

ON ESTIMATION OF PATRAMETERS IN POWER-NORMAL DISTRIBUTION

Goto, Masashi
SHIONOGI Kaiseki Center, SHIONOGI & CO., LTD.

Inoue, Toshiaki
SHIONOGI Kaiseki Center, SHIONOGI & CO., LTD.

Tsuchiya, Yoshihide
SHIONOGI Kaiseki Center, SHIONOGI & CO., LTD.

<https://doi.org/10.5109/13355>

出版情報 : Bulletin of informatics and cybernetics. 21 (1/2), pp.41-53, 1984-03. Research
Association of Statistical Sciences

バージョン :

権利関係 :



ON ESTIMATION OF PARAMETERS IN POWER-NORMAL DISTRIBUTION

By

Masashi GOTO*, Toshiaki INOUE* and Yoshihide TSUCHIYA*

Abstract

Data analysed in medical and biomedical research are often found to have skewed distributions, so transformations are applied to the data in order to render them more tractable by ordinary normal theory analysis. The power-normal distribution, proposed by Goto, Uesaka and Inoue [4], is a distribution which specifies original observations before Box-Cox transformation. In this paper, we mention the maximum likelihood method to estimate the parameters of the power-normal distribution, and consider the sampling behaviours of the maximum likelihood estimates. And we show, through simulation experiment, that biases of the transforming parameter, resulted from neglecting the probability proportional constant term, do not seem to cause any troubles in case the aim is to attain approximate normality of the transformed observations, but they do have considerable influences on the estimates of the parameters of near-normal distribution after the transformation.

1. Introduction

The power-normal transformation of a positive random variable Y is defined as

$$Z(\lambda) = \begin{cases} \frac{Y^\lambda - 1}{\lambda} & ; \lambda \neq 0 \\ \log Y & ; \lambda = 0, \end{cases} \quad (1.1)$$

where λ is called the transforming parameter [2]. The power-normal distribution (PND) is defined as the distribution having the probability density function (p.d.f.)

$$g(y; \lambda, \mu, \sigma) = \frac{y^{\lambda-1}}{A(k)\sqrt{2\pi\sigma}} \exp \left\{ -\frac{(z(\lambda) - \mu)^2}{2\sigma^2} \right\}, \quad (1.2)$$

where λ , μ and σ are parameters of the PND, and $A(k)$ is the probability proportional constant term of the PND given by

$$A(k) = \begin{cases} \Phi(k) & : \lambda > 0 \\ 1 & : \lambda = 0 \\ \Phi(-k) & : \lambda < 0, \end{cases}$$

* SHIONOGI Kaiseki Center, SHIONOGI & CO., LTD., 22-41, 1-chome, Izumicho, Suita City, Osaka, Japan.

where $\Phi(k)$ is cumulative distribution function (c.d.f.) of standard normal distribution [4].

The original object of the power-normal transformation is to transform the distribution of the observed data into normal distribution with mean μ and variance σ^2 . However, strictly speaking, the transformed distribution becomes near-normal distribution truncated at right hand side or left hand side depending on the sign of the transforming parameter [5]. Then, k becomes a standardized truncation point of the near-normal distribution, given by a function of λ , μ and σ . And the term $1-A(k)$ is truncated probability. Thus, to estimate the parameters of the PND, namely λ , μ and σ^2 , evaluation of the magnitude of $A(k)$ plays very important role.

In ordinary power-normal transformation, λ was estimated by maximum likelihood (ML) method assuming $A(k)=1$ implicitly, and performances of the transformation only were evaluated from aspects of simplicity of structure for the assumed model, constancy of error variance, near-normality and independence of the transformed observations. Thus, there occur no trouble when the aim of the transformation is to attain approximate normality. However, if the aim is to identify the PND on the original scale of the observations, there occur some troubles. In other words, in order to obtain the consistent estimators of λ , μ and σ^2 by ML method, it is necessary to solve the ML equations taking the magnitude of $A(k)$ into account.

The object of this study is to apply ML method to the estimation of parameters of the PND, and to evaluate, through simulation experiment, the influences of $A(k)$ on the ML estimates of parameters when the magnitude of $A(k)$ is counted for and not.

In our simulation, we fixed 3 different magnitude of $A(k)$, 6 values of transforming parameter λ and 3 values of τ (that coincides with coefficient of variation after the transformation), then generated 50 samples corresponding to prefixed each sample size, 10, 20, 50, 100 and 200, and examined the sampling behaviours of the ML estimates for these samples.

In Section 2, we indicate some difficult points for ordinary ML method when the magnitude of $A(k)$ is counted for, and propose conditional ML method to get rid of these difficulties. The design of simulation experiment to be examined and the way to evaluate the results of the examination are described in Section 3, and the results are summarized in Section 4. Interpretation and discussion of the results are given in Section 5. In Section 6, we investigate some relative performances of other estimation methods, including Jackknifing methods, in relation to the discussion of Section 5. In Section 7, we conclude the study and note some remained problems.

2. Parameter Estimation by ML Method

If n observations y_1, y_2, \dots, y_n have the PND with parameters λ , μ and σ^2 independently, the log likelihood function is given by

$$\begin{aligned} \log L(y; \lambda, \mu, \sigma) = & -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (z_i(\lambda) - \mu)^2 \\ & + (\lambda - 1) \sum_{i=1}^n \log y_i - n \log A(k). \end{aligned} \quad (2.1)$$

The procedure to estimate the parameters λ , μ and σ^2 , setting $A(k)=1$, by ordinary ML method is referred to M_1 -method. In other words, this M_1 -method is one which has been proposed by Box and Cox [2]. On the other hand, the procedure to estimate those parameters allowing for magnitude of $A(k)$ is called M_2 -method for which an algorithm is proposed later in this section.

Since the assumption of $A(k)=1$ in M_1 -method means that $\{z_i\}$ given by (1.1) have exact normal distribution, the estimates of μ and σ^2 are given as mean and variance of $\{z_i\}$ ($i=1, 2, \dots, n$). However, $\{z_i\}$ have in fact near-normal distribution. Thus, denoting estimates of μ and σ in M_1 -method by μ^* and σ^* , μ^* and σ^* are not consistent estimators of μ and σ , respectively. On the other hand, M_2 -method is an estimation procedure for which the transformed data is assumed to have exactly truncated-normal distribution. For M_2 -method, the estimates of μ and σ are obtained conditionally on the truncation point (k) not independent of μ and σ . The estimation procedures for M_2 -method are based on Newton-Raphson iteration method accompanied by a grid approach. 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:

Step 1. Set $a_0=0.5$ and $a_1=1.0$.

Step 2. Divide the interval $[a_0, a_1]$ by $m+1$ knots A_1, A_2, \dots, A_{m+1} where A_i are calculated by

$$A_i = A_1 + \frac{A_{m+1} - A_1}{m}(i-1),$$

$$A_1 = a_0 \quad \text{and} \quad A_{m+1} = a_1.$$

Step 3. Select $\hat{\lambda}_1$ for initial value of $\hat{\lambda}_l$.

Step 4. Putting $A(k)=A_i$, and obtain \hat{k}_l by solving equation

$$k_l = \begin{cases} \Phi^{-1}(A_i); & \lambda > 0 \\ -\Phi^{-1}(A_i); & \lambda < 0. \end{cases}$$

Step 5. Estimate μ_l and σ_l by

$$\hat{\mu}_l = \begin{cases} \bar{z} - \frac{z_0}{z_0 + \hat{k}_l} \left(\bar{z} + \frac{1}{\hat{\lambda}_l} \right); & \hat{\lambda}_l > 0 \\ \bar{z} - \frac{z_0}{z_0 - \hat{k}_l} \left(\bar{z} + \frac{1}{\hat{\lambda}_l} \right); & \hat{\lambda}_l < 0 \end{cases}$$

and

$$\hat{\sigma}_l^2 = \begin{cases} s^2 + \frac{z_0}{z_0 + \hat{k}_l} \left(\bar{z} + \frac{1}{\hat{\lambda}_l} \right)^2; & \hat{\lambda}_l > 0 \\ s^2 - \frac{z_0}{z_0 - \hat{k}_l} \left(\bar{z} + \frac{1}{\hat{\lambda}_l} \right)^2; & \hat{\lambda}_l < 0, \end{cases}$$

respectively, where $z_0 = \Phi(k)/A_0$, $s^2 = \sum_{i=1}^n (z_i - \bar{z})^2$, $\bar{z} = \sum_{i=1}^n z_i/n$, and $z_i = (y_i^2 - 1)/\lambda$.

Further, $\hat{\lambda}_l$ is the estimate of λ in the l -th iteration.

Step 6. Estimate $\hat{\lambda}_{l+1}$ using $\hat{\lambda}_l$, $\hat{\mu}_l$, $\hat{\sigma}_l$ and \hat{k}_l by equation

$$\hat{\lambda}_{l+1} = \hat{\lambda}_l + \left[\frac{\partial \log L}{\partial \lambda} / \left(-\frac{\partial^2 \log L}{\partial \lambda^2} \right) \right] \Big|_{\lambda = \hat{\lambda}_l, \mu = \hat{\mu}_l, \sigma = \hat{\sigma}_l, k = \hat{k}_l},$$

where

$$\begin{aligned}\frac{\partial \log L}{\partial \lambda} &= -\sum_{i=1}^n f(y_i) \frac{\partial f(y_i)}{\partial \lambda} + \sum_{i=1}^n \log y_i - \frac{n}{A(k)} \frac{\partial A(k)}{\partial \lambda}, \\ -\frac{\partial^2 \log L}{\partial \lambda^2} &= \sum_{i=1}^n \left\{ \left(\frac{\partial f(y_i)}{\partial \lambda} \right)^2 + f(y_i) \frac{\partial^2 f(y_i)}{\partial \lambda^2} \right\} \\ &\quad - \frac{n}{A(k)} \left\{ \frac{1}{A(k)} \left(\frac{\partial A(k)}{\partial \lambda} \right)^2 - \frac{\partial^2 A(k)}{\partial \lambda^2} \right\}, \\ f(y_i) &= \frac{y_i}{\lambda \sigma} - k, \\ \frac{\partial f(y_i)}{\partial \lambda} &= \frac{1}{\lambda^2 \sigma} \{ (\log y_i^2 - 1) y_i^2 + 1 \}, \\ \frac{\partial^2 f(y_i)}{\partial \lambda^2} &= \frac{1}{\lambda^3 \sigma} \{ (\log y_i^2)^2 - 2(\log y_i^2 - 1) y_i^2 - 2 \}\end{aligned}$$

and

$$\frac{\partial A(k)}{\partial \lambda} = \begin{cases} -\frac{\Phi(k)}{\lambda^2 \sigma}; & \lambda > 0 \\ \frac{\Phi(k)}{\lambda^2 \sigma}; & \lambda < 0 \end{cases} \quad \frac{\partial^2 A(k)}{\partial \lambda^2} = \begin{cases} \frac{k-2\lambda\sigma}{\lambda^4 \sigma} \Phi(k); & \lambda > 0 \\ -\frac{k+2\lambda\sigma}{\lambda^4 \sigma} \Phi(k); & \lambda < 0. \end{cases}$$

- Step 7. If $|\hat{\lambda}_{l+1} - \hat{\lambda}_l| > \varepsilon$, replace l by $l+1$ and return to step 4, where ε is a constant to judge convergence. In our simulation experiment ε is set at 0.01.
- Step 8. Calculate the log-likelihood from (2.1) using $\hat{\lambda}_{l+1}$, $\hat{\mu}_{l+1}$, $\hat{\sigma}_{l+1}$ and \hat{k}_{l+1} .
- Step 9. If $i < m$, replace i by $i+1$ and return to Step 3.
- Step 10. Choose A_j that gives the maximum log likelihood in knots A_1, \dots, A_{m+1} .
- Step 11. If $|a_0 - a_1| > \delta$, replace $a_0 = A_{j-1}$ and $a_1 = A_{j+1}$ and return to Step 2, where δ is a constant to judge convergence and A_{j-1} and A_{j+1} are lower and upper adjacent knots of A_j ($j=2, 3, \dots, m$), respectively.
- Step 12. Stop the iteration, and denote $\hat{\lambda}_{l+1,j}$, $\hat{\mu}_{l+1,j}$ and $\hat{\sigma}_{l+1,j}$ by $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\sigma}$, respectively.

3. Design of Simulation Experiment and the Way to Evaluate the Results

The log likelihood functions of M_1 -method and M_2 -method are non-linear with respect to the parameter λ , so the estimate of λ can not be presented explicitly. And the performances of the transformation can not be evaluated without observations.

Thus, we performed some simulation experiments in which main aims were;

- (1) to confirm the transformed data, using λ^* estimated by M_1 method, satisfy normality regardless the magnitude of truncation or sample size,
- (2) to evaluate how larger the biases of λ^* , μ^* and σ^* estimated by M_1 -method in fact are than those of $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\sigma}$ estimated by M_2 -method, and
- (3) to check how $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\sigma}$ are influenced by the magnitude of truncation for M_2 -method.

In the simulation, we first fixed the value of k , λ and τ , where τ is defined σ/μ

and coincides with coefficient of variation after the transformation. Three levels of k ($k=1, 2$ and 3) were defined corresponding to the values of $1-A(k)=15.87, 2.88$ and 0.14% . The values of λ were determined in 6 ways corresponding to the 6 typical shapes of the PND (Fig. 1), namely $\lambda>1$ (J-shaped distribution), $\lambda=1$ (truncated normal distribution), $c<\lambda<1$, $\lambda=c$ (exponential distribution), $0<\lambda<c$ (exponential distribution) and $\lambda<0$ (L-shaped distribution), where $c=4/(k^2+4)$. And 3 values of τ ($\tau=2, 4$ and 16) were selected. The selected values of k, λ and τ are shown in Tab. 1 and Tab. 2.

Table 1. The truncation point (k) and truncated probability

truncation point k	truncated probability $1-A(k)$	c $4/(k^2+4)$	τ
1	0.1587	0.8	2
2	0.0288	0.5	4
3	0.0014	0.3077	16

Table 2. Selected truncation points (k) and the transforming parameters (λ)

k	$\lambda>1$	$\lambda=1$	$c<\lambda<1$	$\lambda=c$	$0<\lambda<c$	$\lambda<0$
1	2.0	1.0	0.90	0.80	0.40	-1.0
2	2.0	1.0	0.75	0.50	0.25	-1.0
3	2.0	1.0	0.60	0.3077	0.20	-1.0

Incidentally, once the values of k, λ and τ are given, μ and σ are given by

$$\mu = \frac{1}{\lambda(k\tau-1)},$$

$$\sigma = \frac{\tau}{\lambda(k\tau-1)},$$

respectively.

Then, we specified sample size in 5 ways, namely $n=10, 20, 50, 100$ and 200 , and generated 50 samples, using c.d.f. of the PND for all combinations of k, λ, τ and n .

To evaluate the influences of k and n on the estimates of λ, μ and σ^2 for M_1 -method and M_2 -method, we calculated mean bias (MB);

$$\frac{1}{50} \sum_{i=1}^{50} \hat{\theta}_i^{(j)} - \theta^{(j)}, \quad j=1, 2, 3$$

and mean squared error (MSE);

$$\frac{1}{50} \sum_{i=1}^{50} (\hat{\theta}_i^{(j)} - \theta^{(j)})^2, \quad j=1, 2, 3$$

based on 50 samples, where $\theta^{(1)}=\lambda$, $\theta^{(2)}=\mu$ and $\theta^{(3)}=\sigma$.

Then, considering $\{k\}$, $\{\lambda\}$, and $\{\tau\}$ as factors of variation on the MB and the MSE, we evaluated the effect of those factors by 3-way ANOVA.

For M_1 -methd, transforming 50 samples for all combinations of $\{k\}$, $\{\lambda\}$, $\{\tau\}$, and $\{n\}$ using the estimated λ^* , we calculated skewness and kurtosis to check normality of these sample distributions after the transformation. Then, the number of samples

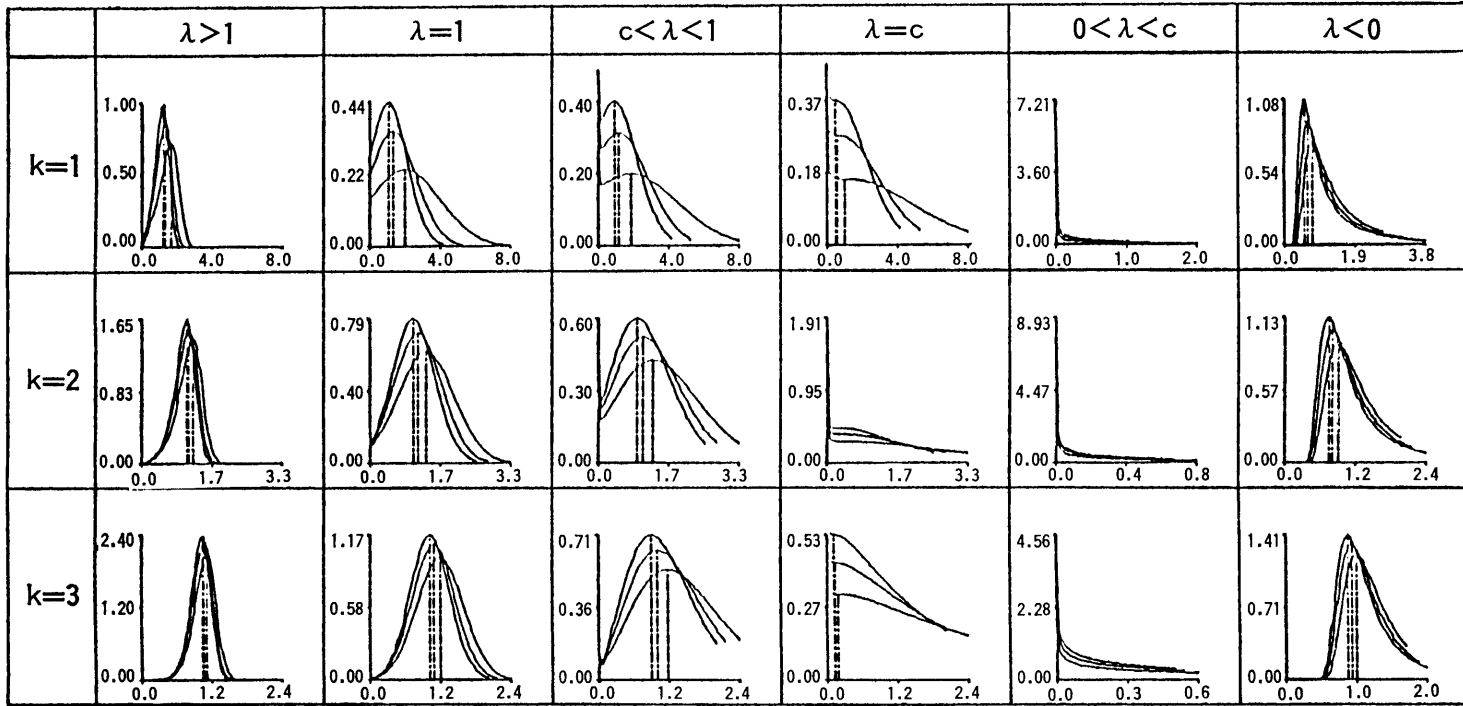


Figure 1. Typical shapes of the power-normal distribution

which were declared to be significant at level 0.05 by the normality tests on skewness and kurtosis in the 50 samples were counted. And we applied 3-way ANOVA with factors k , λ and τ to the number.

4. Results of the Simulation Experiment

4.1 The case of M_1 -method

Results of the normality test on the skewness for the original and the transformed data are shown in Tab. 3. It was seen from the table that normality after the transformation was rejected for a few samples in the cases $n=10$ and $n=20$, but no sample was rejected in the cases $n=50$, 100 and 200.

The ANOVA based on the number of samples which were declared to be significant in each 50 samples, showed that the effects of $\{k\}$ and $\{n\}$ were significant at level 0.05 except for the range $0 < \lambda < c$. For ranges $\lambda < 0$, $\lambda = c$ and $c < \lambda < 1$, the proportions of variation explained (PVE) by $\{n\}$ were from 80 to 90% and by $\{k\}$ were from 10 to 20%. On the contrary, for ranges $\lambda = 1$ and $1 < \lambda$ the PVE by $\{n\}$ were from 11 to 43% and by $\{k\}$ were from 35 to 55%.

Normality test on the kurtosis of the transformed data did not show uniform tendencies on the counts of samples which were declared to be significant at level 0.05 in each 50 samples, depending on the ranges of λ . Normality was rejected about 5 to 10 times for ranges $\lambda = 1$ and $\lambda > 1$. The ANOVA on the counts showed that the effects of $\{k\}$ and $\{n\}$ were significant at level 0.05 for ranges $\lambda = 1$ and $\lambda > 1$. For other ranges only variation of $\{n\}$ was significant at level 0.05.

The behaviors of the estimates λ^* , μ^* and σ^* showed similar tendencies for $n=10$, 20, 50, 100 and 200. Namely, the MB of λ^* were relatively large for all the values of $\{n\}$ in the case $k=1$ and $\lambda > 1$ (the maximum MB was -1.1376 for $n=10$ and $\tau=16$), and were relatively small in the case $k=3$ and $0 < \lambda < c$ (the minimum MB was -0.0080 for $n=200$ and $\tau=2$).

On the other hand, the MB of μ^* and σ^* were relatively large in the case $k=1$ and $0 < \lambda < c$ and were small in the cases $k=3$, $\lambda > 1$ and $k=3$, $\lambda < 0$. The maximum MB of μ^* and σ^* were 0.5131 ($k=1$, $n=10$, $\tau=16$) and -2.2872 ($k=1$, $n=10$, $\tau=2$), and the minimum MB were -0.0015 ($k=3$, $n=200$, $\tau=4$) and -0.0044 ($k=3$, $n=200$, $\tau=16$). Particularly, the MB of the estimates λ^* , μ^* and σ^* are shown in Tab. 4 for $n=200$.

The ANOVA based of the MB of λ^* showed that the effects of $\{k\}$ and $\{n\}$ were significant at level 0.05 for all the ranges of λ , and the interaction effect between $\{k\}$ and $\{n\}$ also was significant except for the range $c < \lambda < 1$. The PVE by sources $\{k\}$, $\{n\}$ and their interaction $\{k \times n\}$ for each range of λ were 56.4%, 14.5%, 11.6%, for $\lambda > 1$, 65.1%, 11.3%, 5.9%, for $\lambda = 1$, 79.7%, 4.2%, 1.8%, for $0 < \lambda < 1$, 80.3%, 7.7%, 7.7%, for $\lambda = c$, 80.5%, 7.2%, 2.4%, for $0 < \lambda < c$, and 65.9%, 13.6%, 5.4%, for $\lambda < 0$, respectively.

Table 3. The number of samples in 50 which were declared significant by the normality test on skewness at level 0.05 (B; Before the transformation, A; After the transformation).

λ	k	τ	n=10		n=20		n=50		n=100		n=200	
			B.	A.	B.	A.	B.	A.	B.	A.	B.	A.
$\lambda > 1$	1	2	13	0	9	0	4	0	6	0	14	0
		4	12	1	9	0	5	0	6	0	11	0
		16	11	1	8	0	6	0	8	0	17	0
	2	2	14	4	13	0	28	0	37	0	44	0
		4	14	1	17	0	21	0	37	0	47	0
		16	16	1	16	0	17	0	39	0	48	0
$\lambda = 1$	1	2	14	1	21	0	23	0	41	0	49	0
		4	15	1	21	0	26	0	37	0	49	0
		16	22	0	7	0	27	0	40	0	50	0
	2	2	16	1	4	0	4	0	9	0	21	0
		4	10	1	4	0	6	0	17	0	9	0
		16	14	3	6	0	6	0	8	0	18	0
$c < \lambda < 1$	1	2	4	1	9	0	4	0	3	0	2	0
		4	11	0	11	0	2	0	3	0	3	0
		16	14	0	9	0	7	0	3	0	6	0
	2	2	25	0	23	0	35	0	49	0	50	0
		4	22	0	18	0	34	0	44	0	50	0
		16	17	1	18	0	33	0	46	0	50	0
$\lambda = c$	1	2	7	0	11	0	23	0	39	0	50	0
		4	12	1	16	0	31	0	41	0	50	0
		16	7	0	21	0	27	0	43	0	48	0
	2	2	17	0	17	0	30	0	44	0	50	0
		4	14	0	24	0	23	0	45	0	50	0
		16	8	0	17	0	26	0	42	0	48	0
$0 < \lambda < c$	1	2	18	0	21	0	45	0	50	0	50	0
		4	20	0	28	0	40	0	49	0	50	0
		16	28	1	27	0	44	0	49	0	50	0
	2	2	23	1	38	0	46	0	50	0	50	0
		4	29	1	28	0	50	0	50	0	50	0
		16	29	0	32	0	46	0	50	0	50	0
$\lambda < 0$	1	2	36	2	42	0	50	0	50	0	50	0
		4	33	0	45	0	50	0	50	0	50	0
		16	30	1	43	0	50	0	50	0	50	0
	2	2	45	0	47	0	50	0	50	0	50	0
		4	44	0	47	0	50	0	50	0	50	0
		16	45	2	46	0	50	0	50	0	50	0
$\lambda < 0$	1	2	45	0	50	1	50	0	50	0	50	0
		4	44	0	48	1	50	0	50	0	50	0
		16	44	1	49	0	50	0	50	0	50	0
	2	2	44	0	49	0	50	0	50	0	50	0
		4	42	1	49	0	50	0	50	0	50	0
		16	45	0	50	1	50	0	50	0	50	0
$\lambda < 0$	1	2	48	1	49	0	50	0	50	0	50	0
		4	46	0	49	0	50	0	50	0	50	0
		16	40	0	49	0	50	0	50	0	50	0
	2	2	40	1	47	0	50	0	50	0	50	0
		4	40	1	46	0	50	0	50	0	50	0
		16	44	1	48	0	50	0	50	0	50	0
$\lambda < 0$	1	2	32	0	43	0	49	0	50	0	50	0
		4	33	0	43	0	49	0	50	0	50	0
		16	36	0	44	0	50	0	50	0	50	0
	2	2	40	1	47	0	50	0	50	0	50	0
		4	40	1	46	0	50	0	50	0	50	0
		16	44	1	48	0	50	0	50	0	50	0
$\lambda < 0$	1	2	32	0	43	0	49	0	50	0	50	0
		4	33	0	43	0	49	0	50	0	50	0
		16	36	0	44	0	50	0	50	0	50	0
	2	2	40	1	47	0	50	0	50	0	50	0
		4	40	1	46	0	50	0	50	0	50	0
		16	44	1	48	0	50	0	50	0	50	0

Table 4. Mean biases of the estimates of λ , μ and σ for M_1 -method and M_2 -method

		k=1			k=2			k=3		
		$\tau=2$		$\tau=16$	$\tau=2$		$\tau=4$	$\tau=2$		$\tau=16$
		$\lambda > 1$	$\lambda = 1$	$c < \lambda < 1$	$\lambda = c$	$0 < \lambda < c$	$\lambda < 0$	$\lambda > 1$	$\lambda = 1$	$c < \lambda < 1$
λ	M_1	-0.8743	-0.8644	-0.8734	-0.4597	-0.4525	-0.1117	-0.4507	-0.4507	-0.1365
		-0.4461	-0.4430	-0.4299	-0.2190	-0.2321	-0.0690	-0.0613	-0.0613	-0.0490
		-0.3970	-0.3997	-0.3999	-0.1593	-0.1626	-0.0202	-0.1626	-0.1626	-0.0445
		-0.3559	-0.3464	-0.3348	-0.1093	-0.1081	-0.0019	-0.1107	-0.1107	-0.0195
		-0.1749	-0.1738	-0.1685	-0.0508	-0.0521	-0.0080	-0.0516	-0.0516	-0.0118
μ	M_2	0.4609	0.4414	0.4501	0.2487	0.2525	0.1135	0.2601	0.2601	0.0975
		-0.2554	-0.2412	-0.2429	-0.1594	-0.0953	-0.0111	-0.1259	-0.1259	-0.0489
		-0.1297	-0.1140	-0.1239	-0.0657	-0.0396	-0.0343	-0.0511	-0.0511	-0.0065
		-0.1020	-0.1163	-0.1088	-0.0434	-0.0481	-0.0104	-0.0478	-0.0478	-0.0104
		-0.0968	-0.0902	-0.0935	-0.0194	-0.0295	-0.0038	-0.0275	-0.0275	0.0037
σ	M_1	-0.0501	-0.0461	-0.0430	-0.0084	-0.0170	-0.0030	-0.0123	-0.0123	-0.0030
		0.1399	0.1335	0.1618	0.0539	0.0753	0.0479	0.0433	0.0433	0.0619
		0.0384	0.0894	0.0837	-0.0044	-0.0037	-0.0060	-0.0022	-0.0022	-0.0031
		0.0755	0.1570	0.1603	-0.0137	-0.0002	-0.0108	-0.0038	-0.0038	-0.0040
		0.0889	0.1878	0.1697	-0.0021	-0.0035	-0.0079	-0.0090	-0.0090	-0.0093
σ	M_2	0.0805	0.1985	0.2190	-0.0148	-0.0187	-0.0150	-0.0127	-0.0127	-0.0061
		0.1958	0.3754	0.4131	-0.0367	0.0120	-0.0152	-0.0118	-0.0118	-0.0323
		-0.0903	-0.1412	-0.1582	-0.0019	0.0067	-0.0028	-0.0079	-0.0079	-0.0028
		-0.0016	-0.0013	-0.0016	-0.0019	0.0019	-0.0016	-0.0014	-0.0014	-0.0013
		-0.0092	-0.0071	-0.0076	-0.0062	0.0068	-0.0053	-0.0054	-0.0054	0.0042
σ	M_1	-0.0041	-0.0046	0.0040	-0.0037	-0.0036	-0.0047	-0.0027	-0.0027	0.0047
		-0.0068	-0.0057	-0.0057	-0.0058	-0.0061	-0.0064	-0.0056	-0.0056	-0.0051
		-0.0191	-0.0197	-0.0174	-0.0138	-0.0131	-0.0066	-0.0121	-0.0121	-0.0138
		0.0186	0.0156	0.0125	0.0127	0.0136	0.0161	0.0136	0.0136	-0.0161
		-0.4502	-0.2286	-0.1484	-0.0356	-0.0206	-0.0034	-0.0141	-0.0141	-0.0042
σ	M_2	-0.9015	-0.4623	-0.2920	-0.0764	-0.0432	-0.0009	-0.0281	-0.0281	-0.0040
		-0.9924	-0.5085	-0.3287	-0.0991	-0.0626	-0.0162	-0.0415	-0.0415	-0.0102
		-1.1414	-0.5725	-0.3655	-0.1286	-0.0893	-0.0242	-0.0625	-0.0625	-0.0238
		-2.2782	-1.1433	-0.7454	-0.2712	-0.1855	-0.0405	-0.1450	-0.1450	-0.0312
		-0.0750	-0.1495	-0.2252	0.0012	-0.0118	-0.0009	-0.0229	-0.0229	-0.0044
σ	M_1	-0.0096	-0.0060	-0.0076	-0.0061	-0.0066	-0.0062	-0.0064	-0.0064	-0.0052
		-0.0216	-0.0216	-0.0167	-0.0267	-0.0169	-0.0224	-0.0214	-0.0214	-0.0225
		-0.0258	-0.0152	-0.0186	-0.0149	-0.0154	-0.0108	-0.0111	-0.0111	-0.0156
		-0.0284	-0.0164	-0.0119	-0.0121	-0.0136	-0.0150	-0.0155	-0.0155	-0.0148
		-0.0188	-0.0251	-0.0272	-0.0391	-0.0164	-0.0155	-0.0145	-0.0145	-0.0134
σ	M_2	0.2343	0.2057	0.2941	0.2852	0.1711	0.1746	0.1646	0.1646	0.1708

The ANOVA on the MB of μ^* and σ^* showed almost similar results as the ANOVA on one of λ^* . The results of the ANOVA based on the MSE of λ^* , μ^* and σ^* were quite the same as those based on the MB. These results are summarized in Tab. 5.

4.2 The case of M_2 -method

In the cases $n=10$ and $n=20$, no estimates $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\sigma}$ could be obtained. Further, in the cases $n=50$ and $n=100$ the estimates could not be obtained for a few samples. And in the case $n=200$ only, all the estimates could be obtained. Thus, we summarize the results only for $n=200$.

Table 5. Mean square errors of the estimates of λ , μ and σ for M_1 -method and M_2 -method

		$k=1$			$k=2$			$k=3$		
		$\tau=2$		$\tau=4$	$\tau=2$		$\tau=4$	$\tau=2$		$\tau=4$
		$\lambda > 1$	$\lambda = 1$	$0 < \lambda < 1$	$\lambda = c$	$0 < \lambda < c$	$\lambda < 0$	$\lambda > 1$	$\lambda = 1$	$0 < \lambda < 1$
λ	M_1	0.7771	0.2007	0.1593	0.1284	0.0311	0.2157	0.2261	0.2389	0.0515
		0.7722	0.1873	0.1616	0.1135	0.0287	0.0681	0.0472	0.0580	0.0218
		0.7575	0.1589	0.1212	0.0307	0.0307	0.0701	0.0288	0.0332	0.0062
		0.743	0.1578	0.1135	0.0307	0.0307	0.0701	0.0435	0.0336	0.0007
μ	M_1	0.0782	0.0204	0.0134	0.0123	0.0031	0.0224	0.0095	0.0096	0.0002
		0.0725	0.0189	0.0140	0.0113	0.0024	0.0300	0.0069	0.0059	0.0009
		0.0630	0.0092	0.0048	0.0020	0.0004	0.0134	0.0023	0.0023	0.0002
		0.0630	0.0092	0.0048	0.0020	0.0004	0.0134	0.0023	0.0023	0.0002
σ	M_1	0.0041	0.0142	0.0193	0.0207	0.1227	0.0103	0.0004	0.0005	0.0001
		0.0291	0.0319	0.0547	0.1901	0.0266	0.0285	0.0011	0.0011	0.0001
		0.0092	0.0092	0.0092	0.0092	0.0092	0.0092	0.0001	0.0001	0.0001
		0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
σ	M_2	0.0006	0.0007	0.0005	0.0011	0.0039	0.1753	0.0006	0.0006	0.0002
		0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0006	0.0006	0.0006
		0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0006	0.0006	0.0006
		0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0006	0.0006	0.0006
σ	M_1	0.2035	0.8149	0.9886	1.3061	5.2098	0.0664	0.0021	0.0021	0.0001
		0.0526	0.2147	0.2599	0.3293	1.3158	0.0232	0.0003	0.0003	0.0003
		0.0865	0.1088	0.1348	0.1866	0.5592	0.0517	0.0013	0.0013	0.0003
		0.0865	0.1088	0.1348	0.1866	0.5592	0.0517	0.0013	0.0013	0.0003
σ	M_2	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
		0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
		0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
		0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

The MB of $\hat{\lambda}$, which were -0.2554 ($\tau=1$), -0.2412 ($\tau=4$), and -0.2429 ($\tau=16$), were relatively large for $k=1$ and $\lambda > 1$. For $k=3$ and $0 < \lambda < c$, they were 0.0033 ($\tau=2$), -0.0005 ($\tau=4$), and -0.0030 ($\tau=16$), so relatively small. Generally speaking, the MB of $\hat{\lambda}$ decreased with k increased, and they were especially small for the range $0 < \lambda < c$. Differences of the MB of $\hat{\lambda}$ among $\{\tau\}$ were generally slight except for $k=3$ and $\lambda=1$, in this case they moved -0.0391 ($\tau=1$) to -0.0065 ($\tau=16$).

The MB of $\hat{\mu}$ and $\hat{\sigma}$ were almost invariant for $k=1, 2$ and 3 . The maximum MB of $\hat{\mu}$ was 0.0197 for $k=1$, $0 < \lambda < c$ and $\tau=4$. And the MB of $\hat{\mu}$ were relatively large for $\lambda < c$, compared to other ranges of λ . Generally their values were small. On the other hand, the maximum MB of $\hat{\sigma}$ was 0.2941 for $k=1$, $\lambda < 0$ and $\tau=16$. And the MB

of $\hat{\sigma}$ were especially large for the range $\lambda < 0$.

The MSE of $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\sigma}$ showed similar tendencies as we can see from Tab. 5. The MSE of $\hat{\lambda}$ decreased with k increased. They were relatively small for range $0 < \lambda < c$. Those of $\hat{\mu}$ and $\hat{\sigma}$ were almost invariant with the changes of k . However, they varied with the changes of λ and they were especially large for the range $\lambda < 0$.

For M_2 -method, as described above, in the case $n=200$ only all the estimates $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\sigma}$ could be obtained. So we evaluate the effects of $\{k\}$ and $\{\tau\}$ on the MB and the MSE of $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\sigma}$ by two-way ANOVA.

The ANOVA on the MB of $\hat{\lambda}$ showed the effect of $\{k\}$ was significant at level 0.05 and its PVE was more than 90%, while the effect of $\{\tau\}$ was not significant. No effects of $\{k\}$ and $\{\tau\}$ were significant by the ANOVA on the MB of $\hat{\mu}$ and $\hat{\sigma}$.

The ANOVA on the MSE of $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\sigma}$, showed the same tendencies as on the MB.

5. Interpretation of the Results

Satisfactory fulfillment of normality of the transformed data were confirmed by normality test on skewness and kurtosis, even when the data had been power-normally transformed using λ^* having biases. Skewness were particularly improved by the transformation, while kurtosis were not always improved depending on the distribution shapes of the original data. These results suggest that even when the λ^* has bias, the transformed data satisfy at least symmetry for most cases.

From the results for M_1 -method, it was seen that most influencing factors on the MB and the MSE of estimates λ^* , μ^* and σ^* were k and n . Especially, the influences of k on the MB and the MSE were remarkable. For small k , the estimates λ^* , μ^* and σ^* had large biases and the degree decreased with k increased. We think the results confirm, as was expected, that consistent estimators can not be obtained by M_1 -method. The influences of τ on the MB and the MSE of μ^* and σ^* were small, and on applying M_1 -method, it may be possible not to allow for τ .

It was seen that the MB and the MSE of λ^* and $\hat{\lambda}$ both increased with k increased, and for small k , the differences of them between λ^* and $\hat{\lambda}$ were remarkable. These results were found for all ranges of λ . On the other hand, the MB and the MSE of $\hat{\mu}$ and $\hat{\sigma}$ were smaller than those of μ^* and σ^* regardless the values of τ except for range $\lambda < 0$. Further, influences of k on the MB and the MSE of μ^* and σ^* were significant at level 0.05, while those of $\hat{\mu}$ and $\hat{\sigma}$ were not. These results suggest that effect of bias reduction of M_2 -method are witnessed except for $\lambda < 0$.

The fact that for $\lambda < 0$ the MB and the MSE of $\hat{\mu}$ and $\hat{\sigma}$ were larger than those of μ^* and σ^* is very interesting, reflecting the dependence of the performances of M_2 -method on the distributional shapes.

For M_2 -method the estimates of parameters could not be obtained in some cases especially for small n . Main reason of this fact is considered that for M_2 -method the magnitude of $A(k)$ is counted for and consequently, correlations among parameters become higher, so the information matrix becomes unstable.

6. Discussion

It is considered that in case the aim of transformation is to attain near-normality of the transformed data, estimation of λ by M_1 -method is of enough use for n considerably large. Even for small n , the distribution of the transformed data seems to be symmetric for most cases (See [6] for symmetry-oriented transformation). However, on discussing behaviours of the estimates μ^* and σ^* (for small k), it is necessary to diagnose the magnitude of k , since μ^* and σ^* have biases dependent on k . For large n , the value of k estimated using λ^* , μ^* and σ^* satisfies approximately $A(k)=1$.

Inoue and Goto [7] have tried to obtain jackknife estimate of λ in M_1 -method to avoid these troubles. The aim of jackknifing was to reduce the biases of λ^* , μ^* and σ^* . The results were as follows; namely for $k=2$ and $k=3$ bias reduction was remarkable, but for $k=1$ the reduction was not so, compared to estimates of ordinary M_1 -method. Although the influence of τ on the estimate of λ was not so large for ordinary M_1 -method, the effect of bias reduction by jackknifing λ^* increased with τ increased. These results suggest that it is better to apply jackknifing of λ^* in M_1 -method in order to obtain stable estimates of λ , μ and σ for small n , on the reason that some effect of bias reduction can at least be expected (Also see [3]).

Further, in case the aim is to identify the PND, namely to obtain consistent estimators of the parameters λ , μ and σ^2 , it is recommended to apply M_2 -method. For M_2 -method, requirement of a prior specification of $A(k)$ raises some difficulties in actual application, since it's almost impossible to predict the magnitude of $A(k)$ in advance. Thus, we set up $A(k)$ in an appropriate range and then must find the magnitude of $A(k)$ by trials and errors iteration. Further, for M_2 -method the correlations among parameters become higher than for M_1 -method. Consequently, the second derivative of information matrix may be unstable particularly for small n , and in some cases, convergence can not be attained in the iteration procedure. To avert these troubles, it is necessary to stabilize the information matrix by some devices, including ridge adjustment, or to apply some data-adaptive procedures, including jackknife, cross-validation and Bootstrap (See also [1] and [8] for analytical investigation of sampling behaviours of the estimate of λ).

7. Concluding Remarks

The conclusions of this paper are summarized as follows:

- (1) Near-normality was attained by transforming the original data using λ^* for large n . For small n , symmetry at least was attained by the transformation for most cases.
- (2) The MB and MSE of λ^* , μ^* and σ^* were most influenced by k .
- (3) The MB and the MSE of λ^* were relatively small for range $0 < \lambda < c$, and relatively large for range $\lambda > 1$, while those of μ^* and σ^* were relatively small for range $\lambda > 1$ and relatively large for range $0 < \lambda < c$.
- (4) The MB and the MSE of $\hat{\lambda}$ were influenced by k , namely, they decreased with k increased, while those of $\hat{\mu}$ and $\hat{\sigma}$ were almost invariant with the changes of k .

- (5) The MB and the MSE of $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\sigma}$ were smaller than those of λ^* , μ^* and σ^* , especially for small k , except for $\lambda < 0$. In the case $\lambda < 0$, the MB and the MSE of $\hat{\mu}$ and $\hat{\sigma}$ were larger than those of μ^* and σ^* .

The remained problems are;

- (1) evaluation of the influences of distributional shapes on the biases of estimates of parameters,
- (2) consideration of the result that the MB and the MSE of the estimates were larger for M_2 -method than for M_1 -method for $\lambda < 0$,
- (3) stabilization of the information matrix for small n in M_2 -method, and
- (4) close examination of the influences of k on estimates for M_2 -method.

Acknowledgement

We would like to express our appreciation to Professor Choichiro Asano of Kyushu University who was interested in our study and recommended us to contribute this paper. We also would like to thank Professor Takeaki Nagai of Oita University for giving us many helpful suggestions.

References

- [1] BICKEL, P. J. and DOKSUM, K. A.: *An analysis of transformations revisited*. J. Amer. Statist. Assoc., **76** (1981), 296-311.
- [2] BOX, G. E. P. and COX, D. R.: *An analysis of transformations*. J. Roy. Statist. Soc., **B26** (1964), 211-243.
- [3] CRESSIE, N. A. C.: *Transformations and the jackknife*. J. Roy. Statist. Soc., **B43** (1981), 177-182.
- [4] GOTO, M., UESAKA, H. and INOUE, T.: *Some linear models for power transformed data*. Invited presentation at the 10th International Biometric Conference at Brazil. August, (1979), 6-10, (Res. Rep. No. 93, Res. Instit. Fund. Infor. Sc., Kyushu University.)
- [5] GOTO, and Inoue, T.: *Some properties of power-normal distribution*. Bull. Biometric. Soc. Japan., **1**, 28-54.
- [6] HINKLEY, D. V.: *On power transformations to symmetry*. Biometrika, **62** (1975), 101-111.
- [7] INOUE, T. and GOTO, M.: *Performances of power normal transformation; Parameter estimation and behaviours of the estimates of parameters* (in Japanese), Proceedings of The 49th Conference of Japan Statistical Society, (1981), 113-114.
- [8] UESAKA, H. and GOTO, M.: *Some properties of Box-Cox estimates of transforming parameters in power-normal distribution*. Bull. Biometric. Soc. Japan., **3** (1982), 13-34.

Communicated by Ch. Asano

Received August 10, 1983