ON TRANSFORMATIONS USED IN THE ANALYSIS OF VARIANCE

By J. H. Curtiss

Cornell University

1. Introduction. Transformations of variates to render their distributions more tractable in various ways have long been used in statistics [12, chapter XVI]. The present extensive use of the analysis of variance, particularly as applied to data derived from designs such as randomized blocks and Latin squares, has placed new emphasis on the usefulness of such transformations. In the more usual significance tests associated with the analysis of variance, it is assumed *a priori* that the plot yields are statistically independent normally distributed variates which all have the same variance, but which have possibly different means. The hypotheses to be tested are then concerned with relations among these means. But in practice, it sometimes seems appropriate to specify for each variate a distribution in which the variance depends functionally upon the mean; moreover, in such cases, the specification is generally not normal. For example, when the data is in the form of a series of counts or percentages, a Poisson exponential or binomial specification may seem in order, and the variance of either of these distributions is functionally related to the mean of the distribution. Before applying the usual normal theory to such data, it is clearly desirable to transform each variate so that normality and a stable variance are achieved as nearly as possible.

Various transformations have been devised to do this, and a number of articles explaining the nature and use of these transformations have recently been published.[1] However, the available literature on the subject appears to be mainly descriptive and non-mathematical. The object of this paper is to provide a general mathematical theory (sections 2 and 3) for certain types of transformations now in use. In the framework of this theory we shall discuss in particular the square root and inverse sine transformations (section 4), and also several logarithmic transformations (section 4 and section 5).

2. General theory. As it arises in the analysis of variance, the problem of stabilizing a variance functionally related to a mean may be stated as follows: Suppose X is a variate whose mean $\mu = E(X)$ is a real variable with a range S of possible values, and whose standard deviation $\sigma = \sigma_X = \sigma(\mu)$ is a function of μ not identically constant. Required, to find a function $T = f(X)$ such that both $f(X)$ and $\sigma_T^2 = E\{[T - E(T)]^2\}$ are functionally independent of μ for μ on S. $\left(\text{By "functionally independent," we mean that } \frac{\partial f}{\partial \mu} \equiv 0, \text{ and } \frac{\partial \sigma_T^2}{\partial \mu} \equiv 0 \text{ for } \mu \text{ on } S.\right)$

[1] See references [1], [2], [3], [4], [5], [6], [13], [16].

The following line of argument is adopted in certain of the references mentioned above ([1], [2], [3], [4]): From the relation $dT = f'(X)dX$, we deduce as an approximation by some sort of summation process that $\sigma_T = f'(\mu)\sigma(\mu)$. Setting this expression equal to a constant, say c, we obtain $f'(\mu) = c/\sigma(\mu)$, so $f(x)$ is an indefinite integral of $c/\sigma(x)$. The roughness of the approximation used here is only too apparent.[2] For example, if X is normally distributed, then the variance of $T = X^2$ as given by the approximation is $4\sigma^2\mu^2$, while actually it is $4\sigma^2\mu^2 + 2\sigma^4$.

Indeed, it is easily seen that in important special cases the problem of stabilization as above stated could have no solution other than the trivial one in which T is identically constant on the set of points of increase of the d.f.[3] of X. For instance, if X has a Poisson exponential distribution, then the identity $E[\{f(X) - E[f(X)]\}^2] \underset{\mu}{\equiv} c$, or $E\{[f(X)]^2\} \underset{\mu}{\equiv} c + \{E[f(X)]\}^2$, becomes

$$\sum_{k=0}^{\infty} [f(k)]^2 \frac{e^{-\mu}\mu^k}{k!} \underset{\mu}{\equiv} c + \left[\sum_{k=0}^{\infty} [f(k)] \frac{e^{-\mu}\mu^k}{k!}\right]^2, \qquad \mu > 0.$$

Expanding both sides in powers of μ, we need only equate the coefficients of the zero-th power of μ on each side to find that $[f(0)]^2 = c + [f(0)]^2$, which implies that $c = 0$ and hence that $f(0) = f(1) = f(2) = \cdots$. A similar demonstration can be given for the case in which X has a binomial distribution with a fixed number of values of the variate.

As to the problem of choosing $T = f(X)$ so that its distribution is exactly normal, we can observe immediately that a single-valued function $f(X)$ will never transform a variate X with a discrete distribution into a variate with a continuous one. On the other hand, any variate X with a continuous d.f. $F(x)$ can be transformed into a normally distributed variate T by the transformation $T = f(X)$ defined by the equation

$$F(X) = \int_{-\infty}^{T} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\, dt.$$

However, aside from the practical difficulty of solving this equation for T, the resulting function $T = f(X)$ will not generally be functionally independent of the mean of X.

These considerations lead us to seek asymptotic solutions to the problems of normalization and stabilization. Such solutions are considered in the next section.

3. Asymptotic theorems. In the remainder of this paper, we shall suppose that the distribution of X depends on a parameter n which is to tend somehow to

[2] Tippett [14] says: "This derivation is not mathematically sound, and the result is only justified if on application it is found to be satisfactory."

[3] i.e., distribution function. For any given one-dimensional variate X we shall denote the probability or relative frequency assigned to a set R by $P(R)$. The d.f. of the variate then is the point function $F(x) = P(X \leq x)$. This function is sometimes called the cumulative frequency function of X.

infinity. The mean $\mu = \mu_n$ of X, with range S_n, will in general depend upon n (although by this we do not mean to exclude the case in which μ_n is constant for all values of n), and perhaps will depend also on some further independent parameters, which we shall denote collectively by θ, with range Σ. We shall seek a variate $T = f(X)$, in which $f(X)$ is functionally independent of μ and of the parameters θ for μ on S_n, θ on Σ, and such that the distribution of $f(X) - f(\mu_n)$ tends as $n \rightarrow \infty$ to a normal distribution, while $\lim_{n\to\infty}\sigma_T^2 = c^2$, where c^2 is an absolute constant. It is implied here that in case the additional parameters θ are present, the function $f(X)$ may depend non-trivially on n; but if n is the only parameter on which the distribution of X depends, then $f(X)$ must be functionally independent of n.

A solution to the problem just proposed is given in certain cases by Theorems 3.1 and 3.2 below, which are suggested by the heuristic reasoning of the second paragraph of section 2.

THEOREM 3.1. *Let $\psi_n(x)$ be a non-negative function of x and n, defined almost everywhere and integrable*[4] *with respect to x over any finite interval of the x-axis for each $n > 0$. Let*

$$T = f(X) = \int_a^X \psi_n(x)\,dx,$$

where a is an arbitrary constant. Let $F_n(y)$ be the d.f. of the variate $Y = (X - \mu_n)\psi_n(\mu_n)$. Suppose further that a continuous d.f. $F(y)$ exists such that $\lim_{n\to\infty}F_n(y) = F(y)$ for all values of y. Then either one of the following two conditions is a sufficient condition for the d.f. $H_n(w)$ of the variate $W = f(X) - f(\mu_n)$ to tend uniformly to $F(w)$, $-\infty < w < \infty$:

(a) *To each w for which $0 < F(w) < 1$, there corresponds for all n sufficiently large at least one root $x = x_n$ to the equation*

$$\int_{\mu_n}^{x} \psi_n(u)\,du = w, \tag{3.1}$$

and this root x_n has the property that

$$\lim_{n\to\infty}(x_n - \mu_n)\psi_n(\mu_n) = w. \tag{3.2}$$

(b) *For all n sufficiently large, $\psi_n(\mu_n) > 0$, and $\lim_{n\to\infty}q_n(w) = 1$ uniformly in any closed finite subinterval of the open interval defined by $0 < F(w) < 1$, where*

$$q_n(w) = \frac{\psi_n(w[\psi_n(\mu_n)]^{-1} + \mu_n)}{\psi_n(\mu_n)}. \tag{3.3}$$

To prove this theorem we shall first suppose that condition (a) is satisfied. Let w_1 and w_2 be the end points of the open interval (possibly infinite) defined by $0 < F(w) < 1$. If w lies in this interval, and if n is large enough for the root x_n in (3.1) to exist, then from the monotonic character of $\int_{\mu_n}^{X} \psi_n(x)\,dx$ we can

[4] "Integrable" here means absolutely integrable in the sense of Lebesgue.

infer that

$$
\begin{aligned}
H_n(w) &= P[f(X) - f(\mu_n) \leqq w] = P\left[\int_{\mu_n}^{X} \psi_n(x)\,dx \leqq w\right] \\
&= P(X \leqq x_n) = P[Y \leqq (x_n - \mu_n)\psi_n(\mu_n)] \qquad (3.4)\\
&= F_n[(x_n - \mu_n)\psi_n(\mu_n)].
\end{aligned}
$$

Since $F(w)$ is continuous, $\lim_{n\to\infty} F_n(w) = F(w)$ *uniformly* on any finite or infinite interval of values of w, as is well known.[5] Therefore $\lim_{n\to\infty} F_n(w_n) = F(w)$ if $\lim_{n\to\infty} w_n = w$. Thus from (3.2) and (3.4), we find that $\lim_{n\to\infty} H_n(w) = F(w)$ for $w_1 < w < w_2$.

If $w' \leqq w_1$, and $w_1 < w'' < w_2$, then $0 \leqq H_n(w') \leqq H_n(w'') = F(w'') + [H_n(w'') - F(w'')]$. We can make the right hand member of this relation less than any given positive number ϵ by first choosing w'' so that $F(w'') < \frac{1}{2}\epsilon$ (it will be remembered that $F(w)$ is a continuous d.f., and $F(w_1) = 0$) and then choosing n so large that the quantity in square brackets is also less than $\frac{1}{2}\epsilon$ in absolute value. Thus $\lim_{n\to\infty} H_n(w') = 0$. Similarly if $w' \geqq w_2$, we can show that $\lim_{n\to\infty} H_n(w') = 1$. Hence $\lim_{n\to\infty} H_n(w) = F(w)$ for all w, and it follows that the limit is uniform on any finite or infinite interval of values of w.

We shall now show that condition (a) in the theorem is a consequence of condition (b). The result follows at once from the following simple lemma:

LEMMA. *If $\gamma_n(w)$ is a non-negative function integrable over any finite interval of values of w; and if $\lim_{n\to\infty} \gamma_n(w) = 1$ uniformly in any finite closed subinterval of an interval $w_1 < w < w_2$, then for every value of w in this interval there exists for all n sufficiently large a solution $y = y_n$ of the equation $\int_0^y \gamma_n(z)\,dz = w$, and the solution y_n has the property that $\lim_{n\to\infty} y_n = w$.*

For it is clear that if w satisfies the inequality $w_1 < w < w_2$, and if $\eta > 0$ be chosen so that $w_1 < w - \eta < w + \eta < w_2$, then for all n sufficiently large,

$$\int_0^{w-\eta} \gamma_n(z)\,dz \leqq w \leqq \int_0^{w+\eta} \gamma_n(z)\,dz.$$

Thus for each n sufficiently large, there exists a root y_n of the equation $\int_0^y \gamma_n(z)\,dz = w$, and furthermore, this root satisfies the inequality $w - \eta \leqq y_n \leqq w + \eta$. Since η is arbitrarily small, the proof of the lemma is complete.

To apply the lemma, we make the change of variables $z = (u - \mu_n)\psi_n(\mu_n)$ in the integral in (3.1), which reduces it to the form

$$\int_0^y q_n(z)\,dz, \qquad y = (x - \mu_n)\psi_n(\mu_n), \tag{3.5}$$

and the conclusion that (a) is implied by (b) now follows at once.

---

[5] See [7], Theorem 11, pp. 29–30; also [8].

We add the remark that the uniformity of the limit of $q_n(z)$ in condition (b) may be replaced by the condition that for each closed finite sub-interval there exists a function $q(w)$ which dominates $q_n(w)$ for all n sufficiently large.

Our second theorem, which is stated in the terminology and notation of Theorem 3.1, is concerned with the limit of the variance of $T = f(X)$. From the mere fact that the distribution of W tends to a limiting form, it by no means follows that the mean and variance of the distribution of W approach those of the limiting form, as may be shown by trivial examples. Thus additional hypotheses on $\psi_n(x)$ and on the behavior of the distribution of Y become necessary.

THEOREM 3.2. *Let T (or $f(X)$), Y, $F_n(y)$ and $F(y)$ be defined as in Theorem 3.1. Let the mean and variance of the distribution defined by $F(y)$ exist and have respective values 0 and c^2. Then the following three conditions, taken together, are sufficient that*

$$\lim_{n\to\infty}[E(T) - f(\mu_n)] = 0, \tag{3.6}$$

$$\lim_{n\to\infty}\sigma_T^2 = c^2: \tag{3.7}$$

(i) *$E(Y^2)$ exists for $n > 0$, and $\lim_{n\to\infty}E(Y^2) = c^2$.*

(ii) *Condition (b) of Theorem 3.1 holds.*

(iii) *$f(Y[\psi_n(\mu_n)]^{-1} + \mu_n) - f(\mu_n) = O \mid Y \mid$ uniformly in n as $\mid Y \mid \to \infty$.*

As a preliminary step in the proof, we observe that (i) and the relations $\lim_{n\to\infty}F_n(y) = F(y)$, $c^2 = \int_{-\infty}^{+\infty} y^2\, dF(y)$, imply that the improper integral $\int_{-\infty}^{+\infty} y^2\, dF_n(y)$ converges uniformly in n for $n > 0$. As the integrand is positive, the following result is equivalent to the uniform convergence of the integral: *For every $\epsilon > 0$, there exist numbers A_1 and A_2, $A_1 < A_2$, such that for all n sufficiently large,*

$$\left(\int_{-\infty}^{A_1} + \int_{A_2}^{\infty}\right) y^2\, dF_n(y) < \epsilon.$$

To prove this, we write

$$\left(\int_{-\infty}^{A_1} + \int_{A_2}^{\infty}\right) y^2\, dF_n(y) = [E(Y^2) - c^2]$$
$$+ \left[\int_{A_1}^{A_2} y^2\, dF(y) - \int_{A_1}^{A_2} y^2\, dF_n(y)\right] + \left[c^2 - \int_{A_1}^{A_2} y^2\, dF(y)\right].$$

We first choose A_1 and A_2 so that the last bracket here is less than $\frac{1}{2}\epsilon$ in absolute value. By condition (i), the first bracket approaches zero as n tends to infinity, and the Helly-Bray theorem [10, p. 15] states that the second bracket also approaches zero as n tends to infinity, so for all n sufficiently large, the sum of the first two brackets is in absolute value less than $\frac{1}{2}\epsilon$.

It is important to notice that we can always choose A_1 and A_2 in the above demonstration so that $A_1 > w_1$, $A_2 < w_2$, where w_1 and w_2 are as usual the endpoints of the interval defined by $0 < F(w) < 1$.

To continue with the proof of the theorem, we remark that by a change of variables similar to the one used to derive (3.5), the function $W = f(X) - f(\mu_n)$ may be expressed as a function of Y in the following manner:

$$W = \int_{\mu_n}^{X} \psi_n(x)\, dx = \int_0^Y q_n(w)\, dw = Q_n(Y),$$

where $q_n(w)$ is given by (3.3). In terms of W, (3.6) and (3.7) become, respectively,

$$\lim_{n\to\infty} E(W) = 0, \tag{3.8}$$

$$\lim_{n\to\infty} \{E(W^2) - [E(W)]^2\} = c^2, \tag{3.9}$$

and these are the equations which we now establish.

Conditions (ii) and (iii) obviously imply that $\lim_{n\to\infty} Q_n(y) = y$ uniformly in any finite closed subinterval of the interval $w_1 < y < w_2$, and that a constant M exists such that $|Q_n(y)| \leqq M|y|$ for all n. If $E(Y^2)$ exists, so will $E(Y)$. Now

$$\begin{aligned} E(W) &= \int_{-\infty}^{+\infty} Q_n(y)\, dF_n(y) \\ &= \int_{-\infty}^{+\infty} Q_n(y)\, dF_n(y) - \int_{-\infty}^{+\infty} y\, dF_n(y) \\ &= \left(\int_{-\infty}^{A_1} + \int_{A_2}^{\infty}\right) [Q_n(y) - y]\, dF_n(y) + \int_{A_1}^{A_2} [Q_n(y) - y]\, dF_n(y), \end{aligned}$$

where $w_1 < A_1 < A_2 < w_2$. Therefore

$$|E(W)| \leqq \left(\int_{-\infty}^{A_1} + \int_{A_2}^{\infty}\right)(M + 1)\,|y|\, dF_n(y) + \int_{A_1}^{A_2} |Q_n(y) - y|\, dF_n(y).$$

From the uniform convergence of $\int_{-\infty}^{+\infty} y^2 dF_n(y)$, proved above, we can conclude that the pair of improper integrals in this inequality can be made less than an arbitrary $\frac{1}{2}\epsilon > 0$ by proper choice of A_1 and A_2. The third integral approaches zero, by the general Helly-Bray Theorem [10, p. 16], and so becomes less than $\frac{1}{2}\epsilon$ for all n sufficiently large. Thus we have established (3.8). To show that (3.9) is true, we have merely to prove that $\lim_{n\to\infty} E(W^2) = c^2$. Since $E(Y^2) = \int_{-\infty}^{+\infty} y^2 dF_n(y)$, we may write

$$E(W^2) - c^2 = \int_{-\infty}^{+\infty} \{[Q_n(y)]^2 - y^2\}\, dF_n(y) + [E(Y^2) - c^2].$$

The integral may be shown to approach zero by the argument used in the case of $E(W)$, and the required result then follows from condition (i) of the theorem. The proof is now complete.

The sufficient conditions in Theorem 3.2 can be modified in various more or less obvious ways. The existence of the limiting d.f. $F(y)$ was essentially used in the proof only to secure the uniform convergence of $\int_{-\infty}^{+\infty} y^2 dF_n(y)$. Condition (ii) can again be modified along the lines suggested at the end of the proof of Theorem 3.1. Condition (iii) was used only to secure the uniform convergence of the integral $\int_{-\infty}^{+\infty} [Q_n(y)]^2 dF_n(y)$.

For later reference, we shall supplement Theorems 3.1 and 3.2 with the following simple result, which is practically self-evident.

THEOREM 3.3. *Let the distribution of a variate Y depend upon a parameter n, let $F_n(y)$ be the d.f. of Y, and let $F(y)$ be a continuous d.f. with the property that $\lim_{n\to\infty} F_n(y) = F(y)$. Let a_n be a function of n such that $\lim_{n\to\infty} a_n = a \neq 0$. Then the d.f. of the variate $Z = a_n Y$ tends as $n \to \infty$ to the d.f. $F(z/a)$ if $a > 0$, and to the d.f. $1 - F(z/a)$ if $a < 0$. If the variance of Y exists and tends to c^2 as $n \to \infty$, then the variance of $a_n Y$ tends to a^2c^2 as $n \to \infty$.*

If $F(y)$ is the d.f. of a reduced normal distribution, i.e.,

$$F(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}t^2}\, dt,$$

then $F(z/a)$ is also the d.f. of a normal distribution with mean zero and variance a^2. More generally, any affine transformation of a normal variate yields another normal variate.

4. Applications. The theorems of the preceding section have the effect of referring the properties of the distribution of the transformation $T = f(X)$ of Theorem 3.1 back to those of the distribution of a related variate Y. In the applications given in the present section, we shall let $\psi_n(\mu_n)$ be proportional to the reciprocal of the standard deviation of X. The theorems of section 3 state in this case that if the reduced, or standardized, distribution of X approaches a limiting form, then under certain circumstances, the distribution of $f(X) - f(\mu_n)$ will approach a similar limiting form, and σ_T^2 will approach a quantity independent at least of n. In the applications considered here, the reduced distribution of X will always approach the reduced normal distribution.

(I) **The square root transformation for a variate with a Poisson exponential distribution.** *Let X have a Poisson exponential distribution with parameter n. If α is an arbitrary constant, and if*

$$T = f(X) = \begin{cases} \sqrt{X + \alpha}, & X \geqq -\alpha \\ 0 \quad , & X < -\alpha \end{cases} \tag{4.1}$$

then the distribution of $T - \sqrt{n + \alpha}$ *tends as* $n \to \infty$ *to a normal distribution which has mean zero and variance* $\frac{1}{4}$, *and* $\lim_{n\to\infty}\sigma_T^2 = \frac{1}{4}$. For $\mu_n = n$, $\sigma_X = \sqrt{n}$, and it is well known[6] that the distribution of the reduced variate $(X - n)/\sqrt{n}$ tends to the reduced normal distribution as $n \to \infty$. By Theorem 3.3, the distribution of the variate

$$Y = \frac{X - n}{2\sqrt{n + \alpha}} = \frac{1}{2}\cdot\sqrt{\frac{n}{n + \alpha}}\cdot\frac{X - n}{\sqrt{n}},$$

will tend to normality as $n \to \infty$, and the variance of Y will tend to the value $\frac{1}{4}$, which is also the variance of the limiting distribution. Setting

$$\psi_n(x) = \begin{cases} \dfrac{1}{2\sqrt{x + \alpha}}, & x > -\alpha \\ 0\quad, & x \leqq -\alpha, \end{cases}$$

we obtain from $T = f(X) = \int_{-\alpha}^{X} \psi_n(x)dx$ the formula given in (4.1). To prove the statement in italics, we must show that conditions (ii) and (iii) of Theorem 3.2 are satisfied. We have, assuming $n > -\alpha$,

$$q_n(w) = \begin{cases} \left(1 + \dfrac{2w}{\sqrt{n + \alpha}}\right)^{-\frac{1}{2}}, & w > -\frac{1}{2}\sqrt{n + \alpha} \\ 0\quad, & w \leqq -\frac{1}{2}\sqrt{n + \alpha}, \end{cases}$$

so clearly (ii) is satisfied. Also,

$$\begin{aligned} W &= f(Y[\psi_n(\mu_n)]^{-1} + \mu_n) - f(\mu_n) \\ &= \begin{cases} \sqrt{2Y\sqrt{n + \alpha} + n + \alpha} - \sqrt{n + \alpha}, & Y > -\frac{1}{2}\sqrt{n + \alpha} \\ -\sqrt{n + \alpha}, & Y \leqq -\frac{1}{2}\sqrt{n + \alpha}, \end{cases} \end{aligned}$$

from which it follows at once that $| W | < 2 | Y |$ for all Y, and so (iii) is satisfied.

The degree of approximation involved in the equation $\lim_{n\to\infty}\sigma_T^2 = \frac{1}{4}$ has been investigated numerically by Bartlett [1] for values of n from .5 to 15.0 in the cases $\alpha = 0$ and $\alpha = \frac{1}{2}$. He found that the variance of $\sqrt{X + (\frac{1}{2})}$ is considerably closer to the limit $(\frac{1}{4})$ for $1 \leqq n \leqq 10$ than is the variance of $\sqrt{X}$. At $n = 15$, the variance of $\sqrt{X}$ is .256, and that of $\sqrt{X + (\frac{1}{2})}$ is .248.

The question of the degree of convergence to normality and of the possibility of selecting an optimum value of α remain open. By expanding the function $\sqrt{X + \alpha}$ in a Taylor series about $X = n$ with remainder in the form due to Schlömilch, it is possible to derive as accurate an estimate of $| \sigma_T^2 - (\frac{1}{4})|$ as may

[6] See (e.g.) [9].

be desired. A rough result easily obtainable by this method is that $|\sigma_T^2 - (\frac{1}{4})| \leqq 3/(4n), n > 0$.

(II) **The square root transformation for a variate with a Γ distribution.** *Let X have a distribution whose density function is of the following type:*

$$\varphi(x) = \begin{cases} 0 & x \leqq 0 \\ Kx^{\frac{1}{2}n-1}e^{-hx}, & x \geqq 0, h > 0. \end{cases} \tag{4.2}$$

If α is an arbitrary constant, and if

$$T = f(X) = \begin{cases} \sqrt{X + \alpha}, & X \geqq -\alpha \\ 0 \quad , & X < -\alpha, \end{cases} \tag{4.3}$$

then the distribution of $T - \sqrt{(n/2h) + \alpha}$ *tends as* $n \to \infty$ *to a normal distribution which has mean zero and variance* $1/4h$, *and* $\lim_{n\to\infty}\sigma_T^2 = 1/(4h)$. For $\mu_n = n/(2h)$, $\sigma_X = \sqrt{n}/(h\sqrt{2}) = \sqrt{\mu_n/h}$. The distribution of the reduced variate tends to normality as $n \to \infty$,[7] so that of the variate

$$Y = \frac{x - \mu_n}{2\sqrt{\mu_n + \alpha}} = \frac{1}{2}\sqrt{\frac{n}{nh + 2h^2\alpha}} \cdot \frac{x - \mu_n}{\sqrt{\mu_n/h}}$$

tends to normality also with limiting variance $1/(4h)$. Setting

$$\psi_n(x) = \begin{cases} \dfrac{1}{2\sqrt{x + \alpha}}, & x > -\alpha \\ 0 \quad , & x \leqq -\alpha, \end{cases}$$

we obtain T in (4.3) from the relation $T = \int_{-\alpha}^{X} \psi_n(x)\,dx$. The work of verifying that the conditions of Theorem 3.2 are satisfied is the same as in the case of the Poisson exponential distribution treated above, and will not be repeated.

For example, if s^2 denotes the variance of a random sample of $n + 1$ observations drawn from a normal parent distribution with variance σ^2, then it is well known that $(n + 1)s^2$ is distributed according to (4.2) with $h = 1/(2\sigma^2)$. We thus can deduce the further facts, also well known, that the distribution of $\sqrt{n+1}\,s - \sigma\sqrt{n}$ tends to normality, and that the variance of $s\sqrt{n+1}$ approaches the limiting value $\frac{1}{2}\sigma^2$. If n is an integer and $h = \frac{1}{2}$, the distribution defined by (4.2) is called a χ^2 distribution with n degrees of freedom, and the variate is often denoted by χ^2. Our conclusion in this case is that the distribution of $\sqrt{2\chi^2} - \sqrt{2n}$ tends to a normal one with zero mean and unit variance. From this result and the fact that $\sqrt{2n-1} - \sqrt{2n} = O(n^{-\frac{1}{2}})$, it follows immediately that $\sqrt{2\chi^2} - \sqrt{2n-1}$ has the same limiting distribution as $\sqrt{2\chi^2} - \sqrt{2n}$. This result,[8] due to Fisher, is familiar to all users of his table of the probability levels of χ^2.

[7] See (e.g.) [9].

[8] For a discussion of the degree of convergence involved here, see [9].

(III) **The inverse sine transformation for a binomial variate.** *Let X have a binomial relative frequency distribution with parameter p and the n values* 0, $1/n$, $2/n, \cdots, n/n$. *If α is an arbitrary constant, and if*

$$(4.4)\qquad T = f(X) = \begin{cases} \sqrt{n}\sin^{-1}\sqrt{X+\frac{\alpha}{n}} & ,-\frac{\alpha}{n} \leqq X \leqq 1-\frac{\alpha}{n} \\ 0, \quad X < -\frac{\alpha}{n}, \quad X > 1-\frac{\alpha}{n}, \end{cases}$$

where T is measured in radians, then the distribution of $T - \sqrt{n}\sin^{-1}\sqrt{p+(\alpha/n)}$ tends as $n \to \infty$ to a normal distribution which has mean zero and variance $\frac{1}{4}$, and $\lim_{n\to\infty}\sigma_T^2 = \frac{1}{4}$. For here, $\mu_n = p$, and $\sigma_X^2 = pq/n$, where $q = 1 - p$; and the familiar DeMoivre-Laplace theorem states that the distribution of the reduced variate $\sqrt{n}(X - p)/\sqrt{pq}$ will tend to normality as $n \to \infty$. Hence by Theorem 3.3 the distribution of

$$(4.5)\qquad Y = \frac{\sqrt{n}(X-p)}{2\sqrt{\left(p+\frac{\alpha}{n}\right)\left(q-\frac{\alpha}{n}\right)}}$$

will tend to normality with a limiting variance of $\frac{1}{4}$, which is also the variance of the limiting distribution. Setting

$$\psi_n(x) = \begin{cases} \dfrac{\sqrt{n}}{2\sqrt{\left(x+\frac{\alpha}{n}\right)\left(1-x-\frac{\alpha}{n}\right)}}, & -\frac{\alpha}{n} < x < 1-\frac{\alpha}{n} \\ 0 & x \leqq -\frac{\alpha}{n}, x \geqq 1-\frac{\alpha}{n}, \end{cases}$$

we obtain (4.4) from the integral

$$T = \int_{-\alpha/n}^{X} \psi_n(x)\,dx.$$

In proving the conditions (ii) and (iii) of Theorem 3.2 are satisfied, we shall assume for simplicity that $\alpha = 0$. We find that

$$q_n(w) = \begin{cases} \left(1 + 2w\dfrac{q-p}{\sqrt{npq}} - \dfrac{4w^2}{n}\right)^{-\frac{1}{2}}, & -\frac{1}{2}\sqrt{\frac{np}{q}} < w < \frac{1}{2}\sqrt{\frac{nq}{p}} \\ 0 \qquad, & w \leqq -\frac{1}{2}\sqrt{\frac{np}{q}}, \quad w \geqq \frac{1}{2}\sqrt{\frac{nq}{p}}, \end{cases}$$

so obviously (ii) is satisfied. From the Law of the Mean in the form due to Schlömilch, we have

$$(4.6)\qquad \begin{aligned} W &= \sqrt{n}\sin^{-1}\sqrt{p + 2\sqrt{\frac{pq}{n}}\,Y} - \sqrt{n}\sin^{-1}\sqrt{p} \\ &= 2Y\left[\frac{1-\theta}{\left(1+2\theta\sqrt{\frac{q}{np}}\,Y\right)\left(1-2\theta\sqrt{\frac{p}{nq}}\,Y\right)}\right]^{\frac{1}{2}}, \\ &0 < \theta < 1, \qquad -\frac{1}{2}\sqrt{\frac{np}{q}} < Y < \frac{1}{2}\sqrt{\frac{nq}{p}}. \end{aligned}$$

The denominator of the coefficient of $2Y$ here is a quadratic function of Y with a negative coefficient of Y^2, and so must assume its least value in the Y range indicated in (4.6) at one end or the other of the range. From this it is readily seen that the coefficient of $2Y$ is actually always less than unity. For values of Y outside the range, the second member of (4.6) indicates that $W = O(\sqrt{n}) = O(Y)$. Hence (iii) is satisfied, and the proof of the statement in italics is complete for the case $\alpha = 0$. The more general case presents no important new difficulties.

In practice, it is often convenient to express X as a percentage. This merely has the effect of multiplying Y in (4.5) by 100. We find in this case that $\sqrt{n} \sin^{-1}\sqrt{X + 100\alpha/n} - \sqrt{n} \sin^{-1}\sqrt{100p + 100\alpha/n}$ has a distribution approaching normality, and $\sigma_T \to 50$ instead of $\frac{1}{2}$.

Bartlett [1] gives numerical results in the cases $n = 10$, $\alpha = 0$ and $n = 10$, $\alpha = \frac{1}{2}$, which indicate that perhaps the choice $\alpha = \frac{1}{2}$ is more suitable if the estimated p is near 0 or 1, but the choice $\alpha = 0$ is preferable if the estimated p lies between .3 and .7. However, there seems to be no good reason to believe that these conclusions should be valid for other values of n. The question of an optimum α, and of the degree of convergence to normality remain open. We note in passing that the latter problem could doubtless be profitably studied by combining the methods of proof of Theorem 3.1 with the results of Uspensky [15, pp. 129–130] on the degree of approximation of the reduced binomial d.f. to the normal d.f.

IV. **Other transformations of a binomial variate.** *Let X have a binomial relative frequency distribution with the parameter p and the n values $0, 1/n, 2/n, \cdots, n/n$.*

(a) *If*

$$T = f(X) = \begin{cases} \sqrt{n} \sinh^{-1} \sqrt{X} = \sqrt{n} \log\left(\sqrt{X} + \sqrt{1 + X}\right),^{9} & X \geqq 0 \\ 0, & X < 0, \end{cases}$$

then the distribution of $T - \sqrt{n} \sinh^{-1} \sqrt{p}$ tends as $n \to \infty$ to a normal distribution which has mean zero and variance $q/(4 + 4p)$, and $\lim_{n\to\infty}\sigma_T^2 = q/(4 + 4p)$.

(b) *If*

$$T = f(X) = \begin{cases} \sqrt{n} \log X, & X > 0, \\ 0, & X \leqq 0, \end{cases}$$

then the distribution of $T - \sqrt{n} \log p$ tends as $n \to \infty$ to a normal distribution which has mean zero and variance q/p, and $\lim_{n\to\infty}\sigma_T^2 = q/p$.

(c) *If*

$$T = f(X) = \begin{cases} \dfrac{1}{2}\sqrt{n} \log \dfrac{X}{1 - X}, & 0 < X < 1, \\ 0, & X \leqq 0,\ X \geqq 1, \end{cases}$$

[9] All logarithms in this paper are to the base e.

then the distribution of $T - \frac{1}{2}\sqrt{n}\log\frac{p}{1-p}$ *tends as* $n \to \infty$ *to a normal distribution which has mean zero and variance* $1/(4pq)$, *and* $\lim_{n\to\infty}\sigma_T^2 = 1/(4pq)$.

Since the limiting variance of each of these transformations involves the parameter p, they are not to be regarded as solutions of the problem of asymptotic variance stabilization proposed at the beginning of section 3, although it is perhaps of some interest that their distributions become asymptotically normal.

In case (a), $f'(x) = \sqrt{n}/(2\sqrt{x^2 + x})$, $x > 0$. Setting $\psi_n(x) = f'(x)$, $x > 0$, and $\psi_n(x) = 0$, $x \leqq 0$, we obtain

$$(4.7) \qquad Y = (X - p)\psi_n(p) = \frac{\sqrt{n}(X-p)}{\sqrt{pq}} \cdot \frac{\sqrt{q}}{2\sqrt{1+p}},$$

and this variate obviously has the limiting distribution ascribed to $T - \sqrt{n}\sinh^{-1}\sqrt{p}$ in the statement in italics. The truth of that statement now follows by an argument similar to that used in the case of the inverse sine transformation.

If p is allowed to vary with n in such a way that $\lim_{n\to\infty} np = \infty$, it is known that the reduced distribution of X will still tend to normality.[10] If we suppose that $\lim_{n\to\infty} p = 0$, but $\lim_{n\to\infty} np = \infty$, we find from Theorem 3.3 that the limiting distribution of Y in (4.7) will be normal with mean zero and variance $\frac{1}{4}$, and that $\sigma_Y^2 \to \frac{1}{4}$. It is easily verified that the conditions (ii) and (iii) of Theorem 3.2 are still satisfied, so we find that the limiting distribution of $[\sqrt{n}\sinh^{-1}\sqrt{X} - \sqrt{n}\sinh^{-1}\sqrt{p}]$ is normal, with mean zero and variance $\frac{1}{4}$, and $\sigma_T^2 \to \frac{1}{4}$. However, since n is now the only independent parameter, we cannot here regard the transformation $T = \sqrt{n}\sinh^{-1}\sqrt{X}$ as a solution of the problem of variance stabilization, because the variate T depends explicitly upon n.

If in case (b) we proceed as in case (a), we obtain as the analogue of (4.7) the formula

$$Y = (X - p)\psi_n(p) = \frac{\sqrt{n}(X-p)}{\sqrt{pq}}\sqrt{\frac{q}{p}},$$

and this variate has the limiting distribution ascribed to $T - \sqrt{n}\log X$ in the statement in italics. It now turns out that although condition (ii) of Theorem 3.2 is satisfied, condition (iii) is not satisfied. We are then faced with the problem of proving directly that the improper integral

$$\int_{-\sqrt{n}}^{+\infty} [\sqrt{n}\log(p + py/\sqrt{n}) - \sqrt{n}\log p]^2\, dF_n(y)$$

converges uniformly.[11] The trouble occurs only at the lower limit of integration, and may be resolved by first integrating by parts, then dividing the range

[10] See (e.g.) [9].

[11] See the remarks following the proof of Theorem 3.2.

$(-\sqrt{n}, A_1)$ into two ranges $(-\sqrt{n}, -\log n)$ and $(-\log n, A_1)$, and then applying Uspensky's results [15, pp. 129–130], on the degree of approximation involved in the DeMoivre-Laplace theorem.

Case (c) may be handled in a similar manner.

5. The logarithmic transformation. We shall suppose throughout this section that X is a variate whose mean μ_n and standard deviation σ in the relation $\sigma = k_n(\mu_n + \alpha)$, where α is an arbitrary constant, $k_n > 0$, and $\lim_{n\to\infty} k_n$ exists and is finite. If k_n is constant for all n, say $k_n = k > 0$, and if we use the heuristic argument of the second paragraph of section 2 to attempt to find a transformation which will stabilize the variance of X at k^2, we arrive at the function $T = \log (X + \alpha)$, $X > -\alpha$. It is the purpose of this section to study the asymptotic properties of this transformation.

The theory of such a transformation differs in certain important respects from that of the transformations considered in sections 3 and 4. For one thing, our starting point in the study of each transformation considered in section 4 was the fact that although $P(X < 0) = 0$, nevertheless the reduced distribution of X tended to normality as $n \to \infty$. But in the present case, if X is a variate such that $P(X \leqq -\alpha) = 0$, then the corresponding reduced variate $Y = (X - \mu_n)/[k_n(\mu_n + \alpha)]$ has a d.f. $F_n(y)$ such that $F_n(-1/k_n) = 0$. Thus if $\lim_{n\to\infty} k_n = k > 0$, the limiting distribution of Y, if it exists, must have a d.f. $F(y)$ such that $F(-1/k - 0) = 0$. Therefore the limiting distribution of Y can never be normal if $k > 0$.

Moreover (in contrast to the situation in Theorem 3.1) if the reduced variate Y does have a limiting distribution, the variate

$$(5.1)\quad W = \frac{1}{k_n} \log (X + \alpha) - \frac{1}{k_n} \log (\mu_n + \alpha) = \int_{\mu_n}^{X} \frac{1}{k_n(u + \alpha)}\, du, \quad X > -\alpha$$

may have a limiting distribution which is not the same as that of Y. More specifically, we have the following result:

THEOREM 5.1. *Let $P(X \leqq -\alpha) = 0$, let $\lim_{n\to\infty} k_n = k \geqq 0$, let $F_n(y)$ be the d.f. of the reduced variate*

$$Y = \frac{X - \mu_n}{k_n(\mu_n + \alpha)},$$

and let $H_n(w)$ be the d.f. of the variate W given by (5.1). If a continuous d.f. $F(y)$ exists such that $\lim_{n\to\infty} F_n(y) = F(y)$ for all y, then

$$\lim_{n\to\infty} H_n(w) = \begin{cases} F\left[\dfrac{e^{kw} - 1}{k}\right], & k > 0 \\ F(w), & k = 0. \end{cases}$$

The proof is simpler than the statement; essentially we have only to notice that

$$H_n(w) = P\left[-\frac{1}{k_n} < Y \leqq \frac{e^{k_n w} - 1}{k_n}\right]$$
$$= F_n\left[\frac{e^{k_n w} - 1}{k_n}\right], \qquad -\infty < w < \infty,$$

and apply the reasoning used above in connection with (3.4).

From the study of the distribution of T, we now turn for a moment to the question of the limit if σ_T^2. Here the situation is more consistent with the results of section 3.

THEOREM 5.2. *Under the hypotheses of Theorem* 5.1 *and under the additional conditions that the improper integral* $\int_{-\infty}^{0} w^2\, dH_n(w) \left(or \int_{-1/k_n}^{0} k_n^{-2}[\log(1 + k_n y)]^2\, dF_n(y)\right)$ *converges uniformly in n and that* $\int_{-\infty}^{+\infty} y^2\, dF(y) = 1 = E(Y^2)$, *the following relations hold:*

$$\text{(5.2)} \qquad \lim_{n\to\infty} E(W) = \begin{cases} \displaystyle\int_{-1/k}^{\infty} \frac{1}{k} \log(1 + ky)\, dF(y), & k > 0, \\ 0 \qquad , & k = 0, \end{cases}$$

$$\text{(5.4)} \qquad \lim_{n\to\infty} E(W^2) = \begin{cases} \displaystyle\int_{-1/k}^{\infty} \frac{1}{k^2} [\log(1 + ky)]^2\, dF(y), & k > 0 \\ 1 \qquad , & k = 0 \end{cases}$$

The variance σ_T^2 of the variate $T = \log(X + \alpha)$ is related to these mean values by the equation $\sigma_T^2 = k_n^2\{E(W^2) - [E(W)]^2\}$. Thus if $F(y)$ is independent of any unknown parameters θ, and if k is positive and is presumed to have the same value for all variates in any given problem, then the transformation $T = \log(X + \alpha)$ is seen to yield an asymptotic stabilization of the variance under the conditions of Theorem 5.2. If $k = 0$, we find from either Theorem 5.2 or the proof of Theorem 5.2 that $T = \log(X + \alpha)$ converges stochastically to $\log(\mu_n + \alpha)$.

The proof of Theorem 5.2 is similar to that of Theorem 3.2 and will be omitted.

Theorem 5.1 raises the following question: Just what limiting distribution must Y have if $k > 0$, in order that the distribution of W tend to normality? To answer this, we shall note the following simple non-asymptotic result:

THEOREM 5.3. *A necessary and sufficient condition that X have a continuous distribution with density function*

$$\text{(5.4)} \quad \varphi(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi \log(k^2 + 1)}} \dfrac{1}{x + \alpha} \\ \qquad \times \exp\left[\dfrac{-\left(\log \dfrac{(x + \alpha)\sqrt{k^2 + 1}}{u + \alpha}\right)^2}{2 \log(k^2 + 1)}\right], & x > -\alpha \\ 0 \qquad , & x \leqq -\alpha \end{cases}$$

for which $\sigma_X = k(\mu + \alpha)$, *is that the variate* $T = \log(X + \alpha)$ *have a normal distribution with mean* $\log(\mu + \alpha) - \log\sqrt{k^2 + 1}$ *and variance* $\log(k^2 + 1)$.

The proof may be given by a routine change of variables.[12] It is to be noticed that the heuristic argument of the second paragraph of section 2 would lead to the incorrect result that the variance of T was k^2 instead of $\log(k^2 + 1)$. In case $k = 1$, the mean and variance of T are respectively $\log(\mu + \alpha) - .347$ and $.693$. If the transformation $T = \log_{10}(X + \alpha)$ is used, the new mean is $\log_{10}(\mu + \alpha) - \log_{10}\sqrt{k^2 + 1}$ and the new variance is $.189 \log(k^2 + 1)$, which for values of k near zero has the approximate value $.189k^2$.[13]

If X is distributed according to (5.4), the density function $F'(y)$ of the corresponding reduced variate $Y = (X - \mu)/[k(\mu + \alpha)]$ is

$$(5.5)\quad F'(y) = \begin{cases} \dfrac{k}{\sqrt{2\pi \log(k^2 + 1)}} \cdot \dfrac{1}{1 + ky} \\ \qquad \times \exp\left[-\dfrac{\{\log[(1 + ky)\sqrt{k^2 + 1}]\}^2}{2\log(k^2 + 1)} \right], & y > -\dfrac{1}{k} \\ 0 & y \leqq -\dfrac{1}{k}. \end{cases}$$

The d.f. of the variate $W = k^{-1}[\log(X + \alpha) - \log(\mu + \alpha)]$ is $F[(e^{kw} - 1)/k]$, and, of course, the distribution of W is normal with mean $-k^{-1}\log\sqrt{k^2 + 1}$, and variance $k^{-2}\log(k^2 + 1)$. These are the respective values of the integrals in (5.2) and (5.3).

If now the distribution of X depends on a parameter n in such a way that as $n \to \infty$, the distribution of the corresponding reduced variate $Y = (X - \mu_n)/[k_n(\mu_n + \alpha)]$ tends to the distribution given by (5.5), it follows from the above remarks and from Theorem 5.1 that the variate W given by (5.1) has a normal limiting distribution. Furthermore, under the uniform convergence condition of Theorem 5.2, it follows that σ_T^2 tends to the value $\log(k^2 + 1)$, where $T = \log(X + \alpha)$.

These facts provide a sound mathematical basis for the use of the logarithmic transformation, which has had a long history of empirical success in problems of normalization [12, chapter XVI] and stabilization ([6], [16]). When it appears from a reasonably large number of observations on a variate (which is essentially bounded from below) that the standard deviation of the variate is proportional to the mean, then a possible specification for the variate is a distribution of the form (5.4); or, at least for large values of μ, it may be assumed that the distribution of the reduced variate is given by (5.5). Then the variate $T = \log(X + \alpha)$, where $-\alpha$ is any number less than the lower bound of X, will be exactly or approximately normally distributed with a variance independent of the value of μ.

Since (5.4) is only one of an infinity of various different types of distribution

---

[12] Finney [11] has considered the problem of efficiently estimating the variance of the X of Theorem 5.3 in the case $\alpha = 0$. (The actual density function (5.4) appears nowhere in his paper.)

[13] Given (without explanation) by Cochran [6, p. 165].

in which the mean and standard deviation are proportional, the user of a logarithmic transformation in the analysis of variance should always apply tests for departure from normality to the observed distribution of T values. From the point of view of specification, the situation here would seem to be less reassuring than in the cases considered in section 4. While it is true that the Poisson exponential distribution is only one of many types of distribution in which the variance and mean are equal, nevertheless the specification of a Poisson distribution can generally be preceded by a fairly strong chain of *a priori* inductive reasoning. This would not seem to be the case in the specification of (5.4). Theorems 5.1 and 5.2 furnish some grounds for a suspicion that the logarithmic transformation may possibly be more successful in stabilizing the variance than in normalizing the data. The burden of proof, however, lies with the experimenter.[14]

REFERENCES

[1] M. S. Bartlett, "The square root transformation in the analysis of variance," *Jour. Roy. Stat. Soc. Suppl.*, Vol. 3 (1936), pp. 68–78.
[2] Geoffrey Beall, "The transformation of data from entomological field experiments so that the analysis of variance becomes applicable," *Biometrika*, Vol. 32 (1942), pp. 243–262.
[3] C. I. Bliss, "The transformation of percentages for use in the analysis of variance," *Ohio Jour. Science*, Vol. 38 (1938), pp. 9–12.
[4] A. Clark and W. H. Leonard, "The analysis of variance with special reference to data expressed as percentages," *Jour. Amer. Soc. Agron.*, Vol. 31 (1939), pp. 55–56.
[5] W. G. Cochran, "The analysis of variance when experimental errors follow the Poisson or binomial laws," *Annals of Math. Stat.*, Vol. 9 (1940), pp. 335–347.
[6] W. G. Cochran, "Some difficulties in the statistical analysis of replicated experiments," *Empire Jour. Expt. Agric.*, Vol. 6 (1938), pp. 157–175.
[7] H. Cramér, *Random Variables and Probability Distributions*, Cambridge, 1937.
[8] J. H. Curtiss, "A note on the theory of moment generating functions," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 430–433.
[9] J. H. Curtiss, "Convergent sequences of probability distributions," *Am. Math. Monthly*, Vol. 50 (1943), pp. 94–105.
[10] G. C. Evans, *The Logarithmic Potential*, New York, 1927.
[11] D. J. Finney, "On the distribution of a variate whose logarithm is normally distributed," *Jour. Roy. Stat. Soc. Suppl.*, Vol. 7 (1941), pp. 155–161.
[12] Arne Fisher, *The Mathematical Theory of Probabilities*, Second edition, New York, 1930.
[13] R. A. Fisher and F. Yates, *Statistical Tables*, London, 1938.
[14] L. H. C. Tippett, "Statistical methods in textile research. Part 2, Uses of the binomial and Poisson distributions," *Shirley Inst. Mem.*, Vol. 13 (1934), pp. 35–72.
[15] J. V. Uspensky, *Introduction to Mathematical Probability*, New York, 1937.
[16] C. B. Williams, "The use of logarithms in the interpretation of certain entymological problems," *Annals of Appl. Biol.*, Vol. 24 (1937), pp. 404–414.

[14] A transformation closely related to the logarithmic one is $T = k^{-1} \sinh^{-1}(kX)^{\frac{1}{2}}$, where k is an estimate of the Charlier coefficient of disturbancy of a Poisson distribution. This transformation has recently been studied from an empirical point of view by Beall [2]; it was suggested by the heuristic argument of section 2 applied to the case in which $\sigma^2 = \mu + k\mu^2$. Beall presents evidence that for the particular data which he considered, the transformation seemed to stabilize the variance and normalize. A mathematical theory would follow the lines laid down above in the case of $T = \log(X + \alpha)$.