Southern Illinois University Carbondale **OpenSIUC**

Conference Proceedings

Educational Psychology and Special Education

2013

Characterization of the Power Method Transformation through the Method of Percentiles

Jennifer Koran Southern Illinois University Carbondale

Todd C. Headrick Southern Illinois University Carbondale, headrick@siu.edu

Tzu Chun Kuo Southern Illinois University Carbondale

Follow this and additional works at: http://opensiuc.lib.siu.edu/epse_confs Published in *JSM Proceedings*, Statistical Computing Section. Alexandria, VA: American Statistical Association. 3927-3936.

Recommended Citation

Koran, Jennifer; Headrick, Todd C.; and Kuo, Tzu Chun, "Characterization of the Power Method Transformation through the Method of Percentiles" (2013). *Conference Proceedings*. Paper 1. http://opensiuc.lib.siu.edu/epse_confs/1

This Article is brought to you for free and open access by the Educational Psychology and Special Education at OpenSIUC. It has been accepted for inclusion in Conference Proceedings by an authorized administrator of OpenSIUC. For more information, please contact opensiuc@lib.siu.edu.

A Characterization of the Power Method Transformation through the Method of Percentiles

Jennifer Koran¹, Todd C. Headrick ¹, Tzu Chun Kuo ¹
¹ Section on Statistics and Measurement, Department EPSE, 223 Wham Bldg. Southern Illinois University Carbondale, Carbondale, IL USA 62901-4618

Abstract

This paper derives a standard normal based power method polynomial transformation for Monte Carlo simulation studies, approximating distributions, and fitting distributions to data based on the method of percentiles. The proposed method is used primarily when (i) conventional estimators such as skew and kurtosis are unknown or (ii) data are unavailable but percentiles are known (e.g., standardized test score reports). The proposed transformation also has the advantage that solutions to polynomial coefficients are available in simple closed form and thus obviates numerical equation solving. The Monte Carlo results presented in this study indicate that the estimators based on the method of percentiles are substantially superior to their corresponding conventional product-moment estimators in terms of relative bias.

Key Words: Monte Carlo, power method, pseudo-random numbers, simulation, method of percentiles

1. Introduction

Robust approaches, particularly bootstrap methods, are gaining popularity for their ability to produce accurate inferences in distributions with virtually any distributional shape. One problem that arises in the application of the traditional bootstrap approach is when individual observations from which to resample are not available. In these situations the researcher may be restricted to the use of descriptive distributional statistics commonly released to the public, such as means, standard deviations, and percentiles.

When descriptive statistics are available, parametric bootstrap alternatives may be applied (see for example, Culpepper, 2013; Tong & Bentler, 2013). However, the parametric bootstrap must still address issues with the non-normality of the distribution. Typically, such approaches rely on knowledge of the skewness and kurtosis of the distribution (see for example, Headrick, 2011), which may not be included in public reports. One way of getting around this problem is to use a parametric approach based on percentiles rather than moments. Percentiles may be more commonly included in public reports than measures of skew or kurtosis. Indeed, Karian and Dudewitz (1999) introduced a percentile method for the Generalized Lambda Distribution (GLD), which can be used in a parametric bootstrap approach. Two problems remain with the GLD percentiles approach, however. First, closed-form solutions are not available, thus requiring computationally-intensive numerical solutions (Karian & Dudewitz, 1999). Second, "depending on the precise values of [a skew function] and [a kurtosis function], as many as four solutions may exist in just one region" (Karian & Dudewitz, 2011, p.180).

Thus, it is advantageous to have a parametric bootstrap method based on percentiles for the purpose of simulating or modeling nonnormal distributions. Further, such a method would be most useful if it included unique, closed-form solutions.

In view of the above, the present aim is to introduce a power method transformation for non-normal distributions using the method of percentiles. Specifically, the method is based on the median, inter-decile range, left-right tail-weight ratio (a skew function), and the tail weight factor (a kurtosis function) (Karian & Dudewicz, 2011, pp. 172-173). Solutions are obtained in closed-form, and when solutions exist, those solutions are unique.

The rest of the paper is outlined as follows. In Section 2, a summary of the power method transformation for conventional moments is provided. In Section 3, the method of percentiles is introduced, and the equations are developed for a power method transformation through the method of percentiles. In Section 4, the boundary conditions on the percentile-based power method are derived. In Section 5, a Monte Carlo simulation investigating the ability of the percentile-based power method to model various non-normal distributions is used to compare the results of the proposed method to power method results achieved using conventional moments. In Section 6, a brief conclusion is provided.

2. Preliminaries for the Power Method Transformations

2.1 General Considerations

The power method (PM) polynomial transformation based on conventional moments or the proposed method of percentiles considered herein can be generally expressed as (Headrick, 2010, pp. 12-13)

$$p(Z) = \sum_{i=1}^{m} c_i Z^{i-1}$$
 (2.1)

where Z is a standard normal random variable with probability density function (pdf) and cumulative distribution function (cdf)

$$f_Z(z) = \varphi(z) = (2\pi)^{-\frac{1}{2}} \exp\{-z^2/2\},$$
 (2.2)

$$f_Z(z) = \varphi(z) = (2\pi)^{-\frac{1}{2}} \exp\{-z^2/2\},$$

$$F_Z(z) = \Phi(z) = \int_{-\infty}^z \varphi(u) du, -\infty < z < +\infty.$$
(2.2)

Setting m = 4 (or m = 6) in (2.1) gives the third-order Fleishman (1978) (or fifth-order Headrick, 2002) class of PM distributions. The shape of p(Z) in (2.1) is contingent on the values of the coefficients c_i , which are determined by conventional moment matching techniques or through the method of percentiles described in the subsequent section.

In order for (2.1) to produce a valid pdf requires that the PM transformation be a strictly monotone increasing function. This requirement implies that an inverse function exists (p^{-1}) . As such, the cdf associated with (2.1) can be expressed as

$$F_{p(Z)}(p(z)) = \Phi(z). \tag{2.4}$$

Differentiating (2.4) with respect to z will yield the PM pdf as

$$f_{p(Z)}(p(z)) = \frac{\varphi(z)}{p'(z)} \tag{2.5}$$

where p'(z) > 0. Equations (2.4) and (2.5) are the general forms of the cdf and pdf for both power methods discussed herein – conventional moment and percentiles.

Presented in the next subsection is a review of the conventional moment based family of PM distributions.

2.1 The Conventional Moment Based Fleishman Third-order Power Method

The coefficients c_i for (2.1) that determine the shape of a third-order Fleishman (1978) PM are computed using a moment-matching that involves the conventional measures of the mean (α_1) , variance (α_2) , skew (α_3) , and kurtosis (α_4) . Specifically, the c_i are determined by simultaneously solving the following system of equations (e.g. Headrick, 2010, Eqs. 2.18–2.21, p. 15)

$$\alpha_1 = 0 = c_1 + c_3 \tag{2.6}$$

$$\alpha_2 = 1 = c_2^2 + 2c_3^2 + 6c_2c_4 + 15c_4^2 \tag{2.7}$$

$$\alpha_3 = 8c_3^3 + 6c_2^2c_3 + 72c_2c_3c_4 + 270c_3c_4^2 \tag{2.8}$$

$$\alpha_4 = 3c_2^4 + 60c_2^2c_3^2 + 60c_3^4 + 60c_2^3c_4 + 936c_2c_3^2c_4 + 630c_2^2c_4^2 + 4500c_3^2c_4^2$$
 (2.9)
$$+ 3780c_2c_4^3 + 10395c_4^4 - 3.$$

for specified values of α_3 and α_4 and where α_1 and α_2 are standardized to zero and one, respectively.

In general, a standardized non-normal third-order conventional moment based PM distribution will have a valid pdf iff $0 < c_2 < 1$, $0 < c_4 < 0.258199$, and $c_3^2 - 3c_2c_4 < 0$ (Headrick, 2010, pp. 16-21). Figure 1 gives examples of valid PM pdfs with their corresponding conventional parameters of skew and kurtosis and coefficients.

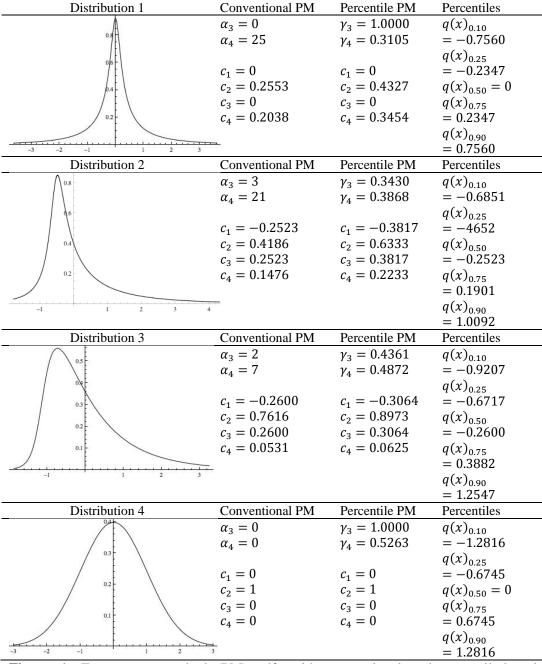


Figure 1: Four power method (PM) pdfs with conventional and percentile-based parameters of skew (α_3) , kurtosis (α_4) , left-right tail-weight ratio (γ_3) , and tail-weight factor (γ_4) , and their corresponding polynomial coefficients. Selected percentiles of each distribution are also shown.

3. The Percentile Based Power Method

The percentiles (θ_u) associated with a conventional moment based PM pdf can be obtained by making use of the PM cdf in (2.4). As such, we can define the following location, scale, and shape parameters as in Karian and Dudewicz (2011, pp. 172-173)

$$\gamma_1 = \theta_{0.50} \tag{3.1}$$

$$\gamma_2 = \theta_{0.90} - \theta_{0.10} \tag{3.2}$$

$$\gamma_3 = \frac{\theta_{0.50} - \theta_{0.10}}{\theta_{0.90} - \theta_{0.50}} \tag{3.3}$$

$$\gamma_4 = \frac{\theta_{0.75} - \theta_{0.25}}{\gamma_2} \tag{3.4}$$

where (3.1)–(3.4) are the (i) median, (ii) inter-decile range, (iii) left-right tail-weight ratio and (iv) tail-weight factor, respectively. The parameters in (3.1)–(3.4) are defined to have the restrictions

$$-\infty < \gamma_1 < +\infty, \ \gamma_2 \ge 0, \ \gamma_3 \ge 0, \ 0 \le \gamma_4 \le 1$$
 (3.5)

where a symmetric distribution implies that $\gamma_3 = 1$.

The derivation of a percentile based system of PM pdfs begins by substituting the standard normal distribution percentiles (z_u) into polynomials of the form in (2.1) and using (3.1)–(3.4) gives

(3.6)

$$\gamma_1 = p(z_{0.50})$$

$$\gamma_2 = p(z_{0.90}) - p(z_{0.10}) \tag{3.7}$$

$$\gamma_3 = \frac{p(z_{0.50}) - p(z_{0.10})}{p(z_{0.90}) - p(z_{0.50})} \tag{3.8}$$

$$\gamma_4 = \frac{p(z_{0.75}) - p(z_{0.25})}{\gamma_2} \tag{3.9}$$

where $z_{0.50}=0$, $z_{0.90}=1.281\cdots$, $z_{0.75}=0.6744\cdots$ from the standard normal distribution. Note from symmetry that $z_{0.10}=-z_{0.90}$ and $z_{0.25}=-z_{0.75}$. The explicit forms of (3.6)–(3.9) are

$$\gamma_1 = c_1 \tag{3.10}$$

$$\gamma_2 = 2c_2 z_{0.90} + 2c_4 z_{0.90}^3 \tag{3.11}$$

$$\gamma_3 = 1 - \frac{2c_3 z_{0.90}}{c_2 + c_3 z_{0.90} + 2c_4 z_{0.90}^2} \tag{3.12}$$

$$\gamma_4 = \frac{2c_2 z_{0.75} + 2c_4 z_{0.75}^3}{2c_2 z_{0.90} + 2c_4 z_{0.90}^3} \tag{3.13}$$

Simultaneously solving for the coefficients in (3.10)–(3.13) gives the convenient closed-form expressions

$$c_1 = \gamma_1 \tag{3.14}$$

$$c_2 = \frac{\gamma_2 \left(\gamma_4 z_{0.90}^3 - z_{0.75}^3\right)}{2z_{0.90}^3 z_{0.75} - 2z_{0.90} z_{0.75}^3} \tag{3.15}$$

$$c_3 = \frac{\gamma_2 (1 - \gamma_3)}{2(1 + \gamma_3) z_{0.90}^2} \tag{3.16}$$

$$c_4 = -\frac{\gamma_2(\gamma_4 z_{0.90} - z_{0.75})}{2z_{0.90}^3 z_{0.75} - 2z_{0.90} z_{0.75}^3}$$
(3.17)

Estimates of the parameters γ_1 , γ_2 , γ_3 , γ_4 for a PM distribution based on the percentiles $p(z_u)$ in (3.6)–(3.9) for a sample of size n can be determined by finding the j and j+1 integer values, and their corresponding expected values of the order statistics $E[p(Z)_{j:n}]$ and $E[p(Z)_{j+1:n}]$, by making use of the following equation (Headrick & Pant, 2012; Johnson, Kotz, & Balakrishnan, 1994)

$$E[p(Z)_{j:n}] = \frac{n!}{(i-1)!(n-j)!} \int_{-\infty}^{+\infty} p(z) \, \varphi(z) \{\Phi(z)\}^{j-1} \{1 - \Phi(z)\}^{n-j} dz$$
 (3.18)

such that

$$E[p(Z)_{j:n}] \le p(z_u) \le E[p(Z)_{j+1:n}] \tag{3.19}$$

and subsequently solve the equation

$$p(z_u) = (v)E[p(Z)_{j:n}] + (1 - v)E[p(Z)_{j+1:n}]$$
(3.20)

for $0 \le v \le 1$. Thus, an estimate of $p(z_u)$ can then be obtained based on the order statistics of a sample as $p(z_u) \cong p(Z_u) = (v)p(Z)_{j:n} + (1-v)p(Z)_{j+1:n}$.

4. Boundary Conditions for Percentile Based Power Method Pdfs

The restriction that p'(z) > 0 in (2.5) implies that a set of solved coefficients may not necessarily produce a valid pdf. To determine if a third-order polynomial produces a valid pdf we first set the quadratic equation p'(z) = 0 and subsequently solve for z as

$$z = \frac{-c_3 \pm (c_3^2 - 3c_2c_4)^{\frac{1}{2}}}{3c_4}. (4.1)$$

A set of solved coefficients will produce a valid pdf if the discriminant $c_3^2 - 3c_2c_4$ in (4.1) is negative. That is, the complex solutions for z must have non-zero imaginary

parts. As such, setting $c_3^2 = 3c_2c_4$ will yield the point where the discriminant vanishes and thus real-valued solutions exist to the equation p'(z) = 0.

Standardizing the inter-decile range in (3.2) to the unit normal distribution ($\gamma_2 = 2z_{0.90}$) and solving for c_4 gives

$$c_4 = \frac{1 - c_2}{z_{0.90}^2}. (4.2)$$

Substituting the right-hand side of (4.2) into (3.3) and (3.4) and setting $c_3 = \pm (3c_2c_4)^{\frac{1}{2}}$, yields

$$\gamma_3 = \frac{1 - 3(c_2 - 1)c_2 \mp 2\sqrt{3}z[-(c_2 - 1)c_2/z_{0.90}^2]^{\frac{1}{2}}}{1 + 3c_2(c_2 - 1)}$$
(4.3)

$$\gamma_4 = \frac{z_{0.75}^3}{z_{0.90}^3} + \frac{c_2 z_{0.75}}{z_{0.90}} - \frac{c_2 z_{0.75}^3}{z_{0.90}^3}.$$
(4.4.)

Inspection of (4.3) indicates that for real values of γ_3 to exist then we must have $c_2 \in [0,1]$ and thus from (4.2) $c_4 \in [0,1/z_{0.90}^2]$. Using (4.3) and (4.4) the graphs of the region for valid third-order power method pdfs is given in Figure 2 along with the minimum and maximum values of γ_3 and γ_4 . In summary, a valid standardized non-normal third-order pdf will have the properties of (i) $0 < c_2 < 1$, (ii) $0 < c_4 < 0.608875$, and (iii) $c_3^2 - 3c_2c_4 < 0$.

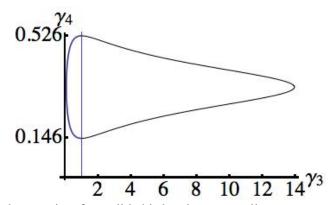


Figure 2: Boundary region for valid third-order percentiles power method pdfs in the left-right tail-weight ratio (γ_3) and tail-weight factor (γ_4) plane. Valid pdfs exist in the region inside the boundary, and a reference line at $\gamma_3 = 1$ indicates symmetric distributions. The lower and upper bounds of γ_3 are 0.072 and 13.928, respectively. The lower and upper bounds of γ_4 for symmetric distributions are 0.146 and 0.526, respectively.

5. The Simulation and Monte Carlo Study

To evaluate the proposed percentiles power method, comparisons among the proposed percentiles and conventional product-moment based procedures are subsequently described. Specifically, the distributions in Figure 1 are used as a basis for a comparison.

In terms of the simulation, a Fortran algorithm was written for each method to generate 25,000 independent sample estimates for the specified parameters of: (i) conventional skew (α_3) and kurtosis (α_4) and (ii) left-right tail-weight ratio (γ_3) and tail-weight factor (γ_4) . All estimates were based on sample sizes of n=25 and n=750. The formulae used for computing estimates of $\alpha_{3,4}$ were based on Fisher's k-statistics i.e. the formulae currently used by most commercial software packages such as SAS, SPSS, Minitab, and so on, for computing indices of skew and kurtosis (where $\alpha_{3,4}=0$ for the standard normal distribution). The formulae used for computing estimates of $\gamma_{3,4}$ were (3.3)-(3.4).

Bias-corrected accelerated bootstrapped median estimates, confidence intervals (C.I.s), and standard errors were subsequently obtained for the estimates associated with the parameters ($\alpha_{3,4}$, $\gamma_{3,4}$) using 10,000 resamples via the commercial software package Spotfire S+ (TIBCO Software, 2008). If a parameter (P) was outside its associated bootstrap C.I., then an index of relative bias (RB) was computed for the estimate (E) as: RB = ((E - P)/P) × 100. Where a bootstrap C.I. contained the value of the parameter, the small amount of associated bias was considered negligible and thus was not reported. The results of the simulation are reported in Tables 1-4.

Table 1: Skew (α_3) and Kurtosis (α_4) results for the Conventional PM. Sample size of n=25.

Dist	Parameter	Estimate	95% Bootstrap C.I.	Standard Error	Relative Bias %
1	$\alpha_3 = 0$	-0.0223	-0.0497,0.0045	0.013660	
	$\alpha_4 = 25$	4.4560	4.4011,4.5261	0.030200	-82.18
2	$\alpha_3 = 3$	1.5750	1.5579,1.5911	0.008122	-47.50
	$\alpha_4 = 21$	3.6960	3.6452,3.7525	0.027010	-82.40
3	$\alpha_3 = 2$	1.2780	1.2677,1.2893	0.005561	-36.10
	$\alpha_{4} = 7$	1.5230	1.4849,1.5662	0.020430	-78.24
4	$\alpha_3 = 0$	0.0034	-0.0038,0.0103	0.003626	
	$\alpha_4 = 0$	-0.1786	-0.1906,-0.1678	0.005579	

Table 2: Skew (α_3) and Kurtosis (α_4) results for the Conventional PM. Sample size of n = 750.

Dist	Parameter	Estimate	95% Bootstrap C.I.	Standard Error	Relative Bias %
1	$\alpha_3 = 0$	2.562	2.5383, 2.5823	0.01117	-26.5
	$\alpha_4 = 25$	22.15	21.6873, 22.6698	0.24850	-81.5
2	$\alpha_3 = 3$	2.180	2.1668, 2.1944	0.00697	-12.3
	$\alpha_4 = 21$	13.36	13.0936, 13.6467	0.14100	-49.7
3	$\alpha_3 = 2$	-0.0051	-0.0265, 0.0163	0.01100	
	$\alpha_{4} = 7$	18.57	18.2203, 18.9412	0.18330	-53.4
4	$\alpha_3 = 0$	1.54	1.5246, 1.5539	0.00743	-15.8
	$\alpha_4 = 0$	12.91	12.6537, 13.1903	0.13610	-45.0

Table 3: Left-right tail-weight ratio (γ_3) and tail-weight factor (γ_4) results for Percentiles PM. Sample size of n = 25.

Dist	Parameter	Estimate	95% Bootstrap C.I.	Standard Error	Relative Bias %
1	$\gamma_3 = 1.0000$	1.0050	0.9942, 1.0154	0.005348	
	$\gamma_4 = 0.3105$	0.3208	0.3191, 0.3227	0.000947	
2	$\gamma_3 = 0.3430$	0.3466	0.3438, 0.3497	0.001485	1.04
	$\gamma_4 = 0.3868$	0.3972	0.3954, 0.3993	0.000983	2.70
3	$\gamma_3 = 0.4361$	0.4472	0.4444, 0.4501	0.001464	2.53
	$\gamma_4 = 0.4872$	0.4960	0.4943, 0.4980	0.001003	1.80
4	$\gamma_3 = 1.0000$	0.9978	0.9912, 1.0045	0.003380	
	$\gamma_4 = 0.5263$	0.5294	0.5279, 0.5310	0.000801	

Table 4: Left-right tail-weight ratio (γ_3) and tail-weight factor (γ_4) results for Percentiles PM. Sample size of n = 750.

Dist	Parameter	Estimate	95% Bootstrap C.I.	Standard Error	Relative Bias %
1	$\gamma_3 = 1.0000$	1.0000	0.9978, 1.0020	0.001062	
	$\gamma_4 = 0.3105$	0.3108	0.3105, 0.3112	0.000171	0.11
2	$\gamma_3 = 0.3430$	0.3432	0.3426, 0.3438	0.000308	
	$\gamma_4 = 0.3868$	0.3873	0.3869, 0.3877	0.000203	0.14
3	$\gamma_3 = 0.4361$	0.4359	0.4353, 0.4364	0.000287	
	$\gamma_4 = 0.4872$	0.4874	0.4870, 0.4877	0.000189	
4	$\gamma_3 = 1.0000$	1.0000	0.9991, 1.0014	0.000539	
	$\gamma_4 = 0.5263$	0.5264	0.5261, 0.5267	0.000159	

6. Conclusion

One of the primary advantages that the percentiles power method has over conventional moment based estimators is that they can be applied in situations in which raw data and higher order (third, fourth) moments are unavailable. Inspection of the simulation results in Tables 1-4 verifies that the percentiles power method results are superior to conventional moment based estimators. In every instance the percent relative bias was substantially smaller or negligible for the percentiles power method estimator as compared to the conventional moment based estimator. In summary, the proposed percentiles power method is an attractive alternative to the traditional conventional-moment based power method. In particular, the percentiles power method has distinct advantages when heavy-tailed distributions are of concern.

References

Culpepper, S. A. (2013). The reliability and precision of total scores and IRT estimates as a function of polytomous IRT parameters and latent trait distribution. *Applied Psychological Measurement*, 37, 201-225.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, 43, 521-532.

Headrick, T. C. (2002). Fast fifth-order polynomial transforms for generating univariate and multivariate nonnormal distributions. *Computational Statistics & Data Analysis*, 40(4), 685-711.

Headrick, T. C. (2010). Statistical simulation: power method polynomials and other transformations. Chapman & Hall.

- Headrick, T. C. (2011). A characterization of power method transformations through L-moments. *Journal of Probability and Statistics*, 2011.
- Headrick, T. C., & Pant, M. D. (2012). A method for simulating nonnormal distributions with specified L-Skew, L-Kurtosis, and L-Correlation. *ISRN Applied Mathematics*, 2012.
- Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous Univariate Distributions* (Vol. 1). New York: Wiley.
- Karian, Z. A., & Dudewicz, E. J. (1999). Fitting the generalized lambda distribution to data: A method based on percentiles. *Communications in Statistics-Simulation and Computation*, 28, 793-819.
- Karian, Z. A., & Dudewicz, E. J. (2011). *Handbook of fitting statistical distributions with R*. Boca Raton FL: CRC Press.
- TIBCO Software. (2008). TIBCO Spotfire S+ 8.1 for Windows. Palo Alto, CA: Author.
- Tong, X., & Bentler, P. M. (2013). Evaluation of a new mean scaled and moment adjusted test statistic for SEM. *Structural Equation Modeling*, 20, 148-156.