The Annals of Statistics
1995, Vol. 23, No. 5, 1587–1593

TESTS FOLLOWING TRANSFORMATIONS[1]

BY HANFENG CHEN

Bowling Green State University

Chen and Loh showed that the Box–Cox transformed two-sample t-test is more powerful than the ordinary t-test under Pitman alternatives where the location shifts appear in the untransformed scale. In this article, we prove that Chen and Loh's result also holds for a general family of transformations. An upper bound on the asymptotic relative efficiency (ARE) is obtained. In addition, we investigate bounds on the ARE under Pitman location shift alternatives in the transformed scale. We find that when the estimate for λ is consistent, a lower bound on the ARE is the reciprocal of Fisher information of the standard transformed distribution. This lower bound is close to 1 for commonly used symmetric distributions.

1. Introduction. It has been common practice to transform or reexpress data so that the transformed data are more appropriate and convenient for statistical analysis. Considering a two-sample problem, Chen and Loh (1992) recently investigated the asymptotic testing power properties of the Box–Cox transformed t-test. They showed that the transformed t-test is asymptotically more powerful than the ordinary t-test under Pitman alternatives where the location shifts appear in the original scale. The present article continues to study the problem in the following two aspects: (1) A general family of transformations $h(x; \lambda)$, $\lambda \in \Lambda$, is considered, where h is a specific function and Λ is a subset (usually an interval) of the real line; (2) the same problem is restudied under Pitman location shift alternatives in the transformed scale.

In analysis of the transformed data, one may report and interpret inferential results either in the original scale on which variables are measured or in the transformed scale, or in both scales (why not?), depending on goals of investigation and popularity or scientific implications of the scales. Therefore, theoretical justifications in the context of either scale for a statistical test based on the transformed data are desirable and meaningful.

Necessity of study for a general family of transformation is clear, since it has been evident that in practice the Box–Cox power transformation family is merely one of many popular transformation families [e.g., the modulus transformations proposed by John and Draper (1980), sinh transformations by

Received January 1992; revised September 1994.
[1]Research supported by a BGSU Faculty Research Committee Basic Grant for 1992.
AMS 1991 *subject classifications*. Primary 62F05, 62G20; secondary 62F03, 62E20, 62F35.
Key words and phrases. Asymptotic relative efficiency, Box–Cox power transformation, Fisher information, Pitman alternative, two-sample problem.

Johnson (1949), the folded power transformations by Mosteller and Tukey (1977), etc.].

Section 2 sets up the framework of the study and introduces some notation. Section 3 discusses bounds on the asymptotic relative efficiency (ARE) of the t-test following a general transformation against the ordinary t-test under Pitman alternatives in the original scale. It will be shown that Chen and Loh's (1992) result also holds for a rather general family of transformations. In addition, an upper bound is obtained and a necessary and sufficient condition to reach the lower bound 1 is established. Section 4 deals with Pitman location shift alternatives in the transformed scale. A lower bound on the asymptotic relative efficiency (ARE) of the transformed t-test is found to be the reciprocal of Fisher information of the standardized transformed distribution.

2. Model description and notation. Suppose $(X_1, \dots, X_{n_1})$ and $(Y_1, \dots, Y_{n_2})$ are two independent samples observed on the variables X and Y, respectively. Let $n = n_1 + n_2$ be the combined sample size with $n_1/n \to \pi \in (0, 1)$ as n_1 and $n_2 \to \infty$. Suppose $h(x; \lambda)$, $\lambda \in \Lambda$, is a general family of increasing transformations of x chosen to make the distributions of $h(X; \lambda)$ and $h(Y; \lambda)$ (nearly) normal or symmetric. The function h here is assumed specific, but the $\lambda \in \Lambda$ is an unknown transformation parameter, where Λ is a subset (usually an interval) of the real line. This general framework for the analysis of transformed data was first suggested by Bickel and Doksum (1981) while they addressed some concerns with analysis of the transformed data.

Suppose that the null hypothesis to be tested is

$$H_0: Y \text{ and } X \text{ have the same distribution,} \tag{1}$$

or equivalently, for some $\lambda_a \in \Lambda$,

$$H_0: h(Y; \lambda_a) \text{ and } h(X; \lambda_a) \text{ have the same distribution.} \tag{2}$$

Two expressions (1) and (2) of H_0 are presented here to motivate different setups of alternatives. To study the asymptotic power properties of tests for H_0, we consider two different sequences of Pitman-type local alternatives corresponding to the null hypotheses (1) and (2), respectively, as follows:

$$H_1: Y \text{ and } X + cn^{-1/2} \text{ have the same distribution}$$

and

$$H_2: h(Y; \lambda_a) \text{ and } h(X; \lambda_a) + cn^{-1/2} \text{ have the same distribution.}$$

The alternative H_1 means that the two populations follow a location-shift model in the original scale. The alternative H_2 describes the situation in which a transformation is successful in establishing a location-shift model in the transformed scale.

Suppose that the test statistic for H_0 is the transformed t-statistic, that is, Student t-test based on the transformed data,

$$t(\hat{\lambda}) = (n_1 n_2/n)^{1/2}\big[\overline{Y}(\hat{\lambda}) - \overline{X}(\hat{\lambda})\big]/s(\hat{\lambda}),$$

where $\hat{\lambda}$ is an estimate of the transformation parameter λ, $\bar{Y}(\hat{\lambda}) = n_2^{-1}\Sigma_1^{n_2} h(Y_j; \hat{\lambda})$, similarly, $\bar{X}(\hat{\lambda})$ is the sample mean of the transformed X-sample and $s^2(\hat{\lambda})$ is the pooled sample variance of the two transformed samples.

Throughout the paper, we assume the following conditions: (1) The index set Λ is compact. (2) Under H_0, $n^{1/2}(\hat{\lambda} - \lambda_0) = O_P(1)$ for some constant $\lambda_0 \in \Lambda$. (3) Let $h(x; \lambda)$ have derivatives with respect to (wrt) both variables x and λ. The derivative wrt x is denoted by $h'(x; \lambda)$ and that wrt λ denoted by $g(x; \lambda)$. Furthermore, the functions h and g are assumed to be equicontinuous in λ for $x \in S_i$, where S_i are a sequence of measurable sets such that $P(X \in \bigcup_{i=1}^{\infty} S_i) = 1$ under the null hypothesis. (4) There is a function $W(x)$ such that (i) $EW(X) < \infty$ and (ii) $h^2(x; \lambda) < W(x)$ and $|g(x; \lambda)| < W(x)$ for all $\lambda \in \Lambda$. All the transformation families mentioned in Section 1 satisfy the assumptions above.

3. ARE under H_1. The following theorem provides a formula for the limiting mean of $t(\hat{\lambda})$ under H_1.

THEOREM 1. *Suppose that under H_0, X has the pdf $f(x)$ with $\int f'(x)\,dx = 0$ and Fisher information $I(f) = E\{f'(X)/f(X)\}^2$ being finite and positive. Then under H_1, $t(\hat{\lambda})$ has an asymptotic normal distribution $N(\mu_1, 1)$, where the limiting mean is given by*

$$\mu_1 = c[\pi(1-\pi)]^{1/2} E[h'(X; \lambda_0)] / \{\mathrm{Var}[h(X; \lambda_0)]\}^{1/2}.$$

PROOF. First, using a similar idea to Doksum and Wong (1983) and Rubin's (1956) theorem, we prove that under the null hypothesis,

$$t(\hat{\lambda}) - t(\lambda_0) = o_P(1). \tag{3}$$

Write $n^{1/2}[\bar{Y}(\hat{\lambda}) - \bar{X}(\hat{\lambda})] = n^{1/2}[\bar{Y}(\lambda_0) - \bar{X}(\lambda_0)] + n^{1/2}(\hat{\lambda} - \lambda_0)[\bar{Y}'(\xi) - \bar{X}'(\xi)]$, where $\bar{X}'(\lambda) = n_1^{-1}\Sigma g(X_i; \lambda)$, $\bar{Y}'(\lambda) = n_2^{-1}\Sigma g(Y_j; \lambda)$ and ξ is between $\hat{\lambda}$ and λ_0. Applying Rubin's (1956) theorem to the independent sums $\bar{Y}'(\lambda)$ and $\bar{X}'(\lambda)$, we get that with H_0-probability 1, $\hat{Y}'(\lambda) - \bar{X}'(\lambda) \to 0$, uniformly in $\lambda \in \Lambda$. So $\bar{Y}(\xi) - \bar{X}(\xi) \to 0$ a.s. and since $n^{1/2}(\hat{\lambda} - \lambda_0) = O_P(1)$, $n^{1/2}[\bar{Y}(\hat{\lambda}) - \hat{X}(\hat{\lambda})] = n^{1/2}[\bar{Y}(\lambda_0) - \bar{X}(\lambda_0)] + o_P(1)$. Applying Rubin's theorem again, we have that with H_0-probability 1, $s^2(\lambda)$ converges uniformly in $\lambda \in \Lambda$ to $\sigma^2(\lambda) = \mathrm{Var}[h(X; \lambda)]$ and that the limiting function $\sigma^2(\lambda)$ is continuous in λ. Since $\hat{\lambda} - \lambda_0 = o_P(1)$, it follows that $s^2(\hat{\lambda}) = \sigma^2(\hat{\lambda}) + o_P(1) = \sigma^2(\lambda_0) + o_P(1)$. Thus

$$t(\hat{\lambda}) - t(\lambda_0) = n^{1/2}[\bar{Y}(\lambda_0) - \bar{X}(\lambda_0)]\{1/[\sigma(\lambda_0) + o_P(1)] - 1/s(\lambda_0)\} = o_P(1),$$

and then (3) follows from $n^{1/2}[\bar{Y}(\lambda_0) - \bar{X}(\lambda_0)] = O_p(1)$ and $s(\lambda_0) = \sigma(\lambda_0) + o_P(1)$ under the null hypothesis.

Now the theorem can be proved by using Le Cam's third lemma [Le Cam and Yang (1990)]. Let $l_n = \Sigma \log[f(Y_j - cn^{-1/2})/f(Y_j)]$. It suffices to prove that under the null hypothesis, $(t(\hat{\lambda}), l_n)$ is asymptotically jointly bivariate normal $(\theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \sigma_{12})$ with $\theta_2 = -\sigma_2^2/2$, $\theta_1 + \sigma_{12} = \mu_1$ and $\sigma_1^2 = 1$. By (3), $(t(\hat{\lambda}), l_n)$ and $(t(\lambda_0), l_n)$ are equivalent. The theorem then follows by verifying in a standard argument that $(t(\hat{\lambda}), l_n)$ has the asymptotic bivariate normal distribution [see, e.g., Hájek and Šidák (1967)]. □

From Theorem 1, the ARE of the transformed test $t(\hat{\lambda})$ against the ordinary t-test under H_1 is given by

$$e\{t(\hat{\lambda}), t|H_1\} = \{Eh'(X; \lambda_0)\}^2 \operatorname{Var}(X)/\operatorname{Var}[h(X; \lambda_0)].$$

The ARE has Pitman's interpretation of asymptotic efficiency, that is, $e\{t(\hat{\lambda}), t|H_1\}$ can be regarded as the ratio of t and $t(\hat{\lambda})$'s sample sizes at which the two tests achieve the same asymptotic power. Consequently, if $e\{t(\hat{\lambda}), t|H_1\} > 1$, then for n large, $t(\hat{\lambda})$ has greater power than t.

In order to transform to normality, following Box and Cox's (1964) idea, one may select a λ from Λ by the method of maximum likelihood estimation (MLE). The desired likelihood function is

$$L(\nu_1, \nu_2, \sigma^2, \lambda) = \text{const. } \sigma^{-n} \exp\{-A(\nu_1, \nu_2, \lambda)/(2\sigma^2)\} \Pi h'(X_i; \lambda) \Pi h'(Y_j; \lambda),$$

where $A(\nu_1, \nu_2, \lambda) = \Sigma[h(X_i; \lambda) - \nu_1]^2 + \Sigma[h(Y_j; \lambda) - \nu_2]^2$. For fixed λ, L is maximized when $\nu_1 = \overline{X}(\lambda)$, $\nu_2 = \overline{Y}(\lambda)$ and $\sigma^2 = s^2(\lambda)$. Then the MLE of λ is defined as the maximizer of $L(\overline{X}(\lambda), \overline{Y}(\lambda), s^2(\lambda), \lambda)$, that is,

$$\begin{aligned} L(\overline{X}(\lambda), \overline{Y}(\lambda), s^2(\lambda), \lambda) &= \text{const. } s^{-n} \Pi h'(X_i; \lambda) \Pi h'(Y_j; \lambda) \\ &= \text{const.} \{J_n(\lambda)\}^{-n/2}, \end{aligned}$$

where $J_n(\lambda) = s^2(\lambda) \exp\{-2z(\lambda)\}$ with $z(\lambda) = n^{-1}[\Sigma \log h'(X_i; \lambda) + \Sigma \log h'(Y_j; \lambda)]$. Note that the MLE of λ appears to be the minimizer of $J_n(\lambda)$. Under the null hypothesis, $J_n(\lambda) \to J_0(\lambda)$, where

$$J_0(\lambda) = \operatorname{Var}[h(X; \lambda)] \exp\{-2E \log h'(X; \lambda)\}. \tag{4}$$

Then the MLE of λ converges a.s. to the minimizer of $J_0(\lambda)$ under certain regularity conditions. When h is the Box–Cox power transformation, Chen and Loh (1992) prove that if $\hat{\lambda}$ is the MLE, then $e\{t(\hat{\lambda}), t|H_1\} \geq 1$. The following theorem confirms that this result holds in general. Moreover, an upper bound for the ARE is obtained and a necessary and sufficient condition to attain the lower bound 1 is established.

THEOREM 2. *Suppose that there exists* $\lambda_c \in \Lambda$ *such that* $h(x; \lambda_c)$ *is a linear transformation of* x. *If* $E \log h'(X; \lambda)$ *is finite and the limit* λ_0 *of* $\hat{\lambda}$ *is a minimizer of* $J_0(\lambda)$, *then*

$$1 \leq e\{t(\hat{\lambda}), t|H_1\} \leq I(f_0),$$

where $f_0(x)$ *is the pdf of* $(X - EX)/\{\operatorname{Var}(X)\}^{1/2}$ *under the null hypothesis. The equality in the lower bound side is attained if and only if* $h(x; \lambda_0)$ *is a linear transformation of* x.

PROOF. Since λ_0 is the minimizer of $J_0(\lambda)$ defined in (4) and since $\lambda_c \in \Lambda$ is such that $h(x; \lambda_c)$ is linear in x, we see that $J_0(\lambda_0) \le J_0(\lambda_c) = \mathrm{Var}(X)$. Thus we have

$$e\{t(\hat{\lambda}), t|H_1\} \ge \{Eh'(X; \lambda_0)/\exp\{E \log h'(X; \lambda_0)\}\}^2. \tag{5}$$

Noting that e^x is strictly convex, using Jensen's inequality yields that the right-hand side of (5) is at least 1 and equal to 1 if and only if $h'(X; \lambda_0)$ is constant, completing the proof for the lower bound 1.

To verify the upper bound, by integration by parts, we see that

$$[Eh'(X; \lambda_0)]^2 = \{E[h(X; \lambda_0) f'(X)/f(X)]\}^2.$$

Applying the Schwarz inequality gives

$$\{E[h(X; \lambda_0) f'(X)/f(X)]\}^2 \le \mathrm{Var}[h(X; \lambda_0)] I(f).$$

Thus, the upper bound follows by $I(f)\mathrm{Var}(X) = I(f_0)$. □

Noting the fact $t(\lambda_c) = t$, the requirement for the existence of λ_c seems to be natural and reasonable, as we desire to compare $t(\lambda)$ with t.

4. ARE under H_2. Now we consider the alternative H_2. Express $h(X; \lambda_a) = \mu + \sigma\varepsilon$, where ε is a random variable with mean 0 and variance 1. Then under H_2, $h(Y; \lambda_a) = \mu + cn^{-1/2} + \sigma\varepsilon$, in distribution. Assume ε has the pdf φ.

THEOREM 3. *Suppose that $\int \varphi'(x)\,dx = 0$ and $I(\varphi) = E[\varphi'(X)/\varphi(X)]^2$ is positive and finite. Then under H_2, $t(\hat{\lambda})$ and t have the normal limiting distributions $N(m_1, 1)$ and $N(m_1, 1)$, respectively, where the limiting means m_1 and m_2 are given by*

$$m_1 = c[\pi(1-\pi)]^{1/2} E\left\{\frac{h'(X; \lambda_0)}{h'(X; \lambda_a)}\right\} / \{\mathrm{Var}[h(X; \lambda_0)]\}^{1/2}$$

and

$$m_2 = c[\pi(1-\pi)]^{1/2} E[1/h'(X; \lambda_a)] / [\mathrm{Var}(X)]^{1/2},$$

provided that $E[h'(X; \lambda_0)/h'(X; \lambda_a)]$ and $E[1/h'(X; \lambda_a)]$ are finite.

The proof of the theorem is similar to that of Theorem 1. For details, see Chen (1992). By Theorem 3, the ARE of the transformed t-test $t(\hat{\lambda})$ against the ordinary t-test under H_2 is

$$e\{t(\hat{\lambda}), t|H_2\} = \left\{\frac{\mathrm{Var}(X)}{\mathrm{Var}[h(X; \lambda_0)]}\right\}\left\{\frac{E[h'(X; \lambda_0)/h'(X; \lambda_a)]}{E[1/h'(X; \lambda_a)]}\right\}^2.$$

When φ is symmetric, Hinkley (1975) and Taylor's (1985) estimators for λ are consistent, that is, $\lambda_0 = \lambda_a$. In this case, $E[h'(X; \lambda_0)/h'(X; \lambda_a)] = 1$ and $e\{t(\hat{\lambda}), t|H_2\}$ is simplified.

THEOREM 4. *In addition to the conditions in Theorem* 3, *assume that* $\lambda_0 = \lambda_a$. *Then*

$$e\{t(\hat{\lambda}), t|H_2\} \geq \tau = 1/I(\varphi).$$

PROOF. When $\lambda_0 = \lambda_a$, we have

$$e\{t(\hat{\lambda}), t|H_2\} = \{\mathrm{Var}(X)/\mathrm{Var}[h(X;\lambda_0)]\}\{E[1/h'(X;\lambda_0)]\}^{-2}.$$

By integration by parts and the Schwarz inequality, we have

$$\begin{aligned}&\{E[1/h'(X;\lambda_0)]\}^{-2}\\ &\quad = \{-\sigma^{-1}E\{X\varphi'[(h(X;\lambda_0)-\mu)/\sigma]/\varphi[(h(X;\lambda_0)-\mu)/\sigma]\}\}^2\\ &\quad \leq \sigma^{-2}\,\mathrm{Var}(X)I(\varphi).\end{aligned}$$

The theorem follows by noting that $\sigma^2 = \mathrm{Var}[h(X;\lambda_0)]$. □

COROLLARY 1. *If* $\lambda_0 = \lambda_a$ *and* $h(X;\lambda_a)$ *is normally distributed, then*

$$e\{t(\hat{\lambda}), t|H_2\} \geq 1, \tag{6}$$

and the equality is attained if and only if $h(x;\lambda_a)$ *is a linear transformation of* x.

PROOF. When φ is the pdf of standard normal, $I(\varphi) = 1$ and (6) follows. The equality holds if and only if $\varphi'[(h(X;\lambda_0) - \mu)/\sigma]/\varphi[(h(X;\lambda_0) - \mu)/\sigma]$ is linear in X, and the condition reduces to that $h(X;\lambda_0)$ is linear in X. □

Chen (1992) presents some tables containing the lower bound τ for commonly used symmetric φ such as the Student t_ν and contaminated normal. It appears that the lower bound τ is close to 1 for the commonly used symmetric distributions.

Acknowledgments. The author is grateful to an Associate Editor and the referees for many constructive suggestions that led to great improvement of the paper.

REFERENCES

BICKEL, P. J. and DOKSUM, K. A. (1981). An analysis of transformations revisited. *J. Amer. Statist. Assoc.* **76** 296–311.

BOX, G. E. P. and COX, D. R. (1964). An analysis of transformation (with discussion). *J. Roy. Statist. Soc. Ser. B* **26** 211–246.

CHEN, H. (1992). Tests following transformations to normal or location-shift models. Technical Report 92-12, Dept. Mathematics and Statistics, Bowling Green State Univ.

CHEN, H. and LOH, W. Y. (1992). Bounds on AREs of tests following Box–Cox transformations. *Ann. Statist.* **20** 1485–1500.

DOKSUM, K. A. and WONG, C-W. (1983). Statistical tests based on transformed data. *J. Amer. Statist. Assoc.* **78** 411–417.

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests.* Academia, Prague.

HINKLEY, D. V. (1975). On power transformations to symmetry. *Biometrika* **62** 101–111.
JOHN, J. A. and DRAPER, N. R. (1980). An alternative family of power transformations. *J. Roy. Statist. Soc. Ser. C* **29** 190–197.
JOHNSON, N. L. (1949). Systems of frequency curves generated by methods of translocation. *Biometrika* **36** 149–176.
LE CAM, L. M. and YANG, G. (1990). *Asymptotics in Statistics: Some Basic Concepts*. Springer, New York.
MOSTELLER, F. and TUKEY, J. W. (1977). *Data Analysis and Regression*. Addison-Wesley, Reading, MA.
RUBIN, H. (1956). Uniform convergence of random functions with applications to statistics. *Ann. Math. Statist.* **27** 200–203.
TAYLOR, J. M. G. (1985). Power transformations to symmetry. *Biometrika* **72** 145–152.

DEPARTMENT OF MATHEMATICS AND STATISTICS
BOWLING GREEN STATE UNIVERSITY
BOWLING GREEN, OHIO 43403