Statistics and Probability Letters 79 (2009) 1621–1629

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Variance stabilizing transformations of Poisson, binomial and negative binomial distributions

Yu Guan *

Department of Statistics, Zhejiang Forestry University, Lin'an Zhejiang, 311300, China

ARTICLE INFO

Article history:
Received 4 February 2008
Received in revised form 5 April 2009
Accepted 7 April 2009
Available online 3 May 2009

ABSTRACT

Consider variance stabilizing transformations of Poisson distribution $\pi(\lambda)$, binomial distribution $B(n, p)$ and negative binomial distribution $NB(r, p)$, with square root transformations for $\pi(\lambda)$, arcsin transformations for $B(n, p)$ and inverse hyperbolic sine transformations for $NB(r, p)$. We will introduce three terms: critical point, domain of dependence and relative error. By comparing the relative errors of the transformed variances of $\pi(\lambda)$, $B(n, p)$ and $NB(r, p)$, and comparing the skewness and kurtosis of $\pi(\lambda)$, $B(n, p)$ and $NB(r, p)$ and their transformed variables, we obtain some better transformations with domains of dependence of the parameters. A new kind of transformation $(n + \frac{1}{2})^{1/2} \sin^{-1}(\frac{2Y-n}{n+2a})$ for $B(n, p)$ is suggested.

1. Introduction

Let Y be a random variable with mean $E(Y) = \mu$ and variance $\mathrm{Var}\,(Y)$. If the relationship between μ and $\mathrm{Var}\,(Y)$ was known, we could use this information to find a variance stabilizing transformation $Z = T(Y)$ such that $\mathrm{Var}\,(Z) \approx C^2$(constant). $T(Y)$ is expanded at the point $Y = \mu$ into a Taylor series: $T(Y) = T(\mu) + T'(\mu)(Y - \mu) + o(Y - \mu)$. By solving the differential equation $T'(\mu) \approx C\mathrm{Var}\,(Y)^{-1/2}$, we obtain $T(Y)$. In Proposition 1, we list some well-known transformations (see Chatterjee et al. (2000) and Montgomery (2005)).

Proposition 1. (1) *If* $Y \sim \pi(\lambda)$ *(Poisson distribution),* $T(Y) = \sqrt{Y}$ *with* $\mathrm{Var}\,(\sqrt{Y}) \approx \frac{1}{4}$.
(2) *If* $Y \sim B(n, p)$ *(binomial distribution),* $T(Y) = \sqrt{n}\sin^{-1}\sqrt{Y/n}$ *with* $\mathrm{Var}\,(\sqrt{n}\sin^{-1}(\sqrt{Y/n})) \approx \frac{1}{4}$.
(3) *If* $Y \sim NB(r, p)$ *(negative binomial distribution), its frequency function is* $P(Y = k) = \binom{k+r-1}{r-1}p^r(1-p)^k (k = 0, 1, 2, \ldots, r > 1)$, *and then* $T(Y) = \sqrt{r}\sinh^{-1}\sqrt{Y/r}$ *with* $\mathrm{Var}\,(\sqrt{r}\sinh^{-1}\sqrt{Y/r}) \approx \frac{1}{4}$.

Bartlett (1936, 1947) first introduced variance stabilizing transformations, and proposed the transformation $\sqrt{Y + 0.5}$ for $Y \sim \pi(\lambda)$. Anscombe (1948) showed that $\sqrt{Y + 3/8}$ is the most nearly constant variance transformation for $Y \sim \pi(\lambda)$ when Y has a larger mean λ, and $\sin^{-1}\sqrt{\frac{Y+3/8}{n+3/4}}$ for $Y \sim B(n, p)$ similarly. Freeman and Tukey (1950) suggested combined transformations $(\sqrt{Y} + \sqrt{Y+1})$ for $Y \sim \pi(\lambda)$ and $(n + \frac{1}{2})^{1/2}[\sin^{-1}\sqrt{\frac{Y}{n+1}} + \sin^{-1}\sqrt{\frac{Y+1}{n+1}}]$ for $Y \sim B(n, p)$. The method of combined transformation was generalized to $NB(r, p)$ by Laubscher (1961). Thacker and Bromiley (2001) and Bromiley and Thacker (2002) investigated the effects of stabilizing transformations on a Poisson distributed quantity and a binomial distributed quantity. Uddin et al. (2006) presented a necessary condition for a variance stabilizing transformation to be an

* Tel.: +86 057163721336.
E-mail address: guanyu@zjfc.edu.cn.

0167-7152/$ – see front matter
doi:10.1016/j.spl.2009.04.010

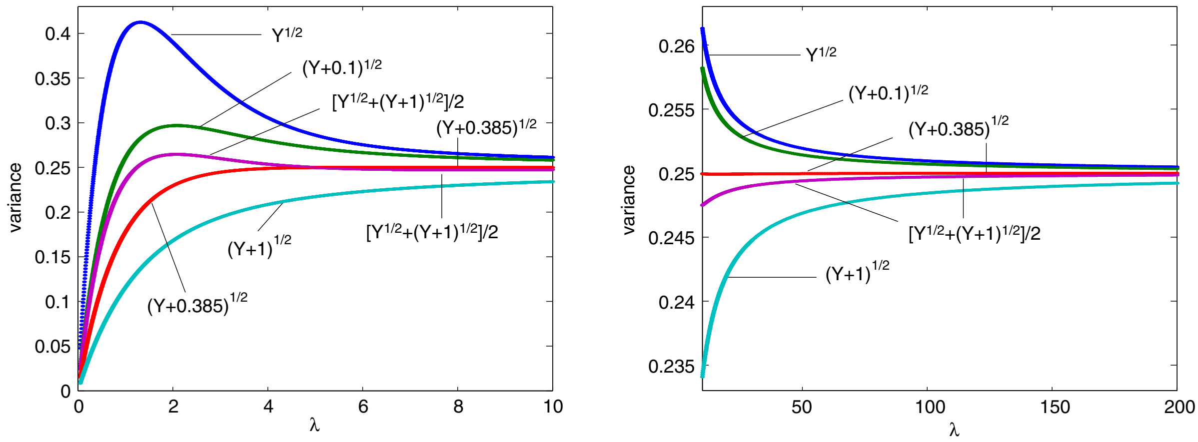

Fig. 1. Five transformed variance curves on $\pi(\lambda)$ for $\lambda \in [0.5, 10]$ (left panel) and $\lambda \in [10, 200]$ (right panel).

approximate symmetrizing transformation. For more general data, the Box–Cox transformation is often used (see Box and Cox (1964), Box and Cox (1982) and Yamamura (1999)).

For $Y \sim \pi(\lambda)$, Anscombe (1948) showed that $\mathrm{Var}\,(\sqrt{Y+a}) = \frac{1}{4}\{1 + \frac{3-8a}{8\lambda} + O(\lambda^{-2})\} \approx \frac{1}{4}$ for large mean λ and any constant $a \geq 0$. On the other hand, it is clear that $\lim_{\lambda\to 0} \mathrm{Var}\,(\sqrt{Y+a}) = 0$ for any constant $a \geq 0$. This implies that $\sqrt{Y+a}$ is only a local variance stabilizing transformation. There is the same case for binomial variables and negative binomial variables. Our task is to find those transformations such that variances of transformed variables change less in larger domains of dependence, i.e. domains of parameters.

Since exact formulas for transformed variances could not be obtained, numerical methods are applied to calculate these variances of various transformed random variables. By virtue of mass numerical computation, we compare fluctuations of these variances and obtain some better transformations with their domains of dependence. The selection criterion is less fluctuation of variances with respect to a larger domain of dependence of parameters. In this paper, we will compute parameters λ, p and a to the third place of decimals.

In Section 2, we study transformation $\sqrt{Y+a}$ and its combined transformations for $Y \sim \pi(\lambda)$, and introduce three concepts: left critical point, domain of dependence and relative error, to describe the stabilization of a transformation. In Section 3, we show that a new kind of transformation $(n+\frac{1}{2})^{1/2}\sin^{-1}(\frac{2Y-n}{n+2a})$ is equivalent to $(n+\frac{1}{2})^{1/2}\sin^{-1}\sqrt{\frac{Y+a}{n+2a}}$ for $Y \sim B(n, p)$. We also give the limit relation between $NB(r, p)$ and $\pi(\lambda)$, and investigate three transformations for $NB(r, p)$ in Section 4. In the last section, the skewness and kurtosis of these three kinds of variables and variables transformed by various transformations are computed and compared. Some better variance stabilizing transformations with domains of dependence of parameters on $\pi(\lambda)$, $B(n, p)$, $NB(r, p)$ are suggested.

2. Poisson distribution

Proposition 2. *Let $Y \sim \pi(\lambda)$ and $a_1 > a_2 \geq 0$; then* $\mathrm{Var}\,(\sqrt{Y+a_1}) < \mathrm{Var}\,(\sqrt{Y+a_2})$.

Proof. Let $\delta = a_1 - a_2 > 0$ and $p_k = P(Y=k) = \frac{\lambda^k}{k!}e^{-\lambda}$ $(k = 0, 1, 2, \ldots)$. We have $E[(\sqrt{Y+a_i})^2] = \lambda + a_i$ $(i = 1, 2)$, and

$$\begin{aligned}\mathrm{Var}\,(\sqrt{Y+a_1}) - \mathrm{Var}\,(\sqrt{Y+a_2}) &= \delta - \left(\left[\sum_{k=0}^{+\infty}\sqrt{k+a_1}\,p_k\right]^2 - \left[\sum_{k=0}^{+\infty}\sqrt{k+a_2}\,p_k\right]^2\right)\\ &< \delta - \sum_{k=0}^{+\infty}\delta\,p_k^2 - 2\sum_{m>n\geq 0}\delta\,p_m p_n = \delta - \delta\left(\sum_{k=0}^{+\infty}p_k\right)^2 = 0. \end{aligned} \tag{1}$$

Here, $\sqrt{(m+a_1)(n+a_1)} - \sqrt{(m+a_2)(n+a_2)} > \delta$, since $[\sqrt{(m+a_1)(n+a_1)}]^2 - [\sqrt{(m+a_2)(n+a_2)} + \delta]^2 = \delta\,(m + a_2 + n + a_2) - 2\delta\sqrt{(m+a_2)(n+a_2)} > 0$. $\square$

Remark. Proposition 2 tells us that $\mathrm{Var}(\sqrt{Y+a})$ is a monotone decreasing function of a. Note that there is only one constant a' at most such that $\mathrm{Var}\,(\sqrt{Y+a'}) = 0.25$ for each λ.

By numerical computation, we obtain the following (see Fig. 1 and Table 1).

Table 1
Variances under transformation by $\sqrt{Y+a}$.

a	λ_M	$V_M - 0.25$	λ_m	$V_m - 0.25$	λ_L	$V_{1000} - 0.25$	Δ
0	1.319	.1625	1000	$.0^{5}939$	0.352	$.0^{5}939$	.6496
0.100	2.086	.0468	1000	$.0^{4}688$	0.898	$.0^{4}688$	.1871
0.200	2.786	.0203	1000	$.0^{4}438$	1.450	$.0^{4}438$	.0811
0.300	3.835	$.0^{2}651$	1000	$.0^{4}188$	2.347	$.0^{4}188$	.0260
0.350	4.769	$.0^{2}223$	1000	$.0^{5}627$	3.237	$.0^{5}627$	$.0^{2}889$
0.375	5.624	$.0^{3}623$	1000	$.0^{7}162$	4.206	$.0^{7}162$	$.0^{2}251$
0.380	5.890	$.0^{3}354$	27.660	$-.0^{4}239$	4.558	$-.0^{5}123$	$0^{2}151$
0.384	6.158	$.0^{3}148$	16.751	$-.0^{4}745$	4.915	$-.0^{5}224$	$.0^{3}890$
0.385	6.237	$.0^{4}983$	15.417	$-.0^{4}911$	5.021	$-.0^{5}249$	$0^{3}757$
0.386	6.321	$.0^{4}495$	14.331	$-.0^{3}119$	5.135	$-.0^{5}274$	$.0^{3}634$
0.387	6.413	$.0^{5}170$	13.422	$-.0^{3}128$	5.261	$-.0^{5}289$	$.0^{3}520$
0.500	1000	$-.0^{4}313$	.0289	.0415	.0705	.1458	.3588

Table 2
Some better variance stabilizing transformations on $\pi(\lambda)$.

a	λ_L	Δ	Δ_5	Δ_4	Δ_3	Δ_2	Δ_1	$\Delta_{0.5}$
0.385	5.021	$.0^{3}189$	$.0^{3}792$	$.0^{2}731$	.0220	.0823	.2899	.5373
0.375	4.206	$.0^{3}623$	$.0^{2}251$	$.0^{2}376$	.0194	.0782	.2851	.5356
(0, 1)	1.189	.0690	.0103	.0251	.0501	.0690	.1183	.3885
(0, 1.25)	1.328	.0464	.0267	.0267	.0332	.0464	.1190	.3791
(0, 1.3)	1.354	.0429	.0302	.0302	.0307	.0429	.1190	.3772
(0, 0.75)	1.027	.1072	.0252	.0459	.0791	.1072	.1157	.3984
(0.01, 1)	1.687	.0278	$.0^{2}557$	.0139	.0254	.0278	.1552	.4174
(0.02, 1)	2.007	.0158	$.0^{2}357$	$.0^{2}928$	.0156	.0161	.1720	.4307
(0, 1, 0.8)	0.973	.1127	.0191	.0402	.0763	.1121	.1127	.3844

(1) When $a \in [0, 0.375]$, there exists a special point $\lambda_0(a)$ such that Var $(\sqrt{Y(\lambda)+a})$ is a monotone increasing function of $\lambda \in (0,\ \lambda_0(a)]$ and a monotone decreasing function of $\lambda \in (\lambda_0(a), +\infty)$.
(2) When $a \in [0.376,\ 0.392]$ and the positive parameter λ increases, Var $(\sqrt{Y(\lambda)+a})$ at first shows monotone increase, then monotone decrease, then monotone increase, and so on.
(3) When $a \geq 0.393$, Var $(\sqrt{Y(\lambda)+a})$ is a monotone decreasing function of $\lambda \in (0, +\infty)$.

According to the characteristic of Var $(\sqrt{Y(\lambda)+a})$ curves, for $a \leq 0.392$ we introduce a concept 'left critical point' λ_L which satisfies the following conditions:

(1) There exist both a maximum and a minimum of Var $(\sqrt{Y(\lambda)+a})$ for $\lambda \in (\lambda_L, 1000]$, but no maximum and no minimum of Var $(\sqrt{Y(\lambda)+a})$ for $\lambda \in (0, \lambda_L)$.
(2) Var $(\sqrt{Y(\lambda_L)+a}) \geq$ Var $(\sqrt{Y(\lambda_m)+a}) >$ Var $(\sqrt{Y(\lambda_L - 0.001)+a})$.
(3) Var $(\sqrt{Y(\lambda_m)+a}) = \min\{\text{Var}\,(\sqrt{Y(\lambda)+a}) | \lambda \in (\lambda_L, 1000]\}$, Var $(\sqrt{Y(\lambda_M)+a}) = \max\{\text{Var}\,(\sqrt{Y(\lambda)+a}) | \lambda \in (\lambda_L, 1000]\}$.

When $\lambda \in [\lambda_L,\ 1000]$, the fluctuation of Var $(\sqrt{Y(\lambda)+a})$ is less. When $\lambda \in (0, \lambda_L)$ decreases, Var $(\sqrt{Y(\lambda)+a})$ decreases violently. Interval $[\lambda_L, 1000]$ or $[\lambda_L, +\infty)$ is called the 'domain of dependence' of parameter λ.

In order to describe the fluctuation of variances precisely, we define the 'relative error' of the variance for parameter $\theta \in \Theta$ as follows:

$$\Delta = \frac{\max\limits_{\theta_1,\theta_2 \in \Theta} \{|\text{Var}\,(\theta_1) - \text{Var}\,(\theta_2)|\}}{\text{Var}^*}. \tag{2}$$

Here Var* is the limit value of Var (θ) $(\theta \in \Theta)$. For $\pi(\lambda)$, $B(n, p)$ and $NB(r, p)$, by Proposition 1 their Var* are all 1/4.

In Table 1, $V_M =$ Var $\sqrt{Y(\lambda_M)+a}$, $V_m =$ Var $\sqrt{Y(\lambda_m)+a}$, $V_\lambda =$ Var $\sqrt{Y(\lambda)+a}$. The last column denotes the relative error Δ for $\lambda \in [\lambda_L, 1000]$. The last row denotes the transformation $\sqrt{Y+0.5}$; its variance is a monotone increasing function of $\lambda > 0$, and this means that $\sqrt{Y+0.5}$ is not a better transformation. When $\lambda \geq 5$, the left critical point of $\sqrt{Y+0.385}$ is the most proximal to 5, and thus $\sqrt{Y+0.385}$ is almost the best variance stabilizing transformation of $\pi(\lambda)$ for $\lambda \geq 5$. And so is $\sqrt{Y+3/8}$ for $\lambda \geq 4$.

Now we discuss the combined transformation $\frac{b_1\sqrt{Y+a_1}+b_2\sqrt{Y+a_2}}{b_1+b_2}$ $(b_1+b_2 \neq 0)$. After much calculation, we are convinced that the combined transformation $\frac{b_1\sqrt{Y+a_1}+b_2\sqrt{Y+a_2}}{b_1+b_2}$ might be seen as a compromise between $\sqrt{Y+a_1}$ and $\sqrt{Y+a_2}$. For example, $\frac{\sqrt{Y}+\sqrt{Y+1}}{2}$ improves the variance stabilization of $\sqrt{Y}$ and $\sqrt{Y+1}$ extremely.

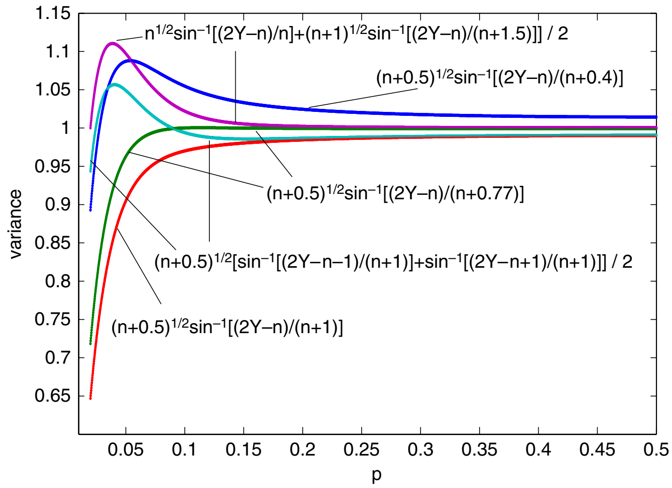

Fig. 2. Five transformed variance curves on $B(n,\ p)$ for $n = 50$ and $p \in [0.02,\ 0.5]$.

Table 2 lists some better transformations and combined transformations for a Poisson variable, where a, $(a_1,\ a_2)$ and $(a_1,\ a_2,\ b)$ respectively denote the transformations $\sqrt{Y+a}$, $\frac{\sqrt{Y+a_1}+\sqrt{Y+a_2}}{2}$ and $\frac{\sqrt{Y+a_1}+b\sqrt{Y+a_2}}{1+b}$, and Δ_k represents the relative error of the transformed variances $\lambda \in [k,\ 1000]$ $(k = 5, 4, 3, 2, 1, 0.5)$.

3. Binomial distribution

Proposition 3. *Let $Y \sim B(n, p)$; then $T(Y) = \sin^{-1}(\frac{2Y-n}{n})$ is a variance stabilizing transformation and* $\mathrm{Var}\,(\sin^{-1}(\frac{2Y-n}{n})) \approx 1/n$.

Proof. We have $\frac{dT}{dY} = \frac{2}{n}\{1-(\frac{2Y-n}{n})^2\}^{-1/2} = \frac{1}{\sqrt{n}}\{\frac{Y}{n}(1-\frac{Y}{n})\frac{1}{n}\}^{-1/2}$. Therefore $\mathrm{Var}\,(\sin^{-1}(\frac{2Y-n}{n})) \approx 1/n$. □

Proposition 4. *Let $Y \sim B(n, p)$, $T_1(Y, a) = \sin^{-1}\sqrt{\frac{Y+a}{n+2a}}$, $T_2(Y, a) = \sin^{-1}(\frac{2Y-n}{n+2a})$ and $a \in [0, 1]$. Then:*

(1) $T_1(n-Y, a) = \pi/2 - T_1(Y, a)$, $T_2(n-Y, a) = -T_2(Y, a)$;
(2) $T_2(Y, a) = 2T_1(Y, a) - \pi/2$;
(3) $\mathrm{Var}\,(T_i(n-Y, a)) = \mathrm{Var}\,(T_i(Y, a))$ $(i = 1, 2)$, $\mathrm{Var}\,(T_2(Y, a)) = 4\mathrm{Var}\,(T_1(Y, a))$.

Proof. (1) Considering $\sin^2\{\sin^{-1}(\frac{Y+a}{n+2a})^{1/2}\} + \sin^2\{\sin^{-1}(\frac{n-Y+a}{n+2a})^{1/2}\} = 1$, we have $T_1(n-Y, a) = \frac{\pi}{2} - T_1(Y, a)$. Similarly, $T_2(n-Y, a) = -T_2(Y, a)$.
(2) Considering $\cos(2T_1(Y, a)) = 1 - 2\sin^2(T_1(Y, a)) = -\sin(T_2(Y, a))$, we have $T_2(Y, a) = 2T_1(Y, a) - \pi/2$.
(3) By items (1) and (2), the equalities of item (3) are easily obtained. □

Remark. By Proposition 4, $\sin^{-1}(\frac{2Y-n}{n+2a})$ is a negative symmetrical transformation with axis of symmetry $Y = n/2$ and is equivalent to $\sin^{-1}\sqrt{\frac{Y+a}{n+2a}}$ for the variance transformation for $B(n, p)$. But the former formula is simpler than the latter, so in this paper we suggest $(n+\frac{1}{2})^{1/2}\sin^{-1}(\frac{2Y-n}{n+2a})$ with approximate variance 1 instead of $(n+\frac{1}{2})^{1/2}\sin^{-1}\sqrt{\frac{2Y-n}{n+2a}}$.

It is well-known that $\pi(\lambda)$ can be derived as the limit of $B(n, p)$ as n approaches infinity and p approaches zero in such a way that $np = \lambda$. Therefore, we study $B(n, p)$ by numerical computation like $\pi(\lambda)$. While n and p are fixed, the variance transformed by $(n+\frac{1}{2})^{1/2}\sin^{-1}(\frac{2Y-n}{n+2a})$ is a monotone decreasing function of addend a also. See Fig. 2 and Table 3.

Freeman and Tukey (1950) suggested the combined transformation $(n+\frac{1}{2})^{1/2}[\sin^{-1}\sqrt{\frac{Y}{n+1}}+\sin^{-1}\sqrt{\frac{Y+1}{n+1}}$ for $Y \sim B(n,\ p)$. Laubscher (1961) proposed the transformation $n^{1/2}\sin^{-1}\sqrt{\frac{Y}{n}}+(n+1)^{1/2}\sin^{-1}\sqrt{\frac{Y+3/4}{n+3/2}}]$. Analogously, we can convert these two transformations into $(n+\frac{1}{2})^{1/2}[\sin^{-1}(\frac{2Y-n-1}{n+1})+\sin^{-1}(\frac{2Y-n+1}{n+1})]$ and $[n^{1/2}\sin^{-1}(\frac{2Y-n}{n})+(n+1)^{1/2}\sin^{-1}(\frac{2Y-n}{n+3/2})]$ with approximate variances 4.

Table 3 lists four better transformations $(n+\frac{1}{2})^{1/2}\sin^{-1}(\frac{2Y-n}{n+0.77})$, $(n+\frac{1}{2})^{1/2}\sin^{-1}(\frac{2Y-n}{n+0.75})$, $(n+\frac{1}{2})^{1/2}[\sin^{-1}(\frac{2Y-n-1}{n+1})+\sin^{-1}(\frac{2Y-n+1}{n+1})]$ and $[n^{1/2}\sin^{-1}(\frac{2Y-n}{n})+(n+1)^{1/2}\sin^{-1}(\frac{2Y-n}{n+3/2})]$ with their numerical results, where parameter p is computed to the fourth place of decimals when $n = 1000$.

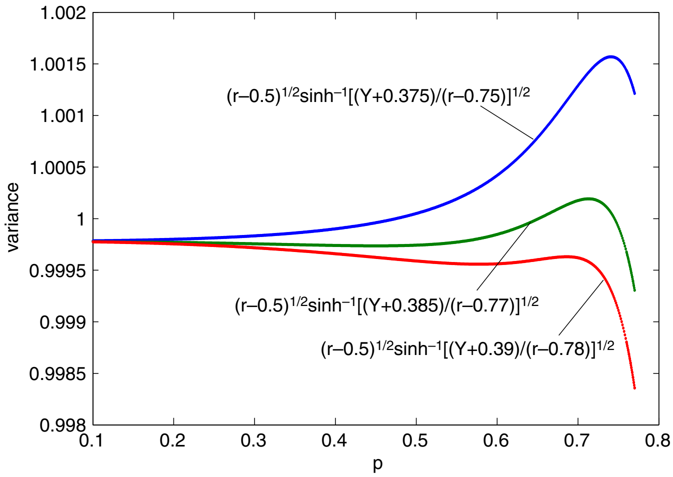

Fig. 3. Three transformed variance curves on $NB(r,\ p)$ for $r = 20$ and $p \in [0.1,\ 0.77]$.

Table 3
Some better variance stabilizing transformations on $B(n, p)$.

a	n	np_L	Δ	Δ_5	Δ_4	Δ_3	Δ_2	Δ_1	$\Delta_{0.5}$
0.385	20	4.08	$.0^{2}122$	$.0^{2}122$	$.0^{2}149$	.0129	.0546	.2700	.5250
0.385	50	4.60	$.0^{2}103$	$.0^{2}103$	$.0^{2}337$	.0181	.0754	.2819	.5334
0.385	100	4.80	$.0^{3}877$	$.0^{3}877$	$.0^{2}418$	.0200	.0789	.2859	.5362
0.385	1000	5.00	$.0^{3}768$	$.0^{3}768$	$.0^{2}498$	.0218	.0820	.2895	.5387
3/8	20	3.70	$.0^{2}231$	$.0^{2}220$	$.0^{2}231$	.0110	.0618	.2656	.5223
3/8	50	3.95	$.0^{2}270$	$.0^{2}270$	$.0^{2}270$	.0158	.0715	.2773	.5303
3/8	100	4.10	$.0^{2}265$	$.0^{2}265$	$.0^{2}299$	.0176	.0748	.2816	.5330
3/8	1000	4.20	$.0^{2}253$	$.0^{2}253$	$.0^{2}368$	.0193	.0779	.2847	.5354
(0, 1)	20	1.12	.0754	$.0^{2}607$	.0192	.0465	.0752	.1050	.3759
(0, 1)	50	1.20	.0709	$.0^{2}822$	.0224	.0484	.0709	.1130	.3835
(0, 1)	100	1.20	.0698	$.0^{2}922$	.0237	.0492	.0698	.1156	.3860
(0, 1)	1000	1.20	.0690	.0102	.0249	.0500	.0690	.1180	.3883
(0, 3/4)	20	0.98	.1122	.0146	.0337	.0702	.1108	.1122	.3883
(0, 3/4)	50	1.05	.1100	.0216	.0419	.0767	.1097	.1103	.3943
(0, 3/4)	100	1.10	.1086	.0235	.0440	.0781	.1086	.1129	.3963
(0, 3/4)	1000	1.10	.1074	.0251	.0457	.0790	.1074	.1153	.3982

4. Negative binomial distribution

Proposition 5. *Let $Y \sim NB(r,\ p)$ and $\lim_{r\to+\infty} r(1-p) = \lambda$ (positive constant); then*

$$\lim_{r\to+\infty} \binom{k+r-1}{r-1} p^r (1-p)^k = \frac{\lambda^k}{k!} e^{-\lambda} \quad \textit{for all} \quad k \geq 0. \tag{3}$$

Proof. When $n \to +\infty$, $n! \approx \sqrt{2\pi}\ n^{n+1/2}e^{-n}$. Then $\binom{k+r-1}{r-1}p^r(1-p)^k \approx \frac{(k+r-1)^{k+r-1/2}e^{-k}}{(r-1)^{r-1/2}} \frac{1}{r^k} \frac{p^r(r(1-p))^k}{k!} = (1+\frac{k}{r-1})^{r-1/2}\, e^{-k}\, (1+\frac{k-1}{r})^k \frac{(r(1-p))^k}{k!} (1-\frac{r(1-p)}{r})^r \approx \frac{(r(1-p))^k}{k!}\, e^{-r(1-p)}$ for large r. □

Remark. According to Proposition 5, $NB(r,\ p) \approx \pi(r(1-p))$ for large r. So there are transformations and geometrical characteristics similar to those for transformed variance curves on $NB(r,\ p)$, like $\pi(\lambda)$ and $B(n,\ p)$, if we just regard $q = (1-p)$ in $NB(r, p)$ as p in $B(n, p)$.

Fig. 3 shows that 0.385 is the best numerical value of a for transformation $(r-\frac{1}{2})^{1/2}\sinh^{-1}\sqrt{\frac{Y+a}{r-2a}}$ on $NB(r, p)$ for $rq \geq 5$, approximately. There are some cases for $\pi(\lambda)$ and $B(n,\ p)$ also.

Anscombe (1948) showed that on $(r-\frac{1}{2})^{1/2}\sinh^{-1}\sqrt{\frac{Y+a}{r-2a}}$ the optimum value of a is $3/8$ when rq is larger and its variance is equal to $\frac{1}{4} + O((rq)^{-2})$. Laubscher (1961) proposed the transformation $r^{1/2}\sinh^{-1}\sqrt{\frac{Y}{r}} + (r-1)^{1/2}\sinh^{-1}\sqrt{\frac{Y+3/4}{n-3/2}}$.

Table 4
Some better variance stabilizing transformations on $NB(r, p)$.

a	r	rq_L	Δ	Δ_5	Δ_4	Δ_3	Δ_2	Δ_1	$\Delta_{0.5}$
0.385	5	3.70	$.0^3229$	–	$.0^3229$	$.0^2567$	.0596	.2914	.5569
0.385	10	4.45	$.0^3408$	$.0^3397$	$.0^2166$	.0143	.0732	.2923	.5491
0.385	20	4.86	$.0^3456$	$.0^3456$	$.0^2330$	.0184	.0783	.2915	.5443
0.385	100*	5.10	$.0^3661$	$.0^3675$	$.0^2472$	.0213	.0816	.2903	.5400
3/8	5	3.475	$.0^3523$	–	$.0^3477$	$.0^2461$	.0563	.2864	.5529
3/8	10	3.92	$.0^2127$	$.0^2121$	$.0^2127$	.0122	.0693	.2871	.5451
3/8	20	4.10	$.0^2149$	$.0^2149$	$.0^2223$	.0160	.0742	.2863	.5405
3/8	100*	4.20	$.0^2233$	$.0^2233$	$.0^2343$	.0188	.0775	.2853	.5366
(0, 3/4)	5	1.075	.0782	–	$.0^2303$	.0267	.0744	.1012	.3913
(0, 3/4)	10	1.06	.0920	$.0^2722$	.0214	.0536	.0912	.1093	.3958
(0, 3/4)	20	1.06	.0982	.0141	.0321	.0653	.0980	.1126	.3973
(0, 3/4)	100*	1.10	.1056	.0231	.0433	.0766	.1056	.1151	.3981

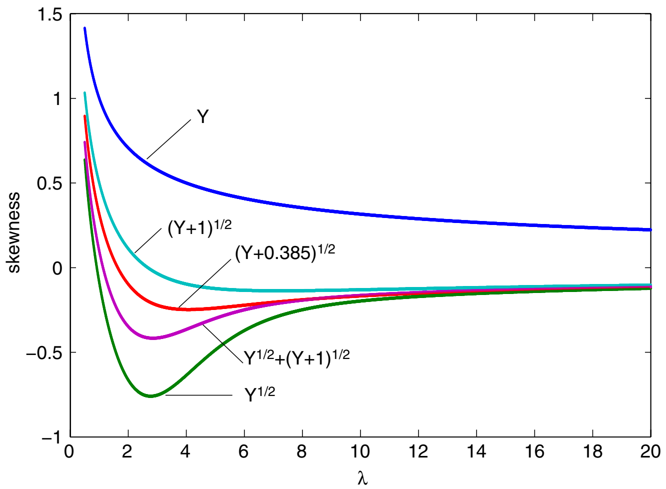

Fig. 4. Five curves showing the skewness of $Y \sim \pi(\lambda)$ and the skewness transformed by four transformations for $\lambda \in [0.5,\ 20]$.

Table 4 shows three better transformations: $(r-\frac{1}{2})^{1/2}\sinh^{-1}\sqrt{\frac{Y+0.385}{r-0.77}}$, $(r-\frac{1}{2})^{1/2}\sinh^{-1}\sqrt{\frac{Y+3/8}{r-3/4}}$ and $r^{1/2}\sinh^{-1}\sqrt{\frac{Y}{r}}+(r-1)^{1/2}\sinh^{-1}\sqrt{\frac{Y+3/4}{r-3/2}}$. The third column expresses the left critical point of $rq = r(1-p)$, with corresponding relative errors in the fourth column. The last six columns denote relative errors of transformed variances for $p \in [0.001,\ 1-k/r]$ ($k = 5, 4, 3, 2, 1, 0.5$) respectively, except for the case $r = 100$ (marked by $*$) with $p \in [0.01,\ 1-k/r]$ ($k = 5, 4, 3, 2, 1, 0.5$) because of a calculated error problem.

5. Skewness, kurtosis and conclusions

Skewness is used as a measure of asymmetry of a random variable about its mean. Kurtosis can be used to detect that a symmetric distribution departs from normality by being heavy-tailed or light-tailed or too peaked or too flat at the center. Utilizing skewness and kurtosis, we study the normality of these transformations.

Let Y denote the random variable $\pi(\lambda)$, or $B(n,\ p)$, or $NB(r,\ p)$. Let $T(Y(a))$ ($a \in [0,\ 1]$) denote the variance stabilizing transformation $\sqrt{Y+a}$ for $\pi(\lambda)$, or $(n+\frac{1}{2})^{1/2}\sin^{-1}(\frac{2Y-n}{n+2a})$ for $B(n,\ p)$, or $(r-\frac{1}{2})^{1/2}\sinh^{-1}\sqrt{\frac{Y+a}{r-2a}}$ for $NB(r,\ p)$. Let $T(Y(0,\ 1))$ denote a combined transformation such as $\sqrt{Y}+\sqrt{Y+1}$ for $\pi(\lambda)$, or $(n+\frac{1}{2})^{1/2}[\sin^{-1}(\frac{2Y-n-1}{n+2})+\sin^{-1}(\frac{2Y-n+1}{n+2})]$ for $B(n,\ p)$, or $r^{1/2}\sinh^{-1}\sqrt{\frac{Y}{r}}+(r-1)^{1/2}\sinh^{-1}\sqrt{\frac{Y+3/4}{r-3/2}}$ for $NB(r,\ p)$.

By comparing the skewness and kurtosis for transformed and not transformed cases, we obtain some conclusions as follows (see Figs. 4–9).

(1) $T(Y(a))$ obviously improves the skewness of primary data. Approximately, when $\lambda \geq 3$ for $\pi(\lambda)$, or $np \in [3,\ n-3]$ for $B(n,\ p)$, or $rq \geq 3$ for $NB(r,\ p)$, $T(Y(a))$ exchanges the skew direction and diminishes its size.
(2) $T(Y(a))$ improves the kurtosis of NB$(r,\ p)$, especially while $a \geq 0.3$. But it has no effect on $\pi(\lambda)$ and $B(n,\ p)$.

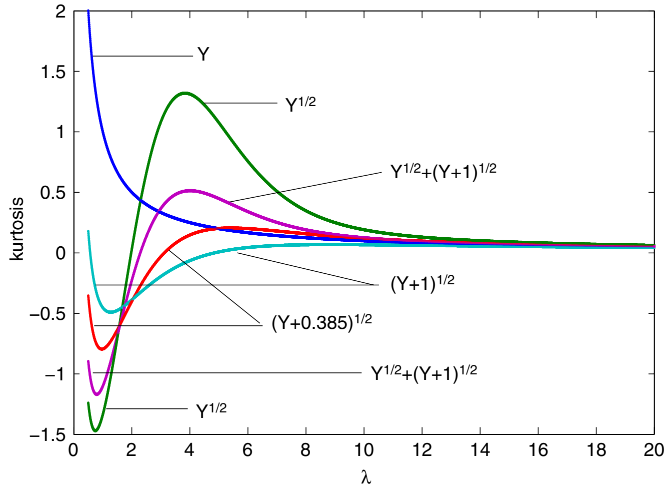

Fig. 5. Five curves showing the kurtosis of $Y \sim \pi(\lambda)$ and the kurtosis transformed by four transformations for $\lambda \in [0.5, 20]$.

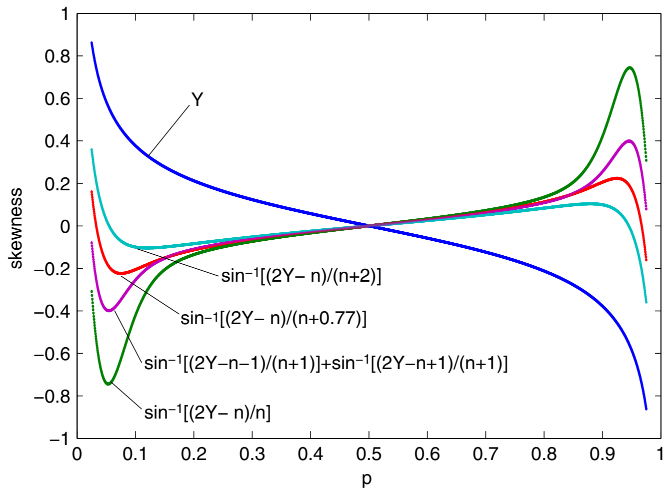

Fig. 6. Five curves showing the skewness of $Y \sim B(n, p)$ and the skewness transformed by four transformations for $n = 50$ and $p \in [0.025, 0.975]$.

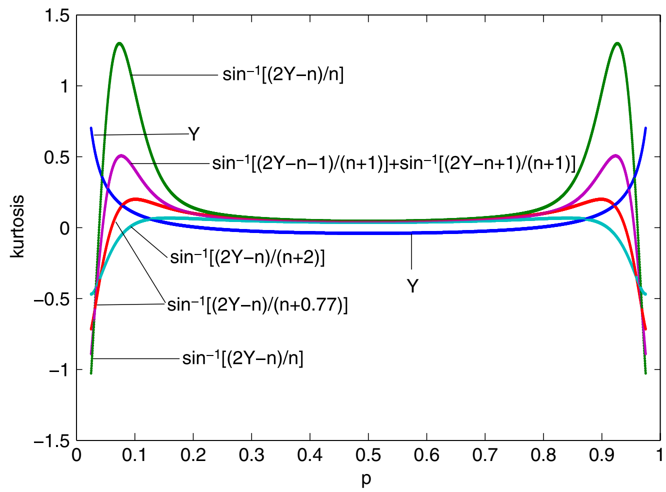

Fig. 7. Five curves showing the kurtosis of $Y \sim B(n, p)$ and the kurtosis transformed by four transformations for $n = 50$ and $p \in [0.025, 0.975]$.

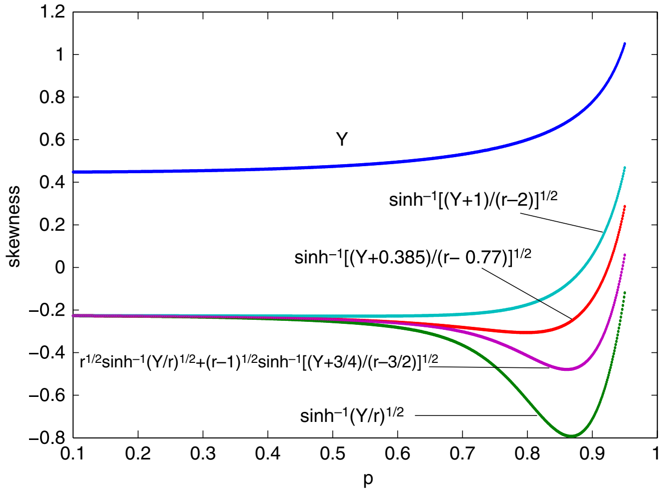

Fig. 8. Five curves showing the skewness of $Y \sim NB(r,\ p)$ and the skewness transformed by four transformations for $r = 20$ and $p \in [0.1,\ 0.95]$.

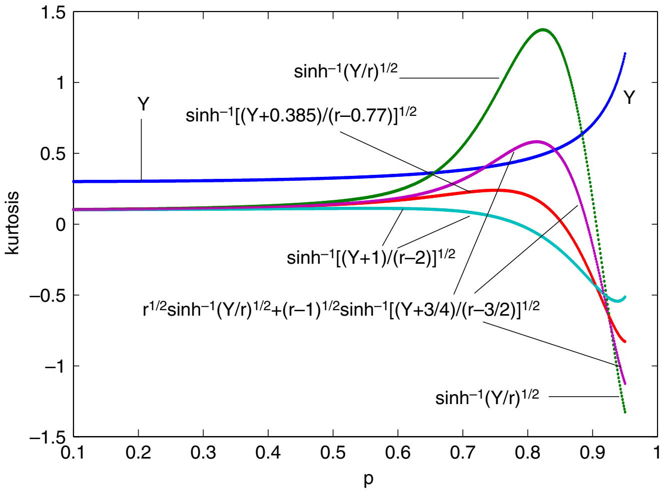

Fig. 9. Five curves showing the kurtosis of $Y \sim NB(n,\ p)$ and the kurtosis transformed by four transformations for $r = 20$ and $p \in [0.1,\ 0.95]$.

(3) Combined transformation $T(Y(0,\ 1))$ behaves near $T(Y(0.15))$ and is not better than $T(Y(0.385))$ or $T(Y(3/8))$ for normalizing random variables.

(4) When $\lambda \geq 10$ or $np \in [10,\ n-10]$ or $rq \geq 10$, the skewness and kurtosis transformed by $T(Y(a))$ are almost independent of a. When $\lambda < 10$ or $np \in (0,\ 10) \bigcup (n-10,\ n]$ or $rq < 10$, the larger a behaves better than the smaller.

In general, $T(Y(0.385))$ and $T(Y(3/8))$ are preferred variance stabilizing transformations for $\pi(\lambda)$, $B(n,\ p)$ and $NB(r,\ p)$ when their means are not less than 3, namely $\sqrt{Y+0.385}$ and $\sqrt{Y+0.375}$ for $\pi(\lambda)$ and $\lambda \geq 3$, $(n+\frac{1}{2})^{1/2}\sin^{-1}(\frac{2Y-n}{n+0.77})$ and $(n+\frac{1}{2})^{1/2}\sin^{-1}(\frac{2Y-n}{n+0.75})$ for $B(n,\ p)$ and $np \in [3,\ n-3]$, and $(r-\frac{1}{2})^{1/2}\sinh^{-1}\sqrt{\frac{Y+0.385}{r-0.77}}$ and $(r-\frac{1}{2})^{1/2}\sqrt{\frac{Y+0.385}{r-0.75}}$ for $NB(r,\ p)$ and $nq \geq 3$. Here the corresponding relative errors of transformed variances are less than 2%. When their means are not less than 5, then $\Delta\{\mathrm{Var}\,(T(Y(0.385)))\}$ is less than 0.1%.

If all the means of the above three distributions are small enough (e.g. ≤ 2) but larger than 0.5, combined transformations are favorable. They are $\sqrt{Y}+\sqrt{Y+1.3}$ and $\sqrt{Y}+\sqrt{Y+1}$ for $\pi(\lambda)$, $(n+\frac{1}{2})^{1/2}[\sin^{-1}(\frac{2Y-n-1}{n+1})+\sin^{-1}(\frac{2Y-n+1}{n+1})]$ for $B(n,\ p)$, and $r^{1/2}\sinh^{-1}\sqrt{\frac{Y}{r}}+(r-1)^{1/2}\sinh^{-1}\sqrt{\frac{Y+3/4}{r-3/2}}$ for $NB(r,\ p)$. When the means are not less than 1 (or 0.5), the relative errors of the transformed variances are less than 12% (or 40%).

Acknowledgement

This research was partially supported by the Open Fund for a Key–Key Silvicultural Discipline of Zhejiang Province Grant 200604.

References

Anscombe, F.J., 1948. The transformation of Poisson, binomial, negative binomial data. Biometrika 35, 246–254.
Bartlett, M.S., 1936. The square root transformation in the analysis of variance. Supplement to the Journal of the Royal Statistical Society 3, 68–78.
Bartlett, M.S., 1947. The Use of Transformations. Biometrics 13, 39–52.
Box, G.E.P., Cox, D.R, 1964. An analysis of transformations. Journal of Royal Statistical Society, B 26, 211–243.
Box, G.E.P., Cox, D.R, 1982. An analysis of transformation Revisited. Journal of the American Statistical Association 77, 177–182.
Bromiley, P.A., Thacker, N.A., 2002. The effects of an arcsin square root transform on a binomial distributed quantity. Tina memo, 2002-007.
Chatterjee, S., Hadi, A.S., Price, B., 2000. Regression Analysis by Example, 3rd ed. Wiley and Son, Inc.
Freeman, M.F., Tukey, J.W., 1950. Transformations related to the angular and the square root. The Annals of Mathematical Statistics 21, 607–611.
Laubscher, N.F., 1961. On stabilizing the binomial and negative binomial variances. Journal of the American Statistical Association 56, 143–150.
Montgomery, D.C., 2005. Design and Analysis of Experiment, 6th ed. Wiley and Son, Inc.
Thacker, N.A., Bromiley, P.A., 2001. The effects of a square root transform on a Poisson distributed quantity. Tina memo, 2001-010.
Uddin, M.T., Noor, M.S, Kabir, A., Ali, R., Islam, M.N., 2006. The transformations of Random variables under symmetry. Journal of Applied Sciences 6, 1818–1821.
Yamamura, K., 1999. Transformation using $(x + 0.5)$ to stabilize the variance of populations. Researches on Population Ecology 41, 229–234.