



Applications of a New Power Normal Family : A Case Study on C_{pk} index

Isogai, Takafumi

(Citation)

神戸大学大学院海事科学研究科紀要, 07<商船・理工論篇>:39-52

(Issue Date)

2010-07

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/81002401>

(URL)

<https://hdl.handle.net/20.500.14094/81002401>



Applications of a New Power Normal Family

- A Case Study on C_{pk} Index -

Takafumi Isogai

Graduate School of Maritime Sciences, Kobe University,
Fukae, Higashinada, Kobe, JAPAN

Abstract

We often come across problems that data sets are positive-valued and skewed, but we want to interpret them through usual means and variances because of their convenience. In this paper we shall examine a case where we have to evaluate usual C_{pk} , the Process Capability Index, for positively skewed data. The value of usual C_{pk} is biased. Thus, new versions of C_{pk} are introduced through a new power normal (Isogai, 1999) and related power normal families. Comparisons are made with normal, our new power normal and related families. Our new power normal family shows its stable performance, but Box-Cox power normal family (Box & Cox, 1964) does not.

Keywords: positive skewness; Process Capability Index C_{pk} ; Box-Cox power normal family

(Received March 28, 2010)

Introduction

In practical fields, such as quality control and biomedical sciences, we often come across non-normal data sets arising from ordinary experimental studies. In spite of non-normality we are usually likely to analyze those data through means and variances under the assumption of normality, because there are no systematic guidelines or knowledge bases about the manners of handling such non-normal data. Therefore, we report some experiences of having dealt with non-normality in typical problems through a new power normal family (Isogai, 1999) as well as other related power normal families.

A new power normal family (shortly, NP family hereafter) has been constructed by Isogai (1999) from the generalized F distribution (see Prentice (1975)) through a Wilson-Hilferty type power transformation of an F random variable. A random variable X of the generalized F distribution is defined by

$$X = \eta \left(F_{2\phi_2}^{2\phi_1} \right)^\gamma \quad (1)$$

for $\eta > 0, \gamma > 0, \phi_1 > 1$ and $\phi_2 > 1$, where $F_{2\phi_2}^{2\phi_1}$ be the random variable of a central F distribution with degrees of freedom $(2\phi_1, 2\phi_2)$. The definition of NP family is that the following random variable

$$\frac{1}{\delta} \left\{ \left(\frac{X}{\eta} \right)^{\frac{\delta}{\sigma}} - \exp(-\delta^2) \right\} \quad (2a)$$

or equivalently

$$\frac{1}{\delta} \left\{ \exp \left(\delta \left(\frac{\log X - \mu}{\sigma} \right) \right) - \exp(-\delta^2) \right\} \quad (2b)$$

is approximately distributed as the standard normal variable for $\eta > 0, \sigma > 0, \mu = \log \eta$ and

$$-\frac{1}{3} < \delta < \frac{1}{3}, \quad (3)$$

where we put

$$\begin{cases} \delta = \delta_1 \delta_2, \\ \sigma = \gamma \delta_2, \\ \delta_1 = \frac{1}{3} \left(\frac{1}{\phi_1} - \frac{1}{\phi_2} \right) / \left(\frac{1}{\phi_1} + \frac{1}{\phi_2} \right), \\ \delta_2 = \left(\frac{1}{\phi_1} + \frac{1}{\phi_2} \right)^{1/2}. \end{cases} \quad (4)$$

In the following, when we need not distinguish between our equivalent models (2a) and (2b), we call them simply NP family (2).

As δ tends to zero, $\log X$ is nearly distributed as the normal distribution with mean μ ($= \log \eta$) and variance σ^2 . When δ is near $-1/3$, the distribution of X is approximated by the extreme value distribution of Type 2 (see Johnson & Kotz (1970)):

$$\Pr(X \leq x) = \exp \left\{ - \left(\frac{x}{\eta} \right)^{-\frac{1}{\sigma}} \right\}$$

for $x \geq 0$. When δ is near $1/3$, the distribution of X is approximated by the Weibull distribution, of which distribution function is given by

$$\Pr(X \leq x) = 1 - \exp \left\{ - \left(\frac{x}{\eta} \right)^{\frac{1}{\sigma}} \right\}$$

for $x \geq 0$.

Up to the present, applicability of our NP family (2), including related notions, has been explored in several situations. For example, a derivation of a simple formula for ‘Median Rank’ and a normal probability plotting method of ‘Sums of Squares’ with respect to multifactor experimental designs (Isogai, 2001); a feasible distribution to describe the maximum pit depth of corrosion pit depth data (Isogai, 2004); a data transformation method to attain joint normality (Isogai, 2005); a practicable distribution to explain a population of ‘Enamel Rater Value’ data coming from an inspection process in manufacturing of aluminum cans (Isogai, 2008).

In this paper we report one case study about non-normal, especially positively skewed data sets. Our topic is ANOVA with respect to the Process Capability Index C_{pk} . In this ANOVA, usual C_{pk} is calculated from positively skewed data, and, as a result, the value of C_{pk} is largely biased. To remedy a bias of usual C_{pk} , we adopt our NP family (2) and define a new version of C_{pk} . When the structure of a linear model becomes more complex, our new power normal

and related families show their stable performance in any conditions, but Box-Cox power normal family (Box & Cox, 1964) does not.

Process Capability Index and a New Power Normal Family

We shall consider the following data, which is based on personal communications. Table 1 shows a list of C_{pk} values coming from an experimental design performed by some Japanese bearing manufactory. Every day they have to adjust the width of a taper bearing in a grinding process. Irregularity in the width of a taper bearing brings undesirable effects in quality of bearings. For a quality inspection they use a nonnegative irregularity characteristic, which measures the extent of irregularity in the width, and takes zero value when there is no irregularity. Until recently a value of C_{pk} with respect to the irregularity characteristic was less than one. To improve C_{pk} they planed to find out suitable operating conditions through a two factorial experimental design. The target value of C_{pk} is greater than 1.33. C_{pk} is defined by

$$C_{pk} = \frac{\text{Upper specified value} - \text{sample mean}}{3 \cdot (\text{standard deviation of the sample})}. \quad (5)$$

Table 1. C_{pk} values of a split-plot design with respect to Irregularity characteristic data.

Levels of R	Levels of A	Levels of B			
		B1	B2	B3	B4
R1	A1	0.86	1.06	2.50	3.77
	A2	0.95	1.37	3.31	4.33
R2	A1	0.79	0.99	2.86	4.01
	A2	0.96	1.22	2.73	3.91

One of factors, denoted by A, indicates two types of a grinding wheel (A_1 : a current one, A_2 : a candidate one), and the other factor, denoted by B, describes four rotation speeds of a grinding wheel (B_1 : 5.4rpm, B_2 : 4.5rpm, B_3 : 3.8rpm, B_4 : 1.9rpm). The first levels of both factors correspond to current operating conditions.

They applied a split-plot design. In their experiment the factor A is treated as a block factor, and B is a plot. They first randomize A, and then randomize B. The whole experiment with all combinations of $A_i B_j$ was replicated two times, which we denote by the factor R (R_1 : the first replication, R_2 : the second replication). In each experiment with $A_i B_j$, 30 irregularity characteristics of bearings were observed and one C_{pk} value was recorded. Scatter plots with respect to main effects are given in Figure 1. Apparently B seems to be significant. The result of ANOVA is given in Table 2.

Table 2 shows that the factor B is only highly significant. They selected the level $A_1 B_3$ as new operating conditions, and under these new conditions an estimate of C_{pk} , that is, an estimate of population mean $\theta(A_1 B_3)$ was $\widehat{\theta(A_1 B_3)} = 2.72$, which is twice as large as the smallest target value 1.33.

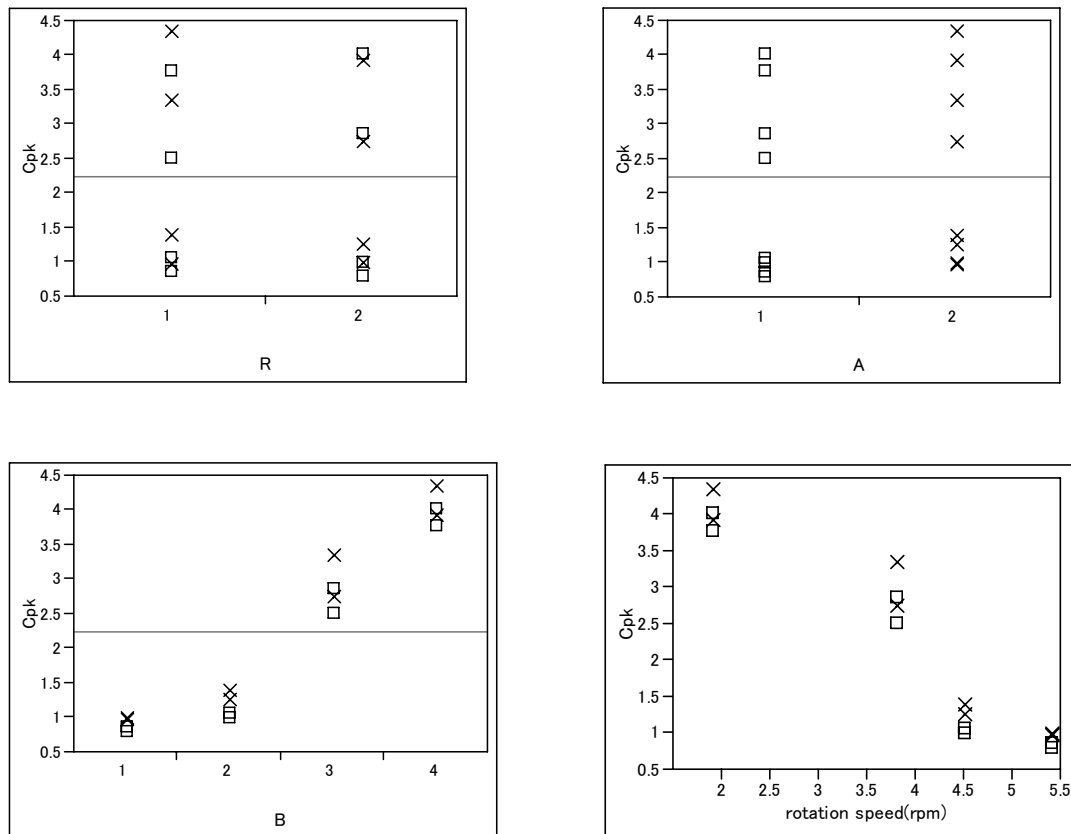


Figure 1. Scatter plots of main effects in Table 1 (A_1 : $\square$, A_2 : $\times$).

Table 2. ANOVA of a split-plot design with C_{pk} values.

Factor	Sum of Sq	DF	Mean Sq	F Value	P Value
R	0.02785	1	0.02785	0.172365	0.749481
A	0.235538	1	0.235538	1.457754	0.440367
E(1)	0.161576	1	0.161576	5.4468	0.0583
B	25.900493	3	8.633498	291.0406	<.0001
AxB	0.025491	3	0.008497	0.2864	0.8339
E(2)	0.177985	6	0.02966		
Total	26.528934	15			

Preliminary analysis

Shapiro-Wilk test for a sample of size 30 in each cell of Table 1 rejected normality of all samples. The measure of skewness, that is, the third order standardized cumulant statistic, detected positive skewness in 11 cells of all 16 cells, and the other cells had negative skewness. Two cells with positive skewness were statistically 5% significant, but the other cells were non-significant. On the whole, our data seems to be positively skewed.

Now, using NP family (2), we shall examine a shape of the distribution with respect to irregularity characteristic. We have one scale (or location) parameter η (or $\mu = \log \eta$) and two shape parameters δ and σ to be estimated. The range of δ is bounded, that is, $-1/3 < \delta < 1/3$.

Thus we need not find out its initial value for iteration in maximum likelihood estimation. For example, we have only to set $\delta_0 = 0.01$ as an initial value. On the other hand, simple estimates of σ and η (or μ) are given by

$$\hat{\sigma}_c = \frac{k_2^{1/2}}{k_1}, \quad (6)$$

$$\hat{\eta}_c = k_1 \text{ (or } \hat{\mu}_c = \log \hat{\eta}_c = \log k_1) \quad (7)$$

where k_1 and k_2 are the first and second order cumulant statistics respectively (for the details, see Isogai (2005)). We can use values of $\hat{\sigma}_c$ and $\hat{\eta}_c$ (or $\hat{\mu}_c$) as initial values for estimation of σ and η (or μ).

For a sample of size 30 in each cell of Table 1, not only $\hat{\sigma}_c$ and $\log k_1$ but also estimates $\hat{\delta}$, $\hat{\sigma}$ and $\hat{\mu}$ with NP family (2) were calculated and shown in Table 3. When an absolute value of $\hat{\delta}$ was greater than 1/3 (that is, $R_1A_2B_4$ and $R_2A_2B_4$ are such cases), we fixed $\hat{\delta} = 0.33$ or $\hat{\delta} = -0.33$ according to the sign of $\hat{\delta}$ and obtained the other estimates $\hat{\sigma}$ and $\hat{\mu}$.

Table 3. Values of various C_{pk} and estimates of δ , μ and σ of NP family.

R	A	B	C_{pk}	$\hat{\sigma}_c$	$\log k_1$	$\hat{\delta}$	$\hat{\sigma}$	$\hat{\mu}$	$C_{pk}(NP)_0$	$C_{pk}(BC)_0$	$C_{pk}(NP)_c$
1	1	1	0.86	0.2206	1.6292	0.129	0.2153	1.6477	0.80	0.80	0.71
1	1	2	1.06	0.3208	1.3779	0.089	0.3178	1.3700	0.85	0.85	0.77
1	1	3	2.50	0.2883	0.9295	0.282	0.2592	1.0067	2.71	2.71	1.52
1	1	4	3.77	0.3628	0.4489	-0.016	0.3653	0.3740	1.49	1.49	1.75
1	2	1	0.95	0.2147	1.6027	-0.309	0.1557	1.4997	0.63	0.63	0.78
1	2	2	1.37	0.3190	1.2431	-0.019	0.3087	1.1865	0.93	0.93	0.94
1	2	3	3.31	0.2347	0.8754	0.260	0.2122	0.9368	4.00	4.00	2.06
1	2	4	4.33	0.3001	0.4906	0.33	0.2386	0.5933	6.97	23.5	2.14
2	1	1	0.79	0.2474	1.6160	0.223	0.2318	1.6668	0.80	0.80	0.65
2	1	2	0.99	0.3093	1.4271	-0.273	0.2530	1.2693	0.62	0.62	0.74
2	1	3	2.86	0.2832	0.8472	0.194	0.2720	0.8888	2.36	2.36	1.69
2	1	4	4.01	0.3005	0.5500	0.327	0.2566	0.6463	5.41	5.41	2.04
2	2	1	0.96	0.1907	1.6422	0.064	0.1868	1.6428	0.86	0.86	0.81
2	2	2	1.22	0.3271	1.2901	0.149	0.3232	1.3097	1.00	1.00	0.86
2	2	3	2.73	0.2790	0.8892	0.088	0.2734	0.8882	1.80	1.80	1.65
2	2	4	3.91	0.4023	0.3364	-0.33	0.2476	0.1027	0.83	0.59	1.68

For non-normal data described by NP family (2), C_{pk} index is defined by

$$C_{pk}(NP) = \frac{1}{3\delta} \left\{ \left(\frac{\text{Upper specified value}}{\eta} \right)^{\frac{\delta}{\sigma}} - \exp(-\delta^2) \right\} \quad (8a)$$

or equivalently, defined by

$$C_{pk}(NP) = \frac{1}{3\delta} \left\{ \exp \left(\delta \left(\frac{\log(\text{Upper specified value}) - \mu}{\sigma} \right) \right) - \exp(-\delta^2) \right\} \quad (8b)$$

with $\mu = \log \eta$. Using estimates $\hat{\delta}$, $\hat{\sigma}$ and $\hat{\mu}$ in each cell of Table 3, values of $C_{pk}(NP)$, denoted by $C_{pk}(NP)_0$ are also given in Table 3.

Furthermore, for a comparative study we shall consider the Box-Cox power normal family (shortly, BC family hereafter):

$$\frac{1}{\tau} \left\{ \frac{X^\lambda - 1}{\lambda} - \xi \right\} \quad (9)$$

for $\tau > 0$ (for the details, see Box & Cox (1964)). The random variable $(X^\lambda - 1)/\lambda$ is approximately distributed as normal with mean ξ and variance τ^2 . For BC family, C_{pk} index is defined by

$$C_{pk}(BC) = \frac{1}{3\tau} \left\{ \frac{\text{Upper specified value}^\lambda - 1}{\lambda} - \xi \right\}. \quad (10)$$

Values of $C_{pk}(BC)$, denoted by $C_{pk}(BC)_0$, are also given in Table 3.

From Table 3 we know that usual C_{pk} apparently has an increasing tendency as levels of the factor B increase. However, $C_{pk}(NP)_0$ and $C_{pk}(BC)_0$ do not have such a clear tendency.

On the other hand, $C_{pk}(NP)_0$ and $C_{pk}(BC)_0$ have almost the same behavior except for two cases: $R_1A_2B_4$ and $R_2A_2B_4$. Especially, with respect to $C_{pk}(NP)_0$, aberrant values of $\hat{\delta}$ in two cases $R_1A_2B_4$ and $R_2A_2B_4$ give us a warning that distributions of data sets have some peculiarities. In fact, we have two-point and three-point distributions in $R_1A_2B_4$ and $R_2A_2B_4$ respectively. Unlike $C_{pk}(NP)_0$, values of $C_{pk}(BC)_0$ are greatly affected by abnormalities in data sets of $R_1A_2B_4$ and $R_2A_2B_4$.

To sum up the above facts, estimation or a behavior of $C_{pk}(NP)_0$ is dissatisfied against our expectation. It seems that one of the reasons is a great variability of estimates $\hat{\delta}$ due to a small sample size 30.

Distribution of $\hat{\delta}$

Here we shall examine distributional properties of $\hat{\delta}$ for small sample cases. We performed simulation studies about maximum likelihood estimation for δ in NP family (2). From NP family (2) with $\delta = 0.01, 0.1, 0.2, 1/3$ and $\sigma = 0.1, 0.3, 0.6, 0.9$ we extracted 1000 random samples of sizes $n = 30$, and drew their boxplots. The results are given in Figure 2.

From Figure 2 we know that, whatever the true value of δ may be, or, even if $\delta = 1/3$, $\hat{\delta}$ moves around the whole parameter space, $-1/3 < \delta < 1/3$, in the small sample case ($n=30$). We also know that the distribution of $\hat{\delta}$ is not seriously affected by a varying value of σ .

Furthermore, in order to examine effects of increasing sample sizes, we extracted 1000 random samples of sizes $n = 30, 50, 100, 200$ from NP family (2) with $\delta = 0.01, 0.1, 0.2, 1/3$ and $\sigma = 0.9$, and drew their boxplots. The results are given in Figure 3. From Figure 3 we know that, as a sample size increases, $\hat{\delta}$ slowly approaches to its true value.

Analysis based on our new power normal family

The simulation result of $\hat{\delta}$ for a sample size $n=30$ suggests that in Table 3 we cannot trust each

value of $\hat{\delta}$ due to its great variability. To avoid this variability, we shall introduce a simple assumption that a sample in each cell of Table 1 has a common shape parameter δ . Then, we can estimate δ , for example, by the mean value of $\hat{\delta}$'s in Table 3. Our result is $\bar{\delta}$ ($= 0.0745$).

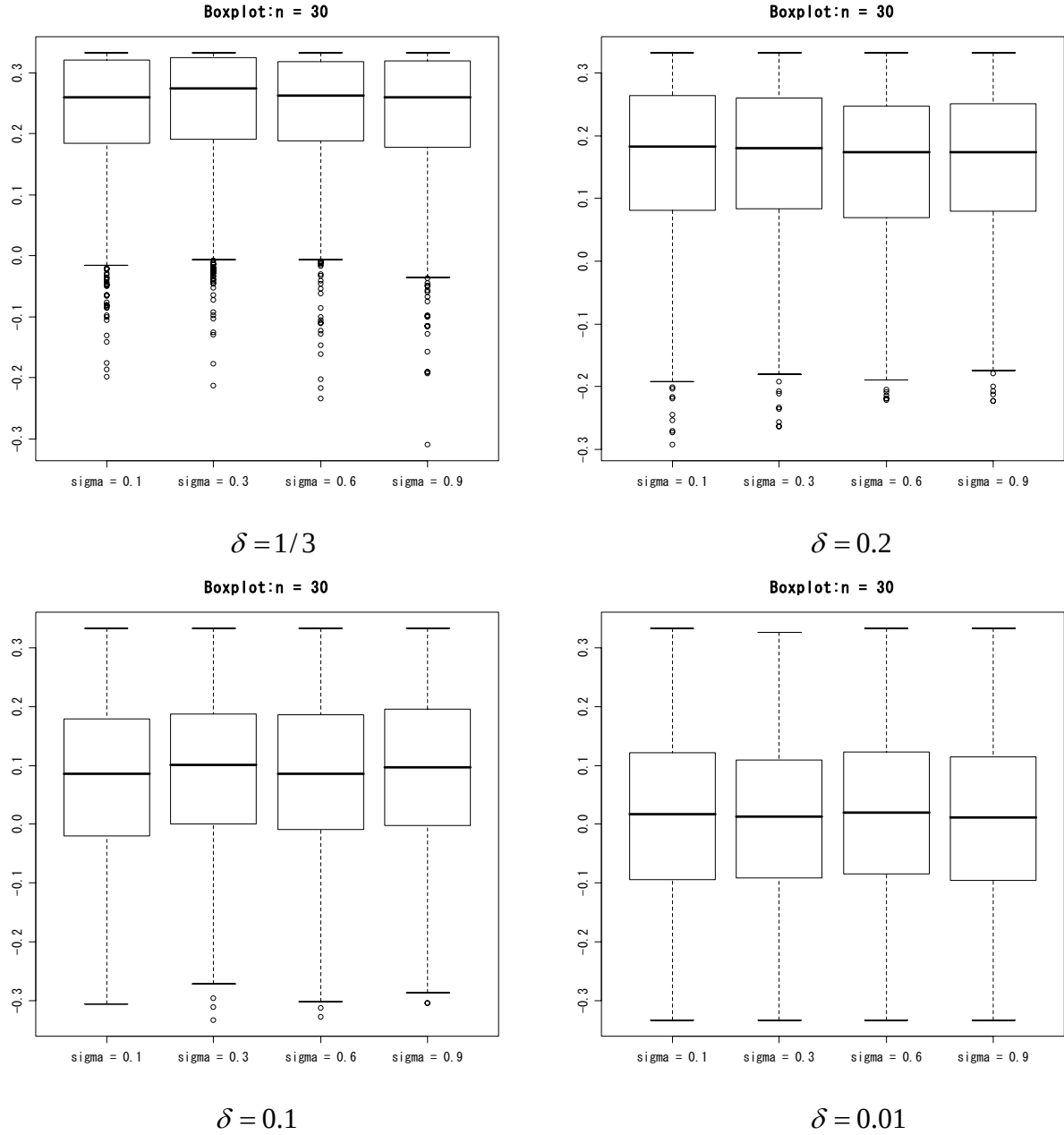


Figure 2. Boxplots of $\hat{\delta}$ with respect to 1000 random samples of size $n = 30$ extracted from NP family with $\delta = 0.01, 0.1, 0.2, 1/3$ and $\sigma = 0.1, 0.3, 0.6, 0.9$.

Now, instead of using estimates $\hat{\delta}$, $\hat{\sigma}$ and $\hat{\mu}$ in each cell of Table 3, we can use $\bar{\delta}$, $\hat{\sigma}_c$ and $\hat{\mu}_c (= \log k_1)$ to obtain $C_{pk}(NP)_c$. The result is also given in Table 3. From Table 3 we know that $C_{pk}(NP)_c$ has a clear increasing tendency as C_{pk} does. ANOVA with respect to $C_{pk}(NP)_c$ gives us the same conclusion as Table 2, that is, the factor B is only highly

significant. A large difference between corresponding values of C_{pk} and $C_{pk}(NP)_c$ is that a value of $C_{pk}(NP)_c$ is clearly smaller than that of C_{pk} .

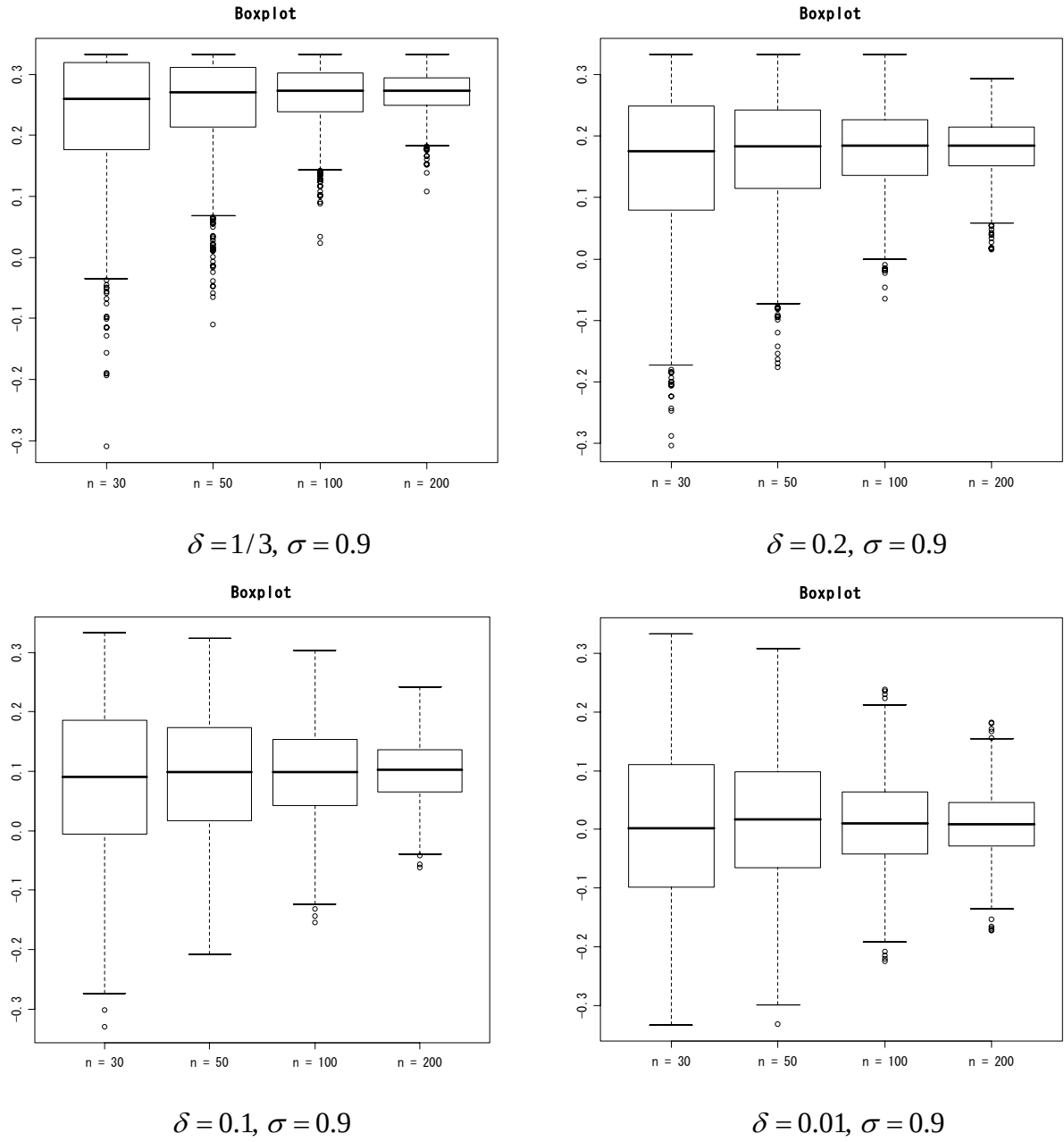


Figure 3. Boxplots of $\hat{\delta}$ with respect to 1000 random samples of size $n = 30, 50, 100, 200$ extracted from NP family with $\delta = 0.01, 0.1, 0.2, 1/3$ and $\sigma = 0.9$.

In order to examine this tendency we shall fit NP family (2) systematically to the original data of Table 1 under the assumption of existence of a common shape parameter δ . As for the other parameters μ and σ , from Table 3 we know that, as the level of the factor B increases, $\log k_1$ has a clear decreasing tendency and $\hat{\sigma}_c$ has a vague increasing one. In Figure 4 we give two scatter plots of $\log k_1$ and $\log \hat{\sigma}_c$ against the rotation speed of a grinding wheel (i.e. the

level of B).

Now we shall consider the following three cases.

Case (1): Main-Effects Model

We suppose main-effects models in μ and $\log \sigma$ of NP family (2b), that is,

$$\begin{cases} \mu = \beta_0 + \beta_1 z_R + \beta_2 z_A + \beta_3 z_B^{(1)} + \beta_4 z_B^{(2)} + \beta_5 z_B^{(3)} \\ \log \sigma = \gamma_0 + \gamma_1 z_R + \gamma_2 z_A + \gamma_3 z_B^{(1)} + \gamma_4 z_B^{(2)} + \gamma_5 z_B^{(3)} \end{cases}, \quad (11)$$

where β_0 and γ_0 are intercept parameters, and the other β_i 's and γ_i 's are parameters of regression coefficients. As for covariates, z_R , z_A , $z_B^{(1)}$, $z_B^{(2)}$ and $z_B^{(3)}$ are defined by

$$z_R = \begin{cases} 1 & \text{(1st level of R)} \\ 0 & \text{(2nd level of R)} \end{cases}, \quad z_A = \begin{cases} 1 & \text{(1st level of A)} \\ 0 & \text{(2nd level of A)} \end{cases}$$

$$z_B^{(1)} = \begin{cases} 1 & \text{(1st level of B)} \\ 0 & \text{(other levels of B)} \end{cases}, \quad z_B^{(2)} = \begin{cases} 1 & \text{(2nd level of B)} \\ 0 & \text{(other levels of B)} \end{cases} \text{ and } z_B^{(3)} = \begin{cases} 1 & \text{(3rd level of B)} \\ 0 & \text{(other levels of B)} \end{cases}.$$

NP family (2) with (11) was fitted to the original data of Table 1. For a comparison we also fitted BC family (9) with the same models in $\xi (= \mu)$ and $\log \tau (= \log \sigma)$ under the assumption of existence of a common power parameter λ . The result is given in Case (1) of Table 4, where SSE means the minimum value of a negative log-likelihood function.

Non-significant estimates are displayed by their zero values. Thus, factors R and A were not significant in both families. We also estimated C_{pk} by using our fitted models, and their values, denoted by $C_{pk}(NP)_1$ and $C_{pk}(BC)_1$, are given in Table 5.

The results of Tables 4 and 5 show that there is no difference between performances of NP and BC families, and again show that values of $C_{pk}(NP)_1$ and $C_{pk}(BC)_1$ are clearly smaller than the corresponding value of usual C_{pk} .

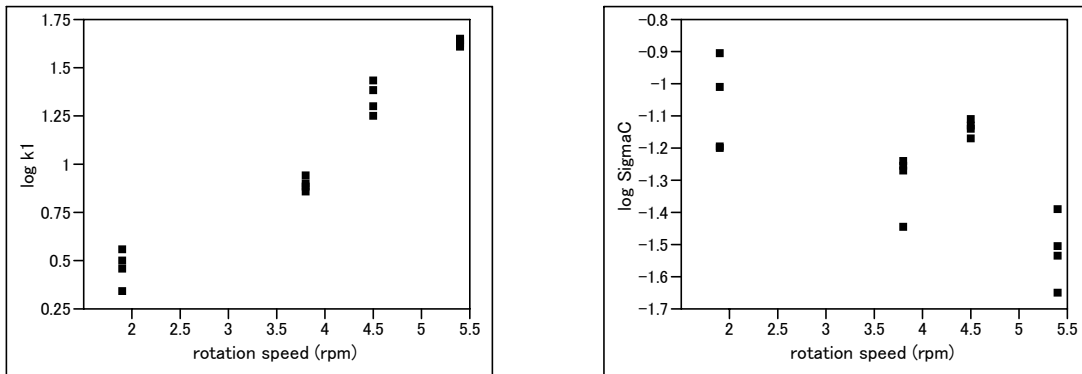


Figure 4. Scatter plots of $\log k_1$ and $\log \hat{\sigma}_c$ against the rotation speed of a grinding wheel.

Case (2): Main-Effects and One Regressor Model

Monotone behaviors with respect to values of $\hat{\beta}_i$ ($i=3,4,5$) and $\hat{\gamma}_i$ ($i=3,4,5$) in Case (1) of Table 4 as well as Figure 4 lead us to suppose a linear regression model with respect to the

rotation speed of a grinding wheel (denoted by x) in μ and $\log \sigma$ of NP family (2b), that is,

$$\begin{cases} \mu = \beta_0 + \beta_1 z_R + \beta_2 z_A + \beta_3 x \\ \log \sigma = \gamma_0 + \gamma_1 z_R + \gamma_2 z_A + \gamma_3 x \end{cases} \quad (12)$$

Table 4. Regression analyses of Irregularity characteristic data with respect to NP , BC and MNP families.

Case (1): Main-Effects Model

NP family	BC family	MNP family
SSE = 583.92	SSE = 584.12	SSE = 584.12
d.f. = 471	d.f. = 471	d.f. = 471
$\hat{\beta}_0 = 0.3736$	$\hat{\beta}_0 = 0.4656$	$\hat{\beta}_0 = 0.4245$
$\hat{\beta}_1 = 0, \hat{\beta}_2 = 0$	$\hat{\beta}_1 = 0, \hat{\beta}_2 = 0$	$\hat{\beta}_1 = 0, \hat{\beta}_2 = 0$
$\hat{\beta}_3 = 1.210$	$\hat{\beta}_3 = 1.853$	$\hat{\beta}_3 = 1.184$
$\hat{\beta}_4 = 0.8890$	$\hat{\beta}_4 = 1.288$	$\hat{\beta}_4 = 0.8832$
$\hat{\beta}_5 = 0.4560$	$\hat{\beta}_5 = 0.5801$	$\hat{\beta}_5 = 0.4393$
$\hat{\gamma}_0 = -1.0194$	$\hat{\gamma}_0 = -0.8573$	$\hat{\gamma}_0 = -1.0395$
$\hat{\gamma}_1 = 0, \hat{\gamma}_2 = 0$	$\hat{\gamma}_1 = 0, \hat{\gamma}_2 = 0$	$\hat{\gamma}_1 = 0, \hat{\gamma}_2 = 0$
$\hat{\gamma}_3 = -0.4895$	$\hat{\gamma}_3 = 0.0216$	$\hat{\gamma}_3 = -0.4870$
$\hat{\gamma}_4 = -0.0865$	$\hat{\gamma}_4 = 0.2900$	$\hat{\gamma}_4 = -0.0891$
$\hat{\gamma}_5 = -0.2506$	$\hat{\gamma}_5 = -0.0458$	$\hat{\gamma}_5 = -0.2344$
$\hat{\delta} = 0.127$	$\hat{\lambda} = 0.429$	$\hat{\lambda} = 0.429$

Case (2): Main-Effects and One Regressor Model

NP family	BC family	MNP family
SSE = 608.56	SSE = 611.38	SSE = 608.65
d.f. = 474	d.f. = 474	d.f. = 474
$\hat{\beta}_0 = -0.4190$	$\hat{\beta}_0 = -0.3532$	$\hat{\beta}_0 = -0.3700$
$\hat{\beta}_1 = 0$	$\hat{\beta}_1 = 0$	$\hat{\beta}_1 = 0$
$\hat{\beta}_2 = 0.0546$	$\hat{\beta}_2 = 0.0422$	$\hat{\beta}_2 = 0.0541$
$\hat{\beta}_3 = 0.3605$	$\hat{\beta}_3 = 0.3362$	$\hat{\beta}_3 = 0.3558$
$\hat{\gamma}_0 = -0.7146$	$\hat{\gamma}_0 = -0.6958$	$\hat{\gamma}_0 = -0.7193$
$\hat{\gamma}_1 = 0, \hat{\gamma}_2 = 0$	$\hat{\gamma}_1 = 0, \hat{\gamma}_2 = 0$	$\hat{\gamma}_1 = 0, \hat{\gamma}_2 = 0$
$\hat{\gamma}_3 = -0.1188$	$\hat{\gamma}_3 = -0.1433$	$\hat{\gamma}_3 = -0.1198$
$\hat{\delta} = 0.0961$	$\hat{\lambda} = -0.0755$	$\hat{\lambda} = 0.312$

(SSE means the minimum value of a negative log likelihood function)

NP family (2) with (12) was fitted to the original data of Table 1. We also fitted BC family (9) with the same models in $\xi (= \mu)$ and $\log \tau (= \log \sigma)$. The result is given in Case (2) of Table 4.

Non-significant estimates are displayed by their zero values. The factor A in μ , i.e. $\hat{\beta}_2$ was 5% significant with respect to NP family, and was not significant with respect to BC family. The other estimates of parameters except for zero estimates were highly significant in both families. Estimated values of regression coefficients and SSE values of Case (2) in Table 4 indicate that both NP and BC families have the same performance in our model fitting.

However, a negative sign of $\hat{\lambda} (= -0.0755)$ in BC family gives us a suspicious impression, because the other estimates of a power parameter have a positive sign. We also estimated C_{pk} by using our fitted models, and their values, denoted by $C_{pk}(NP)_2$ and $C_{pk}(BC)_2$, are given in Table 5.

Table 5. Various C_{pk} values corresponding to regression analyses with respect to NP, BC and MNP families.

R	A	B	C_{pk}	$C_{pk}(NP)_1$	$C_{pk}(BC)_1$	$C_{pk}(NP)_2$	$C_{pk}(BC)_2$	$C_{pk}(MNP)_1$	$C_{pk}(MNP)_2$
1	1	1	0.86	0.82	0.79	0.67	0.60	0.79	0.66
1	1	2	1.06	0.92	0.94	1.06	0.92	0.94	1.05
1	1	3	2.50	1.95	1.90	1.32	1.10	1.90	1.33
1	1	4	3.77	2.11	2.27	1.84	1.40	2.27	1.95
1	2	1	0.95	0.82	0.79	0.75	0.66	0.79	0.75
1	2	2	1.37	0.92	0.94	1.14	0.97	0.94	1.13
1	2	3	3.31	1.95	1.90	1.40	1.15	1.90	1.41
1	2	4	4.33	2.11	2.27	1.91	1.43	2.27	2.03
2	1	1	0.79	0.82	0.79	0.67	0.60	0.79	0.66
2	1	2	0.99	0.92	0.94	1.06	0.92	0.94	1.05
2	1	3	2.86	1.95	1.90	1.32	1.10	1.90	1.33
2	1	4	4.01	2.11	2.27	1.84	1.40	2.27	1.95
2	2	1	0.96	0.82	0.79	0.75	0.66	0.79	0.75
2	2	2	1.22	0.92	0.94	1.14	0.97	0.94	1.13
2	2	3	2.73	1.95	1.90	1.40	1.15	1.90	1.41
2	2	4	3.91	2.11	2.27	1.91	1.43	2.27	2.03

Values of $C_{pk}(NP)_2$ and $C_{pk}(BC)_2$ in Table 5 move in a different way. Though SSE of Case (2) in NP family is larger than that of Case (1), $C_{pk}(NP)_2$ seems to be near $C_{pk}(NP)_1$. On the other hand, $C_{pk}(BC)_2$ seems to deviate from $C_{pk}(BC)_1$ and $C_{pk}(NP)_2$. This discouraging behavior of $C_{pk}(BC)_2$ as well as the negative sign of $\hat{\lambda} (= -0.0755)$ compel us to abandon BC family (9). Thus, we shall construct another new power normal family, which is a modification of NP family (2) and similar to BC family (9) in the setting of a power parameter.

Case (3): Modified New Power Normal Family

We take up a reparameterization of NP family (2a) discussed in Isogai (2005). Replace

$\exp(-\delta^2)$ by one and set $\delta/\sigma = \lambda$, that is, $\delta = \lambda\sigma$ in NP family (2a), then we have a modified new power normal family (shortly, MNP family hereafter)

$$\frac{1}{\sigma} \frac{\left\{ \left(\frac{X}{\eta} \right)^\lambda - 1 \right\}}{\lambda}, \quad (13)$$

which is similar to BC family (9) in the meaning that λ is the same power parameter. However, the other parameters η ($= \exp(\mu)$) and σ are related to population characteristics κ_1 and $\kappa_2^{1/2}/\kappa_1$ respectively.

MNP family (13) with two types of models (11) and (12) was fitted to the original data of Table 1. The result is given in Cases (1) and (2) of Table 4. Non-significant estimates are displayed by their zero values. We also estimated C_{pk} by using our fitted models, and their values, denoted by $C_{pk}(MNP)_1$ for the model (11) and $C_{pk}(MNP)_2$ for the model (12), are given in Table 5.

In Table 4, apparently, estimates of regression parameters of MNP family are very similar to those of NP family. As for an estimate of the power parameter λ , in Case (1) $\hat{\lambda}$ of MNP family is very similar to that of BC family, but very different in Case (2). Regarding various C_{pk} values in Table 5, the behavior of $C_{pk}(MNP)_1$ is the same as that of $C_{pk}(BC)_1$, and the behavior of $C_{pk}(MNP)_2$ is very similar to that of $C_{pk}(NP)_2$. After all, behaviors of various C_{pk} values in Table 5 are very similar except for that of $C_{pk}(BC)_2$. This observation suggests the limits of estimation with respect to BC family.

We summarize our conclusions obtained from Cases (1), (2) and (3).

- (i) In model fitting there seems to be no difference among three families (NP, BC and MNP) in the meaning of SSE.
- (ii) In estimation of C_{pk} , BC family with a regression structure in ξ and $\log \tau$ is less effective than the other families NP and MNP.
- (iii) Usual C_{pk} is inclined to be larger than the true C_{pk} in existence of some non-normality, such as positive skewness.

Some behavior of usual C_{pk} under the existence of positive skewness

To confirm the conclusion (iii) in the preceding subsection, we shall consider a situation that NP family (2) is appropriate for some data but we want to apply Normal family with mean $E[X]$ and variance $V(X)$ to calculate usual C_{pk} of the data, where notations E and V denote Expectation and Variance operators respectively. Our concern is to numerically evaluate a bias between usual C_{pk} and the true value of C_{pk} with respect to NP family (2). Thus, the bias is defined by

$$\begin{aligned} \text{Bias} &= C_{pk} - C_{pk}(\text{NP}) \\ &= \frac{\text{Upper specified value} - E[X]}{3\sqrt{V(X)}} - \frac{1}{3\delta} \left\{ \left(\frac{\text{Upper specified value}}{\eta} \right)^{\frac{\delta}{\sigma}} - \exp(-\delta^2) \right\}. \end{aligned}$$

Furthermore, we are interested in skewness, and so we examine the standardized third-order cumulant $G_3 = E[(X - E[X])^3] / \{V(X)\}^{3/2}$.

We shall deal with four cases: (1) $C_{pk}(NP)=2/3$, (2) $C_{pk}(NP)=1$, (3) $C_{pk}(NP)=4/3$, and (4) $C_{pk}(NP)=3/2$ with fixed $\eta=1$ and fixed σ ($=0.1$ or 0.5). For a given value of $C_{pk}(NP)$ we obtain the corresponding upper specified value from NP family (2) with each δ and σ . In Figure 5 we give a graph of (G_3, Bias) with respect to NP family (2) for $-0.2 \leq \delta \leq 0.2$. The reason why the range of δ is a little narrower than $-1/3 < \delta < 1/3$ is for the security of computational precision.

There is a clear tendency that Bias is positive for positive G_3 and negative for negative G_3 . We also know that increasing σ enlarges Bias as well as skewness G_3 in a positive direction.

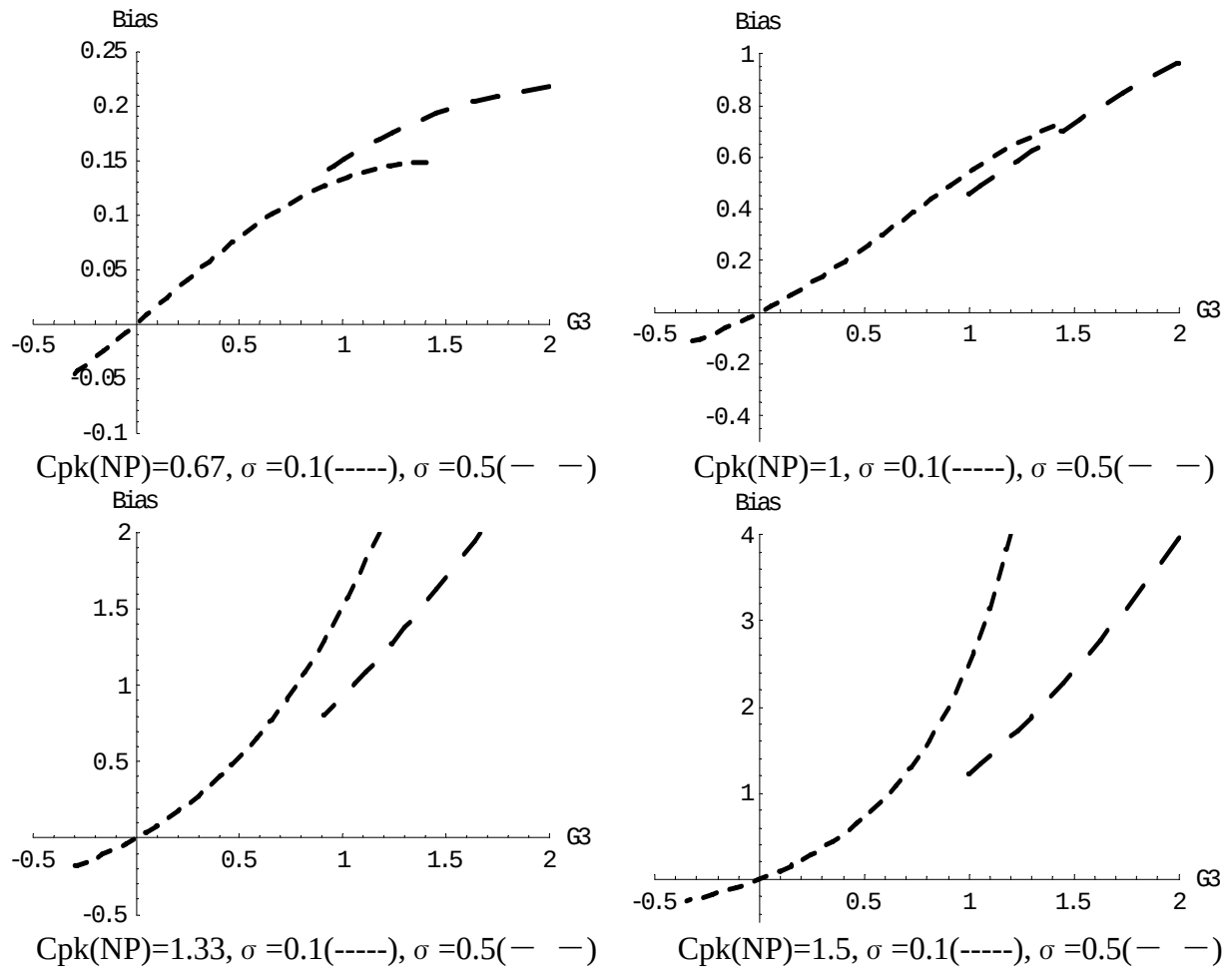


Figure 5. Graphs of (G_3, Bias) with respect to NP family for $-0.2 \leq \delta \leq 0.2$.

Conclusions

In this paper we reported one case study to show the manner of handling non-normal, especially positively skewed data sets that we often come across in application of a normal

family.

The topic was ANOVA with respect to the Process Capability Index C_{pk} . The value of usual C_{pk} is largely biased for positively skewed data. To remedy a bias of usual C_{pk} , we adopted our new power normal family and defined a new version of C_{pk} . In a series of subsequent analyses, our new power normal family and its modified one have shown their stable performance in estimation of parameters as well as C_{pk} , but Box-Cox power normal family has not. This unstable property of Box-Cox family seems to be due to a functional relationship among its parameters which brings high correlation relationships among estimators (Isogai, 2005).

References

- Box, G.E.P. & Cox, D.R. (1964) An analysis of transformations, *Journal of Royal Statistical Society*, B, 26, pp.211-252.
- Isogai, T. (1999) Power transformation of the F distribution and a power normal family, *Journal of Applied Statistics*, 26, pp.355-371.
- Isogai, T. (2001) Applications of power transformation formulas of the F distribution and a power normal family, *Journal of the Japanese Society for Quality Control*, 31, pp.89-104, (in Japanese).
- Isogai, T., Katano, Y. & Miyata, K. (2004) Models and inference for corrosion pit depth data, *Extreme*, 7, pp.253-270.
- Isogai, T. (2005) Applications of a new power normal family, *Journal of Applied Statistics*, 32, pp.421-436.
- Isogai, T., Uchida, H., Miyama, S. & Nishiyama, S. (2008) Statistical modeling of enamel rater value data, *Journal of Applied Statistics*, 35, pp.515-535.
- Johnson, N. L. & Kotz, S. (1970) *Continuous univariate distributions-1*, Wiley-Interscience.
- Prentice, R. L. (1975) Discrimination among some parametric models, *Biometrika*, 62, pp.607-614.