CORRECTION FOR BIAS INTRODUCED BY A TRANSFORMATION OF VARIABLES[1]

By Jerzy Neyman and Elizabeth L. Scott

University of California, Berkeley

1. Introduction. The problem of "normalizing" transformations has two different facets: one is concerned with the identity of transforming functions suitable for variables following a distribution with a particular shape or properties and the other with the nature of the statistics capable of serving as unbiased estimates in cases where a given transforming function appears to be successful. The literature on the first problem is rich (see, for example, [1], [2] and [8]). This paper is concerned with the second problem. Our purpose is to deduce minimum variance unbiased estimates of the effects of experimental treatments expressed in the original units. The solution is obtained for a broad category of transforming functions.

The estimates of treatment effects expressed in the original units are customarily obtained by the inverse transformation of the estimates in transformed units. As is well known [1], [6], [7], this traditional estimate is biased. Occasionally, this bias is important. Further, the bias gains importance when a number of similar estimates of the same effect, obtained from independent sets of observations, are averaged in order to estimate the average effect. The random errors of the particular effects tend to average out but, in general, not the bias.

2. Statement of the problem. Our basic assumption in this paper is that the transformation used in the analysis of an experiment is faultless so that the transformed variables exactly follow normal distributions with some postulated means and with the same unknown variance σ^2. Generically, these normal variables will be denoted by the letter $\xi(\psi)$ where ψ identifies the expectation of the variable concerned. Thus $\xi(\psi)$ is the transformed variable in the experiment. The variable that is directly observable will be denoted by $X(\psi)$. It will be assumed that

$$X(\psi) = f[\xi(\psi)], \tag{1}$$

where f is a strictly increasing function defined for all real values of its argument. Later on, we shall introduce further limitations on f. It will be noticed that f is the inverse of the function used for transforming the observable variable X. For example, with the square root transformation the function f is the square of its argument.

The problem treated is concerned with a particular pair of variables of the

Received August 8, 1959; revised February 26, 1960.

[1] This paper was prepared with the partial support of the Office of Naval Research (Nonr-222-43). This paper in whole or in part may be reproduced for any purpose of the United States Government.

family considered, namely with $\xi(\mu)$ and $X(\mu)$, where μ is the mean of $\xi(\mu)$ and is a well-defined but unknown number. Specifically, we are concerned with estimating

$$\theta \equiv E[X(\mu)]. \tag{2}$$

Our problem arises when the variables $\xi(\mu)$ and $X(\mu)$ are not directly observable. On the other hand, the variables that are observable in the given experiment yield a pair of statistics, $\hat{\mu}$ and S^2, mutually independent and jointly sufficient for μ and σ^2. The first is a normal variable with mean μ and variance $\lambda^2\sigma^2$, where λ^2 is a known number. The second statistic, S^2, is the residual sum of squares and, divided by σ^2, is distributed as χ^2 with a certain number ν of degrees of freedom. Our problem is to devise a function, say $\hat{\theta}(\hat{\mu}, S^2)$ such that its expectation equals θ identically in μ and σ^2. Because of the familiar result of Lehmann and Scheffé [5] that the sufficient system of statistics $(\hat{\mu}, S^2)$ is boundedly complete, it follows that the function $\hat{\theta}(\hat{\mu}, S^2)$ is unique and is the minimum variance unbiased estimate of θ.

Before proceeding to the construction of the estimate $\hat{\theta}(\hat{\mu}, S^2)$ we give two illustrative examples.

3. Example 1: An experiment in randomized blocks. Denote by α and β two unknown parameters capable of assuming values within a certain open set, and by $\xi(\alpha, \beta)$ a normal random variable with expectation

$$E[\xi(\alpha, \beta)] = \alpha + \beta \tag{3}$$

and with a fixed variance σ^2. Correspondingly,

$$X(\alpha, \beta) = f[\xi(\alpha, \beta)]. \tag{4}$$

A randomized block experiment will yield particular values of mn independent random variables $X(\alpha_i, \beta_j)$ for $i = 1, 2, \cdots, m$ and $j = 1, 2, \cdots, n$, with $\sum\beta_j = 0$. Here the β's represent the familiar block effects and α_i stands for the "transformed effect" of the ith treatment in the hypothetical average conditions of the experiment. The analysis of the experiment ordinarily involves the estimation in original units (pounds, inches, number of surviving insects, etc.) of the effect of the ith treatment if it were applied in the average conditions of the experiment. In order that this estimate can be conveniently combined, by averaging, with similar estimates derived from other experiments involving the same treatment, the estimate sought should be unbiased. The quantity to estimate[2] is, then,

$$\theta = E[X(\alpha_i, 0)] = E\{f[\xi(\alpha_i, 0)]\} \tag{5}$$

where $\xi(\alpha_i, 0)$ is a normal variable with an unknown mean $\alpha_i = \mu$ and with

[2] Of course, the definition of the "effect of the ith treatment in the average conditions of the experiment" by means of Formula (5) is not the only possible definition of this concept. An alternative definition might be the average over j of the quantities $E\{f[\xi(\alpha_i, \beta_j)]\}$.

variance σ^2. While $\xi(\alpha_i, 0)$ and $X(\alpha_i, 0)$ are not directly observable, the experiment yields an estimate of μ, namely,

$$\hat{\mu} = \frac{1}{n} \sum_{j=1}^{n} \xi_{ij}, \tag{6}$$

which, according to traditional theory, is normal with mean $\mu = \alpha_i$ and variance

$$\sigma_{\hat{\mu}}^2 = \lambda^2 \sigma^2 = \sigma^2/n. \tag{7}$$

Also, in the usual notation,

$$S^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} (\xi_{ij} - \xi_{i.} - \xi_{.j} + \xi_{..})^2 \tag{8}$$

is the sum of squares of the residuals which, combined with $\hat{\mu}$, forms a sufficient system of statistics for μ and σ^2. It is independent of $\hat{\mu}$, and is distributed as the product of σ^2 by a χ^2 with $\nu = (m-1)(n-1)$ degrees of freedom. Our problem is to devise a function, $\hat{\theta}(\hat{\mu}, S^2)$, which is an unbiased estimate of θ.

4. Example 2: Regression analysis of a randomized cloud seeding experiment. An experiment is performed to check whether the "seeding" of clouds, intended to increase the precipitation in a "target area" T, has an effect. Also, it is intended to estimate the amount of this effect measured in inches of actual precipitation. A certain number $s \geqq 1$ of adjoining areas, presumed to be unaffected by seeding, are used as controls. We shall use the symbol X_i to denote the rainfall from a particular storm falling in the ith control area and the vector symbol $X = (X_1, X_2, \cdots, X_s)$ to denote the precipitation from the same storm in all the controls. For each X we consider the random variable $Y(X)$ representing the target precipitation in conditions when the precipitation in the controls is X and there is no seeding. All these variables are measured in inches.

Now suppose that a storm, with control precipitation equal to X', is seeded and yields Y' inches of rain in the target. In order to estimate the effect of this seeding it is necessary to have an estimate of the rain which would have fallen in the target from the same storm if there were no seeding. In other words, we need an estimate of $E[Y(X')] = \theta$. The hypotheses usually made about the variables X and $Y(X)$ are that, by means of some suitable change of scale, $X_i = f(\xi_i)$, etc., they can be replaced by transformed variables ξ and $\eta(\xi)$, respectively, such that, for each ξ, the variable $\eta(\xi)$ is (approximately) normally distributed with a mean

$$E[\eta(\xi)] = \alpha_0 + \sum_{i=1}^{s} \alpha_i \xi_i, \tag{9}$$

where the α's are unknown constants, and with a variance σ^2 independent of ξ. With these assumptions, the quantity θ to be estimated is

$$\theta = E\{f[\eta(\xi')]\}, \tag{10}$$

where ξ' stands for the transformed value of X'. Denote by μ the expectation of $\eta(\xi')$,

$$\mu = \alpha_0 + \sum_{i=1}^{s} \alpha_i \xi_i', \tag{11}$$

and by $\hat{\mu}$ its minimum variance unbiased estimate obtained from the regression analysis of a random sample of unseeded storms. The same analysis provides the sum S^2 of squares of residuals, which is stochastically independent of $\hat{\mu}$ and, when divided by σ^2, is distributed as χ^2 with a certain number of degrees of freedom ν. The variance of $\hat{\mu}$ is $\lambda^2(\xi')\sigma^2$, where and factor $\lambda^2(\xi')$ depends upon the value of ξ' and, in fact, grows without limit when ξ' diverges from the average $\bar{\xi}$ of the control precipitation from the nonseeded storms used to evaluate $\hat{\mu}$.

Our problem consists in determining a function $\hat{\theta}(\hat{\mu}, S^2)$ such that

$$E[\hat{\theta}(\hat{\mu}, S^2)] \equiv \theta. \tag{12}$$

The difference between Y' and $\hat{\theta}(\hat{\mu}, S^2)$ is the estimated effect of seeding, expressed in inches.

5. Method and auxiliary formulas. Since the normalizing transformations are supposed to amount to a change in scale of measuring the observable random variables, it is natural to assume that the function f determining the observable random variable X in terms of the normal variable ξ is fairly regular. Our method, to be termed the expansion method, of constructing $\hat{\theta}(\hat{\mu}, S^2)$ is limited to the case where (i) $\theta = E\{f[\xi(\mu)]\}$ exists, (ii) f is an entire function, and (iii) the expectation θ may be obtained by taking expectations, term by term, of the Taylor expansion of f, so that

$$\theta = E\{f[\xi(\mu)]\} = f(0) + \sum_{n=1}^{\infty} \frac{1}{n!} f^{(n)} E[\xi^n(\mu)], \tag{13}$$

where $f^{(n)}$ stands for the nth derivative of f evaluated at zero. Then, for each n, we determine a homogeneous combination

$$T_n = \sum_{k=0}^{n} A_{n,k} \hat{\mu}^k S^{n-k} \tag{14}$$

such that

$$E(T_n) = E[\xi^n(\mu)], \tag{15}$$

and show that

$$\hat{\theta}(\hat{\mu}, S^2) = f(0) + \sum_{n=1}^{\infty} \frac{1}{n!} f^{(n)} T_n \tag{16}$$

is the solution of the problem.

Also, for functions f of a particular family, we give an alternative easy method of constructing $\hat{\theta}(\hat{\mu}, S^2)$. Before proceeding we must recall certain formulas and deduce certain bounds.

For every $m = 1, 2, \cdots$, we have

$$E[S^{2m}] = (2\sigma^2)^m \Gamma(\tfrac{1}{2}\nu + m)/\Gamma(\tfrac{1}{2}\nu). \tag{17}$$

Also

$$E[\xi^{2m}(\mu)] = \sum_{k=0}^{m} \frac{(2m)!}{(2k)!(m-k)!} \mu^{2k}(\sigma^2/2)^{m-k}, \tag{18}$$

and it follows that

$$\frac{(2m)!}{m!} (\sigma^2/2)^m \leqq E[\xi^{2m}(\mu)] \leqq \frac{(2m)!}{m!} [(\mu^2 + \sigma^2)/2]^m. \tag{19}$$

Similarly

$$E[\xi^{2m+1}(\mu)] = \mu \sum_{k=0}^{m} \frac{(2m+1)!}{(2k+1)!(m-k)!} \mu^{2k}(\sigma^2/2)^{m-k}. \tag{20}$$

In order to obtain convenient bounds on $E(|\xi^{2m+1}|)$, we notice first that

$$|\mu| \frac{(2m+1)!}{m!} (\sigma^2/2)^m \leqq |E[\xi^{2m+1}(\mu)]| < E|\xi^{2m+1}(\mu)|. \tag{21}$$

Further, by Schwarz' inequality and because of (19),

$$\begin{aligned} E|\xi^{2m+1}(\mu)| &< \{E[\xi^2(\mu)]E[\xi^{4m}(\mu)]\}^{\frac{1}{2}} \\ &\leqq (\mu^2 + \sigma^2)^{\frac{1}{2}} \left(\frac{(4m)!}{(2m)!}\right)^{\frac{1}{2}} [(\mu^2 + \sigma^2)/2]^m. \end{aligned} \tag{22}$$

However, it is easy to see that

$$\frac{m!}{2^m(2m+1)!} \left(\frac{(4m)!}{(2m)!}\right)^{\frac{1}{2}} = \left\{ \prod_{k=1}^{m} \frac{(4k-3)(4k-1)}{(4k+2)^2} \right\}^{\frac{1}{2}} < 1. \tag{23}$$

Consequently, we may write

$$|\mu| \frac{(2m+1)!}{m!} (\sigma^2/2)^m < E|\xi^{2m+1}(\mu)| < \frac{(2m+1)!}{m!} [\mu^2 + \sigma^2]^m, \tag{24}$$

for all m.

6. Term by term evaluation of the expectation of a Taylor series. In this section we use the bounds found in Section 5 in order to prove certain theorems.

THEOREM 1. *In order that the series in the right-hand side of* (13) *be convergent irrespective of the values of* μ *and* σ^2, *it is necessary and sufficient that the radii of convergence of the two series*

$$\sum_{n=0}^{\infty} \frac{1}{n!} f^{(2n)} z^n \quad \text{and} \quad \sum_{n=0}^{\infty} \frac{1}{n!} f^{(2n+1)} z^n \tag{25}$$

both be infinite, so that

$$\lim_{n\to\infty}\left(\frac{1}{n!}\,|f^{(2n)}|\right)^{1/n} = \lim_{n\to\infty}\left(\frac{1}{n!}\,|f^{(2n+1)}|\right)^{1/n} \tag{26}$$
$$= \lim_{n\to\infty}\frac{1}{n}\,(|f^{(2n)}|)^{1/n} = \lim_{n\to\infty}\frac{1}{n}\,(|f^{(2n+1)}|)^{1/n} = 0.$$

It will be noticed that conditions (26) are stronger than the assumption that the Taylor expansion of f,

$$f(\xi) = \sum_{n=0}^{\infty}\frac{1}{n!}f^{(n)}\xi^n, \tag{27}$$

is convergent for all real ξ. In fact, the conditions necessary and sufficient for this to happen may be written as

$$\lim_{n\to\infty}\frac{1}{n^2}\,(f^{(2n)})^{1/n} = \lim_{n\to\infty}\frac{1}{n^2}\,(f^{(2n+1)})^{1/n} = 0. \tag{28}$$

Thus, if the radii of convergence of (25) are infinite, then that of (27) will also be infinite, but the converse is not necessarily true.

In order to prove the therem we simply notice that inequalities (19) and (24) imply that, for all m,

$$\frac{1}{m!}\,|f^{(2m)}|\,(\sigma^2/2)^m \leqq \frac{1}{(2m)!}\,|f^{(2m)}|\,E[\xi^{2m}(\mu)] \leqq \frac{1}{m!}\,|f^{(2m)}|\,[(\mu^2+\sigma^2)/2]^m, \tag{29}$$

$$\frac{1}{m!}\,|\mu f^{(2m+1)}|\,(\sigma^2/2)^m \leqq \frac{1}{(2m+1)!}\,|f^{(2m+1)}E[\xi^{2m+1}(\mu)]| \tag{30}$$
$$\leqq \frac{1}{m!}\,|f^{(2m+1)}|\,[\mu^2+\sigma^2]^m.$$

If we assume that the series (13) is convergent for all values of μ and σ^2, then this will imply that the middle terms in (29) and (30) tend to zero as $m \to \infty$. In turn, this implies that, for all σ^2 and for sufficiently large m,

$$\left(\frac{1}{m!}\,|f^{(2m)}|\right)^{1/m} < \frac{2}{\sigma^2}, \qquad \left(\frac{1}{m!}\,|f^{(2m+1)}|\right)^{1/m} < \frac{2}{\sigma^2}, \tag{31}$$

which is equivalent to (26). On the other hand, if we assume that conditions (26) are satisfied, then the two series (25) are absolutely convergent for all values of the argument and the right inequalities (29) and (30) imply absolute convergence of (13).

THEOREM 2. *Under the conditions of Theorem* 1, *that is, under conditions* (26),

$$\theta = E\{f[\xi(\mu)]\} = f(0) + \sum_{n=1}^{\infty}\frac{1}{n!}f^{(n)}E[\xi^n(\mu)] \tag{32}$$

for all μ *and* σ^2.

In other words, if conditions (26) are satisfied then the expectation of f can be obtained by taking expectations term by term of the Taylor expansion of this function.

Theorem 2 is implied by inequality (24) showing that in the middle part of Formula (30) the expectation of $\xi^{2m+1}(\mu)$ may be replaced by the expectation of the absolute value $|\xi^{2m+1}(\mu)|$.

For convenience of reference we shall adopt the following definition.

DEFINITION. *An entire function f is called of second order if it satisfies conditions* (26).

It will be seen that every indefinitely differentiable function whose derivatives at a particular point are bounded is necessarily a second order entire function. Also, the sum of two second order entire functions is itself a second order entire function.

7. Lack of complete generality of the expansion method. At this point it may be interesting to indicate a purely mathematical formulation of the general problem treated in this paper. This is as follows

For a given positive integer ν, for a given positive number λ^2 and for a given function f, defined on the real line and such that

$$\int_{-\infty}^{+\infty} |f(x)| e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} dx < +\infty \tag{33}$$

for all real μ and for $\sigma > 0$, to determine a function of two arguments $\hat{\theta}(x, y^2)$, independent of μ and σ, such that

$$\begin{aligned} &\lambda 2^{(\nu-2)/2}\sigma^\nu \Gamma(\tfrac{1}{2}\nu) \int_{-\infty}^{+\infty} f(x) e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} dx \\ &\qquad = \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x-\mu)^2/\lambda^2\sigma^2} \int_0^\infty y^{\nu-1} e^{-\frac{1}{2}y^2/\sigma^2} \hat{\theta}(x, y^2)\, dy\, dx, \end{aligned} \tag{34}$$

identically in μ and $\sigma > 0$.

The expansion method provides the solution of this problem when the function f is entire of second order. However, it is easy to construct entire functions f satisfying (33) which are not of the second order. One example is $\exp\{-x^2\}$. To such functions the expansion method is not applicable and we are not certain whether the solution of equation (34) exists.

8. Minimum variance unbiased estimate of the expectation of a second order entire function. From now on we shall deal exclusively with functions $f(\xi)$ which are entire of the second order.

Let $\hat{\mu}$ be a normal variable with expectation μ and variance $\lambda^2\sigma^2$ where λ^2 is a known number. Also, let S^2 be independent of $\hat{\mu}$ and such that S^2/σ^2 is distributed as χ^2 with ν degrees of freedom. Finally, for $n = 0, 1, 2, \cdots$,

$$T_{2n} = \sum_{k=0}^{n} \frac{(2n)!}{(2k)!(n-k)!} \hat{\mu}^{2k}[\tfrac{1}{4}S^2(1-\lambda^2)]^{n-k} \frac{\Gamma(\frac{1}{2}\nu)}{\Gamma(\frac{1}{2}\nu + n - k)} \tag{35}$$

and

$$(36)\quad T_{2n+1} = \sum_{k=0}^{n} \frac{(2n+1)!}{(2k+1)!(n-k)!}\, \hat{\mu}^{2k+1}[\tfrac{1}{4}S^2(1-\lambda^2)]^{n-k}\, \frac{\Gamma(\frac{1}{2}\nu)}{\Gamma(\frac{1}{2}\nu+n-k)}\,.$$

By direct computation it is easy to verify that, for every $m = 1, 2, \cdots$,

$$(37)\qquad E(T_m) \equiv E[\xi^m(\mu)].$$

THEOREM 3. *If f is a second order entire function, then*

$$(38)\qquad \theta(\hat{\mu}, S^2) = f(0) + \sum_{m=1}^{\infty} \frac{1}{n!} f^{(m)} T_m$$

is convergent for all values of $\hat{\mu}$ and S^2 and is an unbiased estimate of $\theta = E\{f[\xi(\mu)]\}$.

Comparing (35) and (36) with (18) and (20), noticing that

$$(39)\qquad \Gamma(\tfrac{1}{2}\nu)/\Gamma(\tfrac{1}{2}\nu+n-k) \leqq (2/\nu)^{n-k}$$

and referring to (19), we find that

$$(40)\qquad |\, T_{2n}\,| < \frac{(2n)!}{n!}\, Y^n$$

and, in a similar manner,

$$(41)\qquad |\, T_{2n+1}\,| < |\, \hat{\mu}\,|\, \frac{2n+1)!}{n!}\, Y^n,$$

with

$$(42)\qquad Y = [\nu\hat{\mu}^2 + S^2(1+\lambda^2)]/2\nu.$$

Because f is a second order entire function, it follows that the series (38) is absolutely convergent for all values of $\hat{\mu}$ and S^2. In order to prove that the expectation of $\hat{\theta}(\hat{\mu}, S^2)$ as defined by (38) can be obtained by taking expectations term by term, it is sufficient to show the convergence of the series obtained from (38) by replacing each T_m by the expectation of its absolute value. This is easily accomplished by noticing that $|\, T_m\,|$ cannot exceed the expression obtained by replacing in formulas (35) and (36) the value of $\hat{\mu}$ by that of $|\,\hat{\mu}\,|$ and $1 - \lambda^2$, which may be negative, by $1 + \lambda^2$. Further computations, similar to those leading to (19) and (24), indicate then that

$$(43)\qquad \sum_{n=1}^{\infty} \frac{1}{n!}\, |\, f^{(n)}\,|\, E\,|\, T_n\,| < +\infty.$$

Because of (37), it follows that $\hat{\theta}(\hat{\mu}, S^2)$ as defined by (38) has the desired property of being an unbiased estimate of θ.

Formula (38) has the advantage of, so to speak, exhausting the method; it provides the minimum variance unbiased estimate of θ whatever the second order entire function f may be. This generality is paid for by the complexity of

the solution provided by (38). In the next section we give a somewhat simpler formula for $\hat{\theta}(\hat{\mu}, S^2)$ which is applicable when the function f satisfies a certain differential equation.

9. Alternative solution applicable to recursive type second order entire functions. We shall say that the function f is of recursive type if it satisfies the second order differential equation

$$f''(x) = A + Bf(x), \tag{44}$$

where A and B are arbitrary constants. However, in order to eliminate the trivial case where f is linear, we shall assume that at least one of the constants differs from zero. It is easy to verify that every recursive function is necessarily a second order entire function. This section will be limited to consideration of recursive type functions f. We shall be particularly interested in the expectation of their Taylor expansion about the point μ. Because the odd central moments of the normal variable are all equal to zero, we shall be concerned only with the derivatives of f of even order. We have, for all n,

$$f^{(2n)}(x) = AB^{n-1} + B^n f(x) \tag{45}$$

and, if $B \neq 0$,

$$\begin{aligned} \theta = E\{f[\xi(\mu)]\} &= f(\mu) + \sum_{n=1}^{\infty} \frac{1}{n!} [AB^{n-1} + B^n f(\mu)](\sigma^2/2)^n \\ &= f(\mu)e^{B\sigma^2/2} + \frac{A}{B}(e^{B\sigma^2/2} - 1). \end{aligned} \tag{46}$$

Alternatively, if $B = 0$, that is, if f is quadratic,

$$\theta = f(\mu) + A\sigma^2/2. \tag{47}$$

Similarly, for $B \neq 0$,

$$E[f(\hat{\mu})] = f(\mu)e^{B\lambda^2\sigma^2/2} + (A/B)(e^{B\lambda^2\sigma^2/2} - 1) \tag{48}$$

and, for $B = 0$,

$$E[f(\hat{\mu})] = f(\mu) + A\lambda^2\sigma^2/2. \tag{49}$$

Eliminating $f(\mu)$ from (46) and (48) and from (47) and (49), we find

$$\theta = e^{B(1-\lambda^2)\sigma^2/2}E[f(\hat{\mu})] + (A/B)[e^{B(1-\lambda^2)\sigma^2/2} - 1] \qquad \text{for } B \neq 0 \tag{50}$$

and

$$\theta = E[f(\hat{\mu})] + A(1 - \lambda^2)\sigma^2/2 \qquad \text{for } B = 0. \tag{51}$$

The last formula indicates that, when $B = 0$, the minimum variance unbiased estimate of θ is given by

$$\hat{\theta}(\hat{\mu}, S^2) = f(\hat{\mu}) + A(1 - \lambda^2)S^2/2\nu. \tag{52}$$

If $B \neq 0$ then, in order to obtain $\hat{\theta}(\hat{\mu}, S^2)$, it is sufficient to determine a function, say $\Phi(aS^2, \nu)$, independent of $\hat{\mu}$, such that its expectation equals $\exp(a\sigma^2/2)$. Taking into account the expansion

$$e^{a\sigma^2/2} = \sum_{n=0}^{\infty} \frac{1}{n!} (a\sigma^2/2)^n, \tag{53}$$

we easily find

$$\begin{aligned} \Phi(aS^2, \nu) &= \sum_{n=0}^{\infty} \frac{1}{n!} \frac{\Gamma(\frac{1}{2}\nu)}{\Gamma(\frac{1}{2}\nu + n)} (aS^2/4)^n, \\ &= \left(\frac{2}{S\sqrt{a}}\right)^{\frac{1}{2}\nu - 1} \Gamma(\tfrac{1}{2}\nu) I_{\frac{1}{2}\nu - 1}(S\sqrt{a}) \end{aligned} \tag{54}$$

where $I_\alpha(x)$ is the Bessel function of imaginary argument.

It follows from (50) that

$$\hat{\theta}(\hat{\mu}, S^2) = \Phi[B(1 - \lambda^2)S^2, \nu][f(\hat{\mu}) + (A/B)] - (A/B), \tag{55}$$

which is the general formula for the minimum variance unbiased estimate of θ corresponding to the case where f is a recursive function. It will be seen that, generally, $\hat{\theta}(\hat{\mu}, S^2)$ is a linear function of the traditional estimate $f(\hat{\mu})$ of θ, with coefficients depending upon S^2, λ^2 and ν. If $\lambda^2 = 1$, that is, if the variance of $\hat{\mu}$ coincides with that of $\xi(\mu)$, then $\hat{\theta}(\hat{\mu}, S^2) \equiv f(\hat{\mu})$. Otherwise $f(\hat{\mu})$ is biased. In the particular case $B = 0$, the correction for bias is additive, as indicated in (52). This makes the square root transformation very convenient (provided, of course, it provides effective normalization!) in dealing with balanced experiments in which the quantities to be estimated are differences of certain averages, the estimates of which all have the same variances.

If $A = 0$ but $B \neq 0$, then the correction for bias in $f(\hat{\mu})$ is multiplicative. Finally, if both A and B differ from zero, we have a combination of a multiplicative and an additive correction.

The importance of bias in the traditional estimate of θ may be evaluated by solving for $E[f(\hat{\mu})]$ equations (50) and (51). We have

$$E[f(\hat{\mu})] = [\theta + (A/B)]e^{-B(1-\lambda^2)\sigma^2/2} - (A/B), \qquad \text{for } B \neq 0 \tag{56}$$

and

$$E[f(\hat{\mu})] = \theta - A(1 - \lambda^2)\sigma^2/2, \qquad \text{for } B = 0. \tag{57}$$

10. Some particular cases. In this section we use the general results of Section 9 to deduce particular formulas. We obtain the minimum variance unbiased estimate of θ and the expectation of $f(\hat{\mu})$ referring to four particular normalizing transformations: (i) the square root transformation, (ii) the logarithmic transformation, (iii) the angular transformation and (iv) the hyperbolic sine transformation.

(i) In the case of the square root transformation, the transformed variable

$$\xi = (X - a)^{\frac{1}{2}} \tag{58}$$

where a is a known constant. We ignore the ambiguity connected with the fact that, for ξ to be a normal variable it must be capable of assuming negative values. The function f is

$$X = f(\xi) = \xi^2 + a. \tag{59}$$

Obviously this is a recursive function with $A = 2$, $B = 0$. Consequently, formula (52) yields directly

$$\hat{\theta}(\hat{\mu}, S^2) = f(\hat{\mu}) + (1 - \lambda^2)S^2/\nu = \hat{\mu}^2 + a + (1 - \lambda^2)S^2/\nu. \tag{60}$$

The bias of the traditional estimate $f(\hat{\mu}) = \hat{\mu}^2 + a$ is obtained from (57), namely,

$$E[f(\hat{\mu})] = \theta - (1 - \lambda^2)\sigma^2. \tag{61}$$

Hence, unless $\lambda^2 \geqq 1$, so that the variance of $\hat{\mu}$ is at least equal to that of $\xi(\mu)$, the use of $f(\mu)$ as an estimate systematically underestimates θ. Furthermore, the better the estimate $\hat{\mu}$, that is, the smaller the value of λ^2, the greater the bias.

(ii) In the case of logarithmic transformation, we have

$$\xi = \log_{10} X \tag{62}$$

and hence

$$X = f(\xi) = 10^{\xi} = e^{m\xi}, \quad \text{say.} \tag{63}$$

Here again the function f is of recursive type with $A = 0$ and $B = m^2$. Formula (55) gives

$$\begin{aligned} \hat{\theta}(\hat{\mu}, S^2) &= \Phi[m^2(1 - \lambda^2)S^2, \nu]f(\hat{\mu}) \\ &= \Phi[m^2(1 - \lambda^2)S^2, \nu]10^{\hat{\mu}}. \end{aligned} \tag{64}$$

Substituting $A = 0$ and $B = m^2$ into (56) we obtain

$$E[f(\hat{\mu})] = E10^{\hat{\mu}} = \theta e^{-m^2(1-\lambda^2)\sigma^2/2}. \tag{65}$$

Thus, with the logarithmic transformation, the bias of the traditional estimate is multiplicative. If the variance of $\hat{\mu}$ is less than σ^2 then the use of $10^{\hat{\mu}}$ will systematically underestimate θ and vice versa. The bias grows with increasing $|1 - \lambda^2|$.

(iii) With angular transformation we have

$$\xi = \arcsin \sqrt{X} \tag{66}$$

and

$$X = f(\xi) = \sin^2 \xi = \tfrac{1}{2}(1 - \cos 2\xi). \tag{67}$$

Here again the function f is of recursive type with $A = 2$ and $B = -4$. Hence, Formula (55) gives

$$\hat{\theta}(\hat{\mu}, S^2) = \Phi[4(\lambda^2 - 1)S^2, \nu]\,(\sin^2 \hat{\mu} - \tfrac{1}{2}) + \tfrac{1}{2}. \tag{68}$$

Substituting $A = 2$ and $B = -4$ in (56) we have

$$E[f(\hat{\mu})] = E[\sin^2 \hat{\mu}] = (\theta - \tfrac{1}{2})e^{2(1-\lambda^2)\sigma^2} + \tfrac{1}{2}. \tag{69}$$

It is seen that, if the variance of $\hat{\mu}$ is less than σ^2, the traditional estimate $\sin^2 \hat{\mu}$ is systematically "too far" from $\frac{1}{2}$. If the true value of $\theta < \frac{1}{2}$, then $\sin^2 \hat{\mu}$ will tend to underestimate θ. Otherwise, if $\theta > \frac{1}{2}$, there will be a tendency to overestimate θ. With $\lambda > 1$ these two tendencies will be reversed.

(iv) The last transformation to be considered here is based on the function

$$X = f(\xi) = \sinh^2 \xi = \tfrac{1}{2}[\cosh 2\xi - 1]. \tag{70}$$

It is of recursive type with $A = 2$ and $B = 4$. Hence, from Formula (55)

$$\hat{\theta}(\hat{\mu}, S^2) = \Phi[4(1 - \lambda^2)S^2, \nu](\sinh^2 \hat{\mu} + \tfrac{1}{2}) - \tfrac{1}{2}. \tag{71}$$

Formula (57) with the indicated values of A and B gives

$$E[f(\hat{\mu})] = E[\sinh^2 \hat{\mu}] = (\theta + \tfrac{1}{2})e^{-2(1-\lambda^2)\sigma^2} - \tfrac{1}{2}. \tag{72}$$

In this case $f(\hat{\mu})$ underestimates or overestimates θ according to whether λ is less or greater than unity.

11. Concluding remarks. (i) Formula (54) defining Φ may seem complicated. In fact, the series on the right converges fairly rapidly so that sufficient accuracy is obtained with only a few terms.

It is easy to check that Formula (54) may be rewritten as follows

$$\Phi(aS^2, \nu) = \sum_{n=0}^{\infty} \frac{(aS^2/2)^n}{n! \prod_{k=1}^{n} (\nu + 2k - 2)} = \sum_{n=0}^{\infty} \frac{(a\hat{\sigma}^2/2)^n}{n! \prod_{k=1}^{n} (1 + \frac{(2k-2)}{\nu}} \tag{73}$$

where

$$\hat{\sigma}^2 = S^2/\nu \tag{74}$$

is the unbiased estimate of σ^2. It will be seen that, for $n > 1$, the absolute value of each term on the right is less than the corresponding term in the expansion of $\exp\{a\hat{\sigma}^2/2\}$. In other words, the series (73) converges faster than the series for the exponential function.

(ii) In some circumstances, the practical statistician may decide to work on the assumption that the vairance σ^2 is known. In order to adjust the formulas deduced in this paper to this case, it is sufficient to replace $\hat{\sigma}^2$ by σ^2 and pass to the limit as $\nu \to \infty$. In particular, this procedure reduces the right hand side of (73) to $\exp\{a\sigma^2/2\}$, which is (53).

(iii) Formulas have been published for correcting the bias introduced by the transformation of variables in some particular cases in [1], [6] and [7], for example. However, these formulas do not agree with ours.

(iv) It is a pleasure to express our indebtedness to the referee who picked up several mistakes in the original text of the paper and, in addition, called our

attention to the important publications [3], [4] and [9], which we overlooked. Among the problems treated in these papers, there is one which is strongly related to ours. In the present notation, this problem consists in finding the function $h(\hat{\mu}, S^2, \nu)$ that is an unbiased estimate of a given function $g(\mu, \sigma^2)$. This problem is treated under the restriction that $\lambda^2 = 1/(\nu + 1)$. Apart from this restriction, in order to reduce our problem to the problem just described, it is sufficient to evaluate the expectation $E\{f[\xi(\mu)]\}$ and to denote the result by $g(\mu, \sigma^2)$. The difference between our results and those in [3], [4] and [9] consists first in a difference in the method and in the conditions of the various theorems: in the earlier papers the conditions are expressed in terms of the function g whereas, in the present paper, they refer to the function f. Also, explicit formulas for the unbiased estimates of θ, as given here, are not contained in the papers quoted.

REFERENCES

[1] F. J. Anscombe, "The transformation of Poisson, binomial, and negative-binomial data," *Biometrika*, Vol. 35 (1948), pp. 246–254.

[2] M. S. Bartlett, "The use of transformations," *Biometrics*, Vol. 3 (1947), pp. 39–52.

[3] Tosio Kitagawa, "The operational calculus and the estimations of functions of parameter admitting sufficient statistics," *Bull. Math. Stat.*, Vol. 6 (1956), pp. 95–108.

[4] A. N. Kolmogoroff, "Unbiased estimates," *Izvestiya Akademii Nauk SSSR, Ser. Mat.*, Vol. 14 (1950), pp. 303–326; see also *Am. Math. Soc. Translation* No. 98.

[5] E. L. Lehmann and H. Scheffé, "Completeness, similar regions and unbiased estimation, Part I," *Sankhyā*, Vol. 10 (1950), pp. 305–340.

[6] H. A. Meyer, "A correction for systematic error occurring in the application of the logarithmic volume equation," *J. Series Penn. Agric. Exp. Stat.*, Paper 1058 (1941), pp. 1–3.

[7] M. H. Quenouille, *The Design and Analysis of Experiment*, Hafner, New York, 1953.

[8] John W. Tukey, "On the comparative anatomy of transformations," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 602–632.

[9] Yasutoshi Washio, Haruki Morimoto, and Nobuyuki Ikeda, "Unbiased estimation based on sufficient statistics," *Bull. Math. Stat.*, Vol. 6 (1956), pp. 69–93.