ON THE COMPARATIVE ANATOMY OF TRANSFORMATIONS[1]

By John W. Tukey

Princeton University

1. Summary. The attention of statisticians has usually been focussed on single transformations, rather than on families of transformations. With a growing appreciation of the advantages of examining the behavior of data or approximations over whole families of transformations (Moore and Tukey [2], Anscombe and Tukey [1]), there arises a need for rationally planned charts for representing families of transformations.

The contributions which (i) the topology of the family and (ii) a definition of the strength of a transformation can make to charting are studied in general and applied to the charting of the simple family of transformations. This family is defined to include all transformations of the form

$$y \text{ is replaced by } z = (y + c)^p$$

and all their limits. It thus includes $z = \log (y + c)$, $z = e^{my}$ and the special case

$$z = \begin{cases} 0, & y = y_{\min}, \\ 1, & \text{otherwise,} \end{cases}$$

where $y_{\min}$ is the least value of y either (i) present in the data or (ii) possible, as well as all linear transformations of these transformations.

Experience having shown that transformations with $p \leqq 1$ are much more frequently useful than any others, the charts developed, presented, and exemplified here are restricted to the part of the simple family—its central region—for which $p \leqq 1$. Separate charts are presented for two cases which should cover most cases which arise in practice:

(1) Where, as with counted data and small counts, the least reasonable value for $y + c = 0$, and this value is likely to occur;
(2) Where $y + c$ is always safely > 0, and the range of y is through not many powers of 10.

2. Introduction. In the statistics of today, transformations seem to have two sorts of uses:

(1) providing approximations for theoretical purposes or general convenience;
(2) aiding in the analysis of data by bending the data nearer the Procrustean bed of the assumptions underlying conventional analyses.

Both are interesting and important. In both the quality of the work of the transformation is often judged by the numerical value of some suitable criterion, and transformations which make the value of this criterion sufficiently near some ideal value are acceptable. (Examples follow below.) If we are to consider any

Received July 28, 1955.

[1] Prepared in connection with research sponsored by the Office of Naval Research.

member of a family of transformations as a potentially useful candidate, we should like to understand how one or more criteria vary over the whole family. If we could write a sufficiently simple expression for each criterion in terms of the parameters of the family, we could perhaps reach this understanding algebraically. But this is almost always impossible. We usually have to evaluate each criterion numerically for each of a number of transformations, and then synthesize these few numerical values into an understanding as best we can. If we can find a suitable chart, we can undoubtedly make much more effective use of these few numerical values. Hence the desire for a satisfactory chart.

As one example, note that intermediate statistical texts often prove that, as the number of degrees of freedom $\rightarrow \infty$, the distribution of χ^2 tends to normality. None, to my knowledge, goes on to remark that the same is true of any fixed rational function of χ^2, and hence, in particular, of the results of applying any of the transforms of the simple family. But, knowing this, we are not surprised to find Fisher using the 1/2th power, or Wilson and Hilferty using the 1/3rd power of χ^2 in seeking for a good normal approximation. Nor does the approximate normality of log χ^2 surprise us very much. If, for a given number of degrees of freedom, we wish to select one transformation of χ^2 to approximate normality, our choice would be greatly eased by charts showing, for example, the deviations of selected percentage points of the approximations from the values appropriate to normality.

Notice how disorganized and unrelated is our information on such approximations. The approximate normalization by transformation of a Poisson distribution of average value λ can be quite well accomplished by two apparently unrelated transformations: $z = \sqrt{y}$ and $z = \log(y + \lambda)$. No connection between these two successful approximations seems to have been recognized, although our definition of the strength of a transformation now suggests a reasonably close connection.

In analyzing data which does not match the assumptions of the conventional methods of analysis, we have two choices [1]. We may bend the data to fit the assumptions by making a transformation. Or we may develop new methods of analysis with assumptions which fit the data in its "original" form somewhat better. If we can find a satisfactory transformation, it will almost always be easier and simpler to use it rather than to develop new methods of analysis. To judge of its satisfactoriness, we need a criterion. The precise nature of the criterion will depend on the situation—and on what ills we are trying to remove by transformation: nonadditivity of effect, nonconstancy of variance, nonnormality of distribution, or what have you.

If we seek to remove nonadditivity of effect, for example, we may take as our criterion the t-value for removable nonadditivity. An example has already been discussed by Moore and Tukey [2]. With a few additional values, this example is used below to illustrate one of the suggested chart forms. In treating other ills, other criteria would of course be involved.

3. Strength and local structure of transformations. The strength of transformations is investigated in general in Sections 5 to 7 and applied to the simple family in Sections 8 to 10. The definition of strength to which we are led is much that the strength of

$$z = y^p$$

is naturally taken as $1 - p$, so that the sequence of transformations

$$\begin{aligned} z &= y, \\ z &= \sqrt{y}, \\ z &= \log y, \\ z &= y^{-1/2}, \\ z &= y^{-1}, \\ &\cdots \end{aligned}$$

is an equally spaced sequence of strength 0, 1/2, 1, 3/2, 2, $\cdots$.

If we compare an arbitrary transformation with these unmodified power transformations, we are led to define its *power strength* as

$$\frac{\log \left[\frac{dz}{dy}\right]_{y_1} - \log \left[\frac{dz}{dy}\right]_{y_2}}{\log (y_2/y_1)}.$$

For $z = y^p$ this definition yields $1 - p$ independent of y_1 and y_2 . In general the result will depend on both y_1 and y_2 , and may be thought of as the strength of the unmodified power transformation which resembles the given transformation at $y = y_1$ and $y = y_2$.

The philosophy underlying our exploration and heavy use of the local properties of a family of transformations is discussed in Sections 11 and 12. The topologies and differential structures involved are discussed in Sections 13 and 14. The resulting techniques are applied to the body of the simple family in Sections 15 to 20, and to the corners (near the identity and near $z = \text{sgn}^+ y$) in Sections 21 to 24. (The detailed conclusions are summarized in Sections 20, 24, and 25.)

The results of these analyses are then applied in Sections 25 to 27 to the design of the charts.

4. The charts. As noticed in the Summary, the charts are confined to only part of the simple family, i.e., to $z = (y + c)^p$ with $p \leqq 1$ and $0 \leqq y + c$ and their limits. The restriction to positive $y + c$ is easily understood. It ensures monotone-increasing, well-defined, real-valued functions. The restriction to $p \leqq 1$ has no such mathematical reason.

Its origin is entirely empirical and is basically rooted in the psychology of data-gatherers. Experience seems to show quite uniformly that when transformation helps empirical data, it is almost always for $p \leqq 1$. This is clearly a statement

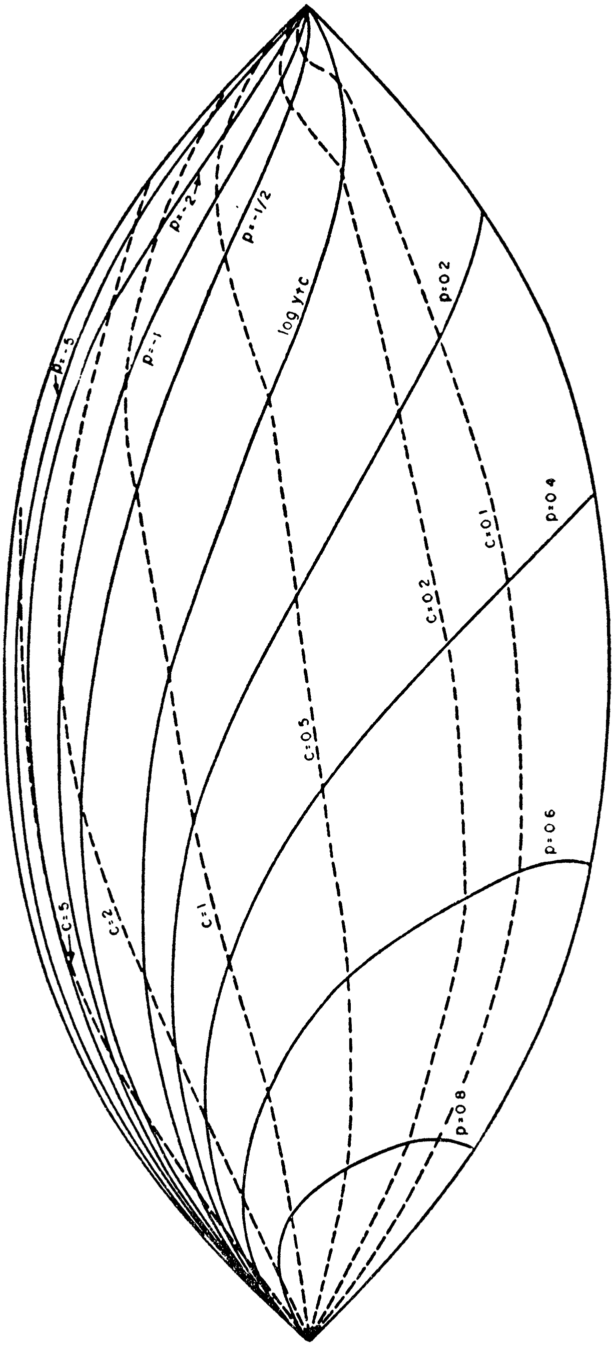

FIG. 1. Basic chart when y is a small count.

about how data-gatherers choose to record (and initially try to analyze) data. For if z is the successfully transformed variable, and if the data-gatherer had chosen to record $w = \sqrt{z}$, then we would have found $z = w^2$ to be successful. The data-gatherer could have done this, but for some reason he does it infrequently. Hence, the restriction to $p \leqq 1$ in the interests of convenience (particularly through the resulting larger scale).

The particular coordinate systems used in these charts are developed in Sections 25 to 27 to meet the requirements developed by studying the topology and strength of the simple family. The choices are by no means unique and do not deserve summarization here. They are believed, however, to be good enough choices to make the charts very useful. (Perhaps even better forms will evolve from practical experience with the presently suggested forms.)

The first chart is intended to apply to counted data with small counts, and to other situations where there is a least value which is not infrequently observed. The chart assumes that the original data has been "coded" (rescaled) so that the least value is zero and that values between 1 and 10 are quite common. The basic chart is shown as Fig. 1. This chart (and other basic charts) can be conveniently used for repeated plotting by a simple device. If a sheet of tracing paper is placed over the chart, and enough of the chart traced to identify locations (for this chart the bounding arcs will suffice), the criterion values can be entered at appropriate points. The tracing paper can then be removed and roughly contoured. (This technique of economizing on special grids was learned from Churchill Eisenhart.) The result of such a tracing for the nonadditivity t in the chinch-bug example discussed by Moore and Tukey [2] is shown as Fig. 2. The numerical entries are values of $10t$ on 20 degrees of freedom, so that the contours correspond roughly to 1%, 5%, and 20% levels.

If a transformation of the simple family renders the effect of treatment and

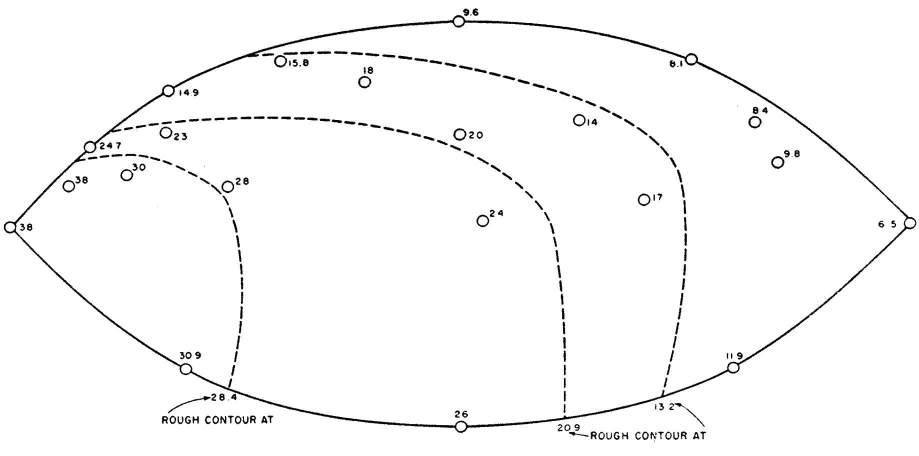

FIG. 2. Use of Fig. 1 to obtain confidence areas. (Entries $10t$, where t measures removable nonadditivity.)

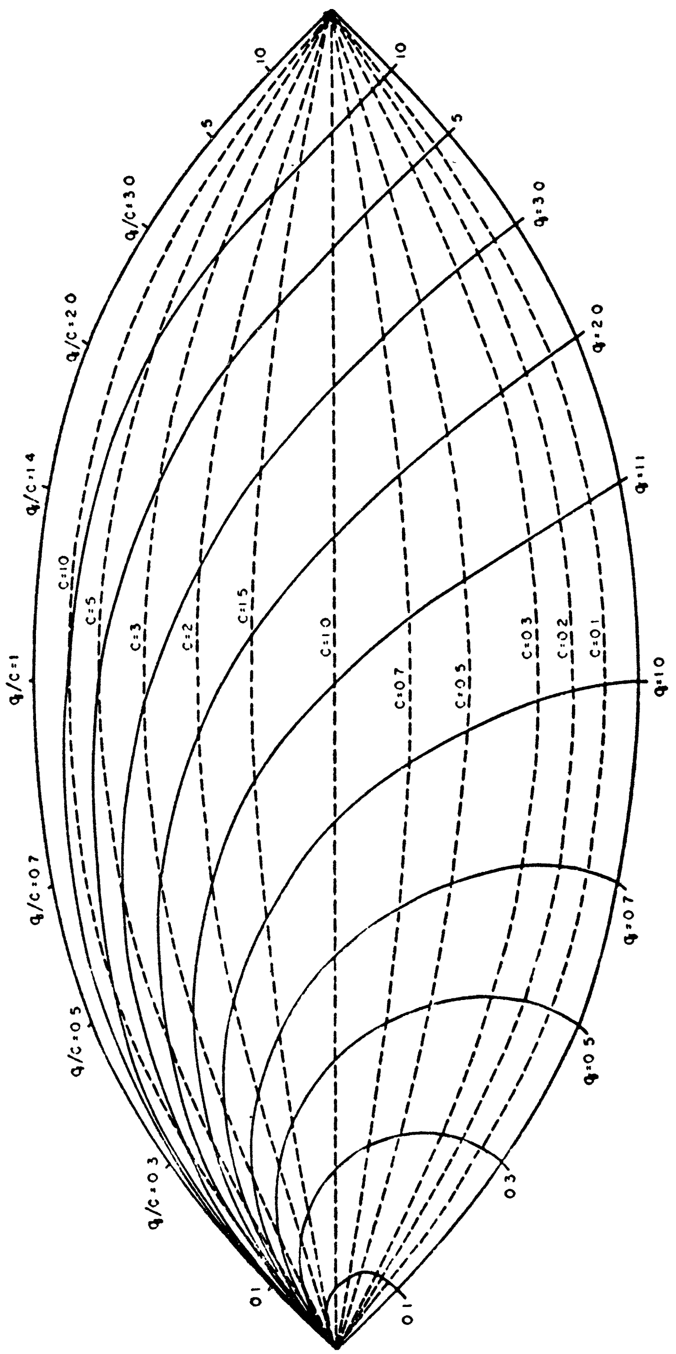

FIG. 3. Basic chart where y is never very large or very small. (Transformation taken as $(y + cy_0{}^{1-qe_0}$.)

replication additive (or at least makes zero the population average of the contrast chosen to indicate removable nonadditivity), then the regions above and to the right of the various contours are confidence regions for this additivity-producing transformation. The confidence coefficients are, of course, 99 %, 95 %, and 80 %. Clearly this single experiment has not narrowed things down too much.

The second chart is intended for use with data ranging through a few powers of 10 and safely away from 0. In using this chart, we think of the simple family in the form

$$z = (y + cy_0)^{1-qe_0},$$

where y_0 and e_0 are constants selected for the particular example. The basic chart is shown as Fig. 3. Again the tracing technique is applicable. In Fig. 4 we show an application to the nonnormality of transformations of χ^2 on 6 degrees of freedom. The values shown are

$$-10 \log_{10} \frac{\text{residual sum of squares}}{\text{mean square for slope}}$$

for the regression of 6 percentage points of χ^2 on the corresponding points for a unit normal. The plot is for $e_0 = 0.5$, $y_0 = 0.6$. The percentage points used were the lower and upper 0.05 %, 0.5 %, and 5 % points. Large entries mean close approximation to normality.

The quality of the approximation can be judged from the case $1 - qe_0 = 0.3$, $c = 0$, for which

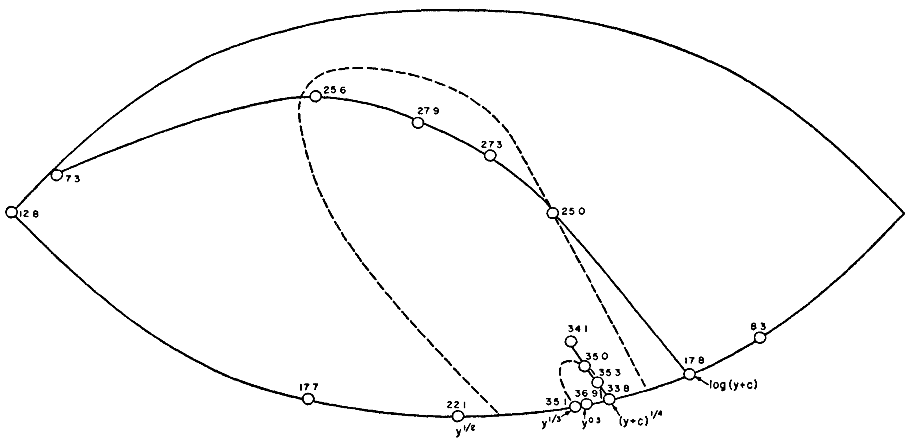

FIG. 4. Use of Fig. 3 to illustrate approximate normality of simple transforms of χ^2_6. (Large entries indicate good approximations. Scaled by $y_0 = 0.6$, $e_0 = 0.5$.)

% point	normality	$3.4300(\chi_6^2)^{0.30} - 5.6490$
lower 0.05%	−3.291	−3.260
lower 0.05%	−2.576	−2.500
lower 5%	−1.645	−1.674
(50%)	(0.000)	(0.023)
upper 5%	1.645	1.685
upper 0.5%	2.576	2.588
upper 0.05%	3.291	3.261

This quality is perhaps worth noting, since the first attempt to provide an example failed in the following way: The attempt was to examine the normality of simple transforms of w_{10}, the range in normal samples of 10, between the upper and lower 0.01% points.

The result was that

$$4.56446(w_{10})^{.487} - 7.824436$$

agrees with a unit normal to within ± 0.01%, which is the apparent accuracy of the available cumulative distribution. The simple family often does quite well in transforming to normality!

I. The Strength of Transformations

5. The purposes of transformations. The analysis of data usually proceeds more easily if

(1) effects are additive;

(2) the error variability is constant;

(3) the error distribution is symmetrical and possibly nearly normal.

The conventional purposes of transformation are to increase the degrees of approximation to which these desirable properties hold. We may think of these as three purposes, but it is often true that a transformation suitable for improving one degree of approximation is also suitable for improving one or both of the others.

The failure of any one of these properties can be recognized by the failure of a difference $y_B - y_A$ to equal another difference $y_D - y_C$, as we shall see shortly. Since $y_B - y_A = y_D - y_C$ is equivalent to $y_C - y_A = y_D - y_B$, we may always assume, and shall assume, that $y_A < y_B \leqq y_C < y_D$. Since the change of scale from y to ay is usually (and wisely) accepted as having no effect on the degree of approximation to any of these three properties, it is better to take the ratio $(y_B - y_A)/(y_D - y_C)$ (or some function of this ratio) as a measure of the failure of this property rather than to take the difference of the differences, $(y_B - y_A) - (y_D - y_C)$, since the latter is not invariant under such changes in scale.

We now observe in what way the differences between $y_B - y_A$ on the one hand and $y_D - y_C$ on the other can exhibit the failure of any one of the three desirable properties.

Nonadditivity of effect occurs when

$$y(u_1, v_2) - y(u_1, v_1) \neq y(u_2, v_2) - y(u_2, v_1),$$

where y is a function of the two variables u and v. Usually $y(u, v)$ represents a typical value (population mean, population median, etc.) of the distribution of values obtained for fixed u and v.

Nonconstancy of variability is exhibited when

$$p(u_1) - q(u_1) \neq p(u_2) - q(u_2),$$

where $p(u_1)$ is the p % point of the distribution of observed values when $u = u_1$ and $q(u_1)$ is the corresponding q % point.

Finally, asymmetry of distribution is exhibited when

$$q(u) - y(u) \neq y(u) - p(u),$$

where $p(u)$ and $q(u)$ are upper and lower percentage points cutting off the same tail areas and $y(u)$ is the median, all of the distribution of values observable for a fixed value of u.

In each case, failure is exhibited by a non-zero quartet. From a theoretical point of view our argument is successfully completed. From an empirical one, there remains a gap. Can we use quartets of individual observed values to study observed data as an indication of failures in the population? If we can, then defining strengths in terms of quartets makes very good empirical as well as theoretical sense.

Clearly if we use observed order statistics to estimate percentage points and medians, we can do this for asymmetry and nonconstancy and, if we are willing to deal with population medians, for nonadditivity as well. Since a transform of a sample mean will not be precisely the mean of the transformed values, we cannot expect exact correspondence for nonadditivity defined for population means. If, however, variability for u and v fixed is not large compared with the changes due to alterations in u and v, this disturbance will be rather unimportant and we can regard the means under specified conditions as pseudo-individual values. Since nonadditivity is most important when such additional variability is not too large, it is almost correct, from a practical point of view, to say that the behavior of quartets of individual or pseudo-individual values describes, empirically and adequately, the apparent failures of additivity, constancy of variance, and symmetry of distribution.

We shall, for reasons which will tend to appear as we proceed, take the logarithm of the ratio of differences

$$\log \frac{y_B - y_A}{y_D - y_C} = \log (y_B - y_A) - \log (y_D - y_C)$$

as the measure of the degree of failure of a quartet to correspond to equal differences—whether nonadditivity of effect, nonconstancy of variance, or asymmetry of distribution is involved. There will usually be many such expressions (many quartets $y_A < y_B \leqq y_C < y_D$) which may be worthy of consideration in a particular problem, and transformations will usually be chosen to make these expressions smaller *as a whole*, balancing a gain on one against a loss on another.

6. The strengths of transformations. How shall we assess the strength of a transformation from y to z? Clearly by the extent to which it alters our measures of dissatisfaction. Given a quartet $y_A < y_B \leqq y_C < y_D$, which is transformed into a quartet $z_A < z_B \leqq z_C < z_D$, we have altered

$$\log \frac{y_B - y_A}{y_D - y_C} \qquad \text{into} \qquad \log \frac{z_B - z_A}{z_D - z_C},$$

and for the initial quartet it is reasonable to measure the strength of the transformation by the change

$$\begin{aligned} k(y_A, y_B; y_C, y_D) &= \log \frac{z_B - z_A}{z_D - z_C} - \log \frac{y_B - y_A}{y_D - y_C} \\ &= \log \frac{(z_B - z_A)(y_D - y_C)}{(z_D - z_C)(y_B - y_A)} \\ &= \log \frac{z_B - z_A}{y_B - y_A} - \log \frac{z_D - z_C}{y_D - y_C}. \end{aligned}$$

The last form of this expression shows that there are natural definitions not only for the results of letting a pair of arguments coalesce

$$k(y_1; y_C, y_D) = k(y_1, y_1; y_C, y_D) = \left[\log \frac{dz}{dy}\right]_{y_1} - \log \frac{z_D - z_C}{y_D - y_C},$$

$$k(y_A, y_B; y_2) = k(y_A, y_B; y_2, y_2) = \log \frac{z_B - z_A}{y_B - y_A} - \left[\log \frac{dz}{dy}\right]_{y_2},$$

but also for the confluent strength arising from the coalescence of both parts

$$k(y_1; y_2) = k(y_1, y_1; y_2, y_2) = \left[\log \frac{dz}{dy}\right]_{y_1} - \left[\log \frac{dz}{dy}\right]_{y_2},$$

all of which involve replacing the difference quotients by their limits, the derivatives of z with respect to y. Clearly we might go one step further, and consider

$$\begin{aligned} k(y) &= \lim_{\Delta \to 0} \frac{k(y; y + \Delta) - k(y, y)}{\Delta} = \lim_{\Delta \to 0} \frac{1}{\Delta} k(y; y + \Delta) \\ &= -\left(\frac{d^2 z}{dy^2}\right) \Big/ \left(\frac{dz}{dy}\right) \end{aligned}$$

as a local measure of the strength of the transformation; however, there are complications which we shall uncover shortly.

With the exception of this last possibility, which we disapprove of, with reasons to be explained, all our measures of the strength of transformations involve at least two arguments (up to as many as four) and the arguments fall into two groups, whose separation (on the y-scale, say) is important. This means that if we face a practical situation, and wish to use a transformation of given strength, we must begin by selecting one or more quartets typical of the data, and then examine the strength of the proposed transformations for that [those] typical quartets.

7. The composition of transformations. Let us consider successive transformations from y to z and from z to w, together with the composed transformation from y to w. The respective measures of strength are, in an obvious notation:

$$k_{yz}(y_A, y_B; y_C, y_D) = \log \frac{z_B - z_A}{y_B - y_A} - \log \frac{z_D - z_C}{y_D - y_C},$$

$$k_{zw}(z_A, z_B; z_C, z_D) = \log \frac{w_B - w_A}{z_B - z_A} - \log \frac{w_D - w_C}{z_D - z_C},$$

$$k_{yw}(y_A, y_B; y_C, y_D) = \log \frac{w_B - w_A}{y_B - y_A} - \log \frac{w_D - w_C}{y_D - y_C}.$$

It is clear that

$$k_{yw}(y_A, y_B; y_C, y_D) \equiv k_{yz}(y_A, y_B; y_C, y_D) + k_{zw}(z_A, z_B; z_C, z_D)$$

but that, since in general

$$k_{zw}(y_A, y_B; y_C, y_D) \neq k_{zw}(z_A, z_B; z_C, z_D),$$

we will have inequality in

$$k_{yw}(y_A, y_B; y_C, y_D) \neq k_{yz}(y_A, y_B; y_C, y_D) + k_{zw}(y_A, y_B; y_C, y_D).$$

Thus strengths correctly calculated are additive, but strengths naively calculated are not. In particular, if w is the same function of z that z is of y, the strengths of the two transformations (applied successively to the same quartet) will not be the same, and moreover, the strength of the composite will not be twice the strength of either.

Thus we see that the concept of strength is a little more subtle than might be supposed.

In particular, an attempt to use $k(y)$ directly fails because by going to the limit *after division by* Δ (which we must do to obtain a finite limit) we have lost the distinction between the original and the transformed quartet. In more detail, and working in terms of confluent strengths and in terms of $z = f(y)$,

$$k_{yw}(y; y + \Delta) = k_{yz}(y; y + \Delta) + k_{zw}(z; z + \delta),$$

where $z + \delta = f(y + \Delta)$, so that, on division by Δ and passage to the limit,

$$\begin{aligned} k_{yw}(y) &= k_{yz}(y) + \frac{df(y)}{dy} k_{zw}(z) \\ &= k_{yz}(y) + \frac{dz}{dy} k_{zw}(z), \end{aligned}$$

and both the shift in argment and the multiplication by dz/dy must be taken into account.

As an example, consider $z = \sqrt{y}$ and $w = \sqrt{z}$, so that $k_{yz}(y) = 1/2y\ k_{zw}(z) = 1/2z$, and $dz/dy = 1/2\sqrt{y}$. We obtain, from the formula above,

$$k_{yw} = \frac{1}{2y} + \frac{1}{2\sqrt{y}}\left(\frac{1}{2z}\right) = \frac{1}{2y} + \frac{1}{2\sqrt{y}}\left(\frac{1}{2\sqrt{y}}\right) = \frac{3}{4y},$$

which is exactly what we would have obtained by direct calculation. Note that we combined $1/2y$ with $1/2z$ to obtain $3/4y$, which is far from the naive answer.

II. Strengths in the Simple Family

8. Confluent strengths of powers. We now return to the simple family, and begin with the confluent strengths of the pure powers, where

$$z = \begin{cases} y^p, & p \neq 0, \\ \log y, & p = 0. \end{cases}$$

Here we have

$$k(y_1\,;\, y_2) = \left[\log \frac{dz}{dy}\right]_{y_1} - \left[\log \frac{dz}{dy}\right]_{y_2},$$

where

$$\log \frac{dz}{dy} = \begin{cases} \log p + (p - 1) \log y, & p \neq 0, \\ -\log y, & p = 0, \end{cases}$$

so that

$$k(y_1\,;\, y_2) = (p - 1) \log (y_1/y_2) \qquad \text{for all } p.$$

Thus the strength is proportional to $p - 1$ for any pair of confluent arguments y_1 and y_2.

If $w = z^q$, and $z = y^p$, we have

$$\begin{aligned} k_{yz}(y_1\,;\, y_2) &= (p - 1) \log (y_1/y_2), \\ k_{zw}(z_1\,;\, z_2) &= (q - 1) \log (z_1/z_2) = (q - 1)\, p \log (y_1/y_2), \\ k_{yw}(y_1\,;\, y_2) &= (pq - 1) \log (y_1/y_2) \end{aligned}$$

where we have used

$$\log (z_1/z_2) = p \log (y_1/y_2).$$

We note that the additive decomposition of strengths is, in terms of the numerical coefficients,

$$pq - 1 = (p - 1) + p\,(q - 1);$$

so that if $p = \frac{1}{2}$, for example, we again get

$$-\tfrac{3}{4} = -\tfrac{1}{2} + (-\tfrac{1}{4}),$$

showing how the second square-rooting had only half the strength of the first.

The special case $q = 0$ (logarithm for the second transformation) offers no difficulty, the decomposition being

$$-1 = (p - 1) + (-p);$$

but an attempt to put $p = 0$ (logarithm for the first transformation) fails, since in this case

$$\log (z_1/z_2) \neq p \log (y_1/y_2)$$

and the power strengths are quite dependent on the particular values of y_1 and y_2 involved. The resulting transformations, such as $\sqrt{\log y}$, are not power transformations and seem to have been relatively little used.

In particular, the sequence of transforms

$$\cdots, \; y, \qquad \sqrt{y}, \qquad \log\, y, \qquad 1/\sqrt{y}, \qquad 1/y\,, \; \cdots$$

are seen to be equally spaced in strength, where a more naive approach might have led to $y, y^{1/2}, y^{1/4}, y^{1/8}, \cdots$ as an equally spaced sequence—a possibility not in accordance with either a careful analysis or empirical experience.

In general, we shall refer to the ratio of $k_{yz}(y_1, y_2)$ to $-\log (y_1/y_2)$ as the *power strength* of a transformation. With this definition, the power strength of y^p is $1 - p$ and the power strengths of the sequence $\sqrt{y}$, $\log y$, $1/\sqrt{y}$, $\cdots$ are $1/2, 1, 3/2, 2, 5/2, \cdots$. (For a transformation which is not a pure power, the power strength will depend on the exact arguments y_1 and y_2.)

9. Confluent strengths in the simple family. We can now go on to consider the confluent strengths of the modified power transformations

$$z = \begin{cases} a_p(y + c)^p, & p \neq 0, \\ \log (y + c), & p = 0, \end{cases}$$

where we are likely to choose the constant a_p for each p so that sgn a_p = sgn p. We have

$$\log \frac{dz}{dy} = (p - 1) \log (y + c) + \text{constant}$$

and hence

$$k(y_1 ; y_2) = (p - 1) \log \frac{y_1 + c}{y_2 + c}.$$

The power strength will be given by

$$\frac{k(y_1 ; y_2)}{-\log (y_1/y_2)} = (1 - p) \frac{\log \left((y_1 + c)/(y_2 + c)\right)}{\log (y_1/y_2)}.$$

If we put $y_1 = y - \delta$, $y_2 = y + \delta$, and expand all logarithms in terms of δ, we obtain

$$(1 - p)\left[\frac{y}{y + c} - \frac{\delta^2}{3}\frac{c(2y + c)}{y(y + c)^3} + \cdots\right],$$

where the first term will usually be controlling, and where, by Rolle's theorem, we may slightly redefine y, keeping it well between y_1 and y_2 so that the first term is exact.

We are now able to give some explanation of the fact, mentioned earlier, that $\sqrt{y}$ and $\log (y + \lambda)$ are both fairly good normalizing transformations for a Poisson variate of average value λ. Their respective power strengths at $y = \lambda$ are exactly

$$(1 - \tfrac{1}{2}) = \tfrac{1}{2}$$

and approximately

$$(1 - 0)\frac{\lambda}{\lambda + \lambda} = \tfrac{1}{2},$$

so that their rough equivalence is assured.

A further appeal to Rolle's theorem shows that nonconfluent strengths are always equal to confluent strengths for points near the center of each pair, and it is easy to convince ourselves that, if ratios of not much more than 10 to 1 are involved for y_B/y_A and y_D/y_C, the confluent strengths give us an adequate picture of the behavior of the modified power transformations. This disposes of most practical situations except one. When y is a count, and zero values actually occur, we obtain infinite ratios for y_B/y_A. A brief inquiry into strengths in this cases does not provide any results of much practical use, which is somewhat unfortunate.

10. Extreme strengths. We have expressed our strengths on a logarithmic basis; this is quite reasonable for weak and moderate transformations, but can easily be misleading for extremely strong transformations in practical applications. For example, suppose that only three different values are ever observed, 0, 1, and 2, and that they have been transformed into 0.0000, 1.0000, and 1.0001.

By our definitions, the transformation carrying them into 0.0000 0000, 1.0000 0000, and 1.0000 0001 would be twice as strong. From most practical points of view this is quite wrong, since both transformations are so very close to the one carrying 0, 1, 2 into 0, 1, 1 (which by our definitions has infinite strength). In practice, infinite strength is *not* infinitely far away.

This means that, in charting, we should pay very considerable attention to the behavior of strengths on our scale where they are small or moderate, much less attention to their behavior where they are large, and should avoid moving infinite strengths to infinity.

III. Local Structures of Familie of Transformations

11. Introduction. It is natural to question the intrusion of topology and differential structure into the present paper. For in choosing transformations for empirical purposes—whether for data analysis or theoretical manipulation—only a limited degree of detail about a transformation is relevant. If our interest is coarse-grained, while the attention of topology is directed to the indefinitely fine grain of arbitrarily small neighborhoods, how can topology help us? There are at least two answers to this question, and both are quite relevant.

First, there is a heuristic principle that the study of behavior in the small often, but not always, leads to results which are helpful at a larger scale. In the present circumstances we may reasonably expect this principle to apply without exception.

Second, we are trying to chart the family of transformations in such a way that, when we plot values of some interesting quantity on the chart, we can use these values as advantageously as possible in:

(1) understanding the relation of the transformation to the quantity;

(2) picking a single useful transformation; or

(3) indicating a region of acceptable transformations.

For any and all of these ends, we should like to have the quantities of interest vary smoothly over the chart.

The finest-grained aspect of smooth variation is continuity. Thus we should at least like to have our quantities continuous functions over the chart. To do this we must respect the topology of the family of transformations in choosing the topology of the chart. This respect need only be one-sided, however, for if

(1) position in the family is a continuous function of position in the chart, and

(2) the quantity is a continuous function of position in the family,

then the quantity will be a continuous function of position on the chart. In a phrase, we must not bring together in the chart things which are apart in the family, although we may, if we wish, keep apart in the chart things which come together in the family.

For all that the second answer leaves us free to pull things apart in the chart wherever and as much as we wish, the first answer suggests that we should restrain such tendencies as much as may be reasonable.

12. Intersection and tangency. Whether one curve intersects (we shall always mean "intersects at a non-zero angle" when we say "intersects"), or is instead tangent to, another curve in the family is a relatively easy question to settle, and one that can usefully be asked near almost any point of the family. The analog of the rule just stated above may be put as follows: If the curves intersect in the family, we cannot allow them to be tangent on the chart; if they are tangent in the family, then we can, if we wish, let them intersect on the chart.

Such considerations may arise in the interior of a family, but are more likely to be important on its boundary. If a family is two-dimensional (as the special family is), the boundary will consist of curves and corners. It is at these corners, which are often singularities in some other sense, that intersection and tangency will be most critical.

At such a corner we should avoid, at all costs, the introduction of new tangencies, and should try to preserve the tangencies of the family whenever this can be done without contravening more important considerations.

We shall have to deal with a least one corner where there are several alternate differential structures. This is likely to force us to eliminate tangencies. For if we have two curves which are tangent in all of these structures, they may not be tangent in the same way ("corresponding" pairs of points, one on each curve, in one structure may not "correspond" in another). In this case, no one sort of tangency in the chart can possibly be satisfactory. We must destroy the tangency and chart the curves as intersecting at a finite angle.

13. Nature of the topology. The discussion above has assumed a topology without describing it. We need to say something more. If we have a family of transformations depending on parameters α, β, γ, $\cdots$,

$$z = z(y; \alpha, \beta, \gamma, \cdots),$$

and if $\alpha = \alpha(t)$, $\beta = \beta(t)$, $\gamma = \gamma(t)$, $\cdots$ defines a curve of transformations

$$z_t = z(y; \alpha(t), \beta(t), \gamma(t), \cdots),$$

when do we say that z_t converges to z_0 as $t \to 0$? With a mild reservation, we say that $z_t \to z_0$ if there exist A_t and B_t so that

$$A_t + B_t z_t(y) \to z_0(y)$$

for every relevant y. Notice two things here: First, we have to introduce the variable linear transformation defined by (A_t, B_t), since we regard the class of all $C + Dz_t$, for C and D fixed, as equivalent and our topology really refers to equivalence classes of transformations. Second, the convergence is for all relevant y, so that a change in the set of y's considered relevant may change the topology. In our present case—the simple family—such changes only occur (i) near the far corner, or (ii) for degenerate cases where at most two values of y are relevant, and all non-degenerate transformations are equivalent. (A transformation is degenerate if all its values for relevant y's are the same. The mild restriction mentioned above is that z_0 should not be degenerate.)

14. Nature of the differential structure. We have also talked of "intersection" and "tangency" without specific basis. We have available a differential structure in which these terms are well defined, but it is rather different from those structures which appear in, say, Riemannian geometry. It does not start with something like a differential form which gives local meaning to ratios of distances and from which local differentiation, tangency, etc., follow. It cannot, since we shall not have local distances. Instead, it starts with the notion of a direction from one transformation to another and proceeds to the definition of a tangent.

Given two transformations, z_1 and z_2, their difference, $z_1 - z_2$, is also a transformation and is, of course, equivalent to any multiple of itself. If we replace z_1 and z_2 by equivalent transformations, say $A + Bz_1$ and $C + Dz_2$, the difference becomes

$$A + Bz_1 - C - Dz_2 = A - C + B(z_1 - z_2) + (B - D)z_2$$
$$= A - C + (B - D)z_1 + D(z_1 - z_2),$$

namely, any member of a one-parameter family of equivalence classes where $z_1 - z_2$ is additively combined with various amounts of z_1 or z_2. The resulting direction is defined, but not too precisely.

If, however, $z_1 = z_t$ and $z_2 = z_0$ where z_t converges to z_0 as $t \to 0$, then we can focus our attention on a particular $A_t + B_t z_t$ whose values converge to those of z_0. If the convergence of the values is sufficiently differentiable for each y, then we have

$$A_t + B_t z_t(y) = z_0 + tz'(y) + O(t^2),$$

where $O(t^2)$ means terms of order t^2 or smaller as $t \to 0$. We may regard z' considered as a transformation, as the derivative of z_t at the point (at the transformation) z_0.

It is well to note that z' is also not uniquely defined. For if we replace B_t by $B_t + tC$, as we may without disturbing the convergence, then

$$A_t + (B_t + tC)z_t(y) = z_0 + t[z'(y) + Cz_0(y)] + O(t^2),$$

and we see that $z' + Cz_0$ is also a "derivative." Except for the inevitable linear transformations this is the most general form, and we may say that the directions at z_0 correspond to the families of transformations of the form

$$A + Bz' + Cz_0,$$

where A, B, and C are arbitrary constants with $B \neq O$.

Two curves

$$x_t = z_0 + tz' + O(t^2)$$

and

$$z_s = z_0 + s\dot{z} + O(s^2)$$

will have the same direction at z_0 if

$$A + B\dot{z} + Cz_0 \equiv z'$$

for the relevant values of y and suitable constants A, B, and C.

Notice the appearance again of the relevant values of y. It will again usually be true that, so long as y takes at least three different values, it will not matter which three. But in exceptional circumstances it will matter, and it is in this way that a corner may receive various differential structures.

The same sort of construction can be extended to higher derivatives. We shall need it in only one special case, and no general discussion seems necessary.

With these general principles and techniques in mind, then, we can proceed to study the simple family.

IV. The Body of the Simple Family

15. Definitions and limiting forms. As previously stated, the *simple family* of transformations consists of all transformations, which can be obtained by compounding linear transformations with one (integral or fractional) power transformation, and of all limiting forms of such transformations.

The effect of an additional linear transformation from z to $A + Bz$ with $B \neq 0$ applied *last* is always regarded as trivial. Transformations which can be trivially converted into each other are equivalent. Thus $z = 3y^2 - 17$ and $z = 0.001 + 0.037y^2$ are regarded as equivalent to each other and to $z = y^2$. This is not the case for an *initial* linear transformation. Thus $z = \log y$ and $x = \log (y + 10\,000\,000)$ are quite different transformations. When we are dealing with a curve of transformations, such as, for example, the curve of pure power transformations with powers between 1 and 0, namely

$$\{y^p, 1 > p > 0\},$$

we shall not distinguish equality from equivalence, and shall freely write such "equations" as

$$\{3y^p, 1 > p > 0\} = \{y^p, 1 > p > 0\} = \{14 - 2y^p, 1 > p > 0\}.$$

We are only concerned with real values, and prefer monotone functions, so that we could take as our definition of power

$$y^p = \operatorname{sgn}(y) \mid y \mid^p,$$

where the signum function, sgn (y), is defined by

$$\operatorname{sgn}(y) = \begin{cases} +1, & y > 0, \\ 0, & y = 0, \\ -1, & y < 0. \end{cases}$$

With this definition,

$$(y^p)^q = y^{pq}$$

but

$$\frac{dy^p}{dy} = py^{p-1}(\operatorname{sgn} y) = py^{p-1}(\operatorname{sgn} y^p) = py^{p-1}(\operatorname{sgn} y^{p-1} = p \mid y \mid^{p-1}.$$

(Differences from other definitions of powers only occur for negative y, and since negative y's occur infrequently in practice, this possible definition should be regarded as precautionary rather than important. We shall not use it explicitly here, merely remarking that its use would make only trivial alterations in our results.)

The limiting forms are associated with the exceptional values of the power,

$-\infty$, 0, and $+\infty$, and with the extreme values, $-\infty$ and $+\infty$, of the constant. For $p \to 0$, we have

$$(y + c)^p = e^{p \log (y+c)} \approx 1 + p \log (y + c),$$

which, after a suitably varying linear transformation, converges to log $(y + c)$. For $p \to \pm\infty$ and c tending to a finite limit, the convergence is to less familiar objects. The transformations involved depend on the particular set of values transformed through the extreme values, $y_{\max}$ and $y_{\min}$, actually present in the data. If we define a function φ_0 by

$$\varphi_0(u) = \begin{cases} 0, & u \neq 0, \\ 1, & u = 0, \end{cases}$$

we then have for $p \to +\infty$ and a suitable choice of A_p and B_p,

$$A_p + B_p(y + c)^p \to -\varphi_0(y - y_{\min}) = -1 + \mathrm{sgn}^+(y - y_{\min}),$$

where the last equality is valid for actually occurring y's, since $y < y_{\min}$ is impossible.

As in the sequel, a "+" on a function indicates that negative values are to be replaced by zero. Hence, in particular,

$$\mathrm{sgn}^+ y = \begin{cases} 0, & y \leqq 0, \\ 1, & y > 0. \end{cases}$$

There remains the case where $p \to \pm\infty$ but p/c tends to a finite limit. (Hence $c \to \pm\infty$, also.) Here

$$(y + c)^p = c^p \left(1 + \frac{y}{c}\right) p = c^p \left[\left(1 + \frac{y}{c}\right)^c\right]^{p/c};$$

and since the expression in [ ] tends to e^y, the whole expression, after a suitably varying linear transformation, tends to

$$e^{my},$$

where $m = \lim p/c$.

Thus the simple family contains:

$$z = (y + c)^p,$$

$$z = \log (y + c),$$

$$z = e^{my},$$

and all linear transforms of these transformations.

16. Normalization. Because (i) they are apparently of the greatest practical importance, and (ii) they determine the remaining cases by symmetry, we shall confine our detailed analysis to transformations with $p \leqq 1$ (and their limits). Consequently, we shall be concerned with least values but not with greatest ones. It will therefore be convenient, for the remainder of this part, for us to fix initial scale of the y's so that the three smallest values of y are 0, 1, and y_2, where $y_2 > 1$. This can be done by a linear transformation, and undone again by another, so that our conclusions are to be changed by at most a linear transformation when we are through.

We shall also assume that the reasonable values of c are such that $y + c$ is non-negative. We shall treat the general case, specializing later to the case where c is positive and bounded away from zero.

17. Tangency and intersection for c small. We now investigate the situation along and near the curve $\{z = y^p\}$ in more detail. We begin by considering $(y + c)^p$ as $c \to 0$, where the value of p is important. If $p < 0$, and temporarily we shall write $p = -k$, then

$$(y + c)^p = \begin{cases} c^p = c^{-k} \to \infty, & y = 0, \\ (y + c)^{-k} \to y^{-k}, & y > 0, \end{cases}$$

and

$$\begin{aligned} 1 - c^k(y + c)^p &= \begin{cases} 0, & y = 0, \\ 1 - \left(\dfrac{c}{y + c}\right)^k, & y > 0 \end{cases} \\ &= \operatorname{sgn}^+(y) - c^k(y^{-k} \operatorname{sgn}^+(y)) + O(c^{k+1}). \end{aligned}$$

As $c \to 0$ for k fixed, the transformation tends to $\operatorname{sgn}^+(y)$ and appears like an additive mixture of this and $y^{-k}\operatorname{sgn}^+(y)$. This additive piece is different for different values of k, so that we learn that the curves of transformations

$$\{(y + c)^p, c \to 0\} = \{1 - c^k(y + c)^p, c \to 0\}, \qquad p < 0,$$

tend to $\operatorname{sgn}^+(y)$, but no two have a common tangent.

For $p = 0$, we must deal with $\log(y + c)$, where

$$\begin{aligned} 1 - \frac{\log(y + c)}{-\log c} &= \begin{cases} 0, & y = 0, \\ 1 - \dfrac{\log^+(y + c)}{-\log c}, & y \geqq 1, \end{cases} \\ &= \operatorname{sgn}^+ y - \frac{\log^+(y + c)}{-\log c} \\ &= \operatorname{sgn}^+ y - \frac{\log^+ y}{-\log c} + O\left(\frac{c}{-\log c}\right), \end{aligned}$$

and hence the curve

$$\{\log(y + c), c \to 0\} = \left\{1 - \frac{\log(y + c)}{-\log c}, c \to 0\right\}$$

also tends to $\operatorname{sgn}^+ y$, but with a different tangent.

There is also another simply described curve with c small (in fact with $c = 0$) which tends to sgn^+y. This is the curve

$$\{y^p, p \to 0\},$$

where we have

$$\begin{aligned} y^p &= \begin{cases} 0, & y = 0, \\ y^p = e^{p \log y}, & y > 0, \end{cases} \\ &= \text{sgn}^+y + p \log^+ y + O(p^2), \qquad y \geqq 0. \end{aligned}$$

We see now that the curves

$$\{y^p, p \to 0\} \text{ and } \{\log(y + c), c \to 0\}$$

are tangent to each other at the transformation ($z = \text{sgn}^+y$), with p corresponding to $1/(-\log c)$.

For $p > 0$, we have

$$\begin{aligned} (y + c)^p &= \begin{cases} c^p, & y = 0 \\ y^p\left(1 + \dfrac{c}{y}\right)^p, & y > 0, \end{cases} \\ &= \varphi_0(y)c^p + y^p + cpy^{p-1} + O(c^2), \end{aligned}$$

so that the curve

$$\{(y + c)^p, c \to 0\}$$

tends to the transformation y^p.

Near y^p we are also interested in

$$y^{p+\delta} = y^p e^{\delta \log y} = y^p + {}^{\delta}y^p(\log y) + O(\delta^2),$$

which is easily seen to have a different tangent than the previous curves at y^p.

18. Tangency and intersection for large c. If c is large,

$$\begin{aligned} (y + c)^p &= c^p\left(1 + \frac{y}{c}\right)^p = c^p \exp[p \log(1 + y/c)] \\ &= \exp[(py/c) + py^2/2c^2 + O(1/c^3)] \\ &= c^p e^{py/c}\left(1 + \frac{py^2}{2c^2} + O\left(\frac{1}{c^3}\right)\right). \end{aligned}$$

Thus, if $c = 1/\epsilon$ and $p = m/\epsilon$ (note that $m < 0$ if p is always < 0), we have

$$c^{-p}(y + c)^p = e^{my} + \frac{\epsilon}{2} my^2 e^{my} + O(\epsilon^3 e^{my}),$$

and the curve

$$\{(y + c)^p, p = mc, c \to \infty\} = \{c^{-p}(y + c)^p, p = mc, c \to \infty\}$$

tends to e^{my}.

We also wish to consider $\{e^{(m+\delta)y}, \delta \to 0\}$ for which

$$\begin{aligned} e^{(m+\delta)y} = e^{my}e^{\delta y} &= e^{my}(1 + \delta y + O(\delta^2)) \\ &= e^{my} + \delta y e^{my} + O(\delta^2 e^{my}). \end{aligned}$$

Thus the curves

$$\{(y + c)^p, p = mc, c \to \infty\} \qquad \text{and} \qquad \{e^{(m+\delta)y}, \delta \to 0\}$$

tend to the same limit, but with different tangents.

We also have to consider $c \to \infty$ and p fixed, for which

$$(y + c)^p = c^p\left(1 + \frac{y}{c}\right)^p = c^p\left(1 + \frac{py}{c} + \frac{p(p-1)}{2c^2}\, y^2 + O\left(\left(\frac{y}{c}\right)^3\right)\right)$$

and

$$\frac{c}{p}\, c^{-p}(y + c)^p - \frac{c}{p} = y + \frac{p-1}{2c}\, y^2 + O\left(\frac{1}{c^2}\right),$$

so that the curves

$$\{(y + c)^p, c \to \infty\} = \left\{\frac{c}{p}\, c^{-p}(y + c)^p - \frac{c}{p}, c \to \infty\right\}$$

all tend to y with a common tangent, a natural parameter along this tangent being $(p - 1)/2c$.

Since e^{my} has been included, we must also consider the case $m \to 0$ where

$$\frac{e^{my} - 1}{m} = y + \frac{m}{2}\, y^2 + O(m^2),$$

so that

$$\{e^{my}, m \to 0\}$$

has the same tangent as the curves just considered, with m corresponding to $(p - 1)/c$.

19. Tangency and intersection for c moderate. If $p \to 1$ with c constant, we learn from

$$(y + c)^{(1+\delta)} = (y + c)e^{\delta \log(y+c)} = (y + c)(1 + \delta \log(y + c) + O(\delta^2))$$

that

$$(y + c)^{(1+\delta)} - c = y + \delta(y + c \log(y + c) + O(\delta^2)$$

and hence that the curves

$$\{(y + c)^p, p \to 1\}$$

tend to y, each with its own tangent.

At the other extreme, where $p = -k$ tends to $-\infty$, we have

$$1 - c^k(y + c)^p = \begin{cases} 0, & y = 0, \\ 1 - \left(1 + \dfrac{1}{c}\right)^{-k}, & y = 1, \\ 1 - \left(1 + \dfrac{1}{c}\right)^{-k}\left(\dfrac{c + 1}{c + y}\right)^k, & y > 1, \end{cases}$$

$$= \text{sgn}^+ y - \left(1 + \frac{1}{c}\right)^{-k} \varphi_0(y - 1) + O\left(\left(\frac{c + 1}{c + y}\right)^k\right);$$

hence the curves

$$\{(y + c)^p, p \to -\infty\} = \{1 - c^k(y + c)^p, p \to -\infty\}$$

tend to $\text{sgn}^+ y$ with a common tangent and a natural parameter which is a multiple of $[1 + (1/c)]^{-k} = [1 + (1/c)]^p$.

We also have, as $m \to -\infty$,

$$1 - e^{my} = \begin{cases} 0, & y = 0, \\ 1 - e^m, & y = 1, \\ 1 - e^m e^{m(y-1)}, & y > 1, \end{cases}$$

$$= \text{sgn}^+(y) - e^m(\varphi_0(y - 1)) + O(e^{m(y-1)}),$$

and we see that the curve

$$\{e^{my}, m \to -\infty\} = \{1 - e^{my}, m \to -\infty\}$$

has the same limit and tangent, with e^m playing the role of $[1 + (1/c)^p$.

Everything is well behaved as $p \to 0$ when $(y + c)^p$ tends to $\log(y + c)$.

20. The resulting picture. If we put together all the individual results of the last three sections, we are forced to the qualitative relation of the curves

$$\{e^{my}, m < 0\},$$

$$\{y,^p\ 1 > p > 0\},$$

$$\{(y + c)^p, p \text{ fixed } (1 > p > 0), 0 < c < \infty\},$$

$$\{\log(y + c), 0 < c < \infty\},$$

$$\{(y + c), p \text{ fixed } (p >), 0 < c < \infty\},$$

$$\{(y + c)^p, c \text{ fixed}, 1 > p > -\infty\}$$

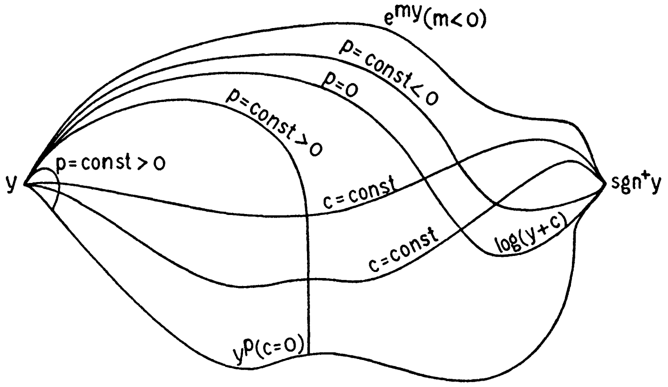

FIG. 5. The topology of the simple family.

shown in Fig. 5. These relations of limit, tangency, and intersection substantially limit the diagrams on which we may reasonably plot the transformations $(y + c)^p$ (with their limits) for $1 \geqq p > -\infty$ and $0 \leqq c < \infty$.

V. THE CORNERS OF THE SIMPLE FAMILY

21. The neighborhood of the near corner. We can try to learn more about the situation near the corners at y and $\mathrm{sgn}^+ y$ by carrying the expansion out to further terms. We begin with the neighborhood of y, where

$$\frac{c}{p}\, c^{-p}(y + c)^p = y + \frac{p-1}{2c}\, y^2 + \frac{(p-1)(p-2)}{6c^2}\, y^3 + O\left(\frac{1}{c^3}\right),$$

$$\frac{e^{my-1}}{m} = y + \frac{m}{2}\, y^2 + \frac{m^2}{6}\, y^3 + O(m^3),$$

$$(y + c)^{1+\delta} - c = y + \delta(y + c) \log (y + c) + \frac{\delta^2}{2}\, (y + c)(\log (y + c))^2 + O(\delta^3),$$

and where, in particular, we may put $c = 0$ in the last expansion.

If, taking account of the signs of $p - 1$ and m in the region we are interested in, we set

$$h = \frac{1-p}{c} = -m,$$

the first two expansions give

$$\frac{c}{p}\, c^{-p}(y + c)^p = y + h\left(-\frac{y^2}{2}\right) + h\left(\frac{1}{c} + h\right)\left(\frac{y^3}{6}\right) + O\left(\frac{1}{c^3}\right),$$

$$\frac{1}{m}\,(e^{my} - 1) = y + h\left(-\frac{y^2}{2}\right) + h(O + h)\left(\frac{y^3}{6}\right) + O\left(\frac{1}{c^3}\right),$$

which are clearly consistent, since $c = \infty$ and $1/c = 0$ are the limiting values which lead to the exponentials.

We know that the situation near y (i.e., near c = anything, $p = 1$) should be represented by an angle of some size, with the curves $\{(y + c)^p, c \text{ fixed}, p \to 1\}$ coming in with separate tangents and the curves $\{(y + c)^p, p \text{ fixed}, c \to \infty\}$ all coming in with a common tangent. It is not unnatural to try to map this into the first quadrant of a (u, v)-plane, with the v-axis playing the role of the common tangent. The natural identification near $(0, 0)$ is, then,

$$v = h,$$

$$u = \frac{h}{c} + h^2.$$

However, experimentation leads to complexities, and reflection shows us that we may reasonably alter the h^2 term, since it is a higher-order function of the h term.

Omitting the h^2 term,

$$u = \frac{h}{c}, \qquad v = h$$

whence

$$c = \frac{v}{u},$$

and the radius of curvature at $(0, 0)$ of the curves p = constant may be found from the parametric form for a circle through $(0, 0)$ tangent to the v-axis, namely,

$$v = R \sin \psi \sim R\psi,$$

$$u = R(1 - \cos \psi) \sim \tfrac{1}{2}R\psi^2,$$

whence

$$R = \frac{(R\psi)^2}{2(R\psi^2/2)} \sim \frac{V^2}{2u} = \frac{h^2}{2h/c} = \frac{hc}{2} = \frac{1 - p}{2}.$$

We thus reproduce the situation around y to this accuracy if we make the following identifications

$$\{\text{curve of constant } c\} = \{\text{ray through origin of slope } c\},$$

$$\{\text{curve of constant } p\} = \{\text{circle with } (0, 0) \text{ and } (1 - p, 0) \text{ as diameter}\}.$$

We then have a map of the simple family which is sound for transformations near the identity.

It is this map which was used by Moore and Tukey [2] and is shown as Fig. 6. In polar coordinates, $\tan \theta = c$ and $r = (1 - p)/\sqrt{1 + c^2}$.

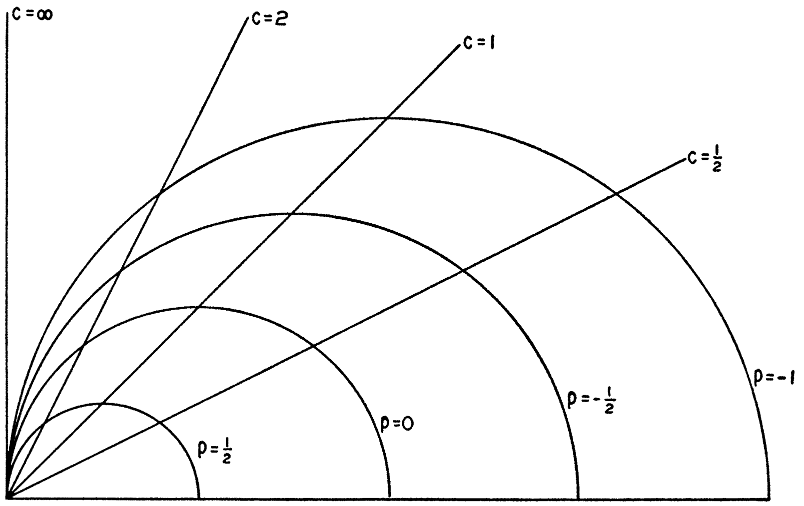

FIG. 6. The neighborhood of the near corner.

22. The neighborhood of the far corner. Now we go to the neighborhood of $\mathrm{sgn}^+ y$, and again carry more terms to find

$$1 - c^k(y+c)^p = \mathrm{sgn}^+ y - \left(1+\frac{1}{c}\right)^{-k} \varphi_0(y-1) - \left(1+\frac{y_2}{c}\right)^{-k} \varphi_0(y-y_2) + \cdots \qquad (p = -k \to -\infty),$$

$$1 - e^{my} = \mathrm{sgn}^+ y - e^m \varphi_0(y-1) - e^{my_2}\varphi_0(y-y_2) + \cdots \qquad (m \to \infty),$$

$$1 - c^k(y+c)^p = \mathrm{sgn}^+ y - \left(\frac{y}{c}\right)^{-k} \mathrm{sgn}^+ y + \cdots \qquad (p = -k < 0, c \to 0),$$

$$1 - \frac{\log (y+c)}{-\log c} = \mathrm{sgn}^+ y - \frac{\log^+ y}{\log c} + \cdots \qquad (c \to 0),$$

$$y^p = \mathrm{sgn}^+ y + p \log^+ y + \cdots \qquad (p \to 0).$$

We again concentrate on the curves which come to a common tangent. (At this corner these are the family $c =$ constant rather than the family $p =$ constant.) This directs our attention to the first of the five formulas and to the quantities

$$s = \left(1+\frac{1}{c}\right)^{-k} \quad \text{and} \quad t = \left(1+\frac{y_2}{c}\right)^{-k},$$

so that we would have

$$\frac{1}{c} = s^k - 1$$

$$y_2 = \frac{t^k - 1}{s^k - 1}.$$

The situation is both complex and dependent on the particular values of y_2. If we are going to obtain results of wide usefulness, we should have to fix, more or less arbitrarily, a single value for y_2.

In any event,

$$\frac{t}{s} = \left[\frac{1 + \frac{y_2}{c}}{1 + \frac{1}{c}}\right]^{-k} = \left[\frac{c + y_2}{c + 1}\right]^{-k} \rightarrow y_2^{-k} \qquad (\text{as } c \rightarrow 0).$$

Thus, as p ranges from 0 toward $-\infty$, the limiting value of t/s ranges from 1 to 0 and we see that we have filled up only one octant.

Moreover, if we calculate coordinates for y^p in what seems to be the same way, namely,

$$s = 1 - (\text{value at } 1) = 1 - 1^p = 0,$$

$$t = 1 - (\text{value at } y_2) = 1 - y_2^p \approx -p \log y_2,$$

we find that

$$\frac{t}{s} = -\infty.$$

Either we must accept a slit of some sort between $\{(y + c)^p, c \rightarrow 0\}$ and $\{y^p, p \rightarrow 0\}$ or we must recognize some difficulty with our procedure.

Actually, closer examination shows us that the coordinate system we are using does not work at all well for y^p. It is set up to work for $1 - c^k(y + c)^p$, which is 0 for $y = 0$, and approaches 1 from below for all $y > 0$, most slowly for the smallest y. On the other hand, y^p is 0 for $y = 0$ and approaches 1 from *above* for all $y \geqq 1$, most slowly for the largest value of y. These are not compatible.

Essentially the only renormalization we can try is to expand $y^p(1 - Cp)$, which leads to

$$\frac{t}{s} = \frac{C - \log y_2}{C},$$

which can never be made to vanish for more than one value of y_2.

We clearly should give up any attempt to get second-order accuracy in our representation around the far corner. Indeed, when we reflect on the effects of varying values of y_2, we see that we are in the situation discussed in Section 6. Not only second-order accuracy, but also first-order accuracy should be given up. We should arrange to have these curves intersect on the chart, even though they are tangent in the family. The far corner is really a singularity!

23. The far corner if $y + c$ is always positive. In view of the essential way in which the far singularity depended on $y + c$ either being actually zero or approaching zero, and in view of the fact that we may often have situations where $y + c$ will not approach zero, it is probably worth while to reconsider our normalization and to look separately at the case where $y + c$ is always positive. It will be convenient to normalize so that $y \geqq 1$, $c \geqq 0$, as we may clearly do.

The only singularity is now in the vicinity of $p = -\infty$, where we have

$$1 - (1 + c)^k(y + c)^p \approx \operatorname{sgn}^+(y - 1) - \left(\frac{1 + c}{y + c}\right)^k \operatorname{sgn}^+(y - 1) \qquad (c \geqq 0)$$

$$1 - \frac{e^{my}}{e^m} \approx \operatorname{sgn}^+(y - 1) - e^{m(y-1)} \operatorname{sgn}^+(y - 1) \qquad (m < 0)$$

and where, therefore, if we adopt as coordinates s and t the deviations of the values from $\operatorname{sgn}^+(y - 1)$ at the next lowest values y_2 and y_3 of y, we find

$$s = \left(\frac{1 + c}{y_2 + c}\right)^k = e^{m(y_2-1)},$$

$$t = \left(\frac{1 + c}{y_3 + c}\right)^k = e^{m(y_3-1)},$$

and we see that the limiting value of t/s is always zero. Hence all the curves for $p \to -\infty$ (and the curve for $m \to \infty$) are tangent. Moreover the final approach to $\operatorname{sgn}^+(y - 1)$ is exponential in speed.

For practical charting, we may either bring all the curves to a common tangent or leave them separate, so long as we pay some attention to the curves $s =$ constant. If we let the curves intersect at finite angles, quantities which depend continuously on the transformation will be continuous on the chart—they will merely tend to be all more and more closely constant on the curves $s = s_0$ as $s_0 \to 0$.

The exact nature of the curves $s = s_0$, as well as the exact nature of the tangency depends on the value of y_2. If we adopted a chart with tangency, we would have to have a separate one for each value of y_2 to avoid introducing discontinuities. If we adopt a chart without tangency, we may hope that the curves $s = s_0$ will not be too badly distorted over a range of values for y_2. Thus it again seems better to use a non-tangent chart.

If y_2 is arbitrarily close to 1, say $y_2 = 1 + \epsilon$, we find

$$s = \left(\frac{1 + c}{y_2 + c}\right)^k = \left(\frac{1}{1 + \dfrac{\epsilon}{1 + c}}\right)^k \approx \exp\left[-k(\epsilon/1 + c)\right],$$

so that $-k/(1 + c) = p/1 + c$ is a natural parameter (which is constant on the curve which is the limits of the curves $s = s_0$ as $y_2 \to 1$). We find that m is equivalent to $p/1 + c$. Thus it may be well to keep the curves

$$\frac{p}{1 + c} = \text{constant}$$

relatively short as $p \to -\infty$.

24. Conclusions. We thus conclude that we need two forms of chart, depending on whether $y + c = 0$ is to be considered reasonable or not. Except near $\mathrm{sgn}^+ y$, both charts are to follow the topology of Fig. 5. Near the identity transformation, both charts should conform to Fig. 6.

Near $\mathrm{sgn}^+ y$, both charts should distort the topology of the family by opening out the angle and making the previously tangent curves meet at finite angles. The details of these openings out are not clearly prescribed. In the case where $y + c$ is safely > 0 we will probably wish to keep the curves $p/(1 + c) =$ constant fairly short. The most obvious difference in the two cases will be that if $y + c$ can vanish, then $(y + c)^p$ approaches $\mathrm{sgn}^+(y + c)$ as $p \to 0$; while if $y + c$ can not vanish, it approaches $\mathrm{sgn}^+(y - y_{\min})$, but only as $p \to -\infty$.

VI. Charting the Simple Family

25. The general case (where $y + c$ can vanish). If we are to be fully prepared for the case where $y + c$ can even be zero, then we will want our chart to have the following characteristics:

(A) In the vicinity of y itself, the topology and differential properties should be those suggested by the analysis of Section 21:

  (1) the points with $0 \leqq c \leqq \infty$, $p \leqq 1$ fill up the vertex of a quadrant; and

  (2) polar coordinates are, roughly, given by $\tan \theta = c$, $r = (1 - p)/\sqrt{1 + c^2}$.

(B) Along the arc $c = 0$, $1 > p > 0$ the spacing of the values should be, roughly, uniform in p as suggested by Section 8.

(C) In the vicinity of $\mathrm{sgn}^+ y$ the topology should either be as suggested by Fig. 5 (and Section 20) or the tangency of the curves with $c =$ constant, $p \to -\infty$ should be replaced by intersection. (See Section 23.)

(D) Along the arc $c = 0$, $1 > p > 0$, the curves $p =$ constant should intersect the arc and not be tangent to it. (See Section 17.)

We have, then, to find a chart form with these properties.

To begin with, we have a figure composed of the vertex of a quadrant, and two curves running out the sides of the quadrant which eventually meet again (at $\mathrm{sgn}^+ y$). The simplest figure which we know with these properties is a quarter sphere, stretching from pole to pole. In the vicinity of each pole we have a quadrant, and the meridians which bound this quadrant meet at the opposite pole. Thus, as a first step, we agree to try taking y as one pole and $\mathrm{sgn}^+ y$ as the other. The arc$\{y^p\}$ becomes a meridian, the arc$\{e^{my}\}$ becomes a meridian 90° removed from this, and the central region fills in the quarter sphere between the two.

Let θ be an angle describing the meridians, so chosen that $\theta = 0°$ for the arc $\{y^p\}$ which has $c = 0$, and $\theta = 90°$ for the arc $\{e^{my}\}$ which corresponds to $c = \infty$. Let ψ be an angle describing the parallels of latitude, with $\psi = 0°$ for the pole representing y, $\psi = 90°$ at the equator, and $\psi = 180°$ at the pole representing $\mathrm{sgn}^+ y$. Our problem is to find analytic representations for θ and ψ in terms of p and c which will give us the desired properties. These we find by successive approximations.

Near the y pole, we want

$$\tan\theta \sim c, \qquad \psi \sim \frac{1-p}{\sqrt{1+c^2}},$$

and when we observe that $1 - p$ ranges up to infinity, it is natural to try

$$\tan\theta = c, \qquad \tan\frac{\psi}{2} = \frac{1-p}{\sqrt{1+c^2}}.$$

But when we realize that while $1 - p$ ranges up to $+\infty$ for $c \neq 0$, it only ranges up to 1 for $c = 0$, it is natural to try, perhaps,

$$\tan\theta = c, \qquad \operatorname{ctn}\frac{\psi}{2} = \frac{c}{1-p} + \frac{1}{1+c}\tan^+\frac{\pi p}{2},$$

where we have (i) replaced $\sqrt{1+c^2}$ by c in order to have the first term ineffective for $c = 0$; (ii) introduced the new term with a factor $1/(1 + c)$ to make it ineffective for $c = \infty$, where the first term $\sim -c/p$, which is appropriate; (iii) gone to the cotangent instead of the tangent in order to make the combination of the two terms easy; and (iv) introduced $\tan^+\pi p/2$ as a function spreading the values of p from 0 to 1 out uniformly along the arc $c = 0$.

This choice of functions gives a rather reasonable chart, but it still fails to meet our requirements in that (i) the curves $\{(y + c)^p, c \to 0\}$ are tangent to $\{y^p\}$, and (ii) near $\operatorname{sgn}^+ y$ we have not begun to obtain the proper local structure, since the curves $c =$ constant intersect, while the curves $p =$ constant are mutually tangent. We can try to get at both difficulties at once by changing to

$$\tan\theta = c + (1-p)\sqrt{c}, \qquad \operatorname{ctn}\frac{\psi}{2} = \frac{c}{1-p} + \frac{1}{1+c}\tan^+\frac{\pi p}{2}.$$

This change pushes the curves $c =$ constant upward, more strongly for lower c and larger $1 - p$, and thus tends to meet both needs. However, the curves $p =$ constant are still mutually tangent to the boundary curve $c = 0$ at $\operatorname{sgn}^+ y$. In order to correct this in a simple way, we need only add a $(-p)^+$ term, which will vanish for $p \geqq 0$. The result is

$$\tan\theta = c + (1-p)\sqrt{c} + (-p)^+, \qquad \operatorname{ctn}\frac{\psi}{2} = \frac{c}{1-p} + \frac{1}{1+c}\tan^+\frac{\pi p}{2}.$$

26. Plotting. We now have represented the central part of the simple family on a quarter-sphere. For practical purposes we require a plot on the plane. The solution adopted here was influenced substantially by the existence of some special equal-area graph paper which the writer obtained from Mount Wilson Observatory during World War II in connection with problems of mutual interest. This graph paper represents a half-sphere area-true on a rectangle and provides two families of spherico-polar coordinates, one with poles at the top and bottom of the rectangle and the other with poles at the centers of the right and left sides. If u, r are rectangular coordinates with $0 \leqq u \leqq \pi$, $-1 \leqq r \leqq +1$

defining the rectangle, then ψ and θ are related by

$$\cos \psi = \sqrt{1 - r^2} \cos u,$$
$$\cos \theta = \sqrt{1 - r^2} \csc \psi,$$
$$\operatorname{sgn} \theta = \operatorname{sgn} r.$$

The combination of this representation of the half-sphere on the plane with the map on the quarter-sphere gives the plot already presented in Fig. 1.

27. The case of moderate $y + c$. In case $y + c$ cannot vanish, we want a plot with a different and simpler topology. The choice

$$\tan \theta = c,$$
$$\tan \frac{\psi}{2} = \frac{q}{\sqrt{1 + c^2}},$$

where the transformation is written in the form

$$(y + cy_0)^{1-qe_0},$$

scaled as to size by y_0 and as to change in exponent by e_0, is the natural transformation near $q = 0$, and meets our requirements everywhere else.

When combined with the same graph paper, the result is as in Fig. 3.

REFERENCES

[1] F. J. Anscombe and J. W. Tukey, "The criticism of transformations," paper presented before the American Statistical Association and the Biometric Society, Montreal, September 12, 1954.

[2] P. G. Moore and J. W. Tukey, "Answer to query 112," *Biometrics*, Vol. 10 (1954), pp. 562–568.