\boldsymbol{\theta}_n^*|Y)\sqrt{n}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)\xrightarrow{d}\mathrm{N}_5(\mathbf{0},\boldsymbol{W}). \tag{4.2.18}$$

Next, results similar to Appendix 5 and Appendix 6 enable us to use the uniform strong law of large number to obtain, almost surely

$$\lim_{n\to\infty}\frac{1}{n}\frac{\partial^2 l_n(\boldsymbol{\theta}|Y)}{\partial\theta_u\partial\theta_v}=\sum_{i=1}^{k}r_i\mathrm{E}_{g^*_i}\left[\frac{\partial^2 l_n(\boldsymbol{\theta}|y)}{\partial\theta_u\partial\theta_v}\right]$$

where the convergence is uniform in θ and the limit function is continuous. Hence, as $\boldsymbol{\theta}_n^*=\gamma_n\hat{\boldsymbol{\theta}}_n+(1-\gamma)\boldsymbol{\theta}_0$ for some $\gamma_n\in(0,1)$ and $\hat{\boldsymbol{\theta}}_n\xrightarrow{a.s.}\boldsymbol{\theta}_0$

$$\left.\frac{1}{n}\frac{\partial^2 l_n(\boldsymbol{\theta}|Y)}{\partial\theta_u\partial\theta_v}\right|_{\theta_n^*}\xrightarrow{a.s.}\sum_{i=1}^{k}r_i\mathrm{E}_{g^*_i}\left[\frac{\partial^2 l_1(\boldsymbol{\theta}_0|y)}{\partial\theta_u\partial\theta_v}\right]$$

as $n\to\infty$ $\forall u,v$, or as $n\to\infty$

$$\frac{1}{n}\nabla^2 l_n(\boldsymbol{\theta}_n^*|Y)\xrightarrow{a.s.}\sum_{i=1}^{k}r_i\mathrm{E}_{g^*_i}\left[\nabla^2 l_n(\boldsymbol{\theta}_0|y)\right]=\boldsymbol{V}^{-1}. \tag{4.2.19}$$

The conclusion is then reached from (4.2.18) and (4.2.19).

It should be noticed that, in order to apply out procedure to the case in which Y is ungrouped, we must know the true probability density function g_i^*. On the other hand, when Y is grouped (and X is either grouped or ungrouped) we do not require the specific functional forms of the g_i^*'s. Moreover, the true probabilities p_{ij}, defined (4.2.7) are consistently estimated by observed frequency $n_{ij}/n_{\cdot i}$.

4.2.3 Example

The following numerical illustration uses Example 17 considered in Section 3.2.3 to show how to fit a simple regression model when both Y and x are given in grouped form. If we consider the observations of miles driven as indirect observations of amount of money spent on gasoline, then the problem of fitting a regression line to the power-transformed variables $Y^{(\lambda_2)}$ (miles driven) and $x^{(\lambda_1)}$ (family income) becomes that of fitting an Engel function to Y and x.

The maximum likelihood estimates are $\hat{\lambda}_1 = 0.3560$, $\hat{\lambda}_2 = 0.3750$, $\hat{\beta}_0 = 0.5032$, $\hat{\beta}_1 = 0.0514$ and $\hat{\sigma} = 1.9004$. The fitting regression line on the power-transformed scale and the corresponding regression line after transforming back to the original scale shown in Figure 4.2.1 and 4.2.2 respectively.

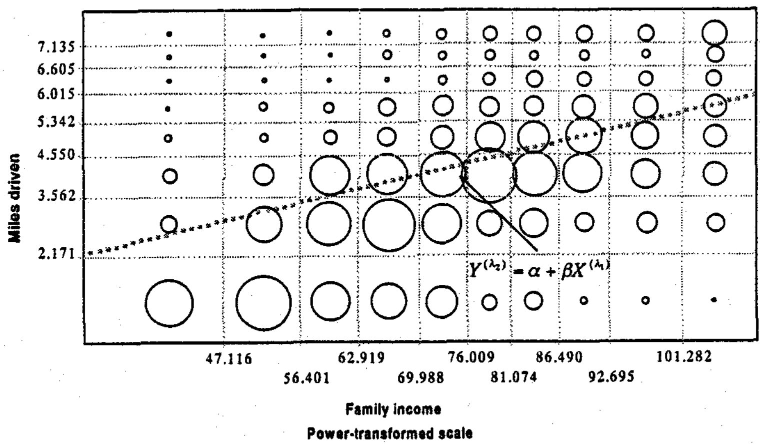

Figure 4.2.1 Example 17: Regression line on the power-transformed scale

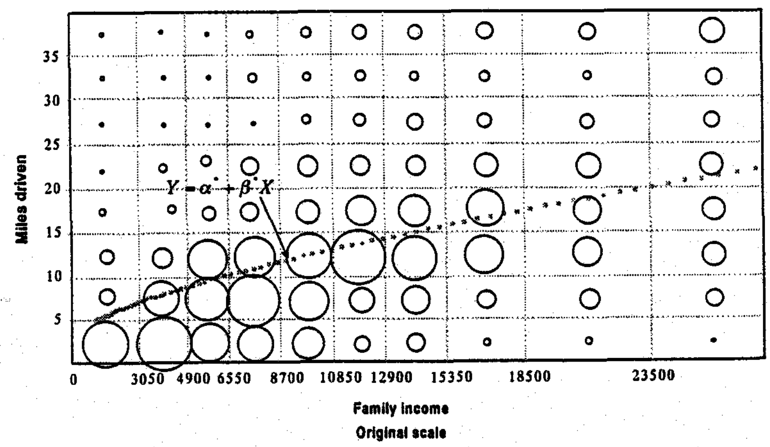

Fig 4.2.2 Example 17: Regression line after transforming back to the original scale

5
Conclusions and Further Developments

5.1 Conclusions

Statisticians have long known that their models do not perfectly mirror reality. In grouped observation problems, the most attention has been fixed on such problem as what to do with the observations which are more log-tailed or prone to outlier, rather than the standard models imply. The motivation of our investigation was to develop the procedure which would give reliable inference in the absent of strong knowledge of progress generating data, and include the standard model based on normal distribution.

We have firstly developed the procedure for fitting the PND to univariate grouped observations. We have considered the most elementary case, in which there was only one variable of interest and its observations were given in a grouped form such as a frequency table. The explicit expressions for the maximum likelihood estimates of parameters were not available in ungrouped observation case, as shown in Goto *et al.* (1983), and the present case was no exception. However, it was possible to gain some insight into our proposed procedure using the criterion of maximization of likelihood, by investigating the asymptotic properties of them. In deriving these properties, we have followed the approach to discrete distributions in Hernandez and Johnson (1981). In their approach, the criterion of minimization of Kullback-Leibler information was used to obtain the parameter estimates. Our procedure showed that the maximum likelihood estimates of parameters had the strong consistency and the asymptotic normality under certain conditions. After having concrete and practical grasp of our procedure through several numerical examples, medium-sized simulation experiment was performed to evaluate (i) the precision of the maximum likelihood estimates of parameters and (ii) the effect of grouping or categorizing on them. For (i), the results of simulation showed that, the precision of maximum likelihood estimate of transforming parameter was not so much influenced by the shape of the PND as the precision of the estimate in ungrouped observation case. Nevertheless, the precision of maximum likelihood estimates of mean and variance was as much influenced by the shape of the PND as the precision of those estimates in ungrouped observation case. However, some numerical examples showed that those effects could be removed by using the normalizing power-transformation, namely the adjusted power-transformation by Jacobian of transformation formula. For (ii), the results of simulation showed that the number of intervals had a great influence on the precision of maximum likelihood estimates and the precision decreased as the number of intervals increased.

The PND was also applied in the mixed observation case, where observation involved ungrouped (exactly specified) and grouped ones. We have particularly focused on the two case of the right/or left censored observations and ungrouped observations available in the tail. Furthermore, it was applied in the observations subjected to an upper constraint such as examination score and percentage or proportion, and in discrete observations.

Next, we have focused on two variables situations and considered two variables grouped in (i) a correlation table, (ii) grouped bivariate regression and (iii) simple linear regression when one or both variables were given in a grouped form and their observations were generated form the BPND. In all of (i), (ii) and (iii), the analogy of the procedure used in univariate grouped observation case showed that the maximum likelihood estimates of parameters had the strong consistency and asymptotic normality under certain conditions. Correlation table and regression were equivalent situation in our viewpoint. Thus, the solution provided for correlation tables was easily modified to include bivariate regression. The same sets f observations were then used to illustrate both procedures.

5.2 Further Developments

Survival times are inherently continuous, and for the most part statisticians model them that way. Yet, times can not often be observed very exactly. For example, in cancer clinical trials, patients may be reexamined only at, say, 3-month intervals, and often the follow-up times themselves are random. If the endpoint of interest is time until detectable tumor recurrence, this quantity is known only up to the approximate 3-month gap between the last negative and the first positive follow-up. Observations of this kind have come to be known as " interval censored", a name that emphasizes the connection with the common problem of right censoring mentioned in Section 2.3.2. Interval and right censoring are not special cases of grouping in narrow sense of the word, because the partition of the sample space in censoring need not be fixed. However, if the censoring time distribution or the distribution of censoring interval limits is not related to survival time, the relevant part of the likelihood is effectively grouped observation likelihood, and these problem can be treated within our framework.

Maximum likelihood estimation has been widely used in grouped observation case as in ungrouped. When models involve an unknown shift or cutoff value, for example in the shifted power-transformation case (Box and Cox, 1964), such problem as likelihood do not have a consistent global maximum likelihood estimates in ungrouped observation case, apparently disappears in grouped observation likelihood in large samples (Atkinson, 1985: Atkinson *et al.*, 1991). One of alternative methods to avoid this problem is the maximum product of spacing (MPS) proposed by Cheng and Amin (1983) and it is interesting to note that their " likelihood" has the functional form of a grouped observation likelihood (Cheng and Iles, 1987: Cheng and Taylor, 1995).

Appendices

Appendix 1

$$\frac{\partial l_n(\boldsymbol{\theta})}{\partial \lambda} = \frac{n}{\sigma}\sum_{i=1}^{k}\frac{\hat{p}_i}{p_{\mathrm{PN}i}(\boldsymbol{\theta})}\left\{\phi(z_i)\frac{\partial y_i^{(\lambda)}}{\partial \lambda} - \phi(z_{i-1})\frac{\partial y_{i-1}^{(\lambda)}}{\partial \lambda}\right\} - \frac{n}{A(\kappa)}\frac{\partial A(\kappa)}{\partial \lambda},$$

$$\frac{\partial^2 l_n(\boldsymbol{\theta})}{\partial \lambda^2} = \frac{n}{\sigma^2}\sum_{i=1}^{k}\hat{p}_i\left[\frac{\phi(z_i)}{p_{\mathrm{PN}i}(\boldsymbol{\theta})}\left\{\sigma\frac{\partial^2 y_i^{(\lambda)}}{\partial \lambda^2} - z_i\left(\frac{\partial y_i^{(\lambda)}}{\partial \lambda}\right)^2\right\}\right.$$
$$\left. - \frac{\phi(z_{i-1})}{p_{\mathrm{PN}i}(\boldsymbol{\theta})}\left\{\sigma\frac{\partial^2 y_{i-1}^{(\lambda)}}{\partial \lambda^2} - z_{i-1}\left(\frac{\partial y_{i-1}^{(\lambda)}}{\partial \lambda}\right)^2\right\} - \left\{\frac{\phi(z_i)}{p_{\mathrm{PN}i}(\boldsymbol{\theta})}\frac{\partial y_i^{(\lambda)}}{\partial \lambda} - \frac{\phi(z_{i-1})}{p_{\mathrm{PN}i}(\boldsymbol{\theta})}\frac{\partial y_{i-1}^{(\lambda)}}{\partial \lambda}\right\}^2\right]$$
$$- \frac{n}{A(\kappa)}\left\{\frac{1}{A(\kappa)}\left(\frac{\partial A(\kappa)}{\partial \lambda}\right)^2 - \frac{\partial^2 A(\kappa)}{\partial \lambda^2}\right\},$$

$$\frac{\partial^2 l_n(\boldsymbol{\theta})}{\partial \lambda \partial \mu} = \frac{n}{\sigma^2}\sum_{i=1}^{k}\hat{p}_i\left[\frac{\phi(z_i) - \phi(z_{i-1})}{p_{\mathrm{PN}i}^2(\boldsymbol{\theta})}\left\{\phi(z_i)\frac{\partial y_i^{(\lambda)}}{\partial \lambda} - \phi(z_{i-1})\partial\frac{\partial y_{i-1}^{(\lambda)}}{\partial \lambda}\right\}\right.$$
$$\left. + \frac{1}{p_{\mathrm{PN}i}(\boldsymbol{\theta})}\left\{z_i\phi(z_i)\frac{\partial y_i^{(\lambda)}}{\partial \lambda} - z_{i-1}\phi(z_{i-1})\frac{\partial y_{i-1}^{(\lambda)}}{\partial \lambda}\right\}\right]$$
$$- \frac{n}{A(\kappa)}\left\{\frac{1}{A(\kappa)}\frac{\partial A(\kappa)}{\partial \lambda}\frac{\partial A(\kappa)}{\partial \mu} - \frac{\partial^2 A(\kappa)}{\partial \lambda \partial \mu}\right\},$$

$$\frac{\partial^2 l_n(\boldsymbol{\theta})}{\partial \lambda \partial \sigma} = \frac{n}{\sigma^2}\sum_{i=1}^{k}\hat{p}_i\left[\frac{1}{p_{\mathrm{PN}i}(\boldsymbol{\theta})}\left\{(z_i^2 - 1)\phi(z_i)\frac{\partial y_i^{(\lambda)}}{\partial \lambda} - (z_{i-1}^2 - 1)\phi(z_{i-1})\frac{\partial y_{i-1}^{(\lambda)}}{\partial \lambda}\right\}\right.$$
$$\left. + \frac{1}{p_{\mathrm{PN}i}^2(\boldsymbol{\theta})}\{z_i\phi(z_i) - z_{i-1}\phi(z_{i-1})\}\left\{\phi(z_i)\frac{\partial y_i^{(\lambda)}}{\partial \lambda} - \phi(z_{i-1})\frac{\partial y_{i-1}^{(\lambda)}}{\partial \lambda}\right\}\right]$$
$$- \frac{n}{A(\kappa)}\left\{\frac{1}{A(\kappa)}\frac{\partial A(\kappa)}{\partial \lambda}\frac{\partial A(\kappa)}{\partial \sigma} - \frac{\partial^2 A(\kappa)}{\partial \lambda \partial \sigma}\right\},$$

$$\frac{\partial l_n(\boldsymbol{\theta})}{\partial \mu} = \frac{n}{\sigma}\sum_{i=1}^{k}\frac{\hat{p}_i}{p_{\mathrm{PN}i}(\boldsymbol{\theta})}\{\phi(z_{i-1}) - \phi(z_i)\} - \frac{n}{A(\kappa)}\frac{\partial A(\kappa)}{\partial \mu},$$

$$\frac{\partial^2 l_n(\boldsymbol{\theta})}{\partial\mu^2} = \frac{n}{\sigma^2}\sum_{i=1}^{k}\hat{p}_i\left[\frac{z_{i-1}\phi(z_{i-1}) - z_i\phi(z_i)}{p_{\mathrm{PN}i}(\boldsymbol{\theta})} - \left\{\frac{\phi(z_{i-1}) - \phi(z_i)}{p_{\mathrm{PN}i}(\boldsymbol{\theta})}\right\}^2\right]$$
$$-\frac{n}{A(\kappa)}\left\{\frac{1}{A(\kappa)}\left(\frac{\partial A(\kappa)}{\partial\mu}\right)^2 - \frac{\partial^2 A(\kappa)}{\partial\mu^2}\right\},$$

$$\frac{\partial l_n(\boldsymbol{\theta})}{\partial\sigma} = \frac{n}{\sigma}\sum_{i=1}^{k}\hat{p}_i\frac{z_{i-1}\phi(z_{i-1}) - z_i\phi(z_i)}{p_{\mathrm{PN}i}(\boldsymbol{\theta})} - \frac{n}{A(\kappa)}\frac{\partial A(\kappa)}{\partial\sigma},$$

$$\frac{\partial^2 l_n(\boldsymbol{\theta})}{\partial\sigma^2} = \frac{n}{\sigma^2}\sum_{i=2}^{k-1}\hat{p}_i\left[\frac{(z_{i-1}^3 - 2z_{i-1})\phi(z_{i-1}) - (z_i^3 - 2z_i)\phi(z_i)}{p_{\mathrm{PN}i}(\boldsymbol{\theta})} - \left\{\frac{z_{i-1}\phi(z_{i-1}) - z_i\phi(z_i)}{p_{\mathrm{PN}i}(\boldsymbol{\theta})}\right\}^2\right]$$
$$-\frac{n}{A(\kappa)}\left\{\frac{1}{A(\kappa)}\left(\frac{\partial A(\kappa)}{\partial\sigma}\right)^2 - \frac{\partial^2 A(\kappa)}{\partial\sigma^2}\right\},$$

$$\frac{\partial^2 l_n(\boldsymbol{\theta})}{\partial\mu\partial\sigma} = \frac{n}{\sigma^2}\sum_{i=1}^{k}\hat{p}_i\left[\frac{(z_{i-1}^2 - 1)\phi(z_{i-1}) - (z_i^2 - 1)\phi(z_i)}{p_{\mathrm{PN}i}(\boldsymbol{\theta})} - \frac{\{\phi(z_{i-1}) - \phi(z_i)\}\{z_{i-1}\phi(z_{i-1}) - z_i\phi(z_i)\}}{p_{\mathrm{PN}i}^2(\boldsymbol{\theta})}\right]$$
$$-\frac{n}{A(\kappa)}\left\{\frac{1}{A(\kappa)}\frac{\partial A(\kappa)}{\partial\mu}\frac{\partial A(\kappa)}{\partial\sigma} - \frac{\partial^2 A(\kappa)}{\partial\mu\partial\sigma}\right\}$$

where

$$\frac{\partial y_i^{(\lambda)}}{\partial\lambda} = \begin{cases}\dfrac{1}{\lambda}\left\{y_i^{\lambda}(\log y_i) - y_i^{(\lambda)}\right\}, & \lambda \neq 0,\\[2mm] \dfrac{1}{2}(\log y_i)^2, & \lambda = 0,\end{cases}$$

$$\frac{\partial^2 y_i^{(\lambda)}}{\partial\lambda^2} = \begin{cases}\dfrac{1}{\lambda}\left[y_i^{\lambda}(\log y_i)^2 - 2\dfrac{1}{\lambda}\left\{y_i^{\lambda}(\log y_i) - y_i^{(\lambda)}\right\}\right], & \lambda \neq 0,\\[2mm] \dfrac{1}{3}(\log y_i)^3, & \lambda = 0,\end{cases}$$

$$\frac{\partial A(\kappa)}{\partial\lambda} = \begin{cases}-\dfrac{\phi(\kappa)}{\lambda^2\sigma}, & \lambda > 0,\\[2mm] 0, & \lambda = 0,\\[2mm] \dfrac{\phi(\kappa)}{\lambda^2\sigma}, & \lambda < 0,\end{cases}$$

$$\frac{\partial^2 A(\kappa)}{\partial\lambda^2} = \begin{cases} \dfrac{2\lambda\sigma - \kappa}{\lambda^4\sigma^2}\phi(\kappa), & \lambda > 0, \\ 0, & \lambda = 0, \\ -\dfrac{\kappa + 2\lambda\sigma}{\lambda^4\sigma^2}\phi(\kappa), & \lambda < 0, \end{cases}$$

$$\frac{\partial^2 A(\kappa)}{\partial\lambda\partial\mu} = \begin{cases} \dfrac{\kappa\phi(\kappa)}{\lambda^2\sigma^2}, & \lambda > 0, \\ 0, & \lambda = 0, \\ -\dfrac{\kappa\phi(\kappa)}{\lambda^2\sigma^2}, & \lambda < 0, \end{cases}$$

$$\frac{\partial^2 A(\kappa)}{\partial\lambda\partial\sigma} = \begin{cases} \dfrac{1-\kappa^2}{\lambda^2\sigma^2}\phi(\kappa), & \lambda > 0, \\ 0, & \lambda = 0, \\ -\dfrac{1-\kappa^2}{\lambda^2\sigma^2}\phi(\kappa), & \lambda < 0, \end{cases}$$

$$\frac{\partial A(\kappa)}{\partial\mu} = \begin{cases} \dfrac{\phi(\kappa)}{\sigma}, & \lambda > 0, \\ 0, & \lambda = 0, \\ -\dfrac{\phi(\kappa)}{\sigma}, & \lambda < 0, \end{cases}$$

$$\frac{\partial^2 A(\kappa)}{\partial\mu^2} = \begin{cases} -\dfrac{\kappa\phi(\kappa)}{\sigma^2}, & \lambda > 0, \\ 0, & \lambda = 0, \\ \dfrac{\kappa\phi(\kappa)}{\sigma^2}, & \lambda < 0, \end{cases}$$

$$\frac{\partial A(\kappa)}{\partial\sigma} = \begin{cases} -\dfrac{\kappa\phi(\kappa)}{\sigma}, & \lambda > 0, \\ 0, & \lambda = 0, \\ \dfrac{\kappa\phi(\kappa)}{\sigma}, & \lambda < 0, \end{cases}$$

$$\frac{\partial^2 A(\kappa)}{\partial\sigma^2} = \begin{cases} \dfrac{2-\kappa^2}{\sigma}\kappa\phi(\kappa), & \lambda > 0, \\ 0, & \lambda = 0, \\ -\dfrac{2-\kappa^2}{\sigma}\kappa\phi(\kappa), & \lambda < 0, \end{cases}$$

$$\frac{\partial^2 A(\kappa)}{\partial\mu\partial\sigma} = \begin{cases} -\dfrac{1-\kappa^2}{\sigma^2}\phi(\kappa), & \lambda > 0, \\ 0, & \lambda = 0, \\ \dfrac{1-\kappa^2}{\sigma^2}\phi(\kappa), & \lambda < 0. \end{cases}$$

Appendix 2

Assume that the parameter space Θ is a compact subset of $\Re^3$. The maximum likelihood estimates of $\boldsymbol{\theta}^{\mathrm{T}} = (\lambda, \mu, \sigma)$ will be defined as that value of $\boldsymbol{\theta} \in \Theta$ which, for a given sample of size n, gives a global maximum for $l_n(\boldsymbol{\theta})$. The value is denoted by $\hat{\boldsymbol{\theta}}_n$.

Let $I_i = [y_{i-1}, y_i)$, $p_i = \int_{I_i} g^*(y)dy$, $\hat{p}_{i,n} = n_i/n$, $p_{\mathrm{PN}i}(\boldsymbol{\theta})$ be as in (2.2.2), and $D_I(\cdot)$ be the indicator function of the set I, that is $D_I(y) = 1$ if $y \in I$ and 0 otherwise.

For each $i = 1, \cdots, k$, consider the sequence of independent and identically distributed $0-1$ random variables $\{D_{I_i}(y_j), j = 1, \cdots, n\}$ with success probabilities $\Pr_{g^*}[D_{I_i}(y) = 1] = p_i$. We write $\Pr_{g^*}(\cdot)$ to emphasize that the probability is determined with respect to the true probability density function g^*.

Since $n_i = \sum_{j=1}^{n} D_{I_i}(y_j)$, it follows from the Strong Law of Large Numbers that as $n \to \infty$

$$\hat{p}_{i,n} \xrightarrow{a.s.} p_i. \tag{A2.1}$$

Let $A = \{\varpi \in \Theta | \hat{p}_{i,n}(\varpi) \to p_i \text{ as } n \to \infty\}$. Then, for $\varepsilon > 0$ and $\varpi \in A$, we obtain,

$$\begin{aligned} \left| \sum_{i=1}^{k} (\hat{p}_{i,n}(\varpi) - p_i) \log p_{\mathrm{PN}i}(\boldsymbol{\theta}) \right| &\le \sum_{i=1}^{k} |\hat{p}_{i,n}(\varpi) - p_i| |\log p_{\mathrm{PN}i}(\boldsymbol{\theta})| \\ &\le \varepsilon \sum_{i=1}^{k} |\log p_{\mathrm{PN}i}(\boldsymbol{\theta})| \end{aligned} \tag{A2.2}$$

for $\forall n \ge \mathrm{N}(\varepsilon, \varpi)$ and $\boldsymbol{\theta} \in \Theta$.

Next, since $y_i^{(\lambda)}$ is monotone increasing in λ and $y_{i-1} < y_i$ for each $i = 1, \cdots, k$, it follows easily that $p_{\mathrm{PN}i}(\boldsymbol{\theta}) > 0$ on Θ. Also, form the continuity of $z_i(\boldsymbol{\theta}) = (y_i^{(\lambda)} - \mu)/\sigma$ for $\boldsymbol{\theta} \in \Theta$, we have that $\log p_{\mathrm{PN}i}(\boldsymbol{\theta})$ is continuous on Θ. This result and Θ being compact by assumption imply the existence of $R_i = \inf_{\theta \in \Theta} \log p_{\mathrm{PN}i}(\boldsymbol{\theta}) > -\infty$. Hence for any $\boldsymbol{\theta} \in \Theta$

$$\sum_{i=1}^{k} |\log p_{\mathrm{PND}i}(\boldsymbol{\theta})| \le R \tag{A2.3}$$

where $R = \sum_{i=1}^{k} |R_i| < \infty$. Therefore, it follows form (A2.2) and (A2.3) that as $n \to \infty$

$$\left|\sum_{i=1}^{k}\hat{p}_{i,n}\log p_{\mathrm{PND}i}(\boldsymbol{\theta})-\sum_{i=1}^{k}p_i\log p_{\mathrm{PND}i}(\boldsymbol{\theta})\right|\xrightarrow{a.s.}0.$$

Appendix 3

Assume that the parameter space Θ is a compact subset of $\Re^3$. The maximum likelihood estimates of $\boldsymbol{\theta}^{\mathrm{T}}=(\lambda,\mu,\sigma)$ will be defined as that value of $\boldsymbol{\theta}\in\Theta$ which, for a given sample of size n, gives a global maximum for $l_n(\boldsymbol{\theta})$ given in (2.2.4). The value is denoted by $\hat{\boldsymbol{\theta}}_n$. Also, suppose that f is a continuous function from Θ to $\Re$, $n^{-1}l_n(\boldsymbol{\theta})\rightarrow f(\boldsymbol{\theta})$ uniformly in $\boldsymbol{\theta}\in\Theta$, $f(\boldsymbol{\theta})$ has a unique global maximum at $\boldsymbol{\theta}=\boldsymbol{\theta}_0$.

To obtain a contradiction, suppose that $\hat{\boldsymbol{\theta}}_n$ does not converge almost surely to $\boldsymbol{\theta}_0$. That is, assume there exists a set B such that for each $\varpi\in B, \hat{\boldsymbol{\theta}}_n(\varpi)\nrightarrow\boldsymbol{\theta}_0$ as $n\rightarrow\infty$, where $\Pr(B)>0$. Let $C=\{\varpi\,|\,n^{-1}l_n(\boldsymbol{\theta})\rightarrow f(\boldsymbol{\theta})\text{ as }n\rightarrow\infty,\text{ uniformly in }\boldsymbol{\theta}\in\Theta\}$ and consider $D=B\cap C$, $\Pr(D)=\Pr(B)>0$.

Since Θ is compact, for each $\varpi\in D$, there exists a subsequence $\{m\}\subset\{n\}$ and limit point $\boldsymbol{\theta}_\varpi$ depending on ϖ such that $\hat{\boldsymbol{\theta}}_m(\varpi)\rightarrow\theta_\varpi$ as $m\rightarrow\infty$, with $\boldsymbol{\theta}_\varpi=\boldsymbol{\theta}_0$. For a fixed $\varpi\in D$, we have

$$\begin{aligned}\left|\frac{1}{m}l_m(\hat{\boldsymbol{\theta}}_m(\varpi))-f(\boldsymbol{\theta}_\varpi)\right|&\le\left|\frac{1}{m}l_m(\hat{\boldsymbol{\theta}}_m(\varpi))-f(\hat{\boldsymbol{\theta}}_m(\varpi))\right|+\left|f(\hat{\boldsymbol{\theta}}_m(\varpi))-f(\boldsymbol{\theta}_\varpi)\right|\\&\le\max_{\theta\in\Theta}\left|\frac{1}{m}l_m(\boldsymbol{\theta}(\varpi))-f(\boldsymbol{\theta}(\varpi))\right|+\left|f(\hat{\boldsymbol{\theta}}_m(\varpi))-f(\boldsymbol{\theta}_\varpi)\right|.\end{aligned}\tag{A3.1}$$

Then, letting $m\rightarrow\infty$ on the right hand side of (A3.1), we get

$$\max_{\theta\in\Theta}\left|\frac{1}{m}l_m(\boldsymbol{\theta}(\varpi))-f(\boldsymbol{\theta}(\varpi))\right|\rightarrow 0\tag{A3.2}$$

because $D=B\cap C$ and $\left|f(\hat{\boldsymbol{\theta}}_m(\varpi))\rightarrow f(\boldsymbol{\theta}_\varpi)\right|\rightarrow 0$ by the continuity of f. Thus, for $\varpi\in D$, $n\rightarrow\infty$

$$\frac{1}{m}l_m(\hat{\boldsymbol{\theta}}_m(\varpi))\rightarrow f(\boldsymbol{\theta}_\varpi).\tag{A3.3}$$

However, by definition of $\hat{\boldsymbol{\theta}}_m$, we have

$$\frac{1}{m}l_m(\hat{\boldsymbol{\theta}}_m)\ge\frac{1}{m}l_m(\boldsymbol{\theta}_0)\tag{A3.4}$$

so that, taking the limit of (A3.4), we obtain $f(\boldsymbol{\theta}_{\varpi}) \geq f(\boldsymbol{\theta}_0)$ for $\varpi \in D$. Since $\Pr(D) > 0$, this contradicts the assumption that $\boldsymbol{\theta}_0$ gives a unique maximum and we conclude that $\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}_0$ as $n \to \infty$ as asserted

Appendix 4

We consider is the result which is sometimes called the uniform strong law of large number and is established by Rubin (1956).

Let $\{Y_i\}_{i=1}^{\infty}$ be a sequence of independently and identically random variables with values in an arbitrary space $\mathcal{X}$. Let Θ be a compact topological space and let f be a complex-valued function on $\Theta \times \mathcal{X}$, measurable in Y for each $\boldsymbol{\theta} \in \Theta$. Let P be the common distribution of the Y_i's. If

(a) there is an integrable v such that $|f(\boldsymbol{\theta}, y)| < v(y)$, $\forall \boldsymbol{\theta} \in \Theta$, $Y \in \mathcal{X}$,

(b) there is a sequence $\{S_i\}_{i=1}^{\infty}$ of measurable sets such that $P(\mathcal{X} - \bigcup_{i=1}^{\infty} S_i) = 0$, and

(c) for each i, $f(\boldsymbol{\theta}, y)$ is equicontinuous in $\boldsymbol{\theta}$ for $Y \in S_i$.

Then, with probability one

$$\frac{1}{n}\sum_{j=1}^{n} f(\boldsymbol{\theta}, y_j) \to \int f(\boldsymbol{\theta}, y) dP(y)$$

as $n \to \infty$, uniformly for $\boldsymbol{\theta} \in \Theta$, and the limit function is continuous.

Appendix 5

Let Θ, the parameter space, defined as $\Theta = \{\boldsymbol{\theta}^{\mathrm{T}} = (\lambda, \mu, \sigma) \,\big|\, |\mu| \leq M, s_1 < \sigma < s_2, a \leq \lambda \leq b$ for some $0 < M, s_1, s_2, b < \infty$ and $-\infty < a < 0\}$.

Let

$$l_1^{(u)}(\boldsymbol{\theta}|y) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2}\left(\frac{y^{(\lambda)} - \mu}{\sigma}\right)^2 + (\lambda - 1)\log y, \tag{A5.1}$$

$$D_{11}(\boldsymbol{\theta}, y) = -\sigma^{-2}, \tag{A5.2}$$

$$D_{12}(\boldsymbol{\theta}, y) = -2\sigma^{-3}(y^{(\lambda)} - \mu), \tag{A5.3}$$

$$D_{13}(\boldsymbol{\theta}, y) = \sigma^{-2} y'^{(\lambda)}, \tag{A5.4}$$

$$D_{22}(\boldsymbol{\theta}, y) = -\sigma^{-2} - 3\sigma^{-4}(y^{(\lambda)} - \mu)^2, \tag{A5.5}$$

$$D_{23}(\boldsymbol{\theta}, y) = 2\sigma^{-3}(y^{(\lambda)} - \mu)y'^{(\lambda)}, \tag{A5.6}$$

$$D_{33}(\boldsymbol{\theta}, Y) = -\sigma^{-2}(y^{(\lambda)} - \mu)y''^{(\lambda)} - \{y'^{(\lambda)}\}^2. \tag{A5.7}$$

Then, $\forall \boldsymbol{\theta} \in \Theta$

$$\left| \frac{\partial^2 l_1^{(u)}(\boldsymbol{\theta}|y)}{\partial\theta_u \partial\theta_v} \right| = |D_{uv}(\boldsymbol{\theta}|y)| \le H_{uv}(y)$$

where

$$y'^{(\lambda)} = \begin{cases} \frac{1}{\lambda}\{y_i^{\lambda}(\log y_i) - y_i^{(\lambda)}\}, & \lambda \ne 0, \\ \frac{1}{2}(\log y_i)^2, & \lambda = 0, \end{cases}$$

$$y''^{(\lambda)} = \begin{cases} \frac{1}{\lambda}\left[y_i^{\lambda}(\log y_i)^2 - 2\frac{1}{\lambda}\{y_i^{\lambda}(\log y_i) - y_i^{(\lambda)}\}\right], & \lambda \ne 0, \\ \frac{1}{3}(\log y_i)^3, & \lambda = 0 \end{cases}$$

and the H_{uv} 's are as follows:

$$H_{11}(Y) = s^{-1},$$

$$H_{12}(y) = 2s_1^{-3}(|y^{(a)}| + |y^{(b)}| + M),$$

$$H_{13}(y) = s_1^{-2}(y'^{(a)} + y'^{(b)}),$$

$$H_{22}(y) = s_1^{-2} + 6s_1^{-4}\left[\{y^{(a)}\}^2 + \{y^{(b)}\}^2 + M\right]$$

$$H_{33}(y) = s_1 H_{12}(y)\left(|y''^{(a)}| + |y''^{(b)}|\right) + 2s_1^{-2}\left[\{y'^{(a)}\}^2 + \{y'^{(b)}\}^2\right].$$

Since $y^{(\lambda)}$, $y'^{(\lambda)}$ and $y''^{(\lambda)}$ are continuous function of (λ, Y), it follows that the D_{uv} 's are continuous in $\boldsymbol{\theta}$ and Y.

Appendix 6

Let Y be a positive random variable and assume that the expected value $E[Y^{2a}]$, $E[Y^{2b}]$, $E[Y^a \log Y]^2$ and $E[Y^b \log Y]^2$ are finite. Let Θ and $H_{uv}(y)$ be defined as in Appendix 5. Then, $E[H_{uv}(y) < \infty]$ $\forall u, v$.

Appendix 7

Let $l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y)$ be the approximation to $D_{I_1 \cup I_k}(y) l_1^{(u)}(\boldsymbol{\theta}|y)$ defined in (2.3.9). Let Θ be as in Appendix 5.

Firstly, $|l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y)| \le |l_1^{(u)}(\boldsymbol{\theta}|y)| \le r(y)$ $\forall \boldsymbol{\theta} \in \Theta$ and $Y > 0$, where, letting $c = \max\{|a|, b\}$

$$r(y) = \frac{1}{2}\log 2\pi + |\log s_1| + |\log s_2| + s_1^{-2}\left[\{y^{(a)}\}^2 + \{y^{(b)}\}^2 + M^2\right] + (c+1)|\log y| - A(\kappa).$$

Since $(\log x)^2 < \{y^{(a)}\}^2 + \{y^{(b)}\}^2$, $\mathrm{E}_{g^*}[Y^{2a}] < \infty$ and $\mathrm{E}_{g^*}[Y^{2b}] < \infty$ imply that $\mathrm{E}_{g^*}[r(y)] < \infty$, so that condition (a) of Appendix 4 is satisfied.

Secondly, Choose the set S_i as follows:

$S_i = [0, y_1] \cup [y_1 + (y_{k-1} - y_1)/(i+1), y_{k-1} - (y_{k-1} - y_1)/(i+1)] \cup [y_{k-1}, y_{k-1} + i], i = 1, 2, \cdots$, so that $\bigcup_{i=1}^{\infty} S_i = [0, \infty)$ and $\Pr_{g^*}[(0, \infty) - \bigcup_{i=1}^{\infty} S_i] = 0$, thus verifying condition (b) of Appendix 4.

Finally, we note that S_i is compact $\forall i$ and $l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y)$ is continuous in both $\boldsymbol{\theta}$ and Y, so that $l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y)$ is uniformly continuous on $\Theta \times S_i$. That is, $|l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y) - l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y_1)| < \eta$ for any $\eta > 0$ whenever $\| (\boldsymbol{\theta}^{\mathrm{T}}, y) - (\boldsymbol{\theta}_1^{\mathrm{T}}, y_1) \|_4 < \delta(\eta)$, where $\| \cdot \|_p$ denotes the Euclidean norm in the p-dimensional space. Thus, by setting $\| \boldsymbol{\theta} - \boldsymbol{\theta}_1 \|_3 < \delta(\eta)$ for $y \in S_i$. Hence, $l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y)$ is equicontinuous in $\boldsymbol{\theta}$ for $y \in S_i$.

Therefore, all the assumptions of Appendix 4 hold and we obtain that, as $n \to \infty$

$$\frac{1}{n}\sum_{j=1}^{n} l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y_j) \xrightarrow{a.s.} \mathrm{E}_{g^*}[l_{1,\varepsilon}^{(u)}(\boldsymbol{\theta}|y)]$$

uniformly in $\boldsymbol{\theta} \in \Theta$.

References

Aigner, D.J. and Goldberger, A.S. (1970). Estimation of Pareto's law from grouped observations. *J. Amer. Staitst. Assoc.*, **65**, 712-723.

Altman, D.G. (1998). Categorizing continuous variables. *Encyclopedia of Biostatistics*, ed. by P. Armitage and T. Colton, Vol. 1, 563-567, John Wiley & Sons.

Akaike, H. (1972). Information theory and an extension of the maximum likelihood principle. *The Second international Symposium on Information Theory*, ed. by B. N. Petrov and F. Csaki, 267-281, Budapest, Akademiai Kiado.

Andrews, D.F., Gnanadesikan, R. and Warner, J.L. (1971). Transformation of multivariate data. *Biometrics*, **27**, 825-840.

Aranda-Ordaz, F.J. (1981). On two families of transformations to additivity for binary response data. *Biometrika*, 68(2), 357-364.

Atkinson, A.C. (1982). Diagnostics, transformations and constructed variables (with discussion). *J. Roy. Statist. Soc.*, **B44**, 1-36.

Atkinson, A.C. (1985). *Plots, Transformation and Regression*. Oxford University Press.

Atkinson, A.C. (1986). Diagnostic tests for transformations. *Technometrics*, **28**, 29-37.

Atkinson, A.C., Perocchi, L.R. and Smith, R.L.(1991). Grouped likelihood for shifted power transformation. *J. Roy. Statist. Soc.*, **B53**, 473-482.

Barnard, G.A. (1967). The use of the likelihood function in statistical practice. *The Fifth Berkley Symp. Mathematical Statistics and Probability*, ed. by L. M. LeCam and J. Neyman, **1**, 27-40.

Beckman, R.J. and Johnson, M.E. (1987). Fitting the Student t distribution to grouped data, with application to a particle scattering experiment. *Technometrics*, **29**, 17-22.

Benn, R.T. and Sidebottom, S. (1976). Algorithm AS 95: Maximum-likelihood estimation of location and scale parameters form grouped data. *Appl. Statist.*, **25**, 88-93.

Bickel, P.J. and Doksum, J.A. (1981). An analysis of transformation revisited. *J. Amer. Statist. Assoc.*, **76**, 296-311.

Bland, M. (1995). *An Introduction to Medical Statistics*. Oxford Medical Publications

Bodin, N.A. (1968). On the theory of grouped sample. *Soviet Math. Dokl.*, **9**, 10-13.

Boardman, T.J. (1973). Estimation in compound exponential failure models when the data are grouped. *Technometrics*, **15**, 271-277.

Borkowf, C.B., Gail, M.H., Carroll, R.J. and Gill, R.D. (1997). Analyzing bivariate continuous data grouped into categories defined by empirical quantiles of marginal distributions. *Biometrics*, **53**, 1054-1069.

Box, G.P.E. and Cox, D.R. (1964). An analysis of transformations (with discussion). *J. Roy. Statist. Soc.*, **B26**, 211-252.

Box, G,P.E and Cox, D.R. (1982). An analysis of transformations revisited, rebutted. *J. Amer. Statist. Assoc.*, **77**, 209-210.

Brown, B.W. Jr. and Hollander, M. (1977). *Statistics: A Biomedical Introduction*. John Wiley & Sons.

Burrigde, J. (1981). A note on maximum likelihood estimation for regression models using grouped data. *J. Roy. Statist. Soc.*, **B43**, 41-45.

Carroll, R.J. (1980). A robust method for testing transformation to achieve approximate normality. *J. Roy. Statist. Soc.*, **B42**, 71-78.

Carroll, R.J. and Ruppert, D. (1981). On prediction and the power transformation family. *Biometrika*, **68**, 609-615.

Carroll, R.J. and Ruppert, D. (1985). Transformation in regression: a robust analysis. *Technometics*, **27**, 1-12.

Chen,G. (1995). Generalized log-normal distributions with reliability application. *Computational Statistics & Data Analysis*, **19**, 309-319.

Cheng, R.C. and Amin, N.A.K. (1983). Estimating parameters in continuous univariate distributions with sifted origin. *J. Roy. Statist. Soc.*, **B45**, 394-403.

Cheng, R.C. and Iles, T.C. (1987). Corrected maximum likelihood in non-regular problems. *J. Roy. Statist. Soc.*, **B49**, 95-101

Cheng, R.C. and Taylor, L. (1995). Non-regular maximum likelihood problems. *J. Roy. Statist. Soc.*, **B57**, 3-44.

Cook, R.D. and Wang, P.C. (1983). Transformations and influential cases in regression. *Technometrics*, **25**, 337-344.

Copas, J.B. (1972). The likelihood surface in the linear functional relationship. *J. Roy. Statist. Soc.*, **B34**, 274-278.

Cox, D.R. (1957). Note on grouping. *J. Amer. Statist. Assoc.*, **52**, 543-547.

Cox, D.R. and Snell, E.J. (1981). *Applied Statistics: Principles and Examples*. Chapman & Hall.

Corner, R.J. (1972). Grouping for testing trends in categorical data. *J. Amer. Statist. Assoc.*, **67**, 601-604.

Cramér, H. (1974). *Mathematical Methods of Statistics.* Princeton University Press.

Cressie, N.A.C. (1978). Removing non-additivity from two-way tables with one observation per cell, *Biometrics*, **34**, 505-513.

Domae, A. and Miyahara, H. (1984). Are the cubic roots of body weights distributed normally ?: processing of health data and fitting of them to a type of distribution by means of micro-computer(in Japanese). *Jpn. J. Behav.*, **11**(2), 1-11.

Daniel, W.W. (1987). *Biostatistics: A Foundation for Analysis in the Health Science*. 4th edition, John Wily & Sons.

Denken, J.G. (1983). Approximating conditional moments of the multivariate normal distribution. *SIAM J. Sci. Statist. Comput.*, 4, 720-732.

Duan, N. (1993). Sensitivity analysis for Box and Cox transformation model: contract parameters. *Biometrika*, **80**(4), 885-897.

Draper, N.R. and Cox. D.R. (1969). On distributions and their transformation to normality. *J. Roy. Statist. Soc.*, B31, 472-476.

Draper, N.R. and Hunter, W.G. (1969). Transformation: some examples revisited. *Technometrics*, **11**, 23-40.

Finney, D.J. (1978). *Statistical Method in Biological Assay*. 3rd edition. Charles Griffin.

Fisher, L.D., and Belle. G.V. (1993). *Biostatistics: A Methodology for the Health Science*. John Wiley & Sons.

Flygare, M.E., Austin, J.A. and Buckwalter, R.M. (1985). Maximum likelihood estimation for the 2-parameter Weibull distribution based on interval-data. *IEEE Trans. Reliability*, **34**, 57-59.

Fryer, J.G. and Pethybeigde, R.J. (1972). Maximum likelihood estimation of a linear regression function with grouped data. *Appl. Statist.*, **21**, 142-154.

Fuchs, C. (1978). On test sizes in linear model for transformed variate. *Technometrics*, 20, 291-299.

Fuchs, C. (1979). Comments on a criterion of transformed proposed by Schlesselman. *J. Amer. Statist. Soc.*, 74, 238-239.

Giesbrecht, F. and Kempthorne, O. (1976). Maximum likelihood estimation in three-parameter lognormal distribution. *J. Roy. Statist. Soc.*, **B38**, 257-264.

Gjeddebaek, N.F. (1949). Contributions to the study of grouped observations: application of the method of maximum likelihood in case of normally distributed observations. *Skandinavisk Aktuarietidskrift*, **32**, 135-159.

Gjeddebaek, N.F. (1956). Contributions to the study of grouped observations II: loss of information cased by grouping if normally distributed observations. *Skandinavisk Aktuarietidskrift*, **39**, 154-159.

Gjeddebaek, N.F. (1957). Contributions to the study of grouped observations III: the distribution of estimates of mean. *Skandinavisk Aktuarietidskrift*, **40**, 20-25.

Gjeddebaek, N.F. (1959a). Contributions to the study of grouped observations IV: some comments on simple estimates. *Biometrics*, **15**, 433-439.

Gjeddebaek, N.F. (1959b). Contributions to the study of grouped observations V: three-class grouping of normal distributions. *Skandinavisk Aktuarietidskrift*, **42**, 194-207.

Gjeddebaek, N.F. (1961). Contributions to the study of grouped observations VI. *Skandinavisk Aktuarietidskrift*, **44**, 55-73.

Gjeddebaek, N.F. (1968).Statistical analysis: grouped observations. *International Encyclopedia of the Social Science*, ed. by D.R. Sills, Macmillan & The Free Press.

Gnanadesikan, R. (1977). *Method for Statistical Data Analysis of Multivariate Observations*. John Wiley & Sons.

Goto, M. (1986). Process of statistical data analysis (in Japanese). *Jpn. J. Behav.*, **13**(2), 48-63.

Goto, M. (1992). Extensive views of power transformations: some recent developments. *Invited paper of Honolulu Conference on Computational Statistics as memorial of the Fifth anniversary JSCS*, JAMS, December, 1-5.

Goto, M. (1995). Double power-transformations and their applications. *Invited paper of International Conference on Statistical Computing for Quality and Productivity Improvement*, Seoul, August, 17-19.

Goto, M. and Hatanaka, S. (1985). Diagnostic of power-normal transformation (in Japanese). *Jpn. J. Appl. Statist.*, **14**(2), 69-80.

Goto, M. and Inoue, T. (1980). Some properties of power-normal distribution. *Bull. Biometric. Soc. Japan*, **1**, 28-54.

Goto, M., Inoue, T. and Nagai, T. (1980). A bivariate power-normal distribution (in Japanese). *Proceedings of the Autumn Meeting of Mathematical Society of Japan*, 77-78.

Goto, M., Inoue, T. and Nagai, T. (1981a). A bivariate power-normal distribution (in Japanese). *Proceedings of the Spring Meeting of Mathematical Society of Japan*, 97-98.

Goto, M., Inoue, T. and Nagai, T. (1981b). A bivariate power-normal distribution (in Japanese). *Proceedings of the Autumn Meeting of Mathematical Society of Japan*, 63-64.

Goto, M., Inoue, T. and Nagai, T. (1982). The bivariate power normal distributions and its applications. *Proceedings of the First Korea and Japan Joint Conference of Statistics*, Okayama, July 16-17.

Goto, M., Inoue, T. and Tsuchiya, Y. (1984). On estimation of parameters in power-normal distribution. *Bull. Inform. Cyber.*, **21**(1-2), 41-53.

Goto, M., Inoue, T. and Tsuchiya, Y. (1986). Power-transformation for ordered categorical data. *Behaviometrika.*, **19**, 17-39.

Goto, M., Inoue, T. and Tsuchiya, Y. (1987). Double power-transformation and its performance: an extensive version of Box-Cox transformation. *J. Japan. Statist. Soc.*, **17**(2), 149-163.

Goto, M., Matsubara, Y. and Tsuchiya, Y. (1983). Power-normal distribution and its applications. *Rep. Stat. Appl. Res.*, JUSE., **30**(3), 8-28.

Goto, M., Uesaka, H. and Inoue, T. (1979). Some linear models for power transformed data. *Invited presentation at the Tenth International Biometric Conference at Brazil*, August, 6-10(Res. Rep. NO.93, *Res. Inst. Fond. Inform. Sc.*, Kyusyu University).

Goto, M. and Yamamoto, S. (1998). The extended power-transformations and their performances. *Submitted.*

Goto, M., Yamamoto, Y. and Inoue. T. (1991). Parameter estimation for power-normal distribution: asymptotic behavior of the estimators (in Japanese). *Bull. Comput. Statist. Jpn.*, **4**(1), 45-60.

Guerrero, V.M. and Johnson, R.A. (1982). Use of the Box-Cox transformation with binary response model . *Biometrika*, **69**(2), 309-314.

Haitovsky, Y. (1973). *Regression Estimation from Grouped Observations*. Griffin.

Haitovsky, Y. (1982). Grouped data. *Encyclopedia of Statistical Science*, Vol. 3, ed. by S. Kotz and N.L. Johnson, 527-526, John Wiley & Sons.

Hamasaki, T (1998). Inference based on grouped observation from bivariate power-normal distribution. *To appear in J. Comput. Statist. Jpn.*

Hamasaki, T. and Goto, M. (1996). Normalizing power-transformation to achieve scale-invariance of observations (in Japanese). *Bull. Comput. Statist. Jpn.*, **9**(1), 37-53.

Hamasaki, T. and Goto, M. (1997). Applications of fitting power-normal distribution to grouped observations. *Proceedings of the Ninth Korea and Japan Joint Conference of Statistics*, Cheju, December 5-6.

Hamasaki, T. and Goto, M. (1998a). Fitting power-normal distribution to grouped observations (in Japanese). *Jpn. J. Behav.*, **25**(1), *Printed.*

Hamasaki, T. and Goto, M. (1998b). Is log-transformation appropriate for a variable following Poisson distribution? (in Japanese). *To appear in Jpn. J. Appl. Statist.*

Hamasaki, T. and Goto, M. (1998c). Some transformation models for ordered categorical data and their applications. *Proceedings of the Nineteen International Society of Clinical Biostatistics*, Dundee, August 24-28

Hamasaki, T. and Goto, M. (1998d). Statistical inference of grouped observations combined with ungrouped observations. *Proceedings of the Third Conference of the Asian Regional Section of the International Association of Statistical Computing*, Manila, December 2-4.

Hamdan, M.A. (1970). The equivalence of tetrachoric and maximum likelihood estimates of ρ in 2×2 tables. *Biometrika*, **57**, 212-215.

Hatanaka, S., Goto, M. and Nagai, T. (1980). Evaluation of some performance of power-normal transformation in multi-sample case (in Japanese). *Proceedings of the Eighth Conference of the Behaviormetric Society of Japan*, 113-114.

Hatanaka, S., Inoue, T. and Goto, M. (1981). Discrimination on power-normal distribution (in Japanese). *Proceedings of the Ninth Conference of the Behaviormetric Society of Japan*, 132-133.

Heitjan, D.F. (1989). Inference from grouped continuous data: a review. *Statistical Science*, **4**(2), 164-183.

Heitjan, D.F. (1998). Grouped data. *Encyclopedia of Biostatistics*, ed. by P. Armitage and T. Colton, Vol. 1, 1784-1785, John Wily & Sons.

Hernandez, F. and Johnson, R.A. (1980). The large sample behavior of transformations to normality. *J. Amer. Statist. Assoc.*, **75**, 855-861.

Hernandez, F. and Johnson, R.A. (1981). Transformation of a discrete distribution to near normality. *Statistical Distributions in Scientific Work*, **5**, 259-270.

Hinkley, D.V. (1977). On quick choice of power transformation. *Appl. Statist.*, **26**, 67-69.

Hinkley, D.V. and Runger, G. (1984). The analysis of transformed data (with discussion). *J. Amer. Statist. Assoc.*, **79**, 302-320.

Holmes, J. (1974). The relative burden of higher gasoline prices. *In Five Thousand American Families: Patterns of Economic Progress*. Vol. 4, Ann Arbor, Institute for Social Research.

Jimura, Y., Kawai, N. and Goto, M. (1996). Bivariate power-normal distribution and its evaluations (in Japanese). *Proceedings of thirteenth Oita Meeting about Statistical Science.*

Johnson, N.L., Kotz, S. and Balakrishnan, N.(1994). *Continuous Univariate Distributions*. John Wiley & Sons.

Kariya, T. (1975). Menarche age distribution estimated by the methods of maximum likelihood on incomplete quantal response data(in Japanese). *Kawasaki Medical J.*, **1**, 75-83.

Kariya, T. (1979). The eruption age distribution of deciduous teeth (in Japanese). *Jpn. J. Appl. Statist.*, **9**(2), 73-81.

Kariya, T. and Akasaka, M(1976). The eruption age distribution of deciduous teeth: maximum likelihood estimation of parameters based on incomplete quantal response data (in Japanese). *Jpn. J. Appl. Statist.*, **5**(1), 3-18.

Kawai, N., Jimura, Y. and Goto, M. (1996). Bivariate power-normal distribution and its applications (in Japanese). *Proceedings of the Tenth conference of Japanese Society of Computational Statistics*, 96-99.

Kempthorne, O. (1966). Some aspects of experimental inference. *J. Amer. Statist. Assoc.*, **61**, 11-34.

Kempthorne, O. and Folks, L. (1971). *Probability, Statistics, and Data Analysis*. Iowa University Press.

Kendall, M.G. and Buckland, W.R. (1982). *A Dictionary of Statistical Terms*. 4th Edition, The Longman Group[Statistics Group of Chiba University (1987). *Kendall toukeigaku yougo jiten*. Maruzen].

Kitagawa, T and Inaba, M.(1979). *Toukeigaku-turon* (in Japanese). 2nd edition, Kyoritsu.

Kulldorff, G. (1962). *Contributions to the Theory of Estimation From Grouped and Partially Grouped Samples*. Almqvist & Wiksell.

Kruscal, J.B. (1968). Statistical analysis: transformation of data. *International Encyclopaedia of the Social Science*, the Macmillan Co. & the Free Press.

Lindley, D.V. (1950). Grouping corrections and maximum likelihood equations. *Proc. Camb. Philos. Soc.*, **46**, 106-110.

Lindsey, J.K. (1995). *Modelling Frequency and Count Data*. Oxford University Press.

Mangasarian, O.L. (1969). *Nonlinear Programming*. McGraw-Hill.

Manley, (1976). Exponential data transformation. *The Statistician*, 25, 37-42.

Mardia, K.V. (1970). *Families of Bivariate Distribution*. Griffin.

Moor, P.G. and Tukey, J.W. (1954) Answer to query 112. *Biometrics*, 10, 562-568.

Morgan, T.M. and Elashoff, R.M. (1986). Effect of categorizing a continuous covariate on the comparison of survival time. *J. Amer. Statist. Assoc.*, **81**, 917-921.

Mori, T. (1975). On an optimal class interval of histogram (in Japanese). *Jpn. J. Appl. Statist.*, **4**(1), 17-24.

Matsubara, Y. and Goto, M. (1979). A censored regression analysis of survival time data based on power-normal distribution (in Japanese). *Jpn. J. Appl. Statist.*, **9**(1), 43-51.

Nabeya, S. (1983). On maximum likelihood for grouped observations (in Japanese). *Jpn. J. Appl. Statist.*, **12**(2), 59-67.

Nagahata. H.(1984). A consideration on the number of class and class boundary based on Fisher information criterion (in Japanese). *Proceedings of the Fifty-second Japanese Statistical Society,* 295-296.

Nelder, J.A. (1968). Weighted regression, quantal response data, and inverse polynomials. *Biometrics*, **24**, 979-985.

Pagono, M. and Gauvreau, K. (1993). *Principles of Biostatistics*. Duxbury.

Pearson, K. (1901) Mathematical contributions to the theory of evaluation VII. On the correlation of characters not quantitatively measurable. *Phil. Trans. of R. Soc. of London,* A195, 1-46.

Pethybridge, R.J. (1975). Maximum likelihood estimation of a linear regression function with grouped data. *Appl. Statist.*, **24**, 28-41.

Pettitt, A.N. (1985). Re-weighted least squares estimation with censored and grouped data: an application of the EM Algorithm. *J. Roy. Statist. Soc.*, **B47**, 253-260.

Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*. John Wiley & Sons.

Rirchie-Scott, A. (1918). The correlation coefficient of a polychoric table. *Biometrika*, **12**, 93-133.

Rubin, H. (1956). Uniform convergence of random functions with application to statistics. *Ann. Math. Statist.*, 27, 200-203.

Sakamoto, Y., Ishiguro, M. and Kitagawa, G. (1986). *Akaike Information Criterion Statistics*. D. Reidel Publishing Company.

Sakia, R.M. (1992). The Box-Cox transformation technique: a review. *The Statistician*, **41**, 169-178.

Schader ,M. and Shmid, F. (1984). Computation of maximum likelihood estimates for μ and σ from a grouped sample of a normal population: a comparison of algorithms. *Statist. Hefete*, **25**, 245-258.

Shclesselman, J. (1971). Power families: a note on the Box and Cox transformation. *J. Roy. Statist. Soc.*, **B33**, 307-311.

Schlesselman, J. (1973). Data transformation in two-way analysis of variances. *J. Amer. Statist. Assoc.*, **68**. 369-383.

Siegel, A.F. and Morgan, C.J. (1996). *Statistics and Data Analysis: An Introduction*. 2nd edition, John Wily & Sons.

Sterling, W.D. (1984). Iteratively reweighted least squares for models with a linear part. *Appl. Statist.*, **33**, 7-17.

Stevens, W.L. (1948). Control by gauging (with discussion). *J. Roy. Statist. Soc.*, **B10**, 54-98.

Stuart, A. and Ord, J.K. (1986). *Kendall's Advanced Theory of Statistics*. Vol.1, 6th edition, Edward Arnold.

Sturges, H. (1926). The choice of a class interval. *J. Amer. Statist. Assoc.*, **21**, 65-66.

Sugiura, N. (1980). Fitting of some distributions to total score of the common first-stage examination for university entrance (in Japanese). *Jpn. J. Appl. Statist.*, **9**(2), 95-107.

Sugiura, N. (1981). Fitting of some distributions to total score of the common first-stage examination for university entrance II (in Japanese). *Jpn. J. Appl. Statist.*, **10**(1), 39-52.

Swan, A.V. (1969). Algorithm AS 16: Computing maximum-likelihood estimates for parameters of the normal distribution form grouped and censored data. *Appl. Statist.*, **18**, 110-114.

Tasaki, T and Goto, M. (1990). On "transformation" in generalized linear models (in Japanese). *Bull. Comput. Statist. Jpn.*, **3**(1), 45-53.

Tallis, G.M. (1967). Approximate maximum likelihood estimates from grouped data. *Technometrics*, **9**, 599-606.

Tallis, G.M. and Young, S.S. (1962). Maximum likelihood estimation of parameters of the normal, log-normal, truncated normal and bivariate normal distribution from grouped data. *Austral. J. Statist.*, **4**, 49-54.

Tukey, J.W. (1957). On the comparative anatomy of transformations. *Ann. Math. Statist.*, **28**, 602-632.

Uesaka, H and Goto, M. (1980). An analysis of clinical chemistry data based on power-normal distribution(in Japanese). *Jpn. J. Appl. Statist.*, **9**(1), 23-33.

Uesaka, H. and Goto, M. (1982). Some properties of Box-Cox estimates of transforming parameters in power-normal distributions. *Bull. Biometric. Soc. Japan*, **3**, 1-22.

Uesaka, H., Botan, Y. and Goto, M. (1981). Fitting of some models in bioassay (in Japanese). *J. Jap. Statist. Soc.*, **11**(1), 88-94.

Uesaka, H. and Goto, M. (1979). On multivariate normality test (in Japanese). *Research Report of the Institute of Mathematical Science*, Kyoto University, NO.345, 64-71.

Winkelmann, R. (1994). *Count Data Models*. Springer-Verlag.

Wolynetz, M.S. (1979a). Algorithm AS 138: Maximum likelihood estimation from confined and censored samples. *Appl. Statist.*, **28**, 185-195.

Wolynetz, M.S. (1979b). Algorithm AS 139: Maximum likelihood estimation in a linear model from confined and censored normal data. *Appl. Statist.*, **28**, 195-206.

Wood, J.T. (1974). An extension of the analysis of transformation of Box and Cox. *Appl. Statist.*, **3**, 278-283.

Yamamoto, S. and Goto, M. (1998). Power-additive transformation and its application: double power-additive transformation. *Submitted*

Yoneda, K. and Uchiyama, M. (1956). Some estimations in the case of relatively large class interval. *Yokohama Math. J.*, **4**, 99-118.

Yule, G.U. and Kendall, M.G. (1968). *An Introduction to the Theory of Statistics*. Griffin.