

Group Reaction Time Distributions and an Analysis of Distribution Statistics

Roger Ratcliff
Dartmouth College

A method of obtaining an average reaction time distribution for a group of subjects is described. The method is particularly useful for cases in which data from many subjects are available but there are only 10-20 reaction time observations per subject cell. Essentially, reaction times for each subject are organized in ascending order, and quantiles are calculated. The quantiles are then averaged over subjects to give group quantiles (cf. Vincent learning curves). From the group quantiles, a group reaction time distribution can be constructed. It is shown that this method of averaging is exact for certain distributions (i.e., the resulting distribution belongs to the same family as the individual distributions). Furthermore, Monte Carlo studies and application of the method to the combined data from three large experiments provide evidence that properties derived from the group reaction time distribution are much the same as average properties derived from the data of individual subjects. This article also examines how to quantitatively describe the shape of reaction time distributions. The use of moments and cumulants as sources of information about distribution shape is evaluated and rejected because of extreme dependence on long, outlier reaction times. As an alternative, the use of explicit distribution functions as approximations to reaction time distributions is considered.

Despite the recent popularity of reaction time research, the use of reaction time distributions for both model testing and model development has been largely ignored. This is surprising in view of the fact that properties of distributions can prove decisive in discriminating among models (Sternberg, Note 1) and can falsify models that quite adequately describe the behavior of mean reaction time (Ratcliff & Murdock, 1976).

Two methods have been used to obtain distributional or shape information. One

method, advocated by Sternberg (1969; Sternberg, Note 2), is to use moments and cumulants to describe distribution shape without assuming any particular reaction time distribution function. A second method, used by Ratcliff and Murdock, is to assume an explicit distribution function and use the parameters of this distribution to provide information about shape. Both these methods are unattractive because they require 5 to 10 times the number of observations usually collected in an experiment. For example, to fit an explicit function such as a gamma distribution to experimentally obtained reaction times, a minimum of about 100 observations per subject per condition are required for reliable convergence of fitting procedures and stability of parameter estimates. Similarly, to obtain stable estimates of higher moments, several thousand observations per condition are typically required. The necessity for a large number of observations becomes a particular problem in experimental endeavors in

This research was supported by Grant APA 146 from the National Research Council of Canada and Grant OMHF 164 from the Ontario Mental Health Foundation to B. B. Murdock, Jr.

I would like to thank Ben Murdock for his support, help, and criticism. I also wish to thank David Andrews for many useful suggestions and Gail McKoon for help in making the article somewhat understandable.

Requests for reprints should be sent to Roger Ratcliff, Department of Psychology, Dartmouth College, Hanover, New Hampshire 03755.

which the test materials used require a great deal of time and effort for construction (e.g., paragraphs; Kintsch, 1974). For such research programs, it would take years to construct enough materials to allow application of either of the two distributional methods.

In the first part of this article, I present a method for combining data from individual subjects to produce group reaction time distributions based on as few as 10 observations per subject cell. To form group distributions, reaction times for each subject are organized in ascending order, and quantiles are calculated. The quantiles are then averaged over subjects to give group quantiles (Vincent averaging; Vincent, 1912). From the group quantiles, a group reaction time distribution can be constructed. This group distribution method averages over individual subjects' data in a way that retains shape information, and this is demonstrated in three ways: First, it is shown that for certain distributional forms (exponential, Weibull, and logistic), Vincent averaging of individual distributions of a particular form with different parameters results in a group distribution of the same functional form. Second, a distribution that has been used to describe reaction time data (Ratcliff & Murdock, 1976) was used in Monte Carlo studies to generate reaction times that were then combined according to Vincent's method. With 20 reaction times per pseudosubject, the group distributions generated by this method have the same form as the distribution used to generate the data. Third, the method was applied to the combined data from three large recognition memory experiments that used the study-test procedure (Ratcliff & Murdock, 1976), with about 120,000 observations in total. It is shown that parameters derived from fitting a distribution function (used by Ratcliff & Murdock, 1976) to the group distribution are the same as averages of the parameters derived from fitting the function to individual distributions.

In the second part of the article, I critically examine the use of moments and cumulants for describing distribution shape. The stability of moment and cumulant estimates is examined first by calculating sampling

standard deviations and second by observing the stability of estimates when outlier reaction times are trimmed from the distribution. In addition, the use of empirical distribution functions to provide information about distribution shape is examined.

The notion of shape can be defined in different ways. Mosteller and Tukey (1977, chap. 1) defined shape as what is left when location (position of the distribution on the abscissa) and scale (the scale on the abscissa) are given up. They showed that shape cannot be defined in terms of the mathematical form of the distribution function. For example, the family of beta density functions have the same functional form, but differ widely in shape (Mosteller & Tukey, 1977, p. 9). However, one of the most striking properties of reaction time distributions is that in the main they all have roughly the same shape, being skewed to the right. (Occasionally normality is claimed for simple reaction time distributions, but this is probably not true [Mosteller, & Tukey, 1977, p. 11].) The group distribution method is concerned with averaging over subjects while preserving distribution shape, which for distribution functions shaped like reaction time distributions often turns out to be much the same as preserving the functional form, as is shown later.

Reaction time distributions have been examined in some detail with respect to specific mathematical models. McGill (1963) provided an excellent summary of work prior to 1963 and presented formal theory for a number of latency models. Green and Luce (1971) have used transform techniques in conjunction with a specific decision model to decompose reaction time distributions into component distributions, and this method of decomposition has been used in testing a neural timing theory (Luce & Green, 1972). Hohle (1965), Snodgrass (1969), and Snodgrass, Luce, and Galanter (1967) have fitted various empirical distributions to choice and simple reaction time data. None of this work, however, provides a general approach to obtaining distributional or shape information.

Before proceeding to a discussion of methods, I briefly illustrate potential uses of distributional information by listing predictions made by four models about distribution shape. First, serial scanning models of item recognition that assume independent and identically distributed comparison stages predict (by the central limit theorem) that as the number of comparison stages increases, the skewness of the reaction time distribution will decrease, and so the distribution will become more normal in shape. Second, the Atkinson and Juola (1973) model of item recognition predicts bimodal reaction time distributions. Third, the multiple observations model for signal detection (Pike, 1973) predicts that when the count criteria increase, mean latency will increase and skewness will decrease. Fourth, the random walk model for item recognition (Ratcliff, 1978) predicts that as the relatedness between probe item and memory item decreases, the mode and mean of the reaction time distribution will diverge. These examples are meant only to indicate the kinds of predictions that models produce and thus the kinds of tests for which distributional analyses prove useful.

Group Reaction Time Distributions

In experimental psychology it is usual to generalize findings across subjects. Often this is done by averaging data over subjects and making inferences based on the group data. Unfortunately, if raw reaction times from several subjects were simply combined to obtain distributional information, then the group distribution would not reflect the shape of the individual distributions. As an illustration, consider two subjects' unimodal reaction time distributions with respective means of 500 msec and 900 msec, each with 100-msec standard deviations. Simply combining the data would give a bimodal distribution, and this would not reflect the unimodal, individual distributions.

If there are enough observations per subject cell, then the best way to obtain distributional information is to derive distributional or shape estimates for each subject cell and then average these estimates over subjects.

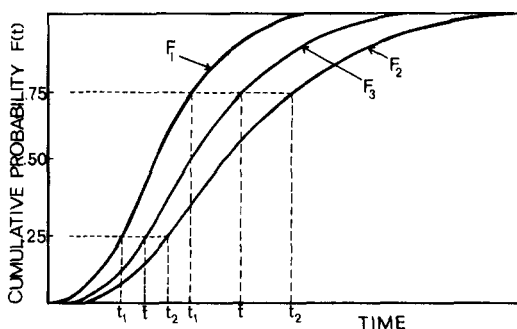


Figure 1. An example of Vincent (1912) averaging applied to cumulative distribution functions; $F_3(t)$ is the Vincent average curve of the curves $F_1(t)$ and $F_2(t)$, and $\bar{t}$ is the average of t_1 and t_2 .

For example, Ratcliff and Murdock (1976) have used the function arising from the convolution of the normal $[N(\mu, \sigma)]$ and exponential $[g(t) = (1/\tau)e^{-t/\tau}]$ distribution functions, $f(t)$, as an empirical summary of the shape of individual subjects' reaction time distributions (see also Hohle, 1965). Generalization was then accomplished by finding the average of each convolution parameter (μ , σ , and τ) across subjects. The expression for the convolution is

$$f(t) = \frac{e^{-[(t-\mu)/\tau] + \sigma^2/2\tau^2}}{\tau(2\pi)^{1/2}} \times \int_{-\infty}^{[(t-\mu)/\sigma] - \sigma/\tau} e^{-y^2/2} dy. \quad (1)$$

In a similar vein, Sternberg (Note 2) found four cumulants of distributions for each subject cell and generalized across subjects by averaging each cumulant across subjects for each cell. However, the usual experiment does not provide the number of observations required for these methods. Distribution information can still be obtained by using the group distribution method.

Group Distribution Method

The method is very similar to the technique devised by Vincent (1912) for plotting learning curves. In Vincent's procedure, each individual's learning curve is divided into equal fractions (number of trials to 10%, 20%, . . .), and performance of subjects at each fraction is summed and then

averaged. Figure 1 shows an example of this "Vincentizing" procedure applied to two cumulative reaction time distributions to produce the average cumulative distribution. In essence, reaction times at a fixed probability level (quantile) from the two distributions are averaged to give the mean quantile reaction time.

The procedure for estimating the sample quantiles is carried out as follows: Each subject's reaction times $T_1, \dots, T_n$ are arranged in ascending order of magnitude: $T_{(1)}, T_{(2)}, \dots, T_{(n)}$, where $T_{(i)}$ is the i th order statistic (David, 1970; Sarhan & Greenberg, 1962). From these ordered reaction times, q sample quantiles are estimated for each subject's data (generally with $q < n$). Each quantile is then averaged across subjects to give a mean $m\%$ sample quantile. In detail, suppose there are n observations for the first subject and one wishes to obtain q quantiles ($q < n$). Then for each subject, each ordered latency $T_{(i)}$ is replaced by q equal latencies, $T_{(i)}$, thereby forming a list that is the length of the product of q and n : $T_{(1)}, T_{(1)}, \dots, T_{(1)}, T_{(2)}, T_{(2)}, \dots, T_{(2)}, T_{(3)}, \dots$. To

calculate the first quantile, the first n latencies are summed and divided by n ; the second quantile is given by the sum of the next n latencies divided by n , and so on. This procedure is equivalent to simple linear interpolation. For example, if there were 14 responses and deciles were to be calculated, the first decile would be given by $(10/14)T_{(1)} + (4/14)T_{(2)}$, the second by $(6/14)T_{(2)} + (8/14)T_{(3)}$, the third by $(2/14)T_{(3)} + (10/14)T_{(4)} + (2/14)T_{(5)}$, and so on. When the quantiles have been calculated for each subject, each quantile is averaged across subjects to give group quantiles.

Group distribution histograms can be constructed by plotting quantiles on the abscissa and then constructing rectangles between adjacent quantiles such that all the rectangles have equal areas, as in Figure 2.

Several points about this method need discussion. First, the group distribution should be thought of as representing the distribution of the average subject, just as average reaction time represents the reaction time of the average subject. Second, order statistics are biased estimators of quantiles

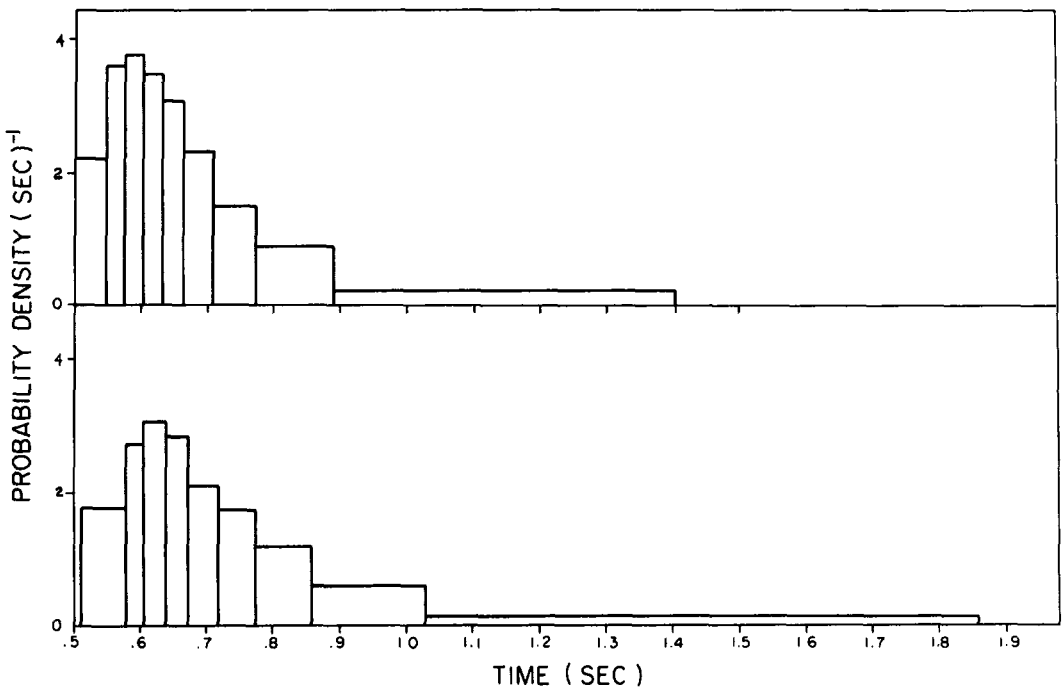


Figure 2. Two sample group reaction time distributions for 10% quantiles. (Data are from the three experiments reported later and represent correct rejections in Output Blocks 1 and 4.)

(David, 1970, chap. 4). However, if there are roughly the same number of observations for each subject cell and if the individual distributions for each subject cell have approximately the same shape, then the group distribution will reflect the same bias as the individual distributions. Third, it is a general problem that the shape of a group curve may not reflect the shape of individual curves. This problem was considered in great detail in the mid-1950s (Bakan, 1954; Estes, 1956; Hayes, 1953; Sidman, 1952; Spence, 1956). The general conclusion reached was that group curves often do not reflect the form of individual curves but that if care is taken group curves can be used to test hypotheses about individual curves. The next two sections examine this problem with respect to the Vincent averaging procedure, and later sections examine the problem with respect to Vincent averaging of reaction time distributions.

Some Exact Results for Vincentized Curves

Estes (1956) considered the problem of averaging learning curves and classified some simple functions into cases in which the functional form is not changed by averaging and cases in which the functional form is changed. Some similar results can be obtained for the Vincentizing procedure. (Note that distributions that differ only by a translation that is shifted along the time axis have the same form under Vincent averaging.) For Vincent averaging, it is necessary to obtain the following relationship: time as a function of cumulative probability (see Figure 1). Consider the exponential distribution. The cumulative probability distribution is given by

$$F(t) = 1 - e^{-t/\tau};$$

$$\therefore t = -\tau \ln [1 - F(t)]. \quad (2)$$

Consider the average of two exponential distributions with parameters τ_1 and τ_2 :

$$\bar{t} = (t_1 + t_2)/2$$

$$= -\frac{(\tau_1 + \tau_2)}{2} \ln [1 - F(t)]. \quad (3)$$

Thus the "Vincentized" average of n exponential distributions is exponential with parameter $\Sigma_{i=1}^n \tau_i/n$. For the Weibull distribution,

$$F(t) = 1 - e^{-(t/\tau)^\gamma},$$

with fixed parameter γ ; the Vincentized distribution is also a Weibull distribution, with parameter $\Sigma_{i=1}^n \tau_i/n$. Similarly, for the logistic distribution,

$$F(t) = 1/[1 + e^{-(t-\alpha)/\beta}];$$

the Vincentized distribution is also logistic, with parameters $\alpha = \Sigma_{i=1}^n \alpha_i/n$ and $\beta = \Sigma_{i=1}^n \beta_i/n$. Although normal distributions will not give exactly a normal distribution when Vincentized, the logistic distribution is a very good approximation to the normal distribution, so that any differences are probably very small. It should be noted that the exponential, Weibull ($\gamma > 1$), and logistic distributions have been postulated (or are very similar to distributions that have been postulated) to represent the distributions of processing stages in various models.

Vincentizing the Gamma Distribution

The gamma distribution has often been used to model reaction time distributions (McGill, 1963). Consider the gamma distribution with parameter 2 (i.e., the convolution of two exponential distributions):

$$F(t) = 1 - e^{-t/\tau}(1 + t/\tau). \quad (4)$$

By following an analysis similar to that presented above (Equations 2 and 3), it can be shown that Vincentized gamma distributions are not members of the gamma family. However, for all practical purposes the difference is negligible. For example, Vincentizing two gamma distributions (Equation 4) with parameters $\tau = 100$ msec and $\tau = 300$ msec gives 165, 276, 405, and 599 msec for the 20th, 40th, 60th, and 80th percentile points, respectively. The gamma distribution with parameter $\tau = 200$ msec has the corresponding points 165, 275, 404, and 599 msec. Thus Vincentizing gamma distributions produces a distribution that is very similar in shape to another gamma distribution.

The examples so far have all considered the Vincentizing of combinations of distributions that differ from one another only in the parameters that have dimensions of time, that is, parameters that represent the duration of some processing stage. There are other parameters that do not represent durations, for example, the number of convolved exponential distributions in the gamma distribution and γ in the Weibull distribution. Vincentizing distributions that vary in these parameters may not produce a distribution that is anything like the average subject's distribution. An extreme example of a distribution with this problem is the beta distribution. Mosteller and Tukey (1977, p. 9), in considering the problems involved in dealing with distribution shape, presented a figure showing the family of beta distributions to illustrate that even distributions belonging to the same family can differ widely in shape. From the graphs presented in Mosteller and Tukey, it can be seen that very serious problems may be involved in averaging across distributions of widely differing shape. To decide whether Vincent averaging will work in cases in which distribution shape varies widely among individual distributions, it is probably best to test the method as above or to perform some Monte Carlo tests as described in the next section.

Some Monte Carlo Studies Using the Convolution Model

The distribution that is the convolution of the normal and the exponential distributions (Equation 1) has been used as an empirical model of reaction time distributions (Ratcliff, 1978; Ratcliff & Murdock, 1976). The fits of the convolution to the data are good enough to make it reasonable to use the convolution in Monte Carlo studies testing the Vincentizing procedure. The Monte Carlo studies are presented to illustrate the use of the Vincentizing procedure under optimal conditions in which the form of the individual distributions is known.

To use the Monte Carlo method it is necessary to generate a random number from the convolution of normal and exponential

distributions. This can be accomplished by simply adding a random number generated from the normal distribution and a random number generated from the exponential distribution. Most computer systems have a random number generator that will produce random numbers between 0 and 1 from a rectangular distribution. Equation 2 can be used to produce exponentially distributed random numbers (with parameter τ) by substituting rectangularly distributed random numbers (*RND*) for $F(t)$. Normally distributed random numbers with mean μ and standard deviation σ can be obtained using the method proposed by Box and Muller (1958), as shown in Equation 5:

$$t = [-2 \ln(RND)]^{1/2} \cos(2\pi RND)\sigma + \mu. \quad (5)$$

Each Monte Carlo study consisted of several experiments (typically 50 to 100). In each experiment, 20 reaction times were obtained from each of 40 pseudosubjects. The 20 reaction times were arranged in ascending order and then averaged across subjects to give group 5% quantiles. The convolution model was then fitted to the set of 5% quantiles (5%, 10%, 15%, . . .) using the maximum likelihood method described in Ratcliff and Murdock (1976). Note that the quantile reaction times are derived from random variables; and so, strictly speaking, the parameter estimates do not have the nice properties of maximum likelihood estimators. However, estimating parameters this way is no worse than estimating parameters by, say, the least squares method, because the quantile reaction times are not independent and the expression being fitted is nonlinear. Results are shown in Table 1.

In general, the parameters μ and τ derived from fits to the Vincentized distribution are very close to the input values (used to generate the pseudodata). However, as τ increases (from .05 to .30), the value of σ (input value = .04) becomes more and more underestimated. This suggests that in any practical use, the value of σ is likely to be underestimated and less reliable than the values of μ and τ . It is interesting to note that the values of s_μ , s_σ , and s_τ are very close to the asymptotic variance estimates for the

Table 1
Monte Carlo Studies for the Convolution Model

Input parameter			Fitted parameter and standard error estimate						No. of experiments
μ	σ	τ	μ	s_μ	σ	s_σ	τ	s_τ	
.50	.04	.05	.5005	.0004	.0371	.0003	.0486	.0005	65
.50	.04	.15	.4996	.0004	.0325	.0004	.1498	.0006	101
.50	.04	.30	.5016	.0008	.0275	.0007	.2955	.0015	58/109 ^a
.50	.10	.50	.5002	.0014	.0749	.0012	.4994	.0024	98
.50	.04	.055-.250 ^b	.4980	.0006	.0314	.0006	.1545	.0012	52

Note. s = the standard error in the mean ($[\sum (X_i - M)^2 / n(n-1)]^{1/2}$).

^a 51 of the 109 experiments terminated with the fitted value of σ equal to zero.

^b The 40 pseudosubjects had different τ s, ranging from .055 to .250 in steps of .005, $M = .1525$.

convolution model presented in Ratcliff and Murdock (1976, Table 2). The last series of Monte Carlo experiments presented in Table 1 used 40 pseudosubjects with different τ values. The value of the average Vincentized τ was almost equal to the average input τ . This result shows that the Vincent-averaging properties of the exponential distribution carry over to parameter τ of the convolution to a good approximation.

These Monte Carlo studies show that application of Vincent's (1912) procedure to the distribution that is the convolution of normal and exponential distributions, a distribution that fits response latency distributions reasonably well, introduces little bias into parameter estimates.

Practical Test of the Group Distribution Method

To provide a stable data base for a practical examination of the method, three experiments (with four subjects per experiment) were combined, giving about 120,000 reaction times in total. The experimental procedure was the study-test recognition memory paradigm. The experiments have been reported as Experiments 2 and 1 in Ratcliff and Murdock (1976) and Experiment 1 in Ratcliff (1978); they are referred to here as Experiments 1, 2, and 3, respectively. A brief description of the study-test procedure is presented here; for further details, consult Ratcliff and Murdock.

In each of the three experiments, a list of study words was presented to the sub-

ject at about one word per sec, followed by a test list containing all the study words plus an equal number of new words in random order. For each word in the test list, the subject had to respond on a 6-point confidence scale ranging from *sure old* to *sure new*. Study lists were 16 words long, and test lists were 32 words long, except in Experiment 1 in which the study list contained 15 words and the test list 30 words. The list words were randomly sampled from the Toronto word pool (Okada, 1971). Repetitions of words were prohibited until at least two lists had intervened. The test list was self-paced, and words stayed in view until a response was made. In Experiment 1 and Experiment 3 rate of presentation of the study lists was varied between .5 sec and 2 sec per item. Effects on mean reaction time were small, on the order of 40 msec. In the following analyses, data from the different presentation-rate conditions are combined; this does not significantly affect distribution shape.

The experimental data are classified into eight cells, four output- or test-position blocks (2-8, 9-16, 17-24, and 25-32) for high-confidence hits, and the same four output-position blocks for correct rejections. (The first output position is excluded because this reaction time is typically several hundred msec slower than other reaction times in the test list.) This division of data gives about 1,200 observations for each of the 96 subject cells (12 subjects $\times$ 8 cells).

It was noted earlier that to test the group distribution method, properties derived from

Table 2

Convolution Model Fits to the Group Reaction Time Distributions and the Average Parameter Values From Fits to Individual Subject Distributions

Output block	Parameter value averaged over subjects			Group distribution value		
	μ	σ	τ	μ	σ	τ
Hit: $T < 5$ sec						
1	492	38	178	488	36	179
2	498	36	200	494	33	196
3	506	37	231	503	35	225
4	517	41	261	507	37	256
Correct rejection: $T < 5$ sec						
1	517	37	213	513	36	216
2	523	40	236	519	38	243
3	526	41	272	521	38	273
4	524	43	300	518	41	302
Hit: $T < 2$ sec						
1	497	41	158	495	39	157
2	502	38	175	499	36	172
3	513	40	192	510	39	186
4	525	45	210	517	42	205
Correct rejection: $T < 2$ sec						
1	523	40	185	520	38	186
2	530	43	200	525	41	202
3	536	45	218	533	44	223
4	538	49	225	532	48	229

Note. Data are truncated at 5 sec and at 2 sec.

the group distribution must be compared with the averages of the properties derived from the individual distributions. The properties chosen for comparison were the parameters of the convolution model, μ , σ , and τ (see Equation 1). Estimates of these parameters were obtained from the group distributions (2% quantiles) for each of the eight cells by fitting the convolution model to the group quantiles. (See Ratcliff & Murdock, 1976, for the maximum likelihood method of fitting.) Estimates of the parameters were obtained from the individual subject distributions by first fitting the convolution to each subject's distribution and then averaging the obtained estimates over subjects. The estimates given by the two procedures can be compared in Table 2: For two conditions each, estimates with latencies longer than 5 sec eliminated and estimates with latencies

longer than 2 sec eliminated. It can be seen that the two procedures give almost identical estimates of the convolution parameters. This supports the claim made earlier that the group distribution provides an unbiased summary of individual data.

Figures 2, 3, and 4 show some sample data. Figure 2 shows group reaction time distributions for correct rejections in Output Block 1 and in Output Block 4. Figure 4 shows group reaction time distributions and fits of the convolution model for hits in all four output blocks. The Figure 2 distributions are based on 20% quantiles, the Figure 4 distributions on 2% quantiles. Figure 3 shows some sample fits of the convolution model to reaction time distributions for individual subjects for hits in Output Block 1. Although the chi-squares are often significant (because the large numbers of observations make the

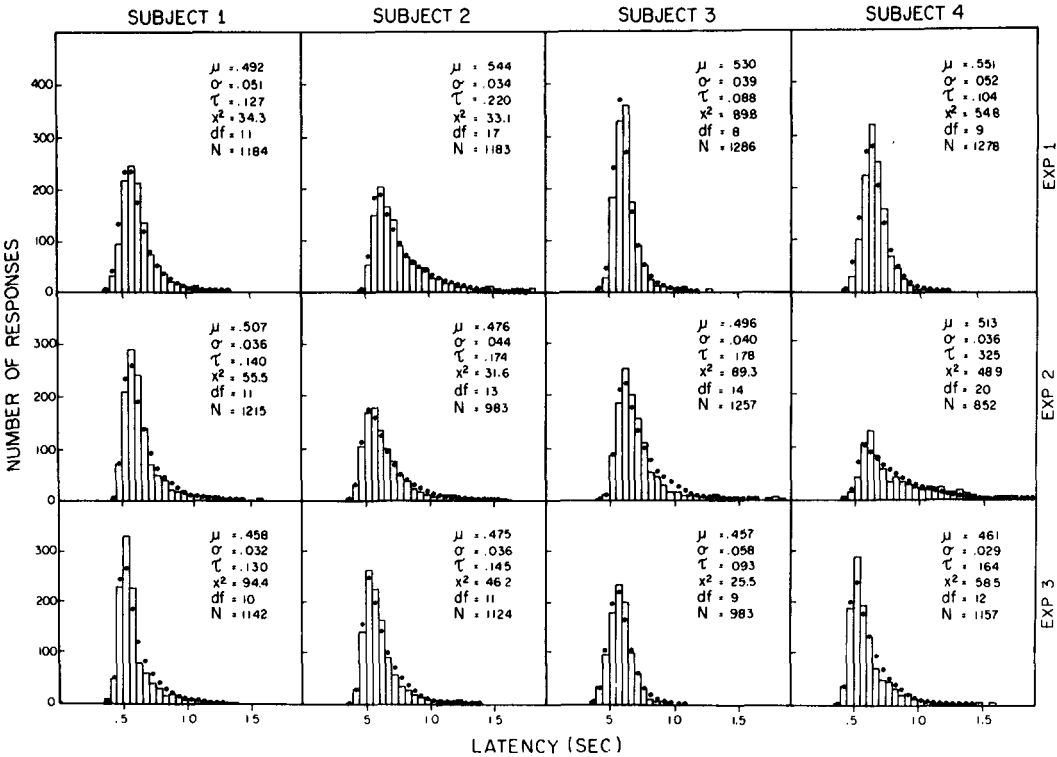


Figure 3. Empirical and fitted latency distributions for hits, Output Block 1. (Exp. 1 = Experiment 2, Ratcliff and Murdock, 1976; Exp. 2 = Experiment 1, Ratcliff and Murdock, 1976; Exp. 3 = Experiment 1, Ratcliff, 1978.)

chi-square a very powerful test), the fits are actually quite good; certainly the convolution captures the overall shape of the distribution. Problems associated with truncation and outliers are discussed in the section entitled Moments and Cumulants.

Probability Mixtures of Distributions and Bimodality

Occasionally a model is developed that predicts bimodal reaction time distributions arising from a probability mixture of processes (e.g., Atkinson & Juola, 1973). The question arises as to whether Vincent averaging across bimodal distributions from individual subjects will produce a bimodal group distribution. In general, the answer is only under conditions in which the proportion of responses in each process is approximately the same across subjects. For example, Figure 5 shows the Vincent average

cumulative distribution function for two distribution functions, each of which is a probability mixture of two processes. One distribution has a 25%–75% combination, and the other has a 75%–25% combination. The resulting group distribution is trimodal (i.e., has four points of inflection) and certainly does not reflect the bimodal nature of the individual distributions. In situations in which bimodality and probability mixtures of processes are expected, it is probably best to collect several hundred latencies per subject condition and investigate the individual latency distributions.

Moments and Cumulants

Moments have been used for many years to determine the shape of frequency curves either through skewness and kurtosis indices or by explicitly determining the frequency curve within Pearson's (cited in Elderton,

1906; Elderton & Johnson, 1969) system. Recently moments and cumulants have been used in the additive factor method for analysis of stage models (Sternberg, 1969; Sternberg, Note 2). In this section, three related problems in the use of moments and cumulants as sources of shape information are discussed. These problems are first that the variance associated with estimates of these measures is extremely large, second that the measures are very sensitive to outliers, and third that the measures give information about a part of the frequency curve that is of little theoretical interest.

To investigate the variability of moments and cumulants, expressions for moments and

cumulants and their standard deviations must be derived. These expressions are derived for an explicit distribution function to allow estimation of numerical values. The convolution of normal and exponential distributions is chosen because it approximates the shape of reaction time distributions.

Moments are defined as follows (Kendall & Stuart, 1969):

$$\mu'_1 = \int_{-\infty}^{\infty} tf(t)dt;$$

$$\mu_i = \int_{-\infty}^{\infty} (t - \mu'_1)^i f(t)dt, \text{ for } i > 1.$$

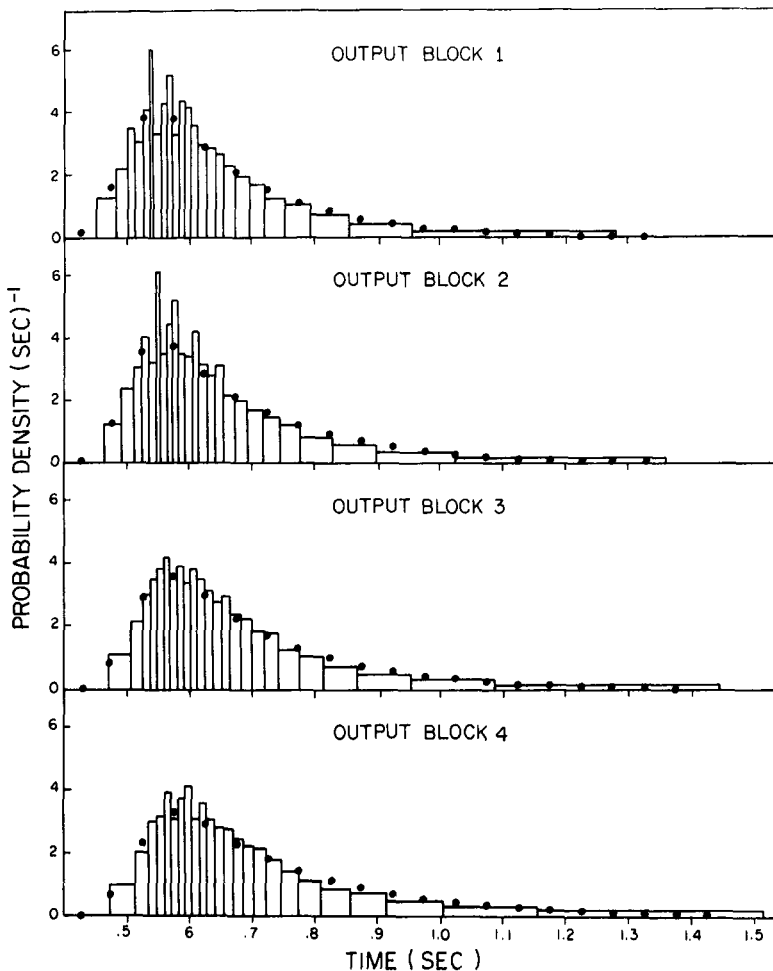


Figure 4. Group reaction time distributions for 2% quantiles for hits together with fits of the convolution model to the group distributions.

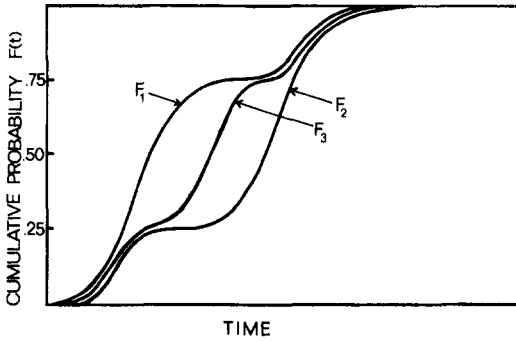


Figure 5. An example of group averaging of two bimodal distributions; $F_1(t)$ has 75% of the probability density in the first peak of the density function, $F_2(t)$ has 25% of the probability density in the first peak, and the resulting Vincent (1912) average distribution $F_3(t)$ is trimodal.

Cumulants are expressed by

$$\kappa_i = \mu_i, \quad \text{for } i \leq 3;$$

$$\kappa_4 = \mu_4 - 3\mu_2^2.$$

The sampling variances of the k statistics k_i (unbiased estimates of the cumulants κ_i) are given by Kendall and Stuart (1969):

$$\text{var}(k_2) = \frac{\kappa_4}{n} + \frac{2\kappa_2^2}{(n-1)},$$

$$\text{var}(k_3) = \frac{\kappa_6}{n} + \frac{9\kappa_4\kappa_2}{n-1} + \frac{9\kappa_3^2}{n-1} + \frac{6n\kappa_2^3}{(n-1)(n-2)},$$

$$\begin{aligned} \text{var}(k_4) = & \frac{\kappa_8}{n} + \frac{16\kappa_6\kappa_2}{n-1} + \frac{48\kappa_5\kappa_3}{n-1} \\ & + \frac{34\kappa_4^2}{n-1} + \frac{12n\kappa_4\kappa_2^2}{(n-1)(n-2)} \\ & + \frac{144n\kappa_3^2\kappa_2}{(n-1)(n-2)} \\ & + \frac{24n(n+1)\kappa_2^4}{(n-1)(n-2)(n-3)}. \end{aligned}$$

Expected values and variances of the k statistics for the convolution of normal and exponential distribution can now be calculated. For the normal distribution, $\kappa_1 = \mu$, $\kappa_2 = \sigma^2$, and $\kappa_i = 0$, for $i > 2$; and for the exponential distribution, $\kappa_i = \tau^i(i-1)!$ To convolve two distributions cumulants are added; so

for the convolution of normal and exponential distributions, $\kappa_1 = \mu + \tau$, $\kappa_2 = \sigma^2 + \tau^2$, $\kappa_3 = 2\tau^3$, and $\kappa_4 = 6\tau^4$. To estimate numerical values for cumulants and sampling variances of cumulants, values in the range of those found in Table 2 are used: $\mu = .5$ sec, $\sigma = .03$ sec, and $\tau = .2$ sec. Also, $\sigma^2 \ll \tau^2$, so that to an accuracy of 1% or 2% it is possible to neglect terms in σ compared with terms in τ . Now some numerical values can be calculated: For $n = 100$, $\kappa_2 = .040 \pm .011$, $\kappa_3 = .016 \pm .012$, and $\kappa_4 = .0096 \pm .0174$; for $n = 1,000$, $\kappa_2 = .040 \pm .004$, $\kappa_3 = .016 \pm .004$, and $\kappa_4 = .0096 \pm .0055$. From these values of cumulants and their estimated sampling standard errors, it can be seen that stability in the third and fourth cumulants is not achieved unless tens of thousands of observations contribute to the estimates. The same kind of instability can be seen in moments if corresponding sampling variances for moments are calculated (Kendall & Stuart, 1969).

The second problem with moments and cumulants is their sensitivity to outliers. There is a practical problem with outlier reaction times in that a proportion of these responses may be spurious, that is, they do not arise from the process under examination. For example, suppose distributional information is being used to evaluate a model that postulates a single retrieval process. Then an eyeblink, a moment's inattention, or a deliberate rest by the subject must be considered spurious for evaluation of the model. The sensitivity of moments and cumulants to these spurious outliers can be demonstrated by examining the effect of truncation on estimates of moments. Table 3 shows values of m'_1 , m_2 , m_3 , and m_4 , estimates of moments for the latency data used earlier (in obtaining group reaction time distributions). The effects of truncation are particularly striking. When 1% to 4% of the slower responses ($2 \text{ sec} < T < 5 \text{ sec}$) are eliminated, mean latency changes by between 20 and 50 msec, variance by a factor of two, and the third and fourth moments by an order of magnitude. Thus, excluding outliers three or more standard deviations above the mean ($m'_1 +$

Table 3
Moments for Latency Data Truncated at 5 sec and 2 sec

Output block	$T < 5 \text{ sec}$					$T < 2 \text{ sec}$				
	m'_1	m_2	m_3	m_4	n	m'_1	m_2	m_3	m_4	n
Hit										
1	6.70	6.98	.91	2.11	13,754	6.54	3.73	.18	.18	13,645
2	6.98	8.93	1.29	3.28	15,158	6.77	4.40	.21	.20	15,011
3	7.37	12.07	2.01	5.53	14,046	7.06	5.05	.25	.25	13,852
4	7.79	14.89	2.38	6.36	12,400	7.36	5.83	.28	.28	12,152
Correct rejection										
1	7.31	9.76	1.44	3.59	13,885	7.08	4.75	.24	.23	13,722
2	7.59	11.54	1.69	4.15	15,118	7.29	5.37	.28	.27	14,890
3	7.98	15.36	2.42	6.13	15,050	7.53	6.03	.30	.29	14,709
4	8.24	20.26	3.73	10.49	13,587	7.63	6.49	.33	.33	13,192

Note. The values for m'_1 are in units of 10 msec, for m_2 in units of 10^4 msec^2 , for m_3 in units of 10^8 msec^3 , and for m_4 in units of 10^{11} msec^4 ; m'_1 is mean latency, m_2 is variance and m_3 and m_4 are the third and fourth moments, respectively.

$3(m_2)^{1/2} < 2 \text{ sec}$) leads to enormous changes in higher moments.

The extreme sensitivity of higher moments to outliers is well-known, and Figure 6 illus-

trates the dependence of moments on tails of the frequency distribution. In Figure 6 is plotted the frequency distribution $f(x) = x^m e^{-1/x} / [\Gamma(m-1)]$ for $m = 10.6$, together

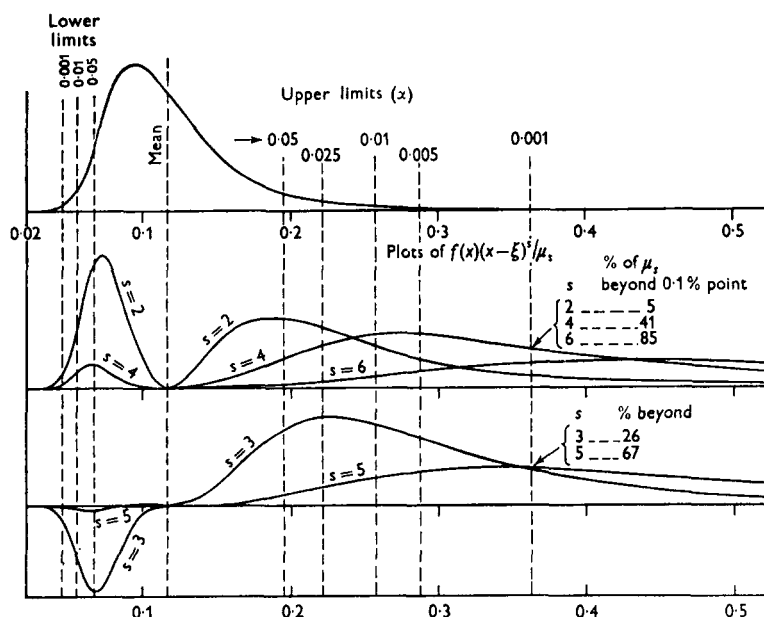


Figure 6. The distribution $f(x) = x^m e^{-1/x} / [\Gamma(m-1)]$, for $m = 10.6$. The lower curves show $f(x) = (x-\xi)^s / \mu_s$, which are the normalized contributions to the moments μ_s , as a function of x , where ξ is the mean. (From "Some Problems Arising in Approximating to Probability Distributions, Using Moments" by E. S. Pearson, *Biometrika*, 1963, 50, 95-112. Copyright 1963 by the Biometrika Trustees. Reprinted by permission.)

Table 4
Values of Skewness $(b_1)^{\dagger}$, Kurtosis b_2 , and Pearson's^a Measures of Skewness $S\kappa_o$ and $S\kappa_e$ for the Latency Data

Output block	$T < 5 \text{ sec}$				$T < 2 \text{ sec}$			
	$(b_1)^{\dagger}$	b_2	$S\kappa_o$	$S\kappa_e$	$(b_1)^{\dagger}$	b_2	$S\kappa_o$	$S\kappa_e$
Hit								
1	4.94	43.4	.447	.744	2.55	12.8	.533	.797
2	4.84	41.2	.475	.820	2.31	10.4	.577	.904
3	4.78	38.0	.478	.836	2.22	9.6	.596	.930
4	4.15	28.7	.492	.903	2.01	8.2	.605	1.008
Correct rejection								
1	4.71	37.7	.483	.794	2.35	10.3	.587	.849
2	4.32	31.2	.495	.828	2.24	9.4	.600	.865
3	4.02	26.0	.503	.860	2.05	8.0	.623	.882
4	4.09	25.6	.480	.873	2.01	7.8	.608	.904

Note. $S\kappa_o = (\text{mean} - \text{mode})/\text{standard deviation}$, where $\text{mean} = m'_1$ from Table 3, $\text{standard deviation} = (m_2)^{\dagger}$ from Table 3, and mode is calculated from the convolution fits for $T < 2 \text{ sec}$ by setting the first derivative of the probability density function to zero and estimating $t [f'(t) = 0]$. $S\kappa_e = 3(\text{mean} - \text{median})/\text{standard deviation}$.

^a Cited in Elderton (1906) and Elderton and Johnson (1969).

with the function $c(x) = (x - \xi)^s f(x) / \mu_s$ for $s = 2, 3, \dots, 6$. Note that

$$\mu_s = \int_{-\infty}^{\infty} (x - \xi)^s f(x) dx,$$

where ξ is the mean. The figure shows clearly the third problem with moments—that the higher moments of an asymmetrical long-tailed distribution depend on the form of the frequency function (and thus outliers) in a region of the tail that may be of no practical interest (Pearson, 1963).

Third and fourth moments are used as indices of skewness and kurtosis through $(\beta_1)^{\dagger} = \mu_3/(\mu_2)^{3/2}$ and $\beta_2 = \mu_4/\mu_2^2$, respectively. Pearson has proposed these alternative measures of skewness: $S\kappa_o = (\text{mean} - \text{mode})/\text{standard deviation}$ and, to avoid the use of the mode, $S\kappa_e = 3(\text{mean} - \text{median})/\text{standard deviation}$ (Kendall & Stuart, 1969). In Table 4 are shown values of $(b_1)^{\dagger}$, b_2 (estimates of $(\beta_1)^{\dagger}$ and β_2), $S\kappa_o$, and $S\kappa_e$ for the latency data used earlier. Note that the estimated value of the mode is rather unstable unless a fitted probability density function can be used to locate the mode (Elderton & Johnson, 1969). Thus, the mode used in the calculation of $S\kappa_o$ was obtained from the convolution fit to the group data (for $T < 2 \text{ sec}$)

by setting the first derivative of the probability density function to zero. The truncated distribution ($T < 2 \text{ sec}$) was chosen because inspection of fits of the convolution to individual subject's histograms indicated that the empirical histograms and fitted models did not differ systematically (see Figure 3).

A rather confusing picture of skewness estimates emerges from Table 4. By using $(b_1)^{\dagger}$ as the estimate of skewness, skewness decreases as output position increases, and skewness is halved by the elimination of 1% to 4% of longer reaction times. On the other hand, by using $S\kappa_o$ and $S\kappa_e$ as measures of skewness, skewness increases as output position increases, and the elimination of outliers results in a change of 10%–20% in $S\kappa_o$ and $S\kappa_e$. From the demonstration in Figure 6 and from the behavior of $(b_1)^{\dagger}$ and $S\kappa$, it must be concluded that the alternative measures of skewness, $(b_1)^{\dagger}$ and $S\kappa$, are concerned with different properties of the distribution function. Which measure should be used depends on whether behavior of the central portion of the distribution function (indicated by $S\kappa$) or behavior of the extreme tail of the distribution function [indicated by $(b_1)^{\dagger}$] is of interest.

I attempted to fit Pearson's (cited in Elderton & Johnson, 1969) system of frequency curves using the moments in Table 3. The curve belongs to Pearson's Type VI class, but the system of frequency curves is not flexible enough to encompass the distributions used in Table 3. The start of a Type VI distribution is at some value, $a > 0$ (Elderton & Johnson, 1969). Calculating the value of a for one set of data in Table 3 gave a value of a around 6 sec, which is beyond the distribution cutoff value. Thus it seems that Pearson's system of frequency curves may not be as flexible as is generally thought (see Patel, Kapadia, & Owen, 1976, for a list of those distributions that belong to Pearson's system and those that do not).

To summarize, moments and cumulants higher than variance have little to offer as sources of shape information about reaction time distributions because of their extreme variability and because they provide information about the extreme tails of the distribution that is of little practical interest. More reasonable sources of shape information are mean, mode, median and standard deviation, together with Pearson's $S\kappa_0$ and $S\kappa_1$ measures of skewness.

A Further Alternative to Moments and Cumulants

Another way to obtain shape information from reaction time distributions is to fit an explicit distribution function and use the parameters of this distribution as a summary of shape. Ratcliff and Murdock (1976) have used the distribution resulting from the convolution of normal and exponential distributions (Equation 1) as an empirical summary of reaction time distributions in memory retrieval paradigms. For simple and choice reaction time paradigms, Snodgrass et al. (1967) have shown that distributions with a rounded mode and exponential tail (e.g., the gamma and so probably the convolution of normal and exponential distributions) are inadequate as descriptions of distribution shape. The distribution they find to give the best fits to their data is the double monomial distribution.

Presenting information about reaction time distributions by providing the parameters of an explicit distribution function (that fits adequately) has the great advantage that it is easy for anyone to reconstruct a distribution (from the formula) that has nearly the same shape as the raw data. This may prove extremely valuable for mathematical modelers who may not wish to invest a large amount of time in obtaining raw data until some initial checks have been carried out. Further examples and discussion of the use of explicit distribution functions as approximations to reaction time distributions can be found in Ratcliff (1978) and Ratcliff and Murdock (1976).

Conclusions and Summary

Information about reaction time distributions can prove very useful in model construction and model testing, but there are few methods available for analysis of distributions. In this article I have presented a method for obtaining group reaction time distributions from experiments in which there are as few as 10 observations per subject cell. The method essentially involves estimating latency quantiles for each subject and then averaging these over the group of subjects. Several distributions were shown to average to give another distribution of the same family with parameters that were the mean of the parameters of the individual member distributions. Several Monte Carlo studies were performed using the distribution that is the convolution of a normal and an exponential distribution, a distribution used to fit reaction time distributions. These studies showed that the parameters derived from the group distributions were the same as the parameters used to generate the individual pseudosubject distributions. Fits of the convolution model to group distributions derived from data combined from three large experiments gave parameters that were almost identical to average parameters from fits to the distributions of individual subjects. The close correspondence between these methods of estimating group averages shows that group distributions provide an excellent summary of

distributional information for the group and do not introduce any systematic bias into the estimate of shape.

Methods of deriving shape information that use moments and cumulants were evaluated, and three major problems were pointed out. First, estimates of the higher moments and cumulants have large standard deviations; for example, 10,000 observations may be needed before the standard deviation on the fourth cumulant is as low as 10% of the size of the fourth cumulant. Second, estimates of moments from data are extremely sensitive to outlier reaction times; the addition of 1% slow responses can change the fourth moment by a power of 10. This problem is particularly severe if an unknown proportion of the slow latencies are spurious, that is, if they are not a result of processes under examination. Third, Figure 6 shows that the third and fourth moments tell one about portions of the distribution that may be of no theoretical interest. It is suggested that the mean and standard deviation together with estimates of median, mode, and Pearson's skewness measures (S_{κ_o} and S_{κ_e}) provide better information about distribution shape. These statistics are adequate, but may not be the most convenient statistics for conceptualizing the distribution or for fitting the distribution to more complex theoretical models. It is further argued that fitting adequate, explicit probability density functions to the observed reaction time distributions may provide more useful summaries of distributional information for researchers involved in mathematical modeling.

Reference Notes

1. Sternberg, S. *Evidence against self-terminating memory search from properties of RT distributions*. Paper presented at the meeting of the Psychonomic Society, St. Louis, Mo. November 1973.
2. Sternberg, S. *Estimating the distribution of additive reaction time components*. Paper presented at the meeting of the Psychometric Society, Niagara Falls, Canada, October 1964.

References

Atkinson, R. C., & Juola, J. F. Factors influencing speed and accuracy of word recognition. In S.

- Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press, 1973.
- Bakan, D. A generalization of Sidman's results on group and individual functions, and a criterion. *Psychological Bulletin*, 1954, 51, 63-64.
- Box, G. E. P., & Muller, M. E. A note on the generation of random normal deviates. *Annals of Mathematical Statistics*, 1958, 29, 610-613.
- David, H. A. *Order statistics*. New York: Wiley, 1970.
- Elderton, W. P. *Frequency curves and correlation*. London: Layton, 1906.
- Elderton, W. P., & Johnson, N. L. *Systems of frequency curves*. New York: Cambridge University Press, 1969.
- Estes, W. K. The problem of inference from curves based on group data. *Psychological Bulletin*, 1956, 53, 134-140.
- Green, D. M., & Luce, R. D. Detection of auditory signals presented at random times: III. *Perception & Psychophysics*, 1971, 9, 257-268.
- Hayes, K. J. The backward curve: A method for the study of learning. *Psychological Review*, 1953, 60, 269-275.
- Hohle, R. H. Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology*, 1965, 69, 382-386.
- Kendall, M. G., & Stuart, A. *The advanced theory of statistics* (Vol. 1, 2nd ed.). London: Griffin, 1969.
- Kintsch, W. *The representation of meaning in memory*. Hillsdale, N.J.: Erlbaum, 1974.
- Luce, R. D., & Green, D. M. A neural timing theory for response times and the psychophysics of intensity. *Psychological Review*, 1972, 79, 14-57.
- McGill, W. J. Stochastic latency mechanisms. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1). New York: Wiley, 1963.
- Mosteller, F., & Tukey, J. W. *Data analysis and regression: A second course in statistics*. Reading, Mass.: Addison-Wesley, 1977.
- Okada, R. Decision latencies in short-term recognition memory. *Journal of Experimental Psychology*, 1971, 90, 27-32.
- Patel, J. K., Kapadia, C. H., & Owen, D. B. *Handbook of statistical distributions*. New York: Dekker, 1976.
- Pearson, E. S. Some problems arising in approximating to probability distributions, using moments. *Biometrika*, 1963, 50, 95-112.
- Pike, R. Response latency models for signal detection. *Psychological Review*, 1973, 80, 53-68.
- Ratcliff, R. A theory of memory retrieval. *Psychological Review*, 1978, 85, 59-108.
- Ratcliff, R., & Murdock, B. B., Jr. Retrieval processes in recognition memory. *Psychological Review*, 1976, 83, 190-214.
- Sarhan, A. E., & Greenberg, B. G. *Contributions to order statistics*. New York: Wiley, 1962.

- Sidman, M. A note on functional relations obtained from group data. *Psychological Bulletin*, 1952, 49, 353-374.
- Snodgrass, J. G. Foreperiod effects in simple reaction time: Anticipation or expectancy. *Journal of Experimental Psychology Monograph*, 1969, 79(3, Pt. 2).
- Snodgrass, J. G., Luce, R. D., & Galanter, E. Some experiments on simple and choice reaction time. *Journal of Experimental Psychology*, 1967, 75, 1-17.
- Spence, K. W. *Behavior theory and conditioning*. New Haven, Conn.: Yale University Press, 1956.
- Sternberg, S. Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, 57, 421-457.
- Vincent, S. B. The function of the vibrissae in the behavior of the white rat. *Behavioral Monographs*, 1912, 1, No. 5.

Received February 23, 1978 ■