**The Stanford Institute for Mathematical Studies in the Social Sciences**

APPLIED MATHEMATICS AND STATISTICS LABORATORIES
STANFORD UNIVERSITY

*Reprint No. 60*

# Some Current Developments in Models of Learning for a Continuum of Responses

PATRICK SUPPES

Reprinted from
*American Institute of Electrical Engineers*
*The Discrete Adaptive Processes Symposium,*
June 1962

# SOME CURRENT DEVELOPMENTS IN MODELS OF LEARNING FOR A CONTINUUM OF RESPONSES*

Patrick Suppes
Departments of Philosophy and Statistics
Stanford University

## 1. INTRODUCTION

Some members of the audience may wonder why a paper that refers to a continuum of responses is part of a symposium on discrete adaptive processes. At least one simple answer to such a query is that the models to be discussed have a discrete time parameter even if the state space is itself continuous. I also think that discrete time processes with continuous states represent an important area of application for engineers interested in adaptive processes. Within experimental psychology models of learning for a continuum of responses have not as yet been developed to the degree that have models restricted to a finite number of responses. However, since 1959 I have been concerned with such models for a continuum of responses and together with several collaborators have made a certain amount of progress in developing and testing a theory about them (Suppes (1959), Suppes (1960), Suppes and Atkinson (1960, Ch. 12), Suppes and Frankmann (1961), Suppes and Zinnes (1961)).

The fundamental theory of these papers has built on stimulus sampling theory as first formulated in a basic paper by Estes (1950). The formulation of the theory to be given here resembles most closely that given for the finite case in Suppes and Atkinson (1960, p. 5). Roughly speaking, the theory runs as follows. On each trial the subject samples a set of stimuli each of which is in a certain state of conditioning. The probability that the subject will make a given response is postulated to be a simple function of the state of conditioning of the stimuli sampled. After a response occurs the subject is given a reinforcement and this reinforcement affects the state of conditioning of the sampled stimuli. The precise mechanism of this change is described below. What is important to note is that theoretically the learning processes are described in terms of the states of conditioning of the stimuli and these states are not directly observable.

The fundamental assumption of the finite theory is that each stimulus is conditioned to exactly one response (in some cases it is desirable to assume at most one response). For a continuum of responses the assumption of conditioning to exactly one response is psychologically unrealistic and mathematically awkward. In Suppes (1960) the assumption was introduced that the conditioning of each stimulus is smeared over a certain interval of responses and the conditioning of any stimulus is represented by a smearing distribution.

Various intuitive arguments can be given to support the representation of conditioning by a smearing distribution. A priori it might seem that the best one would be based on psychophysical facts about the impossibility of making an exact discrimination among a continuum of choices. From this standpoint the smearing distribution would arise from the perceptual inability of the organism to distinguish exactly the given point on the continuum. Unfortunately the empirical estimates of the smearing distribution, which are mentioned below, yield a variance of the distribution which is far in excess of that which can be justified on purely psychophysical grounds. At the present time it seems that a better argument can be put together in terms of ideas of response generalization. I would wish however to emphasize that the derivation of the smearing distribution from more primitive or fundamental psychological ideas has not yet been carried through in detail in any satisfactory form.

Although the applications I shall discuss for stimulus sampling theory as applied to a continuum of responses are all in the context of psychological experiments with adult human subjects, engineers concerned with adaptive processes will be quick to recognize the possible applications of this theory to pattern recognition and related problems. From the standpoint of statistical decision theory it is apparent that the theory does not yield an optimal solution at asymptote to many simple choice problems. But it is also just as apparent that the solution is from a conservative standpoint usually a fairly good one. The mechanisms by which the solution is reached are sufficiently simple that it may be the case they would prove practical for engineers to use in designing adaptive mechanisms. In any case let me proceed now to the statement of the fundamental assumptions of the theory.

## 2. AXIOMS

The axioms are formulated verbally but with some effort to convey a sense of formal precision. It is not difficult to convert them into a mathematically exact form; they fall naturally into three groups. I first state the conditioning axioms, then the sampling axioms and finally the response axioms. In the statement of the axioms I use x for the response variable and z for the parameter of the smearing distribution $K_s(x; z)$ of any stimulus s. Moreover for the cases to be discussed subsequently, and therefore in the axioms, I shall let z be the mode of the smearing distribution. In many applications it is also reasonable to assume that the mode and the mean of the smearing distribution are identical.

### CONDITIONING AXIOMS

C1. For each stimulus s there is on every trial a unique smearing distribution $K_s(x; z)$ on the interval $[a, b]$ of possible

* This research has been supported by the United States Air Force Office of Scientific Research under Contract AF 49(638)-1037.

AIEE Paper

*Reprinted from the American Institute of Electrical Engineers 1962 Joint Automatic Control Conference*

responses such that (a) the distribution $K_s(x; z)$ is determined by its mode z and its variance; (b) the variance is constant over trials for a fixed stimulating situation; (c) the distribution $K_s(x; z)$ is continuous and piecewise differentiable in both variables.

C2. If a stimulus is sampled on a trial, the mode of its smearing distribution becomes, with probability c, the point of the response (if any) which is reinforced on that trial; with probability 1 - c the mode remains unchanged.

C3. If no reinforcement occurs on a trial, there is no change in the smearing distributions of sampled stimuli.

C4. Stimuli which are not sampled on a given trial do not change their smearing distributions on that trial.

C5. The probability c that the mode of the smearing distribution of a sampled stimulus will become the point of the reinforced response is independent of the trial number and the preceding pattern of occurrence of events.

## SAMPLING AXIOMS

S1. Exactly one stimulus is sampled on each trial.

S2. Given the set of stimuli available for sampling on a given trial, the probability of sampling a given element is independent of the trial number and the preceding pattern of occurrence of events.

## RESPONSE AXIOMS

R1. If the sampled stimulus s and the mode z of its smearing distribution are given, then the probability of a response in the interval $[a_1, a_2]$ is $K_s(a_2; z) - K_s(a_1; z)$.

R2 This probability of response is independent of the trial number and the preceding pattern of occurrence of events.

Remarks about these axioms are to be found in the various references already cited. Here I shall restrict myself to a few. Axioms C5, S2 and R2 have a special status as independence of path assumptions. They are the kind of general axioms needed to make the mathematical scheme a workable one. It is primarily on the basis of them that we can prove that for many simple situations models of the theory constitute a discrete-time continuous-state Markov process in the states of conditioning of the stimuli.

The axioms depend essentially upon four concepts, those of a stimulus, a response, a reinforcement and a smearing distribution for the conditioning of the stimulus. These concepts are affected and related by two processes, those of conditioning and sampling. (The precise classification as between concepts and processes is not an important matter. It would be easy to talk about the process of reinforcement rather than the concept of reinforcement.) It is also important to remark that in other versions of the theory it is not always assumed that exactly one stimulus is sampled on each trial. In many cases it is reasonable and desirable to assume that a number of stimuli are sampled. Because this generalization of Axiom S1 introduces additional mathematical complexities, it shall not be considered here.

Before turning to several applications of the theory it will be desirable to state some general theorems for which a certain amount of notation is needed. I have already indicated I shall use x for the response random variable and z for the parameter of the smearing distribution. (In fact, I shall not introduce explicit notations for these random variables but only notations for their values. Thus to be quite precise, x is the value of the response random variable and z is the value of the smearing-parameter random variable.) I use $R_n$ for the distribution of the response random variable on trial n and $r_n$ for the corresponding density. Similarly $G_{s,n}$ is the distribution of the smearing-parameter random variable for stimulus s on trial n with density $g_{s,n}$. Values of the reinforcement random variable are designated by y and the distribution on trial n by $F_n$ with the density $f_n$.

The first general theorem concerns the response density; it relates $r_n(x)$ to the smearing distributions and distributions of the parameters of the smearing distributions on trial n. (Proofs of Theorems 1 and 2 are to be found in Suppes (1960).)

Theorem 1. (General Response Theorem.) If there are N stimuli all equally likely to be sampled, then the mean response density on trial n is given by:

$$r_n(x) = \frac{1}{N} \sum_s \int_a^b k_s(x;z) g_{s,n}(z)dz \quad .$$

The next theorem states the general Markov property of the sequence of smearing-parameter random variables. The restrictions on the reinforcement distribution stated in the hypothesis will be satisfied in our applications to be discussed below.

Theorem 2. (Markov Theorem). If the reinforcement distribution $F(y)$ on trial n is independent of n and depends only on the immediately preceding response on trial n, then the sequence of N-tuples of smearing-parameter random variables is a discrete-type continuous-state Markov process.

The following is an important theorem about the recursion for the density $g_{s,n}$ of the smearing parameter $z_s$ of stimulus s. It strengthens the theorem for N = 1 in Suppes [1960].

Theorem 3. If there are N stimuli all equally likely to be sampled, then

$$g_{s,n+1}(z) = (1 - \frac{c}{N}) g_{s,n}(z) + \frac{c}{N} f_n(z) \quad .$$

Proof. In the proof we need notation for the sampling random variable and the effectiveness-of-conditioning random variable. Let $\sigma$ be the former and $e_i$ the latter, with $e_0$ meaning conditioning was not effective, the probability of which is 1 - c, and $e_1$ meaning that it is. We also use the Dirac delta function. Thus if f(x) is a continuous density defined on the interval (a, b) and $a_0$ is in this interval, then

$$\int_a^b \delta(x - a_0) f(x)dx = f(a_0) \quad .$$

To begin with, we expand $g_{s,n+1}(z)$ in terms of the joint density $j_{n+1}$.

$$g_{s,n+1}(z) = \sum_\sigma \sum_i \int_a^b \int_a^b \int_a^b j_{n+1} \cdot$$

$$(z, e_{i,n}, y_n, x_n, \sigma_n, z_n)dy_n dx_n dz_n \quad .$$

We note first that for $\sigma \neq s$, we may sum out i and integrate, by virtue of Axiom C4 to obtain

$$(1) \qquad g_{s,n+1} = \frac{N-1}{N} g_{s,n}(z) + \frac{1}{N} \sum_i \int_a^b \int_a^b \int_a^b j_{n+1} \cdot$$

$$(z, e_{i,n}, y_n, x_n, s, z_n)dy_n dx_n dz_n \quad .$$

We consider now the second term on the right-hand side. Conditionalizing, we obtain

$$g_{s,n+1}(z) = \frac{N-1}{N} g_{s,n}(z) +$$

$$\frac{1}{N} \sum_i \int_a^b \int_a^b \int_a^b j_{n+1}(z|e_{i,n}, y_n, x_n, s, z_n)$$

$$j_n(e_{i,n}|y_n, x_n, z_n) j_n(y_n x_n z_n)dy_n dx_n dz_n \quad .$$

Now by Axioms C2 and C5, $j_{n+1}(e_{1,n}|y_n, x_n, z_n) = c$ and $j_{n+1}(e_{0,n}|y_n, x_n, z_n) = 1 - c$. When i = 1, we have $z = y_n$ by Axiom C2, and thus we replace $j_{n+1}(z|c_{1,n}, y_n, x_n, s, z_n)$ by $\delta(z - y_n)$, and when i = 0 the corresponding expression by $\delta(z - z_n)$. We then have

$$g_{s,n+1}(z) = \frac{N-1}{N} g_{s,n}(z) + \frac{c}{N} \int_a^b \int_a^b \int_a^b$$

$$\delta(z - y_n) j_n(y_n x_n z_n)dy_n dx_n dz_n$$

$$+ \frac{1-c}{N} \iiint \delta(z - z_n) j_n(y_n x_n z_n)dy_n dx_n dz_n \quad .$$

Integrating out $x_n$ and $z_n$ in the first case and $y_n$ and $x_n$ in the second we obtain

$$g_{s,n+1}(z) = \frac{N-1}{N} g_{s,n}(z) + \frac{c}{N}\int_a^b \delta(z-y_n) f_n(y_n)dy_n$$

$$+ \frac{1-c}{N}\int_a^b \delta(z-z_n) g_{s,n}(z_n)dz_n$$

$$= \frac{N-1}{N} g_{s,n}(z) + \frac{c}{N} f_n(z) + \frac{1-c}{N} g_{s,n}(z)$$

$$= (1 - \tfrac{c}{N})\, g_{s,n}(z) + \tfrac{c}{N} f_n(z) \quad ,$$

which is the desired result.

On the basis of Theorem 3 we may next derive a general recursion for the mean response density $r_n(x)$.

Theorem 4. If there are N stimuli all equally likely to be sampled, then

$$r_{n+1}(x) = (1 - \tfrac{c}{N}) r_n(x) + \tfrac{c}{N}\left[\tfrac{1}{N}\sum_s \int_a^b k_s(x;z) f_n(z)dz\right] .$$

Moreover, if all stimuli have the same smearing distribution $K(x; z)$, then

$$r_{n+1}(x) = (1 - \tfrac{c}{N}) r_n(x) + \tfrac{c}{N}\int_a^b k(x;z) f_n(z)dz \quad .$$

Proof. From Theorem 1 we have

$$r_{n+1}(x) = \tfrac{1}{N}\sum_s \int_a^b k_s(x;z) g_{s,n+1}(z)dz \quad .$$

Applying Theorem 3 to the right-hand side, simplifying and then again applying Theorem 1, we obtain

$$r_{n+1}(x) = \tfrac{1}{N}\sum_s \int_a^b k_s(x;z)[(1 - \tfrac{c}{N}) g_{s,n}(z) + \tfrac{c}{N} f_n(z)]dz$$

$$= (1 - \tfrac{c}{N})[\tfrac{1}{N}\sum_s \int_a^b k_s(x;z) g_{s,n}(z)dz + \tfrac{c}{N}[\tfrac{1}{N}\sum_s \int_a^b k_s(x;z) f_n(z)dz]$$

$$= (1 - \tfrac{c}{N}) r_n(x) + \tfrac{c}{N}\int_a^b k(x;z) f_n(z)dz \quad ,$$

if all smearing distributions are identical.

The first three applications described below have a non-contingent reinforcement procedure, i.e., the distribution F(y) is independent of n and the past. We have at once from preceding theorems the following integral equation for the asymptotic response density.

Theorem 5. If all N stimuli have the same smearing distribution $K(x; z)$ then in the non-contingent case

$$r(x) = \lim_{n\to\infty} r_n(x) = \int_a^b k(x;y) f(y)dy \quad .$$

For detailed examination of the empirical adequacy of the theory in relation to experimental data a good many more consequences of the axioms need to be derived. However, in the present survey the comparison of theory and data must be limited and therefore somewhat superficial.

## 3. APPLICATIONS

Because the first four experiments to be discussed were performed on the same apparatus, it will be useful to begin with a brief description of this apparatus. The subject was told that his task on each trial is to predict where a target is located on the circumference of a fixed circle. In the first three experiments at the end of each trial the "correct" position of the target was shown to the subject; this was the reinforcing event for that trial.

From the subject's standpoint the apparatus appeared as a vertical blue plastic screen six feet square mounted eight inches above the floor. At either side hung heavy white curtains which blocked the subjects' view of the area beyond the screen. A 3/8 in. steel shaft supported horizontally in a bearing mounted on the back of the box projected 21 in. on the subjects' side of the screen. Within the enclosure light sources supported on a counter-balanced arm affixed to the shaft were arranged to project a 1/8 x 3/4 in. bar of red light onto the screen at either 1.25-ft. or 2.5-ft. from the shaft. By turning a knob on the end of a shaft the subject could make his prediction by positioning the light bar at any point on the 2.5 ft. or 5.0-ft. diameter circle about the axis of the shaft (the diameter used varied from group to group and from experiment to experiment).

A 23-in. diameter scale with 400 equally spaced divisions was mounted behind the enclosure to indicate the location of the light. The entire scale could be rotated to change the physical location of the zero point. By turning a knob on his end of the shaft the experimenter could position the light bar to indicate the reinforcement location when so desired.

A. Non-contingent unimodal reinforcement. This experimental application of the theory is reported in detail in Suppes and Frankmann (1961). The experiment was run on the apparatus described above with the simple non-contingent reinforcement distribution being the triangular distribution on the circumference of the circle (see Figure 1). Analytically this distribution is defined on the interval $(0, 2\pi)$ by

$$(1) \qquad f(y) = \begin{cases} \dfrac{y}{\pi^2} \quad , & 0 < y \le \pi \\ \dfrac{1}{\pi^2}(2\pi - y) \quad , & \pi < y \le 2\pi \quad . \end{cases}$$

For the moment let us consider only the asymptotic response distribution $r(x)$. From Theorem 5 we have

$$(2) \qquad r(x) = \int_0^{2\pi} k(x;y) f(y)dy \quad .$$

The experimenter selects f(y) and observes an empirical histogram corresponding to r(x). If r(x) were known analytically, we could under suitable circumstances solve the integral equation (2) for k(x;y). In actual fact, it is more sensible to proceed in a different way. For the circular display it is reasonable to assume that the smearing distribution k(x; y) about a point of reinforcement y is symmetric and therefore k(x; y) = k(t), where t = x - y. We assume further some simple analytical form for k(x; y) in which it has exactly one free parameter, for instance, a uniform distribution with width 2a centered on y, or a symmetric beta distribution with free parameter m. We then use the method of moments to estimate a or m, as the case may be, from the computed variance of the empirical response histogram corresponding to r(x).

Let us see what this approach yields for the experiment in question. The uniform smearing density with parameter a is defined by

$$(3) \qquad k(t) = \begin{cases} \dfrac{1}{2a} \quad , & |t| \le a \\ 0 \quad , & a < |t| \le \pi \quad , \end{cases}$$

where $-\pi \le t \le \pi$ and $0 < a < \frac{\pi}{2}$. (It simplifies matters to restrict the total width of the uniform smearing distribution to less than half the circumference of the circle; it would be surprising to find a larger estimated value than $a = \frac{\pi}{2}$, although in the present experiment a is closer to $\frac{\pi}{2}$ than anticipated a priori.) Combining (1), (2) and (3), we have for the asymptotic response density:

$$(4)\quad r(x) = \begin{cases} \frac{1}{2a\pi^2}(a^2 + x^2), & a < x \le a \\ \frac{x}{\pi^2}, & a < x \le \pi - a \\ \frac{1}{2a\pi^2}(2a\pi - a^2 - (x-\pi)^2), & \pi - a < x \le \pi \end{cases},$$

and for the variance of r(x):

$$(5)\quad \sigma^2(r(x)) = \frac{\pi^2}{6}\left[1 + \left(\frac{a}{\pi}\right)^2\left(2 - \frac{a}{\pi}\right)\right].$$

The symmetric beta smearing density with parameter m is defined by:

$$(6)\quad k(t) = \frac{1}{\pi\beta(m,\frac{1}{2})}\left(1 - \frac{t^2}{\pi^2}\right)^{m-1},$$

for $-\pi < t < \pi$ and $\beta(m, \frac{1}{2})$ the usual beta coefficient. In this case for the asymptotic response density we have

$$(7)\quad r(x) = \frac{1}{\pi^2}\left\{\int_{-\pi}^{x-\pi} k(t)(2\pi - x + t)dt + \int_{x-\pi}^{x} k(t)(x - t)dt + \int_{x}^{\pi} k(t)(t - x)dt\right\}.$$

In order to integrate (7) when (6) is substituted in, we restrict ourselves to integer values of m — this turns out to be no serious handicap. Integrating, then, we obtain for $0 < x < \pi$

$$(8)\quad r(x) = \frac{1}{\pi}\left\{\frac{(\pi^2 - x^2)^m - (\pi^2 - (x-\pi)^2)^m}{m\pi^{2m}\beta(m,\frac{1}{2})} + 1 - \frac{x}{\pi} + \frac{2}{\beta(m,\frac{1}{2})}\sum_{\nu=0}^{m-1}\binom{m-1}{\nu}(-1)^\nu \frac{x^{2\nu+2} - (x-\pi)^{2\nu+2}}{(2\nu+1)\pi^{2\nu+1}}\right\}.$$

(Because r(x) is symmetric about $\pi$, we do not give its form for $\pi \le x < 2\pi$.)

For the variance of r(x) as defined by (8) we have:

$$\sigma^2(r(x)) = \pi^2\left\{\frac{2}{2m+1} + \frac{3}{2(2m+3)(2m+1)} - \frac{2}{\beta(m,\frac{1}{2})}\left[\frac{1}{m(m+1)} - 2\sum_{\nu=0}^{m-1}\binom{m-1}{\nu}(-1)^\nu \frac{\beta(2\nu+3,3)}{2\nu+1}\right]\right\}.$$

In the experiment under discussion two groups of subjects were run for 300 self-paced trials without interruption (the average rate was 6 trials per minute). One group of 30 subjects used the 2.5-ft. circle, and the other group of 30 subjects the 5.0-ft. circle. Using the method of moments, as mentioned earlier, the parameters a and m were estimated for both groups from data on the last 200 trials.

The response histogram for the group of subjects using the 2.5-ft. circle had a mean of $.997\pi$ and a variance of $.224\pi^2$. From this variance it was inferred that $\hat{a} = .475\pi$ and $\hat{m} = 6$ (when m = 6 the predicted variance is $.222\pi^2$, which is sufficiently close to the observed value for the integer value of m not to introduce a serious error). The predicted response densities for $\hat{a} = .475\pi$ and $\hat{m} = 6$ are so nearly identical that only one r(x) is shown in Figure 1, together with f(y) and the empirical histogram based on the last 200 trials. (The maximum difference between the two response densities occurs at x = 0, for which $r_U(x) = .076$ and $r_{\mathcal{D}}(x) = .071$.)

The group of subjects using the 5.0-ft. circle has a response histogram with a mean of $.993\pi$ and a variance of $.210\pi^2$. It was thus inferred that $\hat{a} = .405\pi$ and $\hat{m} = 8$ (when m = 8 the predicted variance is $.211\pi^2$). The predicted response for $\hat{a} = .405\pi$ and the empirical histogram are shown in Figure 2.

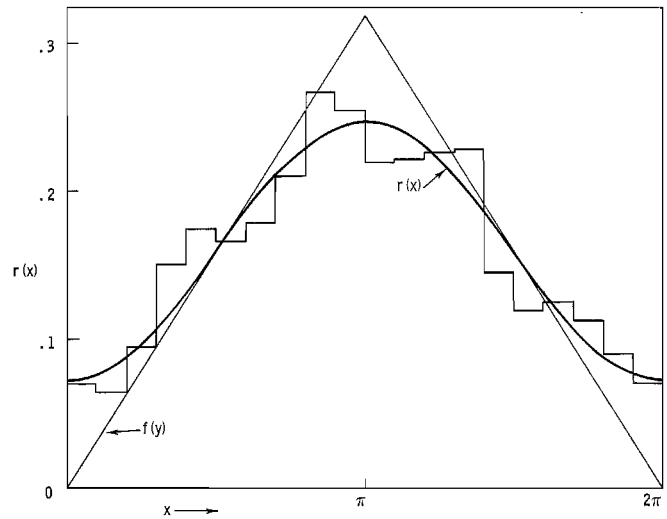


Figure 1

Response histogram for 2.5-ft. circle and fitted response density for last 200 trials (unimodal reinforcement).

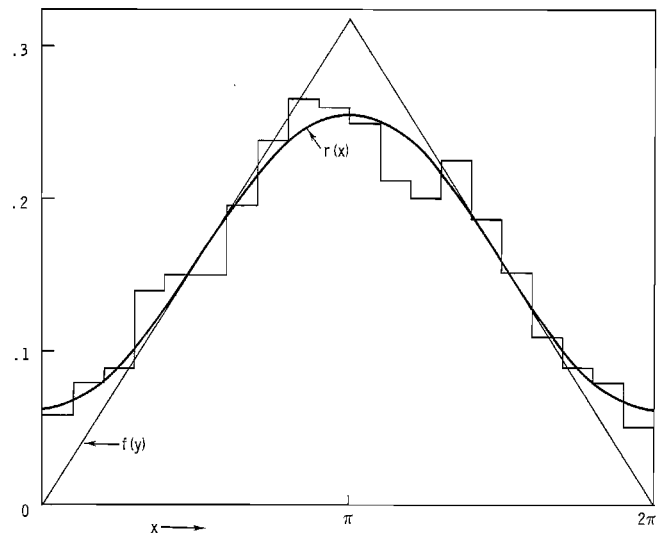


Figure 2

Response histogram for 5.0-ft. circle and fitted response density for last 200 trials (unimodal reinforcement).

Figures 1 and 2 suggest that the response densities fit the empirical asymptotic empirical histograms rather well, and in particular, that they fit them considerably better than does the triangular reinforcement distribution. In order to evaluate this goodness-of-fit hypothesis, the obvious procedure is to apply a $\chi^2$ test, but the difficulty is that the 6,000 observations on which each histogram and fitted response density are based are not statistically independent. Fortunately, it is known that for stationary stochastic processes observations sufficiently spaced in trials are statistically independent. Using this idea, we considered responses on every fifth trial for each subject starting with Trial 102. Standard tests of independence, for example, zero-order versus first-order Markov process, supported the null hypothesis of independence.

Using then the 1,200 approximately statistically independent responses for each circle, the following results were obtained. First, the null hypothesis that the response data fit the reinforcement distribution was tested. On the basis of dividing the circumference of the circle into 30 class intervals of approximately equal frequency for the 2.5-ft. circle data, the rejection of the null hypothesis was decisive. For the 2.5-ft. circle $\chi^2 = 154.21$ with 29 df, and for the 5.0-ft. circle $\chi^2 = 74.02$ with 29 df; both results are highly significant.

Secondly, we tested for each circle the null hypothesis that the response distribution based on the estimated parameter a for the uniform smearing distribution fit the response data. For the 2.5-ft. circle $\chi^2 = 47.56$ with 28 df, for which $P < .02$. For the 5.0-ft. circle $\chi^2 = 30.09$, also with 28 df, for which $P > .30$. The somewhat disappointing result for the 2.5-ft. circle is entirely due to one cell which had an observed frequency of 66, whereas a frequency of 40.68 was predicted. The fit of the remaining 29 cells was quite good, and in general the $\chi^2$ results support the relatively good fit indicated by Figures 1 and 2.

It is interesting to note that, as predicted by the theory, the variance of the response distribution was greater than that of the reinforcement distribution. This general conclusion is opposite to that of statistical decision theory for the rational strategy, asymptotically, is to respond very closely to the mean of the reinforcement distribution and thus to have a variance that is approximately zero. What is particularly noteworthy is that this conclusion about the mean response histogram is also supported by an analysis of individual data. Twenty-seven of the 30 subjects on the 2.5-ft. circle and 22 of the 30 subjects on the 5.0-ft. circle had variances greater than that of the corresponding reinforcement histograms.

Some additional remarks on the appropriate way to approach the problem of analyzing the data more thoroughly are given below.

B. Non-contingent bimodal reinforcement. In this second experiment, as yet unpublished, we used a bimodel non-contingent reinforcement distribution rather than a unimodal one. The distribution actually used was simply what we might call the bimodal triangular distribution, that is, the density from 0 to $\pi$ was an isosceles triangle and the same triangle represents the distribution from $\pi$ to $2\pi$ (see Figure 3). Detailed analysis of this experiment, which was originally run with Raymond W. Frankmann, is now underway in collaboration with Michael Levine and Henri Rouanet. In this experiment 30 subjects were run for 600 trials on the apparatus already described. In Figure 3 are shown the histogram of responses for Trials 301-450, together with the reinforcement distribution, and the response histogram for Trials 451-600. Both of these histograms show the same general tendencies as do Figures 1 and 2. In the first place, the response histograms fit the reinforcement distribution very closely. On the other hand, the response histograms have a greater spread than the reinforcement distribution, which is particularly noticeable in the distribution of the tails. A more detailed analysis of the data will not be given here.

C. Simple discrimination with uniform reinforcement distribution for each region. In this experiment, done in collaboration with Henri Rouanet, a discrimination situation is introduced. In addition to the apparatus described earlier, a panel of four lights (blue, green, red and yellow) is placed in front of the subject, and at the beginning of each trial one of the four lights goes on. Corresponding to each light, there is a different probability distribution for the reinforcement on the circle. The four reinforcement distributions were uniform with two of them equal to $\pi/5$ radians and two of them to $2\pi/5$. The four reinforcement regions were equally spaced on the circle, their relative positions remaining identical but the absolute locations being randomized for different subjects.

The theoretical approach to the discrimination situation used here is described in Suppes (1960). Essentially it consists of treating the learning connected with each region as a separate process, and thus each of the four processes reduces to a special case of the non-contingent theorems stated in Sec. 2.

One of the primary purposes of this experiment was to test more severely the results obtained under (A) and (B) that the variance of the mean response histograms at asymptote is greater than that of the reinforcement distribution. It seems

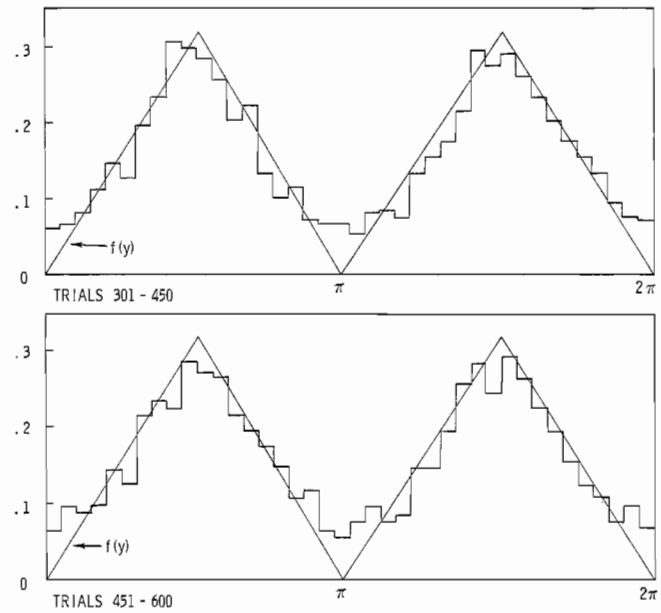


Figure 3
Response histograms for trials 301-450 and trials 451-600 (bimodal reinforcement).

reasonable to suppose that this might be less likely to occur in a situation of the kind just described, because the reinforcement is restricted always to a certain region. It would be natural for subjects eventually to respond to a given discriminating light only in the region of reinforcement.

Forty-four subjects were run for 500 trials, with two interruptions of about six minutes each after the 200th and 350th trial. The average rate was five trials per minute.

Because of the similarity of the data for the two small regions, as well as that of the two large regions, it has been possible to combine the data from the four regions into two regions, which I shall designate the small and large region respectively. In Figures 4 and 5 the empirical histograms for the two regions are shown, as based on the last 300 trials. (Thus, each histogram is based on $44 \cdot 300/2 = 6,600$ observations.)

Because only slightly more than fifty per cent of the individual subjects had a response variance at asymptote on the large region compatible with the theory formulated above, we considered some additional models in which the reinforcement was defined in a different manner. In Model II the effective reinforcement on trial n is defined to be the mean of the experimenter-determined reinforcement on trial n and n - 1. Thus the effective reinforcement distribution f'(y) is the distribution of the mean of two random variables distributed according to the density f(y):

$$f'(y) = 2\int_{-a}^{a} f(x)f(2y - x)dx \quad ,$$

where 2a is the reinforcement interval. In terms of developments already given it is then easily shown that the anticipated recursions in terms of f'(y) hold, for example,

$$g_{s,n+1}(z) = (1 - \tfrac{c}{N})g_{s,n}(z) + \tfrac{c}{N} f'(z) \quad ,$$

and at asymptote

$$r(x) = \int_{-a}^{a} k(x;y)f'(y)dy \quad .$$

The fitted response densities shown in Figures 4 and 5 are for Model II and use the equation just given. The subscript indicates whether a beta or uniform smearing distribution was used. Contrary to the situation described in (A), the two dif-

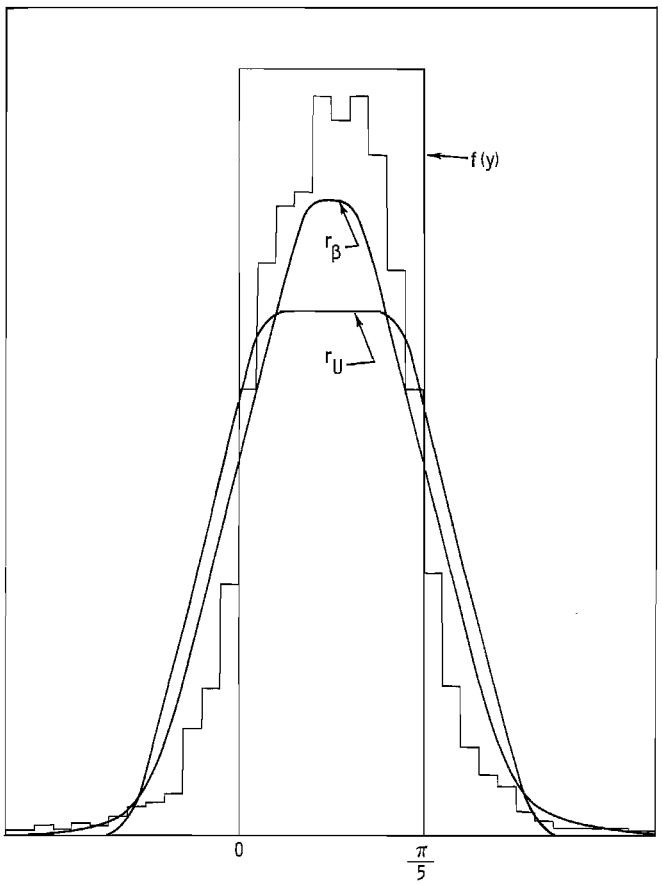


Figure 4

Response histogram for small region and fitted uniform and beta response densities for last 300 trials (simple discrimination experiment).

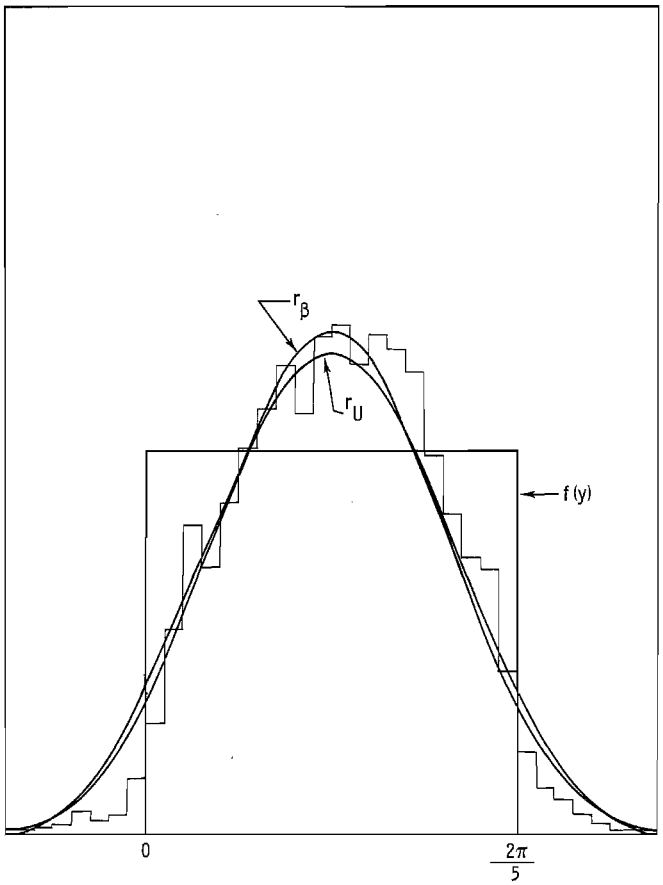


Figure 5

Response histogram for large region and fitted uniform and beta response densities for last 300 trials (simple discrimination experiment).

ferent kinds of smearing distributions are easily discriminated, and it is clear from the figures that the beta distribution fits better for both the small and large regions.

The consideration of other models and the detailed analysis of sequential data will be reported elsewhere.

For the remaining three applications to be presented here, no experimental data will be considered, although experiments have been performed for two of the three cases (not for the application to continuous games).

D. Non-determinate reinforcement. Reinforcement schedules are usually said to be determinate when the subject is completely informed of the correct response at the end of each trial. I summarize briefly here an application to the non-determinate case reported in detail in Suppes and Zinnes (1961). In the prototype experiment subjects are instructed to locate or "hit" an unseen target that is said to be located on the circumference of the circular apparatus already described. If the subject's response lies within a specified distance $\alpha$ of the target point he is informed that he has made a hit, otherwise a miss.

The relations between reinforcement and conditioning are postulated to be the following for this experimental schema. Axiom C2 is interpreted to mean that when a subject is informed that his response yielded a hit then with probability c the mode of the smearing distribution of the sampled stimulus becomes the point of his actual response. Moreover, Axiom C3 is interpreted to cover the case of a miss, i.e., it is postulated that no reinforcement occurs, and thus that there is no change in the smearing distribution of the sampled stimulus. (On general grounds it seems reasonable to challenge this interpretation of a miss as equivalent to no reinforcement, for some information is provided about the location of a target when the subject is informed that he missed. However, of the three main types of models discussed in Suppes and Zinnes (1961), the so-called identity model, generated by the assumptions just stated, fits our present experimental data the best.)

The general character of this application may be illustrated by deriving the expression for r(x), the mean asymptote response density. First of all, let us suppose that the target's position y is determined by a non-contingent distribution f(y), and let $\alpha$ be the maximum distance from y for a hit to be scored. Then the probability of a hit on trial n given response $x_n$ is

$$\pi_1(x_n) = \int_{x_n-\alpha}^{x_n+\alpha} f(y)dy \quad ,$$

and the probability of a miss is $\pi_0(x_n) = 1 - \pi_1(x_n)$. Following the same strategy used in the proof of Theorem 3, we obtain the following identities (for an arbitrary interval (a, b)):

$$g_{s,n+1}(z) = \frac{N-1}{N} g_{s,n}(z) + \frac{c}{N}\left[\int_a^b\int_{x_n-\alpha}^{x_n+\alpha} \delta(z - x_n) j_n(y_n x_n) dy_n dx_n \right.$$

$$\left. + \int_a^b\int_a^b\int_{y\notin(x_n-\alpha, x_n+\alpha)} \delta(z - z_n) j_n(y_n x_n z_n) dy_n dx_n dz_n \right]$$

$$+ \frac{1-c}{N}\int_a^b \delta(z - z_n) g_{s,n}(z_n)dz_n$$

$$= \frac{N-1}{N} g_{s,n}(z) + \frac{c}{N}\left[\int_a^b \delta(z - x)\pi_1(x)r_n(x)dx\right.$$

$$\left. + \int_a^b\int_a^b \delta(z - z_n)\pi_0(x)k(x;z_n)g_{s,n}(z_n)dx\,dz_n\right]$$

$$+ \frac{1-c}{N} g_{s,n}(z)$$

$$= (1 - \tfrac{c}{N})g_{s,n}(z) + \frac{c}{N}\left[\pi_1(z)r_n(z) + g_{s,n}(z)\int_a^b \pi_0(x)k(x;z)dx\right] .$$

Asymptotically this recursion simplifies to

$$g(z) = \left[\frac{\pi_1(z)}{\int_a^b \pi_1(x)k(x;z)dx}\right] r(z) ,$$

where the subscript s has been dropped, because of identical results for all stimuli on the assumption that the smearing distribution is the same for all. From Theorem 1 we then obtain at once that

$$r(x) = \int_a^b k(x;z)g(z)dz$$

$$= \int_a^b k(x;z)H(z)r(z)dz ,$$

where

$$H(z) = \frac{\pi_1(z)}{\int_a^b \pi_1(x)k(x;z)dx}$$

The thing to be noted about this expression for r(x) is that it involves r(z), which means that in every case an integral equation must be solved to find r(x). When the kernel k(x;z)H(z) is degenerate a solution may be found in a simple manner. Some examples are given in Suppes and Zinnes (1961).

E. Continuous games. To return to the simpler framework of determinate reinforcement, we may formulate a non-zero-sum, two-person game as follows. On each play of the game, i.e., on each trial, the players try to hit targets. The players' responses consist of predicting where the target will be on a continuum of possible positions. The position y of player i's target on trial n is given by the conditional density $f^i(y|x_1, x_2)$, where $x_j$ is player j's prediction.

For simplicity of notation let us also assume that each player has exactly one stimulus element to sample, i.e., let us restrict ourselves to a one-element model. The basic tool we use is to exploit the recursion of Theorem 3:

$$g_{n+1}(z) = (1 - c)g_n(z) + cf_n(z) .$$

For each player i, we need to find $f_n^i(y)$. To begin with, we have immediately from our basic assumptions

$$f_n^i(y) = \int_a^b\int_a^b\int_a^b\int_a^b f^i(y|x_1,x_2)k_1(x_1;z_1)\cdot k_2(x_2;z_2)g_n(z_1,z_2)dx_1dx_2dz_1dz_2 .$$

Substituting in the expressions for $f_n^1(y)$ and $f_n^2(y)$ and solving for the asymptotes, which it may be shown exist under very weak assumptions, we obtain the following pair of simultaneous integral equations for $g_1(z)$ and $g_2(z)$.

$$g_i(z) = \int_a^b\int_a^b\int_a^b\int_a^b f^i(z|x_1,x_2)k_1(x_1;z_1)k_2(x_2;z_2)\cdot g(z_1,z_2)dx_1dx_2dz_1dz_2$$

for i = 1, 2. In general we cannot solve this pair of equations because of the appearance of the joint distribution $g(z_1, z_2)$ on the right-hand side. Some solutions for special cases are given in Suppes and Atkinson (1960, Ch. 12). It is interesting to note that the difficulties of solution found here for the one-element model do not arise in the discrete case (see Suppes and Atkinson (1960, Ch. 1), but do arise already in the two-response case for the standard linear model (see Estes and Suppes (1959)).

As far as I know this application of the theory to continuous games has not yet been tested experimentally.

F. Latency distributions. It is somewhat surprising that the recent intensive experimental and theoretical consideration of all-or-none vs. incremental response conditioning has not been matched by a corresponding interest in the conditioning of latency distributions, i.e., the distributions of response times. Writing in 1959, Luce remarks correctly, I think, that the number of relevant papers in the literature is far less than would be expected, a priori. Perhaps part of the explanation of this situation is that the detailed derivation and analysis of latency distributions is a somewhat complicated matter. Some of these difficulties may be by-passed by obtaining results in general form and leaving aside the study of the exact analytical form of the latency distributions.

The all-or-none model can be described in the framework of stimulus sampling theory in the following simple manner. On each trial the subject samples a single stimulus, and a conditioning association between this stimulus and the correct response either exists or does not. If it does, the subject is in the conditioned state C, if not, in the unconditioned state U. In the paired-associate paradigm I wish to consider here the same response is always correct for a given stimulus display, so it is natural to exclude the case of a wrong conditioning association being established, contrary to the situation for experiments with a probabilistic reinforcement schedule. When the subject is in the conditioned state C the correct response is made with probability 1; when in the unconditioned state U, the correct response is made only with the guessing probability p, which is assumed constant and independent of previous responses as long as the subject is in state U.

The learning process in terms of responses is completely described by adding to the above assumptions those stating how the subject moves from the unconditioned to the conditioned state, i.e, from state U to state C. Although a more psychologically primitive formulation may be given, the basic assumption is just that there is a Markov process in these two states with transition matrix

| | C | U |
|---|---|---|
| C | 1 | 0 |
| U | c | 1 - c |

and with the initial probability of state U being 1. Assuming this matrix corresponds essentially to assuming that as long as the subject is in the unconditioned state there is a constant probability c of becoming conditioned on every trial; and once conditioned the subject remains so (according to the model).

It should be clear how to extend this simple model to response latency distributions. We simply postulate that when the subject is unconditioned there is a distribution U(t) of latency, and when he is conditioned there is a distribution C(t). Behaviorally the distribution U(t) corresponds to the guessing probability p for the correct response and C(t) to the probability 1 of a correct response in the conditioned state. It is natural to postulate that U(t) is independent of whether the guessed response is correct or not, just as the probability c of conditioning is independent of this fact.*

* It is not difficult to modify the axioms of Sec. 2 to include this latency model.

The mean probability $p_n$ of a correct response is given by the equation

$$p_n = 1 - (1-p)(1-c)^{n-1} , \tag{1}$$

as is well known from the literature. The corresponding expression for the mean response latency distribution $R_n(t)$ is easily shown by the methods of Sec. 2 to be:

$$R_n(t) = C(t)-(C(t)-U(t))(1-c)^{n-1} , \tag{2}$$

and immediately then for the mean $E(R_n)$:

$$E(R_n) = \bar{C}-(\bar{C}-\bar{U})(1-c)^{n-1} , \tag{3}$$

where $\bar{C}$ is the mean of C(t) and $\bar{U}$ the mean of U(t). It is also clear that an expression just like (3) holds for the $r^{th}$ raw moment of $R_n(t)$.

The expression for the variance of $R_n(t)$ is slightly more complicated but easy to derive in the form of a recursion

$$\sigma^2(R_n) = (1-c)\sigma^2(R_{n-1}) + c\sigma^2(C) + c(1-c)[\bar{C}-E(R_{n-1})]^2 , \tag{4}$$

where $\sigma^2(C)$ is, of course, the variance of the distribution C(t).

Unfortunately the kind of mean learning results represented by (2)-(4) are not too useful, for equations of precisely the same form hold in the incremental linear model. It is an already familiar fact in the literature (e.g., Bower (1961)) that the all-or-none and incremental linear models have the same mean response curve. Fortunately, on the other hand, the statistical techniques developed in Suppes and Ginsberg (1961) for analyzing in detail the specific properties of the all-or-none conditioning model in terms of response probabilities go over with little modification to the analysis of response latencies.

The most important observation about the all-or-none model is this. If we consider only responses prior to the last error, then we should obtain a sequence of Bernoulli trials with correct and incorrect responses distributed binomially with constant parameter p. In Suppes and Ginsberg (1961) four statistical tests are proposed to test this implication about the distribution of responses prior to the last error. From the standpoint of differentiating all-or-none from incremental conditioning, the most important is the test for stationarity. To test the null hypothesis of Bernoulli trials with constant parameter p, it is appropriate to use the statistic

$$X^2 = \sum_{t,i} n(t)\left(\frac{n_i(t)}{n(t)} - \frac{n_i}{N}\right)^2 \Big/ \frac{n_i}{N}$$

where i = 0, 1, $n_i(t)$ is the number of correct (i=1) or incorrect (i=0) responses in block t, n(t) is the total number of responses in block t, $n_i$ is the number of correct (or incorrect) responses summed over all blocks, and N is the total number of responses summed over all blocks. The statistic $X^2$ has a $\chi^2$ distribution with (m-1)(T-1) degrees of freedom, where m is the number of possible responses on any trial, and T is the total number of blocks. (For the usual experiment it is convenient to sum over blocks of trials.) It is at once apparent that this same test may be applied to the distribution of response latencies prior to the last error. It may be applied directly to empirical histograms of response latencies prior to the last error, summed over subjects and the given trials in a block. The null hypothesis is that the histograms are identical, and thus the test of stationarity is in fact a standard test of homogeneity. At a level of less detail we may test the stationarity of the means and variances of the empirical histograms. As in the case of response probabilities, the mean learning curves for means and variances should be horizontal; this implication yields a good graphical test for the all-or-none conditioning of response latencies. The simplicity of this prediction about variances is worth comparing with the relative complexity of the analytical recursion given by (4), which does not restrict the analysis to trials prior to the last error.

For applications (A)-(E) maximum likelihood estimates of the parameters not directly under the control of the experimenter, e.g., the conditioning parameter c or the parameter of the smearing distribution, are not feasible. I shall not attempt to discuss here the methods of estimation that seem both practical and statistically sound for these applications.

Fortunately, for the case of latency distributions in paired-associate experiments the situation is better even if not completely satisfactory. To obtain some concrete even if special results let us assume that the latency distributions U(t) and C(t) are exponential distributions with parameters $\lambda_1$ and $\lambda_2$ respectively. In other words I am postulating that the density u(t) is given by the equation

$$u(t) = \lambda_1 e^{-\lambda_1 t} ,$$

with $\lambda_1 > 0$ and $t > 0$.

Let now $x_n$ be the response random variable that has values 0 or 1 on trial n, with 0 representing an incorrect response and 1 a correct response, and let $t_n$ be the response latency random variable that has as values positive real numbers representing response times. On the assumption that subjects are homogeneous and statistically independent, the likelihood function L may be written as:

$$L(p,c,\lambda_1,\lambda_2) = \prod_{i=1}^{s} f_i(x_1,t_1,\ldots,x_{n_i},t_{n_i},\ldots) ,$$

where s is the number of subjects. Now beyond a certain trial, with probability 1, $x_n$ is always 1. Let $n_j+1$ be the trial number of the last error for subject j. On the assumptions of independence already stated, we then have:

$$L(p,c,\lambda_1,\lambda_2,) = \prod_{i=1}^{s_0} f_i(1,t_1,1,t_2,\ldots)\cdot \prod_{j=1}^{s_1} f_j(x_1,t_1,\ldots,$$

$$x_{n_j},t_{n_j}|0,t_{n_j+1},1,t_{n_j+2},\ldots)\cdot f_j(0,t_{n_j+1},1,t_{n_j+2},\ldots)$$

$$= p^y(1-p)^{n-y}\lambda_1^n e^{-\lambda_1 t} \prod_{i=1}^{s_0}$$

$$f_i(1,t_1,1,t_2,\ldots)\cdot \prod_{j=1}^{s_1} f_j(0,t_{n_j+1},1,t_{n_j+2},\ldots) ,$$

where $s_0$ is the number of subjects making no errors, $s_1 = s - s_0$, $y_j$ is the number of successes before the last error for subject j, $y = \sum_{s_1} y_j$, $n_j$ is the number of responses before the last error for subject j, $n = \sum_{s_1} n_j$, $t_j$ is the sum of the latencies prior to the last error of subject j and $t = \sum_{s_1} t_j$.

Let us define H as the restricted likelihood function concerned only with the data prior to the last error (for those subjects who made at least one error). Maximum likelihood estimates of p and $\lambda_1$ with respect to H are then immediate. (Notice that H is not a function of either c or $\lambda_2$.)

$$H(p,\lambda_1) = p^y(1-p)^{n-y}\lambda_1^n e^{-\lambda_1 t} .$$

Thus

$$\log H(p,\lambda_1) = y \log p + (n-y)\log(1-p) + n\log\lambda_1 - \lambda_1 t ,$$

and

$$\frac{\partial \log H}{\partial p} = \frac{y}{p} - \frac{n-y}{1-p} = 0$$

$$\frac{\partial \log H}{\partial \lambda_1} = \frac{n}{\lambda_1} - t = 0$$

whence

$$\hat{p} = \frac{y}{n}$$

$$\hat{\lambda}_1 = \frac{n}{t} .$$

The difficulty with the full likelihood function L is that we are not able to determine from direct observation of the responses or latencies immediately after the last error whether the subject is in state U or C, for he may be guessing the correct response while still in the unconditioned state. But it should be noted in this connection that if we ignore latencies after the last error, it is possible to compute the maximum-likelihood estimates of p, $\lambda_1$ and c. Using the fact that the distribution of last errors is as follows (Bower, 1961):

$$P(L-k) = \begin{cases} bp & \text{for } k = 0 \\ b(1-p)(1-c)^{k-1} & \text{for } k \geq 1 , \end{cases}$$

where

$$b = \frac{c}{1-p(1-c)} ,$$

we may at once write down the expression for this restricted likelihood function, which we may call $J(p, \lambda_1, c)$.

$$J(p,\lambda_1,c) = p^y(1-p)^{n-y}\lambda_1^n e^{-\lambda_1 t} \frac{c^{s_0} p^{s_0}}{[1-p(1-c)]^{s_0}} \frac{c^{s_1}}{[1-p(1-c)]^{s_1}} \cdot$$

$$(1-p)^{s_1}(1-c)^n$$

$$= \frac{c^s}{[1-p(1-c)]^s} p^{y+s_0}(1-p)^{n+s_1-y}(1-c)^n\lambda_1^n e^{-\lambda_1 t} .$$

Estimates for p, $\lambda_1$ and c may be obtained at once from this last equation. First, we have:

$$\frac{\partial \log J}{\partial p} = \frac{s(1-c)}{1-p+pc} + \frac{y+s_0}{p} - \frac{n+s_1-y}{1-p} = 0$$

$$\frac{\partial \log J}{\partial c} = -\frac{sp}{1-p+pc} + \frac{s}{c} - \frac{n}{1-c} = 0$$

$$\frac{\partial \log J}{\partial \lambda_1} = \frac{n}{\lambda_1} - t = 0 .$$

Solving these three equations we obtain, as before,

$$\hat{\lambda}_1 = \frac{n}{t} ,$$

and c may be simply expressed in terms of p:

$$\hat{c} = \frac{s(1-\hat{p})}{n-y+s_1} .$$

The estimate p may then itself be expressed in terms of a cubic equation all of whose coefficients are functions of the observables n, y, $s_0$, $s_1$, and s.

The estimation procedures just given ignore the problem of estimating $\lambda_2$. It is evident that a reasonably good estimate can be obtained directly from latencies on trials after the last error. A more detailed analysis will not be pursued here. Extensive experimental studies of latencies are now underway at Stanford, and I hope to be able to report on the empirical adequacy of the one-element latency model in the near future.

NOTES

REFERENCES


Bower, G. Application of a model to paired-associate learning. Psychometrika, 26, (1961), pp. 255-280.

Estes, W. K. Toward a statistical theory of learning. Psychological Review, 57, (1950), pp. 94-107.

Estes, W. K. and Suppes, P. Foundations of linear models. Chapter 8 in Studies in Mathematical Learning Theory, edited by R. R. Bush and W. K. Estes, Stanford University Press, 1959, pp. 137-179.

Luce, R. D. Response latencies and probabilities. Chapter 70 in Mathematical Methods in the Social Sciences, 1959, edited by K. J. Arrow, S. Karlin, P. Suppes, Stanford University Press, 1960, pp. 298-311.

Suppes, P. A linear model for a continuum of responses. Chapter 19 in Studies in Mathematical Learning Theory, edited by R. R. Bush and W. K. Estes, Stanford University Press, 1959, pp. 400-414.

Suppes, P. Stimulus sampling theory for a continuum of responses. Chapter 23 in Mathematical Methods in the Social Sciences, 1959, edited by K. J. Arrow, S. Karlin and P. Suppes, Stanford University Press, 1960, pp. 348-365.

Suppes, P. and Atkinson, R. C. Markov Learning Models for Multiperson Interactions. Stanford University Press, 1960. 296 pp.

Suppes, P. and Frankmann, R. Test of stimulus sampling theory for a continuum of responses with unimodal noncontingent determinate reinforcement. Journal of Experimental Psychology, 61, No. 2 (1961), pp. 122-132.

Suppes, P. and Ginsberg, R. A fundamental property of all or none models, binomial distribution of responses prior to conditioning, with application to concept formation in children. Technical Report No. 39, Psychology Series, Institute for Mathematical Studies in the Social Sciences, Applied Mathematics and Statistics Laboratories, Stanford University, September, 1961, 34 pp. (Psychological Review, in press.)

Suppes, P. and Zinnes, J. L. Stochastic learning theories for a response continuum with non-determinate reinforcement. Psychometrika, 26, No. 4 (1961), pp. 373-390.