

AD-A080 886

WISCONSIN UNIV-MADISON DEPT OF STATISTICS
TRANSFORMATION OF A DISCRETE DISTRIBUTION TO NEAR NORMALITY.(U)
NOV 79 F HERNANDEZ, R A JOHNSON
UNIS-08-79-046

F/S 12/1

N00014-78-C-0722

NL

UNCLASSIFIED

1 10
AC
2000 0000



END

DATE

FILMED

3 - 80

DDC

(12) LEVEL II

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN

Madison, Wisconsin

ADA 080886

(15) N 140.14-72-3-0722

(14) UNIS-DS-79-
Technical Report, No. 546
(11) November 1979
(12) 12

(6)

TRANSFORMATION OF A DISCRETE DISTRIBUTION
TO NEAR NORMALITY.

(10)

by
Fabian/Hernández Richard A. Johnson
University of Wisconsin University of Wisconsin

DTIC
ELECTE
S FEB 21 1980 D

B

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

80 2 20 003
400 243

AB

DDC FILE COPY

Transformation of a Discrete Distribution to Near Normality¹

by

Fabian Hernández² Richard A. Johnson
University of Wisconsin University of Wisconsin

0. Abstract

Utilizing an information number approach, we propose an objective method for the normalization of either discrete distributions, or sample counts, by means of a power transformation. Approximations are also given to the original known probabilities. Next, we derive the large sample distribution of our estimate of the power transformation. We compare our methods with the Box-Cox procedure, applied to observed counts, and conclude that their technique often provides good approximations even though their underlying assumption of normality is clearly violated. Two examples illustrate our methods.

1. Introduction and Summary

The transformation or 're-expression' of counts is now common practice for the data analyst. Tukey (1977), page 83, specifically mentions some advantages of transforming counts. Our procedure selects a 'normalizing' transformation from the family

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

considered by Box and Cox. It provides an objective way to determine λ .

¹ This research was sponsored by the Office of Naval Research under Grant No. N00014-78-C-0722.

² Now at Dirección General de Estadística, México.

In Section 2, we discuss criteria by which discrete random variables, or their transformations, may be judged to be nearly normal. Employing the Kullback-Leibler information number, we introduce a transformation technique that applies where the true underlying distribution is known. Also, we develop a discretization of the normal distribution that approximates a given discrete distribution.

Sample analogues, of the methods in Section 2, are developed in Section 3 and their asymptotic distribution derived. These can be used to obtain approximate confidence intervals for the 'best' transformation parameter.

For comparative purposes, we show that the Box-Cox technique leads to sensible results, provided that we add a positive constant. Section 5 includes the re-expression of counts and the normalization of raw test scores.

2. Transformation and Approximation of a Known Discrete Distribution.

2.1 Normal Approximation to Transformation of Smoothed Discrete Random Variables.

Let X be a discrete random variable which takes the value i with probability $P_i = P[X=i]$ for $i \geq 0$. Then $Y = X+U$ is absolutely continuous when U is independent of X and is uniform on $[c, c+1]$ some fixed $c > 0$.

Let Y have p.d.f. $g(\cdot)$, $V = Y(\lambda)$ have p.d.f. $g_\lambda(\cdot)$ and $\phi_{\mu\sigma}(\cdot)$ be the p.d.f. of a normal distribution with mean μ and standard deviation σ . In our search for a transformation, we replace the discrete variable X by the absolutely continuous Y , and then select a power transformation of Y .

Employing the Kullback-Leibler information number between g_λ and $\phi_{\mu\sigma}$ as a measure of closeness, we propose to minimize

$$I[g, \lambda]_{\mu, \sigma} = E_{g, \lambda} \left\{ \log \left(\frac{g(V)}{\phi_{\mu, \sigma}(V)} \right) \right\} \quad (2.1)$$

with respect to μ, σ and λ . Minimizing (2.1) first over μ and σ , we find that the optimal value λ_0 of λ is found by minimizing

$$G(\lambda) = \frac{1}{2} [\log(2\pi) + 1 + E_g (\log[g(X)]) + (1-\lambda) E_g (\log(X)) + \log[V_g(X^{\lambda})]] \quad (2.2)$$

provided that $E_g(X^2)$ and $E_g(\log(X))^2$ are finite. Here,

$V_g(X^{\lambda}) = E[X^{\lambda}(\lambda) - E(X^{\lambda})]^2$ and $E_g(\cdot)$ denotes that the expectation is taken with respect to g .

Proposed Procedure for Transforming a Known Discrete Distribution (2.1).

Replace X by the absolutely continuous random variable $Y = X + U$, where X has the given discrete distribution and U , stochastically independent of X , has a uniform distribution on the interval $[c, c+1]$, for some $c > 0$. To 'normalize' X , we make the p.d.f. of $Y(\lambda)$ 'closest' (in the information number sense) to a normal p.d.f. by minimizing (2.1) with respect to (μ, σ, λ) .

Example 1 Let X have the Poisson distribution with parameter α . Let $Y = X + U$ with $U \sim U[1, 2]$ (i.e. $c=1$). The function $G(\cdot)$ defined in (2.2) becomes

$$\begin{aligned} G(\lambda) = \text{const} - \lambda \sum_{i=0}^{\infty} \{ (1+i) \log[1/(1+i)] + \log(2+i) \} P_0(i, \alpha) - \lambda \\ + \frac{1}{2} \log \sum_{i=0}^{\infty} (1+i)^{2\lambda+1} \{ [1/(1+i)]^{2\lambda+1} - 1 \} P_0(i, \alpha) / (1+2\lambda) \\ - \left(\sum_{i=0}^{\infty} (1+i)^{\lambda+1} \{ [1/(1+i)]^{\lambda+1} - 1 \} P_0(i, \alpha) / (1+\lambda)^2 \right) - \log(|\lambda|) \end{aligned}$$

for $\lambda \neq -\frac{1}{2}, -1$. Here $\text{const} = [\log(2\pi) + 1] / (2 + \alpha \log(\alpha) - 1) + \sum_{i=0}^{\infty} \{ \log(1+i) + (1+i) \log[(2+i)/(1+i)] \} P_0(i, \alpha)$ and $P_0(i, \alpha) = \alpha^i \exp(-\alpha) / i!$. In Figure 1, we plot $G(\lambda)$ vs λ for $\alpha = 2, 3, 4$ and 5.

According to our criterion, the usual variance stabilizing transformation (See Bartlett (1947), p. 41), $\lambda = \frac{1}{2}$, is a reasonable choice since it makes $[G(\lambda) - G(\frac{1}{2})]$ 'small' for these values of α . In Table 2.1, we record λ_0 , the information number of the transformed variable $I[g_{\lambda_0}; \phi_{\mu, \sigma}]$, the information number of the untransformed variable $I[g_1; \phi_{\mu, \sigma}]$, $\mu_V = \alpha \lambda_0$, and $\sigma_V^2 = \alpha + 1/12$. We also include the information number $I[g_{\lambda_0}; \phi_{\mu, \sigma}(h_1)]$ corresponding to the square root transformation.

Table 2.1

Transformation and Comparison of Information Numbers for the Poisson Distribution				
α	λ_0	$I[g_{\lambda_0}; \phi_{\mu, \sigma}]$	$I[g_1; \phi_{\mu, \sigma}]$	$I[g_{\lambda_0}; \phi_{\mu, \sigma}(h_1)]$
2	.36889	.033815	.08104	.035925
3	.48816	.019620	.05047	.019641
4	.54318	.013090	.03572	.013308
5	.57540	.009729	.02753	.013167

2.2. Normal Approximation of Discrete Probabilities

Alternatively we can approximate the probability P_i by a probability q_i obtained from a normal c.d.f. To measure the accuracy of this type of approximation we utilize the Kullback-Leibler information number in its discrete population version (see Kullback (1968) p. 126).

Let $dc(0,1)$. We propose to approximate p_i by

$$q_i \approx q_i(\theta) = \phi\left(\frac{(1-d)(\lambda)}{\sigma}\right) - \phi\left(\frac{(1-d)(\lambda)}{\sigma}\right), \quad i \geq 1. \quad (2.3)$$

where $\phi(\cdot)$ is the c.d.f. of a standard normal distribution.

Here $q_0 = \phi\left(\frac{(1-d)(\lambda)}{\sigma}\right) - \mu$ and if $p_i = 0$ for $i > N$, $q_N = 1 - \phi\left(\frac{(N-1+d)(\lambda)}{\sigma}\right) - \mu$.

Let $P = \{p_i; i \geq 0\}$ and $Q = \{q_i; i \geq 0\}$ with q_i defined in (2.3).

Proposed Procedure for the Normal Approximation of Probabilities (2.2)

Let $dc(0,1)$ be fixed. Minimize the Kullback-Leibler information number

$$I[P;Q] = p_0 \log \left\{ \frac{p_0}{\phi\left(\frac{(1-d)(\lambda)}{\sigma}\right) - \mu} \right\} + \sum_{i=1}^N p_i \log \left\{ \frac{p_i}{\phi\left(\frac{(1+d)(\lambda)}{\sigma}\right) - \mu} \right\} \quad (2.4)$$

with respect to $\theta = (\mu, \sigma, \lambda)$.

Example 2 Let X have a binomial distribution with parameters m

and p and set $q = 1-p$. We take $d = \frac{1}{2}$ for the normal probabilities (2.3).

In Table 2.2, we display the optimal choices μ, σ, λ , that minimize (2.4),

the information number for the 'best' approximating normal probabilities,

$I[P;Q(\theta)]$, and the information number $I[P;R]$ corresponding to the

'usual' normal approximation which sets $\mu = mp$, $\sigma = \sqrt{mpq}$ and $\lambda = 1$. The

results are given for $p \leq 0.5$ since the last two columns are symmetric

about $p = 0.5$.

ACCESSION for	
NTIS	White Section ☒
DDC	Buff Section ☐
UNANNOUNCED	☐
JUSTIFICATION	
BY	
DISTRIBUTION/AVAILABILITY CODES	
Dist.	ALL and/or SPECIAL
A	

Table 2.2

Approximation of Binomial Probabilities, $m = 10$

p	μ	σ^2	λ	$I[P;Q(\theta)]$	$I[P;R]$
0.1	0.74123	0.53867	0.56941	0.00003	.01678
0.2	1.53683	0.75304	0.65421	0.00014	.00813
0.3	2.37057	1.01300	0.74109	0.00032	.00538
0.4	3.39250	1.42022	0.84353	0.00056	.00238
0.5	4.86330	2.22872	0.97668	0.00087	.00115

Notice that the Procedure (2.2) fits the binomial probabilities

better when p is 'small' (or 'large'). This is opposite the

performance of the 'usual' normal approximation. Table 2.3 displays

$P, Q(\theta)$ and R so that comparisons can be made. Since the

probabilities $Q(\theta)$ and R are symmetric functions of p about $p=0.5$,

we display them only for $p \leq 0.5$.

3. Transformation of Counts and Normal Approximation of Observed

Proportions.

3.1 Transformation of Counts

Let X be a discrete random variable taking the value i with

probability $p_i = P[X=i] > 0$, for $i = 0, 1, \dots, N < \infty$. Let $X_1, \dots, X_n$

be i.i.d. as X and denote the frequency of the value i by f_{in} . Set

$I_{(i)} =$ the indicator of $[X_i=i]$ so $f_{in} = \sum_{i=1}^n I_{(i)}(X_i)$ is the frequency of $[X=i]$ and the relative frequency is

$$\hat{p}_{in} = f_{in}/n \quad \text{for } i = 0, 1, \dots, N \quad (3.1)$$

Once having observed $x_1, x_2, \dots, x_n$, we treat these as possible values of a random variable Y_n , where Y_n now takes the value i with probability

Table 2.3

Comparison of Probabilities
 $P = \text{Binomial}$
 $Q(e_n) = \text{Best transformed distribution}$
 $R = \text{Usual normal approximation}$

i	p = 0.1			p = 0.2		
	P	Q(e _n)	R	P	Q(e _n)	R
0	0.34868	0.34862	0.29908	0.10737	0.10824	0.11784
1	0.36742	0.36553	0.40184	0.26843	0.26448	0.22848
2	0.19371	0.19528	0.242157	0.30199	0.30639	0.30737
3	0.05740	0.05695	0.05272	0.20133	0.20181	0.22848
4	0.01116	0.01103	0.00409	0.08808	0.08624	0.09379
5	0.00149	0.00159	0.00011	0.02642	0.02591	0.02122
6	0.00014	0.00018	0.00000	0.00550	0.00578	0.00264
7	0.00001	0.00002	0.00000	0.00079	0.00099	0.00018
8	0.00000	0.00000	0.00000	0.00007	0.00014	0.00000
9	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
10	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

i	p = 0.3			p = 0.4		
	P	Q(e _n)	R	P	Q(e _n)	R
0	0.02825	0.02964	0.04225	0.00605	0.00732	0.01193
1	0.12106	0.11667	0.10806	0.04031	0.03851	0.04136
2	0.23347	0.23442	0.21472	0.12093	0.11783	0.11317
3	0.26683	0.27135	0.26993	0.21499	0.21701	0.20698
4	0.20012	0.20018	0.21472	0.25082	0.25571	0.25311
5	0.10292	0.10036	0.10806	0.20066	0.20079	0.20698
6	0.03676	0.03577	0.03439	0.11148	0.10824	0.11317
7	0.00900	0.00937	0.00691	0.04247	0.04097	0.04136
8	0.00145	0.00185	0.00088	0.01062	0.01109	0.01010
9	0.00014	0.00028	0.00007	0.00157	0.00218	0.00164
10	0.00001	0.00004	0.00000	0.00010	0.00035	0.00019

i	p = 0.5		
	P	Q(e _n)	R
0	0.00098	0.00173	0.00221
1	0.00977	0.01003	0.01122
2	0.04395	0.04171	0.04349
3	0.11719	0.11416	0.11447
4	0.20508	0.20746	0.20452
5	0.24609	0.25196	0.24817
6	0.20508	0.20548	0.20452
7	0.11719	0.11294	0.11447
8	0.04395	0.04196	0.04349
9	0.00977	0.01056	0.01122
10	0.00098	0.00203	0.00221

□

$\hat{p}_{in}$. Next, construct

$$V_n = Y_n + U_n \text{ with p.d.f. } g_n(\cdot) \quad (3.2)$$

where Y_n and U_n are stochastically independent and U_n has a uniform distribution on the interval $[\beta, \beta+1]$, for some fixed $\beta > 0$. Let $W_n = V_n^{(\lambda)}$ have p.d.f. $g_n(\cdot)$ and $\phi_{\mu\sigma}(\cdot)$ be the p.d.f. of a normal distribution with mean μ and standard deviation σ .

Now, suppose we want to transform the data so that they appear to come from a normal distribution. We propose to transform V_n . Instead of the original observations $x_i, i = 1, \dots, n$ and to select a transformation λ , in such a way that the distribution of $V_n^{(\lambda)}$ is 'closest' to a normal distribution in the Kullback-Leibler information number sense. That is, we minimize

$$I[g_n, \phi_{\mu\sigma}] = E \left\{ \log \left[\frac{g_n(V_n)}{\phi_{\mu\sigma}(V_n)} \right] \right\} \quad (3.3)$$

with respect to μ, σ and λ .

Thus, minimizing first with respect to μ and σ , it is easily shown that the optimal value, λ_{en} , of λ is obtained by minimizing

$$G_n(\lambda) = E \left[\log(2\pi) + 1 + \frac{1}{g_n} (\log g_n(V_n)) + (1-\lambda) E [\log(V_n)] \right] \quad (3.4)$$

$$+ 2 \log[V_{g_n}^{(\lambda)}]$$

In summary,

Proposed Procedure for Transforming Counts to Near Normality (3.1)

Having observed $x_1, x_2, \dots, x_n$, introduce the discrete random variable Y_n which takes the value i with probability $\hat{p}_{in} = \frac{1}{n} \sum_{j=1}^n I_{(i)}(x_j)$.

Replace Y_n by the absolutely continuous random variable $V_n = Y_n + U_n$, where U_n is independent of Y_n and is uniform on the interval $[\beta, \beta+1]$.

some fixed $\beta > 0$. In order to 'normalize' the observations, we transform V_n employing Procedure 2.1 $\square$

The transformation λ is selected by minimizing (3.4) with respect to λ .

In order to determine the asymptotic behavior of $\theta_n = (\mu_n, \sigma_n, \lambda_n)'$, the value of θ that minimizes (3.3), we establish the following auxiliary result.

Lemma 3.2. Suppose X is a discrete random variable with $p_i = P[X=i]$, $i=0,1,\dots,M < \infty$. Let $X_1, \dots, X_n$ be independently distributed as X and V_n , defined by (3.2), have p.d.f. $g_n(\cdot)$. Set $V = X+U$, where U is uniform on the interval $[\beta, \beta+1]$ and is independent of X . Let $g(\cdot)$ be the p.d.f. of V and $g_\lambda(\cdot)$ be the p.d.f. of $V = V(\lambda)$. Then, with probability one

$$i) \lim_{n \rightarrow \infty} E_{g_n} [\log(g_n(V_n))] = E_g [\log(g(V))]$$

$$ii) \lim_{n \rightarrow \infty} E_{g_n} [\log(V_n)] = E_g [\log(V)]$$

$$iii) \lim_{n \rightarrow \infty} E_{g_n} [\log(\phi_{\mu_n, \sigma_n}(V_n))] = E_g [\log(\phi_{\mu, \sigma}(V))]$$

Proof i) Since $E_{g_n} [\log(g_n(V_n))] = \sum_{i=0}^M \hat{p}_{in} \log(\hat{p}_{in})$ it follows by (8.1) applications of the Strong Law of Large Numbers.

$$ii) \text{ Again by the Strong Law, } E_{g_n} [\log(V_n)] = \sum_{i=0}^M r_i \hat{p}_{in} \rightarrow \sum_{i=0}^M r_i p_i = E_g [\log(V)].$$

$$iii) \text{ Also } E_{g_n} [\log(\phi_{\mu_n, \sigma_n}(V_n))] = -\frac{1}{2} \log(2\pi\sigma_n^2) - \frac{1}{2} \sum_{i=0}^M n_i(\theta) \hat{p}_{in} \rightarrow -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2} \sum_{i=0}^M n_i(\theta) p_i$$

$$= E_g [\log(\phi_{\mu, \sigma}(V))]; \text{ where } n_i(\theta) = \int \left(\frac{v - \mu}{\sigma} \right)^2 dv. \quad \square$$

Theorem 3.3 Let X, V_n, V and M be as in Lemma 3.2. Set $\theta' = (\theta_1, \theta_2, \theta_3) = (\mu, \sigma, \lambda)$ and suppose that the following conditions are satisfied

i) The parameter space Θ is a compact set given by

$$\Theta = \{(\mu, \sigma, \lambda) : |\mu| \leq M, c \leq \sigma \leq d, a \leq \lambda \leq b; -\infty < 0 < b, c, d, M < \infty\}.$$

ii) $H(\theta) = I(g_\lambda; \phi_{\mu, \sigma})$ has a unique minimum at $\theta_c = (\mu_c, \sigma_c, \lambda_c) \in \Theta$. Then,

1) $\lim_{n \rightarrow \infty} \theta_n = \theta_c$, with probability one.

If we further assume that

iii) θ_c is an interior point of Θ

iv) $V^2 H(\theta_c) = \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} H(\theta) \right) \Big|_{\theta=\theta_c}$, is non-singular.

Then

2) $\sqrt{n}(\theta_n - \theta_c) \xrightarrow{d} N_3(0, VV')$ where $V = [V^2 H(\theta_c)]^{-1}$ and M is defined in (3.5).

Proof According to (3.3) and Lemma (3.2)

$$\begin{aligned} I(g_n; \phi_{\mu_n, \sigma_n}) &= E_{g_n} [\log(g_n(V_n))] + (1-\lambda) E_{g_n} [\log(V_n)] - E_{g_n} [\log(\phi_{\mu_n, \sigma_n}(V_n))] \\ &= -\frac{1}{2} \log(2\pi\sigma_n^2) + E_{g_n} [\log(g_n(V_n))] + (1-\lambda) E_{g_n} [\log(V_n)] + E_{g_n} \left[\log \left(\frac{\mu_n - \mu}{\sigma_n} \right)^2 \right] \end{aligned}$$

→ $I(g_{\lambda}; \phi_{\mu})$, with probability one, where $I(g_{\lambda}; \phi_{\mu})$ is given by (2.1). It is obvious that this convergence is uniform in $\phi \in \Theta$, except possibly for the last term. But

$$E_{\theta_0} \left| \frac{v(\lambda)}{\sigma} - \mu \right|^2 = \sum_{i=0}^{\infty} n_i(\theta) \hat{p}_{in} \rightarrow \sum_{i=0}^{\infty} n_i(\theta) p_i, \text{ uniformly in } \theta$$

because $n_i(\theta) = \int_{i+\beta}^{i+\beta+1} \left(\frac{\lambda}{\sigma} \right)^2 dv$ is uniformly continuous in θ for $\phi \in \Theta$. Since $M(\theta) = I(g_{\lambda}; \phi_{\mu})$ has a unique global minimum at θ_0 , it follows that $\lim_{n \rightarrow \infty} \theta_{0n} = \theta_0$ with probability one.

To establish the asymptotic normality, we introduce $H_n(\theta) = I(g_{n\lambda}; \phi_{\mu})$. A Taylor expansion of $\sqrt{n} \psi_{n\lambda}(\theta_{0n})$ yields

$$\sqrt{n} \psi_{n\lambda}(\theta_{0n}) = \sqrt{n} \psi_{n\lambda}(\theta_0) + \sqrt{n} \psi_{n\lambda}'(\theta_0) [\sqrt{n}(\theta_{0n} - \theta_0)]$$

where $\theta_{0n} = r_{n\lambda} \theta_0 + (1-r_{n\lambda}) \theta_{0n}$ with $r_{n\lambda} \in (0,1)$. Since $\theta_{0n} \xrightarrow{a.s.} \theta_0$, an interior point of Θ , it follows that $\psi_{n\lambda}'(\theta_{0n}) \equiv 0$ for large enough n . Hence $\sqrt{n} \psi_{n\lambda}(\cdot)$ and $\sqrt{n} \psi_{n\lambda}'(\theta_{0n})$ have the same limiting distribution. The Multivariate Central Limit Theorem applied to $\psi_{n\lambda}(\theta)$ yields $\sqrt{n} \psi_{n\lambda}(\theta_0) \xrightarrow{d} N_3(0, W)$ where

$$W = E[C_2(x_1, \theta) C_2(x_1, \theta)] \quad (3.5)$$

with

$$C_1(x_1, \theta) = - \sum_{i=0}^{\infty} \int_{i+\beta}^{i+\beta+1} \left(\frac{v(\lambda)}{\sigma} - \mu \right)^2 dv I_{(i)}(x_1),$$

$$C_2(x_1, \theta) = \frac{1}{\sigma} - \frac{1}{\sigma} \sum_{i=0}^{\infty} \int_{i+\beta}^{i+\beta+1} \left(\frac{v(\lambda)}{\sigma} - \mu \right)^2 dv I_{(i)}(x_1)$$

and

$$C_3(x_1, \theta) = \sum_{i=0}^{\infty} \int_{i+\beta}^{i+\beta+1} \left(\frac{v(\lambda)}{\sigma} - \mu \right) \frac{\partial v(\lambda)}{\partial \lambda} - \log(v) dv I_{(i)}(x_1)$$

Consequently, $\sqrt{n} \psi_{n\lambda}(\theta_0) [\sqrt{n}(\theta_{0n} - \theta_0)] \xrightarrow{d} N_3(0, W)$, and since the (k, z) entry of $\sqrt{n} \psi_{n\lambda}(\theta)$ is of the form $\sum_{i=0}^{\infty} b_{ikz}(\theta) \hat{p}_{in}$, with $b_{ikz}(\cdot)$ uniformly continuous on Θ ,

$$\lim_{n \rightarrow \infty} \sqrt{n} \psi_{n\lambda}'[r_{n\lambda} \theta_0 + (1-r_{n\lambda}) \theta_{0n}] = \sqrt{n} \psi_{n\lambda}'(\theta_0), \text{ in probability.}$$

Furthermore, $\sqrt{n} \psi_{n\lambda}(\theta_0)$ is non-singular, and setting $V = [\sqrt{n} \psi_{n\lambda}(\theta_0)]^{-1}$, the result follows from Slutsky's theorem. $\square$

Remark Theorem 3.3 says that θ_{0n} converges, with probability one, to θ_0 , the value of θ that minimizes the Kullback-Leibler information number between g_{λ} , the p.d.f. of (Y, U) , and a normal p.d.f. Hence, Procedure (2.1) can be interpreted as the infinite-sample analogue of the current technique. $\square$

3.2 A Normal Approximation to Observed Proportions

We want to approximate observed proportions (3.1) by a set of normal probabilities $Q(\theta) = \{q_i(\theta); 0 \leq i \leq M\}$ given by (2.3).

Proposed Procedure to Approximate Observed Proportions (3.4) In order to approximate $\hat{p}_n = (\hat{p}_{1n}, \dots, \hat{p}_{Mn})$, by a collection of normal probabilities (2.3), we minimize the Kullback-Leibler information number

$$I(\hat{p}_n; Q(\theta)) = - \sum_{i=1}^M \hat{p}_{in} \log \left[\frac{\hat{p}_{in}}{q_i(\theta)} \right] \quad (3.6)$$

with respect to θ . $\square$

Theorem 3.5 Let X assume the value 1 with probability p_i for $i = 0, 1, \dots, M$ where M is finite. Let $x_1, \dots, x_n$ be i.i.d. with the same distribution as X and $\hat{p}_{in}$ be the observed proportion of the value 1. Set $\theta' = (\theta_1, \theta_2, \theta_3) = (\mu, \sigma, \lambda)$ and assume that the following conditions are satisfied.

1) The parameter space Θ is as in Theorem 3.3.

ii) $F(\theta) = I(p; Q(\theta)) = - \sum_{i=0}^M p_i \log \left[\frac{p_i}{q_i(\theta)} \right]$ has a unique minimum at θ_0 .

Then, θ_0 , the value which minimizes (3.6), satisfies

1) $\lim_{n \rightarrow \infty} \theta_n = \theta_0$, with probability one.

If we further assume that

iii) θ_0 is an interior point of Θ

iv) $\nabla^2 F(\theta_0) = \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} F(\theta) \right)_{\theta=\theta_0}$ is non-singular.

Then

2) $\sqrt{n}(\theta_n - \theta_0) \xrightarrow{d} N_3(0, VV')$ where $V = [\nabla^2 F(\theta_0)]^{-1}$ and M is defined in (3.7).

Proof 1) According to (3.6)

$$I(\hat{p}_n; Q(\theta)) = - \sum_{i=0}^M \hat{p}_{in} \log(\hat{p}_{in}) - \sum_{i=0}^M \hat{p}_{in} \log(q_i(\theta)) \longrightarrow I(p; Q(\theta)),$$

with probability one. But $q_i(\theta)$ is uniformly continuous in $\theta \in \Theta$, so the convergence is uniform in θ . Consequently, $\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta_0$, with probability one.

2) To establish the limiting normality of $\hat{\theta}_n$, we introduce the function $F(\theta) = I(\hat{p}_n; Q(\theta))$. A Taylor expansion of $\sqrt{n}F(\hat{\theta}_n)$ yields

$$\sqrt{n}F(\hat{\theta}_n) = \sqrt{n}F(\theta_0) + \nabla^2 F(\theta_0) [\sqrt{n}(\hat{\theta}_n - \theta_0)]$$

where $\hat{\theta}_{1n} = t_{n0} + (1-t) \hat{\theta}_n$ with $t \in (0, 1)$. Next, since $\hat{\theta}_n \xrightarrow{d} \theta_0$ an interior point of Θ , $\nabla F(\hat{\theta}_n) \equiv 0$ for large enough n . Hence, $-\sqrt{n}F(\theta_0)$ and $\nabla^2 F(\theta_0) [\sqrt{n}(\hat{\theta}_n - \theta_0)]$ have the same asymptotic distribution. Moreover, the Multivariate Central Limit Theorem applied to $\nabla F(\theta)$ yields

$$\sqrt{n}F(\theta_0) \xrightarrow{d} N_3(0, M)$$

where

$$M = \left[\sum_{i=0}^M \frac{\partial \log q_i(\theta)}{\partial \theta_k} \cdot \frac{\partial \log q_i(\theta)}{\partial \theta_l} p_i \right]. \quad (3.7)$$

Consequently, $\nabla^2 F(\theta_0) [\sqrt{n}(\hat{\theta}_n - \theta_0)] \xrightarrow{d} N_3(0, M)$. It is easily shown that $\nabla^2 F(\theta) \rightarrow \nabla^2 F(\theta_0)$, uniformly, with probability one.

Setting $V = [\nabla^2 F(\theta_0)]^{-1}$ and utilizing Slutsky's theorem

$$V(\nabla^2 F(\theta_0) [\sqrt{n}(\hat{\theta}_n - \theta_0)]) \xrightarrow{d} N_3(0, VV')$$

and hence,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N_3(0, VV'). \quad \square$$

4. A Comparison with the Box-Cox Procedure Applied to Discrete Observations.

In our model $[X=0]$ may have positive probability so, we consider $Y = X + C$ with $c > 0.5$. The Box and Cox (1964) method selects λ by maximizing

$$s_{\max}(\lambda) = -\frac{n}{2}[\log(2\pi) + 1] - \frac{n}{2} \log \left(\sum_{i=1}^n (Y_i - \bar{Y})^2 / n \right) + (\lambda - 1) \sum_{i=1}^n \log(Y_i) \quad (4.1)$$

with respect to λ .

Lemma 4.1 Let Θ the parameter space be restricted as in Theorem 3.3. Then the MLE $\hat{\lambda}_n$ obtained by maximizing (4.1) converges with probability one to λ_0 the value of λ that maximizes

$$\phi(\lambda) = \lambda E_Y[\log(Y)] - \frac{1}{2} \log(E_Y(Y^{2\lambda}) - [E_Y(Y^\lambda)]^2) + \log(|\lambda|) \quad (4.2)$$

Proof Since Θ is compact and Y only assumes a finite number of values, it follows from Rubin (1956) that the probability one convergence of $n^{-1} s_{\max}(\lambda)$ is uniform in λ . Consequently $\hat{\lambda}_n \rightarrow \lambda_0$. $\square$

Let $Y_n = X + U$ where U is independent of X and is uniform on $[c-\frac{1}{2}, c+\frac{1}{2}]$. Procedure 2.1, the large sample limit of Procedure 3.1, selects λ to minimize $G(\lambda)$. Alternatively, the Box-Cox approach, applied directly to the discrete observations, requires the minimization of $\phi(\lambda)$. It can be shown that

$$G_1(\lambda) = -\phi(\lambda) + \text{constant} + \text{Error}$$

where

$$\text{Error} = \lambda \sum_{i=0}^{1+c+h} p_i \int_{1+c-h}^{1+c+h} [\log(1+c) - \log(y) dy]$$

$$- \frac{1}{2} \log \left[\frac{\sum_{i=0}^{1+c+h} p_i (1+c)^{2\lambda} \left[\sum_{j=0}^N p_j (1+c)^j \right]^2}{\sum_{i=0}^{1+c+h} p_i \int_{1+c-h}^{1+c+h} y^{2\lambda} dy - \left[\sum_{i=0}^{1+c+h} p_i \int_{1+c-h}^{1+c+h} y^\lambda dy \right]^2} \right] \quad (4.3)$$

When

$$\left| \int_{1+c-h}^{1+c+h} [\log(y) - \log(1+c)] dy \right| \leq \frac{1}{1+c-h} \quad (4.4)$$

and, for $r=1,2$

$$\left| \int_{1+c-h}^{1+c+h} [y^r - (1+c)^r] dy \right| \leq \frac{|r|}{2} (1+c-h)^{r-1}, r \leq 1 \quad (4.5)$$

are small, we would expect the Error to be small. Consequently the Box-Cox procedure and Procedure (3.1) should give nearly the same answer when the sample size is large.

From (4.4) and (4.5) making c large appears to improve the agreement between the two procedures. The most frequently employed transformations $\sqrt{X+C}$, $\sqrt[4]{X+C}$, $\log(X+C)$ and $(X+C)^{-1}$ satisfy $r \leq 1$.

5. Application to the Re-expression of Counts and Normalization of Test Scores

Example 1 Tukey (1977), page 572, displays the following counts for the duration (in days) of incubation for 1663 eggs of the ridley turtle

days	50	51	52	53	54	55	56	57	58	59	60	61	64	65
number of eggs	77	122	10	321	725	180	162	21	14	6	9	14	1	1

We suggests the re-expression $\sqrt{\text{days}-49}$.

Setting $\beta = 0$, we apply the Proposed Procedure 3.1 to the above data and obtain $\theta_{\lambda,1663} = (3.570, 1.333, 0.815)'$. Moreover, using the limiting distribution of $\theta_{\lambda,n}$ derived in Theorem 3.3, we can establish an approximate confidence interval for λ_n . The estimated standard error of $\lambda_{n,1663}$ is $\sqrt{.015/1663}$. Hence, an approximate 95% confidence interval for λ_n is $(0.697, 0.933)$. Notice that $\lambda = \frac{1}{2}$ is not included in the interval. $\square$

Example 2 Ghiselli (1964), page 78, proposes the use of the square root transformation for the normalization of the 100 test scores.

We set $\beta = 0$. The application of the Proposed Procedure 3.1 yields $\theta_{\lambda,100} = (4.66, 0.54, 0.070)$. Utilizing the limiting distribution of $\theta_{\lambda,n}$ we obtain the estimated standard error of $\lambda_{n,100}$ and then the approximate 95% confidence interval $(-0.30, 0.44)$ for λ_n . Figure 3 presents a comparison of the relative frequency histograms of a) the transformed scores and b) the original scores. We also applied the Box-Cox procedure to the above scores. The estimated power transformation is $\hat{\lambda} = 0.078$ which is in good agreement with the value $\lambda_{n,100} = 0.070$. $\square$

The sample procedures introduced in Section 3, can also be applied to situations where the observations can only be ordered. That is, the observations can be assigned to exactly one of the categories $1, 2, \dots, M$. Here M corresponds to the highest category, $M-1$ to the second highest, etc. With this scoring, we are able to apply our methods to the relative frequencies of the categories.

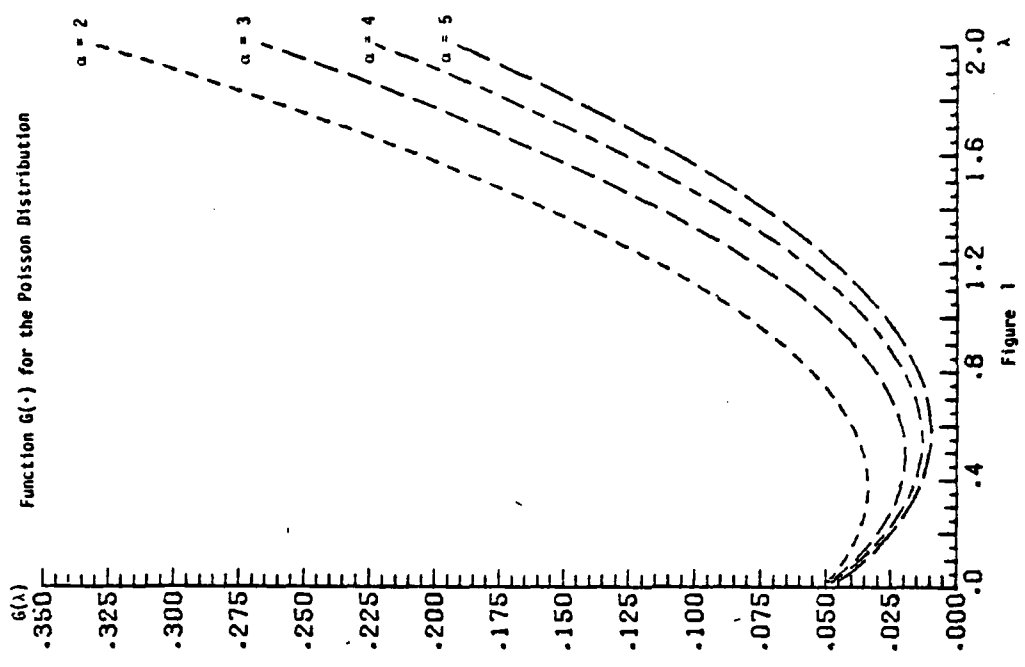
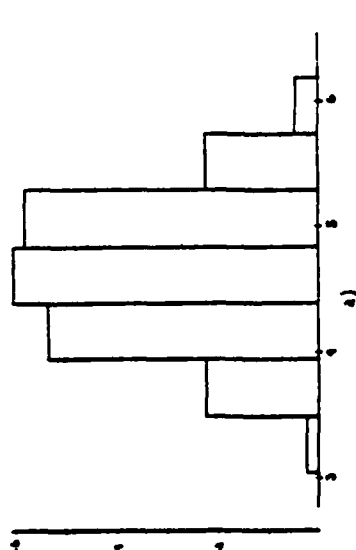


Figure 1

REFERENCES

- Bartlett, M.S. (1947). "The Use of Transformations." Biometrics, 3, 39-52.
- Box, G. E. P. and Cox, D. R. (1964). "An Analysis of Transformations." J. Roy. Statist. Soc., Ser. B, 26, 211-243, discussion 244-252.
- Ghiselli, E. E. (1964). Theory of Psychological Measurement, New York: McGraw-Hill, Inc.
- Kullback, S. (1968). Information Theory and Statistics, New York: Dover Publications, Inc.
- Rubin, H. (1956). "Uniform Convergence of Random Functions with Applications to Statistics." Ann. Math. Statist., 27, 200-203.
- Tukey, J. W. (1977). Exploratory Data Analysis, Reading: Addison-Wesley Publishing Company.

Relative Frequency Histogram of the Transformed Scores



Relative Frequency Histogram of the Original Scores

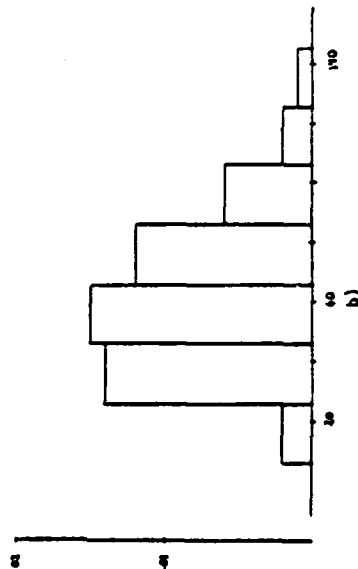


Figure 2

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS USE ONLY COMPLETING FUND
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
Technical Report No. 546		
4. TITLE (and Subtitle) TRANSFORMATION OF A DISCRETE DISTRIBUTION TO NEAR NORMALITY		
5. AUTHOR(s) Fabian Hernández Richard A. Johnson		
6. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics University of Wisconsin Madison, Wisconsin 53706		
7. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research 800 N. Quincy Street Arlington, VA 22217		
8. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		
9. DISTRIBUTION STATEMENT (of this Report) Distribution of this document is unlimited		
10. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
11. SUPPLEMENTARY NOTES		
12. KEY WORDS (Continue on reverse side if necessary and identify by block number) Transformations Discrete distributions		
13. ABSTRACT (Continue on reverse side if necessary and identify by block number) Utilizing an information number approach, we propose an objective method for the normalization of either discrete distributions, or sample counts, by means of a power transformation. Approximations are also given to the original known probabilities. Next, we derive the large sample distribution of our estimate of the power transformation. We compare our methods with the Box-Cox procedure, applied to observed counts, and conclude that their technique often provides good approximations even though their underlying assumption of normality is clearly violated. Two examples illustrate our methods.		

DD FORM 1 JAN 76 1473

EDITION OF 1 NOV 68 IS OBSOLETE

SA 0102-1P-014-6400

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DISTRIBUTION STATEMENT A

Approved for public release

Distribution Unlimited