# THE GENERAL-GAMMA DISTRIBUTION AND REACTION TIMES

by

William J. McGill and John Gibbon

TECHNICAL REPORT NO. 61

November 20, 1963

PSYCHOLOGY SERIES



INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

# The General-Gamma Distribution and Reaction Times*

by

William J. McGill and John Gibbon

* This paper was begun at Lincoln Laboratory, Lexington, Massachusetts during the summer of 1962. It was completed during 1963-64 at the Institute for Mathematical Studies in the Social Sciences, Stanford University where W.J.M. was in residence as a Special Fellow of USPHS. The preparation of this report was supported by Air Force Contract AF 49(638)-1253.

## Abstract

The general-gamma distribution describes input-output times in a multi-stage process consisting of exponential components whose time-constants are all different. The distribution and its unique history are examined. A stochastic process that leads to it is presented. The conditional density (hazard) function is studied as a means for estimating parameters. Finally the multi-stage process model is applied to simple reaction times in an effort to reveal underlying detection and response components.

A system receives an input and grinds it through a series of discrete processing stages, then generates an output. There is no feedback, no parallel processing, no branching, no backward motion of any kind. The description sounds almost too simple to bother with. Yet something like this with a few added restrictions turns out to be a rather well known stochastic process whose behavior is not especially easy to work out. In fact, if we start with just the processing times and an uncertain knowledge of the size of the system, the task of unraveling it proves to be fairly tricky.

Our interest in chain-like stochastic processes is aroused because they offer plausible models for important latency effects. For instance when an animal runs down an alley or learns to delay before responding, the duration of the latent period can be examined as a potential stochastic process. Bush & Mosteller's (1955) treatment of response times in an alley maze illustrates this type of application. Stochastic chains have also been used as models for reaction time; (see Christie & Luce, 1956; Stone, 1960; Restle, 1961; and LaBerge, 1962). It is clear that any latency which is thought to span an intervening and unobserved series of events can be a candidate for our interest, and the stubborn resistance that many of these latency effects offer to theoretical analysis shows the problem is not trivial.

The term "general-gamma distribution" is used in this paper to describe the density function for input-output times in a multi-stage process whose time constants are all different. It is an obvious extension of the ordinary gamma distribution generated by a process with equal time constants. The present paper is an effort to put together what is known about the general-gamma process, and to add a few modest contributions to the literature. It is perhaps strange that a compilation should be deemed necessary, but as we hope to make clear, the general-gamma distribution is one of the strange beasts in the statistical jungle. In the first place it pops up in mutant forms in a number of different scientific fields. As early as 1910 we find it in the theory of radioactive decay; see Bateman (1910) and Jeffreys & Jeffreys (1956). In queuing theory, it is known as the "general-Erlang" distribution; see Morse (1958). There is even an article in statistical learning theory containing a discrete version of the general-gamma distribution; see Miller & McGill (1952) and McGill (1963, p. 348). The intimate connections among the processes described in these papers are not at once evident. Each field is constrained to view its problems in its own way. So we should not be surprised to find that the distribution has been discovered a number of times and that some information about it has not achieved wide circulation.

A second anomaly is that the general-gamma distribution has a very innocent looking functional form that leads many writers to stop as soon as they have developed it. But if we take an empirical latency distribution and ask for a critical test of the likelihood that it comes from a multi-stage process, or if we try to estimate the transition constants of the various stages, a good deal of this innocence is dispelled.

In the present paper we shall first develop the general-gamma distribution as it appears in the literature and outline its background. Then we shall connect it with the stochastic process from which it comes, and investigate the estimation problem by looking at what is known as the "conditional density function" or the "hazard function." Finally an application of the general-gamma process to the analysis of simple reaction times will be considered.

1. Latency Distribution

Imagine a system that drives an input through $k + 1$ stages of processing. Each stage is characterized by a time constant $\lambda_0, \lambda_1, \ldots, \lambda_k$, identifying the stage and indicating (in a loose way) the probability of transition to the next stage. Note that the transition parameters depend on the stage and not on time. Hence an input moves through the processing chain in a series of irregular jumps patterned by the sequence of time constants. Since the within-stage transition probability is fixed, the passage-time $t_i$ through any stage will be governed by an exponential distribution:

$$f(t_i) = \lambda_i e^{-\lambda_i t_i} \;, \qquad \text{where } \lambda_i \text{ and } t_i \geq 0 \;.$$

Suppose a clock starts as the process is entered, and runs until the $k + 1$st transition occurs. The clock will then be said to measure the latency distribution of the process. Evidently the timing can be stopped and reset with each transition between adjacent stages, but many interesting cases include stages that are not easily available for direct examination. Reflexes and reaction times are typical examples. Consequently our interest is centered on the total latency of the process. This latency is obtained from the component passage times:

$$t = \sum_{i=0}^{i=k} t_i ,$$

where $t$ is summed over the $k + 1$ stages each of which is exponentially distributed with an arbitrary time constant. The distribution of $t$ is what we have called the general-gamma distribution. Its density function $f_k(t)$ is found to be:

$$f_k(t) = \sum_{i=0}^{i=k} c_{ik}\lambda_i e^{-\lambda i t} , \qquad (1)$$

where the function is defined for $t \geq 0$ . It is a surprisingly simple result; just a weighted sum of the exponential components. The weights are computed from the time constants:

$$c_{ik} = \frac{1}{\prod_{\substack{j = 0 \\ j \neq i}}^{j = k} \left(1 - \frac{\lambda_i}{\lambda_j}\right)} .$$

Processes with only a single stage (i.e., $k$ equal to zero) are included by defining $c_{oo}$ as unity. The area in the tail of the distribution is found easily by integrating the separate terms of the sum:

$$1 - F_k(t) = \sum_{i=0}^{i=k} c_{ik} e^{-\lambda_i t} . \qquad (2)$$

As we have suggested, this distribution has an interesting history. It is encountered in a number of fields that maintain little contact with one another, and has been rediscovered several times. Part of the history is illuminated by a proof of Eq. 1 based on generating functions that we can develop quite easily.

The moment generating function of any component latency in the process is defined as

$$m_i(\theta) = \int_0^{\infty} e^{\theta t} f(t_i) dt_i \quad ,$$

which, given our definition of $f(t_i)$, is readily shown to be

$$m_i(\theta) = (1 - \theta/\lambda_i)^{-1} \quad .$$

Since the total latency is the sum of $k + 1$ independent random components, its m.g.f. is immediately represented by the product of the component generating functions:

$$M_k(\theta) = \prod_{i=0}^{i=k} (1 - \theta/\lambda_i)^{-1} \quad . \qquad (3)$$

Evidently $M_k(\theta)$ inverts back into $f_k(t)$, but the problem is to see how to do it. It happens that the inversion of transformations in the form of Eq. 3 is a classical problem in engineering applications of the Laplace transform. The inversion is carried out by expanding $M_k(\theta)$ in partial fractions:

$$\prod_{i=0}^{i=k} (1 - \theta/\lambda_i)^{-1} = \sum_{i=0}^{i=k} \frac{c_{ik}}{1 - \theta/\lambda_i} \quad ,$$

where now the problem is to determine the constants, $c_{ik}$ . This expansion is handled as follows. We first change the initial index to $j$ and then multiply through by $1-\theta/\lambda_i$ in order to isolate $c_{ik}$ :

$$\prod_{\substack{j=0\\j\neq i}}^{j=k}(1-\theta/\lambda_j)^{-1} = c_{ik} + (1-\theta/\lambda_i)\sum_{\substack{j=0\\j\neq i}}^{j=k}\frac{c_{jk}}{1-\theta/\lambda_j} .$$

If $\theta$ is set equal to $\lambda_i$ , an explicit expression is found for each $c_{ik}$ in terms of the complete set of time constants. It is then a simple matter to invert

$$M_k(\theta) = \sum_{i=0}^{i=k}\frac{c_{ik}}{1-\theta/\lambda_i}$$

term by term. This yields our Eq. 1 for $f_k(t)$ , and completes the proof.

In electrical engineering, Eq. 1 is called the "Heavyside expansion" of $f_k(t)$ . Starkey (1955, pp. 49-50) uses just this problem as an example to illustrate the inversion of a Laplace transform that is the reciprocal of a polynomial; see also W. L. Smith (1953, pp. 456-458).

We noted in our introduction that Eq. 1 is called the "general-Erlang" distribution in queuing theory; (see Cox & Smith, 1961, pp. 114-116; and Morse, 1958, ch. 5). This forbidding terminology seems to have arisen innocently enough. Erlang was a Danish engineer who became well known after 1917 for his applications of probability to telephone traffic. He is now regarded as a pioneer queue theorist. In any event, Erlang (1920) proposed the gamma distribution as a model for the duration of telephone calls. A. Jensen (1948), writing in a memorial volume dedicated to Erlang, extended the latter's idea to include the more general-gamma process and almost inevitably Erlang's name became attached to it.

Jensen attributed the distribution to Lundberg (1940), and Palm (1946). However a much earlier version was published by Bateman (1910) who derived the weighting constants in a way that is very similar to the one we have outlined. Bateman encountered the multi-stage process in studies of radioactive decay with elements such as uranium. The parent element decays with an appropriate time constant into a new element which has its own time constant, decaying in turn into still another element, and so on. Bateman wanted the probability distribution of the number of radioactive products produced within a given observation interval if the time constants are arbitrary. His solution is outlined by Bharucha-Reid (1960, pp. 295-298) who also shows that it is easily converted into a stochastic process.

We conclude this introduction to the general-gamma distribution by observing that it is quite difficult to visualize the properties of the process with only the information provided by its latency distribution. This is because the weights become intractable when the number of stages exceeds three or four. Consequently it is hard in such cases to see just how the time constants will affect the distribution. We can note from Eq. 2 that

$$\sum_{i=0}^{i=k} c_{ik} = 1 , \qquad k = 1 , 2 , 3 \ldots ,$$

since $F_k(0)$ must be zero. Moreover the plausible notion (which turns out to be correct) that $f_k(0)$ is zero for processes with more than one stage implies

$$\sum_{i=0}^{i=k} c_{ik}\lambda_i = 0 \,, \qquad k = 1 \,, 2 \,, 3 \,\ldots.$$

If these relations seem confusing we might also add that half the weights are negative and that a particular $c_{ik}$ may have almost any value whatever, depending on the relative sizes of the time constants.

2. Orderly Time Constants

In a number of situations that are of considerable practical interest, the latency distribution of the general-gamma process simplifies to a form that is more manageable than the general case. These situations occur when the time constants of the chain form an orderly array. The ordinary gamma distribution (all time constants equal) is an obvious example, but there are others. Consider, for instance, the following process suggested by Jensen (1948):

$$\begin{aligned} \lambda_0 &= \lambda \\ \lambda_1 &= \lambda(1+1/\beta) \\ &\;\;. \\ &\;\;. \\ &\;\;. \\ \lambda_k &= \lambda(1+k/\beta) \;. \end{aligned}$$

In view of its linearly increasing time constants, this array forms a "linear birth process." Substitution in Eq. 1 produces the latency distribution:

$$f_k(t) = \lambda(1+ k/\beta) \binom{\beta + k - 1}{k} e^{-\lambda t}(1 - e^{-\frac{\lambda}{\beta}t})^k \ . \tag{4}$$

The troublesome weighting constants have been eliminated, and the latency distribution is very close to the negative binomial whose properties are well known. Moreover, when $\beta$ is set equal to unity, $f_k(t)$ becomes the particularly simple form of the Yule-Furry process discussed by Feller (1957), Bharucha-Reid (1960), and McGill (1963).

Finally we should note that with $\lambda$ fixed, the time constants in the chain approach each other as $\beta$ increases. If we examine the limit of $f_k(t)$ under this restriction, Eq. 4 approaches the gamma distribution. Hence Eq. 4 provides a very simple way to pass from the general-gamma distribution to the ordinary gamma in which all time constants are equal.

3. Stochastic Process

When time constants refuse to behave nicely, or when they are unknown, another simplification must be found. Another one exists and it is very important indeed. To exhibit it, we begin with the differential equation for a stochastic process given by Feller (1957, p. 402),

$$P'_t(k) = -P_t(k)\lambda_k + P_t(k-1)\lambda_{k-1} \ , \quad \text{for } k \geq 1 \ . \tag{5}$$

Our notation differs slightly from Feller's. The symbol $P_t(k)$ is the probability that $k$ transitions take place in a (fixed) time period $t$, and its derivative with respect to time is $P'_t(k)$. The differential equation reflects the fact that in any small time interval

$\Delta t$ at the end of $t$, the process can find itself at stage $k$ via one of two routes; either it was already in $k$ and remained there with probability $1-\lambda_k \Delta t$, or it was at stage $k-1$ and made the transition upward with probability $\lambda_{k-1}\Delta t$.

The solution of Eq. 5 is well known and closely related to Eq. 1, but we shall carry our development a step further before writing it down. First we note that $P_t(k)$, the probability of $k$ transitions in time $t$, forms a discrete probability distribution with continuous parameters. We shall refer to this as the <u>response distribution</u> in order to keep it distinct from the <u>latency distribution</u> which is also defined for the same process. Since the time constant $\lambda_k$ (governing transitions out of stage $k$) is independent of time, it follows that an important relation exists between the response distribution and the latency distribution, namely

$$f_k(t)dt = P_t(k) \cdot \lambda_k dt . \tag{6}$$

The infinitesimal probability of a latency that stops the clock at time $t$ is constructed from the probability of $k$ transitions in time $t$, i.e., $P_t(k)$, multiplied against the probability that the final transition happens in the infinitesimal interval $dt$ tacked on the end of $t$, i.e., $\lambda_k dt$. In view of the relation between response and latency distributions we now recast Eq. 5 as follows:

$$P_t'(k) = -f_k(t) + f_{k-1}(t) . \tag{5a}$$

Integrating the differential equation yields a new difference equation:

$$\begin{aligned} P_t(k) &= F_{k-1}(t) - F_k(t) , \qquad & k > 0 , \\ P_t(0) &= 1 - F_0(t) , \qquad & k = 0 , \end{aligned} \tag{5b}$$

where, of course, $F_k(t)$ is the cumulative latency distribution. One minor substitution brings us to

$$f_k(t) = \lambda_k[F_{k-1}(t) - F_k(t)] \ , \qquad (5c)$$

for $k > 0$ . Our argument is that Eq. 5c is true of the general-gamma distribution for any proper choice of parameters, and it is a fact that Eq. 1 satisfies Eq. 5c. This claim is verified easily by examining the moment generating functions of both sides. (The m.g.f. of a distribution in "tails" form is no special problem; see McGill, 1963, p. 353.) We find

$$M_k(\theta) = \frac{\lambda_k}{\theta} [M_k(\theta) - M_{k-1}(\theta)] \ .$$

We have a simple recursive relation

$$\frac{M_k(\theta)}{M_{k-1}(\theta)} = (1 - \theta/\lambda_k)^{-1} \ ,$$

and step by step it reduces to the m.g.f. of the general-gamma distribution. Thus the general-gamma distribution is a solution of the difference equation for the stochastic latency process given by Eq. 5c.

The response distribution of the general-gamma process is evidently

$$P_t(k) = 1/\lambda_k \sum_{i=0}^{i=k} c_{ik}\lambda_i e^{-\lambda_i t} \ , \qquad (7)$$

which follows directly from Eqs. 5c and 6. Equation 7 expresses the probability of finding exactly $k$ transitions in a fixed time interval $t$ . The result is given in Bateman (1910), Lundberg (1940), Jensen (1948), Feller (1949), Bartlett (1955, p. 55) Bharucha-Reid (1960, p. 186 and p. 297). It is a generalization of the Poisson distribution.

In the ordinary Poisson process, subscripts are omitted from the time constants, and all stages have identical properties.

A distribution that covers discrete trials instead of continuous time but is otherwise similar to Eq. 7, was published by Woodbury (1949), and by Miller & McGill (1952).

The end product of our development is that we have found a stochastic process in the form of Eq. 5c which is directly applicable to the general-gamma distribution. For instance Eq. 5c makes it clear that

$$f_k(0) = 0 \ ,$$

$$f_k(\infty) = 0 \ ,$$

and Eq. 5a indicates that $f_k(t)$ passes through a unique maximum. Moreover, by adding down Eq. 5b we find:

$$1 - F_k(t) = \sum_{i=0}^{i=k} P_t(i) \ . \qquad (8)$$

In other words the familiar relation between the tail area of a gamma distribution and the Poisson sum turns out to be a consequence of the stochastic process and is a general property of all multi-stage gamma processes. The linear birth process, for example, takes on the following forms; (see Eq. 4):

$$P_t(k) = \binom{\beta + k - 1}{k} e^{-\lambda t}\left(1 - e^{-\lambda\frac{t}{\beta}}\right)^k \ , \qquad (4a)$$

which is a negative binomial distribution; and

$$1 - F_k(t) = \sum_{i=0}^{i=k} \binom{\beta + i - 1}{i} e^{-\lambda t}\left(1 \quad e^{-\lambda\frac{t}{\beta}}\right)^i \ , \qquad (4b)$$

The tail area of the latency distribution in a linear birth process is found by summing the appropriate terms of the negative binomial.

4. Conditional Density Function

Given a latency distribution generated by a multi-stage process of unspecified length, can we operate on it to determine the $\lambda_i$, and to decompose the overall distribution into inferred distributions of parts of the process?

First recall that the time constant, $\lambda_k$, of the last stage may be recovered from

$$\lambda_k = \frac{f_k(t)}{F_{k-1}(t) - F_k(t)} \quad . \tag{9}$$

This is a simple adjustment of Eq. 5c, but it requires knowledge of $F_{k-1}(t)$, and the latter is not easily determined if the length of the chain and its parameters are unknown. A slight generalization of Eq. 9 yields

$$\lambda_k(t) = \frac{f_k(t)}{1 - F_k(t)} \quad . \tag{10}$$

In some areas of statistics Eq. 10 is known as a "hazard function." We call it a conditional density function following Davis (1952), since it does in fact wipe out any considerations that might have produced latencies shorter than those occurring right now or later on. If we reach a particular point in time, and the clock measuring latency is still running, any circumstance that might have led to a shorter latency is irrelevant, provided only that the current probability of the final transition can be stated.

It is easily shown that the tails-distribution of latencies in a multi-stage process may be represented in terms of the conditional density function, namely:

$$1 - F_k(t) = e^{-\int_0^t \lambda_k(\tau)d\tau} . \qquad (11)$$

Many sources establish this result. The proof given by Davis (1952) is especially easy to follow. In any event, since the probabilities in Eqs. 2 and 11 are identical, it must be the case that the conditional density function carries full information concerning the time constants $\lambda_o , \lambda_1 \ldots , \lambda_k$ of the general-gamma process. We can attempt to extract this information by using what we have learned about the stochastic process underlying the general-gamma distribution. First, however, we point out a simple relation between the conditional density function and the ordinary cumulative distribution. The relation is obtained from Eq. 11 by taking the logarithm of both sides and then differentiating. We find:

$$\frac{d \log (1 - F_k(t)}{dt} = -\lambda_k(t) . \qquad (10a)$$

The slope of the logged tails-distribution is the conditional density function; (see Parzen, 1962, p. 168). This is a very simple and powerful graphic aid for viewing latency phenomena. An example based on empirical reaction times and sample approximations of the quantities in Eq. 10a is shown in Fig. 1

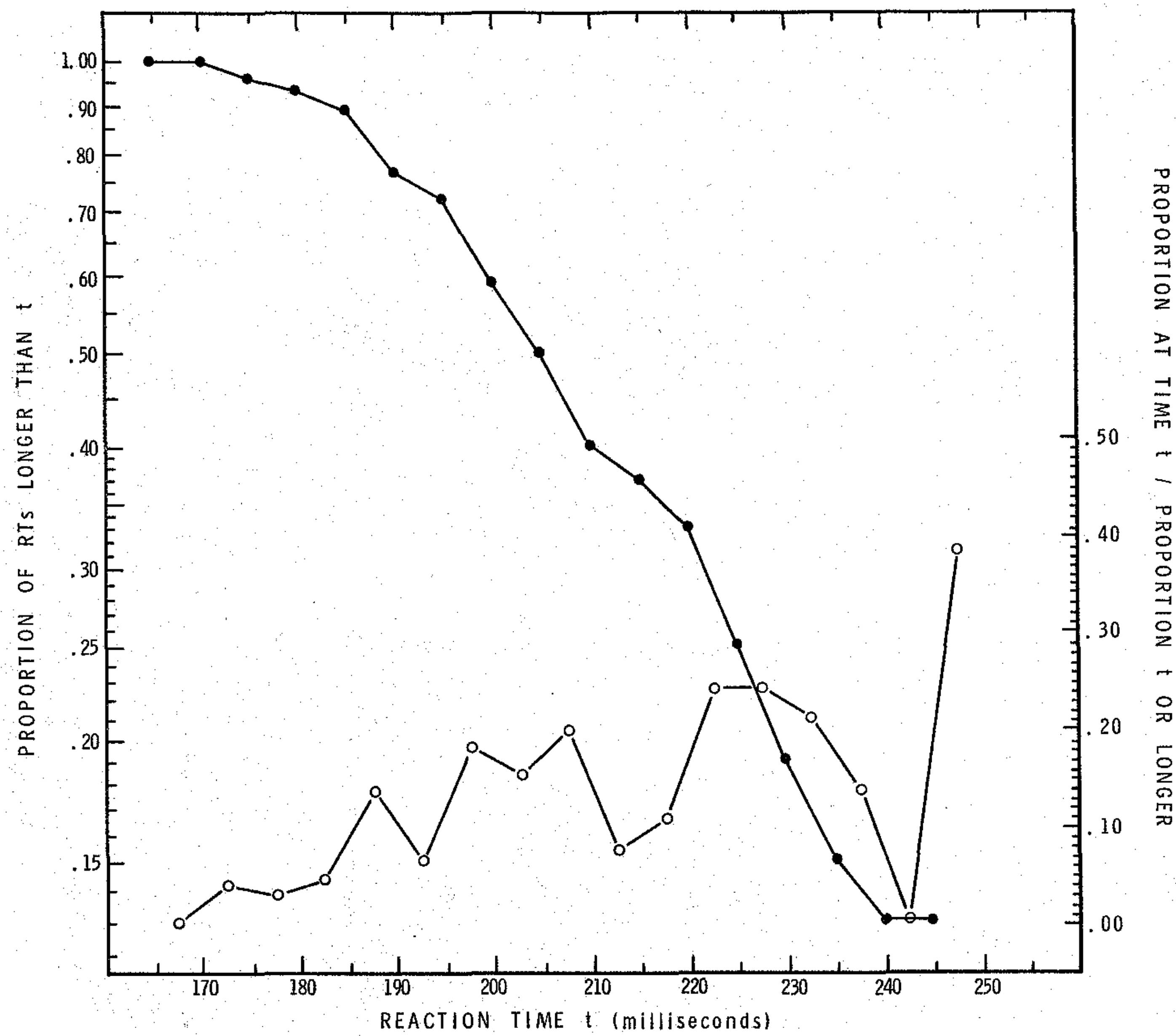


Fig. 1. Frequency distribution of reaction times to brief tones (1000 cps) of medium intensity. Data based on 100 responses of one listener. Right hand ordinate (lower curve) is conditional distribution of the same data. Each point on lower curve is approximately -5 times the slope of the line segment in the distribution directly above it. The factor of 5 is the size of the class interval (5 milliseconds). Discrepancies are grouping errors due to measuring proportion-longer-than-t from front end of class interval instead of middle. Erratic behavior of conditional distribution in this example appears to be due to narrow width of class intervals.

Now we turn to the stochastic-process interpretation of the general-gamma distribution so as to analyze its conditional density function. First we write:

$$\frac{\lambda_k(t)}{\lambda_k} = \frac{P_t(k)}{\sum_{i=0}^{i=k} P_t(i)} . \qquad (12)$$

This follows via substitutions from Eqs. 6 and 8 into Eq. 10, and is a very useful result. The expression on the right hand side of Eq. 12 is familiar from queuing theory; (see Saaty, 1961, ch. 14; Fry, 1928, p. 360). It is the probability that exactly $k$ transitions prior to the last one have occurred before time $t$. So we have shown that the conditional density function at time $t$, divided by the time constant of the last stage, yields the probability that an ongoing process is in its last stage at time $t$. Because of the nature of the multi-stage process, this probability must be small at first and must increase monotonically as elapsed time increases. Long waits can only increase the likelihood that the last stage of the process has been reached. It follows that the conditional density function of the general-gamma distribution is monotonic increasing with time. In view of Eq. 10a this implies a characteristic behavior of the slope of the tails distribution. The behavior is illustrated in Fig. 2 and offers a good test of the conjecture that an empirical latency distribution might have been generated by a multi-stage process.

The complement of Eq. 12, i.e., the probability that the last stage in the ongoing process is not yet reached by time $t$, can be written:

$$1 - \frac{\lambda_k(t)}{\lambda_k} = \frac{1 - F_{k-1}(t)}{1 - F_k(t)} . \qquad (12a)$$

The result comes directly from Eq. 8. In view of our discussion of Eq. 12, the complementary probability must decrease monotonically toward zero as elapsed time increases. The conclusion follows from our interpretation of Eq. 12 or it can be established independently via substitution from Eq. 2 in Eq. 12a. In the latter case we proceed by removing a factor $e^{-\lambda_h t}$ from the numerator and denominator of the right hand side on the supposition that $\lambda_h$ is the smallest time constant of the set. It is then easy to show that

$$\lim_{t \to \infty} \left(1 - \frac{\lambda_k(t)}{\lambda_k}\right) = \begin{cases} 1 - \lambda_h/\lambda_k & \text{for } h < k , \\ 0 & \text{for } h = k . \end{cases} \qquad (13)$$

Moreover, when the smallest time constant happens to be $\lambda_k$, we can show (now via the derivative of Eq. 12a) that the approach to the limit is monotonic. Hence, we have established that the asymptotic slope of the logged tails-distribution is not the time constant of the last stage but the smallest time constant wherever it may be located. Whatever the actual sequence of processing stages, the time constants become rearranged from large (fast) to small (slow) in the latency distribution. We might then just as well treat any process as though the time constants were arranged in order of decreasing magnitude since we can extract them in this way (in reverse order) from the conditional density function.

It should be evident that the foregoing operation on the k+l-stage general-gamma distribution produces an inferred distribution of a subprocess obtained by summing together all $k$ stages other than the slowest. This follows from Eq. 12a. If we find an accurate value for the asymptotic slope of the tails-distribution (or the conditional density function), then we can produce the distribution of the $k$-stage subprocess from the distribution of the whole process. The outcome iterates on down through smaller and smaller subprocesses, so that in general

$$[1 - F_{k-j}(t)] = [1 - F_k(t)] \prod_{i=0}^{i=j-1} \left[1 - \frac{\lambda_{k-i}(t)}{\lambda_{k-i}}\right] \qquad (14)$$

where the time constants are extracted in order of increasing magnitude. In principle the process can be decomposed or "unpeeled" completely in this way, starting with nothing more than the overall latency distribution. The sequence information about time constants is lost but everything else can be recovered. Of course very serious practical difficulties are encountered in applying these results to empirical distributions. Our data usually do not yield solid information about the tail region of the latency distribution. Hence the estimated asymptote of the conditional density function can be seriously in error. This error is propagated into the estimation of the next time constant with the result that the estimation technique finds it increasingly difficult to see back through more than one or perhaps two stages. The errors are very similar to those that Van Liew (1962) ascribes to the "backward projection" method of graphic analysis used in physiology. Estimating parameters is hazardous with anything less than extremely stable distribution data. On the other hand the weakness is counterbalanced by the fact that the smallest (and hence the most significant) time constants may be estimated without committing ourselves in advance to the number of stages in the process.

5. Applications to Simple Reaction Times

Simple reaction times belong to a cluster of relations (such as the $\Delta I/I$ functions) involving stimulus intensity and generally viewed as being determined by sensory mechanisms. Other problems such as foreperiod effects or effects due to instructions, or changing stimulus probabilities, or rewards for fast or slow responding, are deliberately detoured in order to keep the emphasis on sensory processes. But these other problems and the demonstrable effects they have on reaction times, cannot fail to complicate our thinking. The net result is that efforts to reconstruct sensory processes now seem to produce some fairly intricate mechanisms. Few of us consider the substrate of reaction times in terms of a simple neural chain linking S to R. Instead, the process is thought to be organized into a series of complex and functionally distinct stages: receptor activity, stimulus analysis, sharpening, transmission, signal detection, motor response. This conception illustrates the modern blurring of boundary lines between sensation and perception. The blur is an obvious consequence of the nonsensory variables mentioned above and the effects they have on thresholds. The conception is more like a flow chart or a block diagram than like a reflex arc, and, of course, it is built into much of the recent work on signal detection theory in psychophysics; see Green (1960) and Swets, Tanner & Birdsall (1961) for good summaries. The same type of construction is found in recent attempts to portray decision mechanisms underlying choice reaction times; see Christie & Luce (1956); Audley (1960); and Stone (1960).

If we try to analyze the block diagram, one of the most annoying obstructions encountered almost at once is the fact that the response movement is immediately available to observation but not particularly interesting. We want to study detection but we are unable to view the mechanism (from the outside) except by looking back through the response. It was to illuminate this problem that our analysis of the general-gamma process was carried out.

We consider simple r.t.'s to have three principal stages: 1) input, 2) detection, 3) response. A train of impulses flows back from the receptor. It is detected by a suitable mechanism and a response is ordered. Assume that the detection is accomplished by counting impulses. A neural counter sits and waits for the initial barrage of the impulse-train. Nothing is settled until the counter reaches a predetermined criterion count (determined by instructions, stimulus probabilities and pay offs) whereupon a detection state exists and the response is called out. The model may be wrong in its choice of details, but it is undoubtedly wrong in the details suggested. The integration time is essentially endless and allows no detection failures. The system is also noiseless so there can be no false starts. Despite these obvious weakpoints, a simple counter model of stimulus detection might work quite well at high intensities. It has been developed on just about this form by Restle (1961) and LaBerge (1962). Evidently low intensity stimuli take time to detect because they generate slow impulse rates. Hence the detection time of a counter mechanism must be a function of stimulus intensity. Our problem is to see how to submit this general line of reasoning to systematic checking if our investigation is obstructed by the response that terminates the sequence.

The simplest way is to assume that the response movement (or the part of it that is variable) has a characteristic latency independent of stimuli and independent of the detection time. So we apply the theory of the general-gamma process by assuming that at high and medium intensities the response stage has a slow exponential delay while detection and input stages are made up of exponential components with fast time constants that depend on the stimulus. In this way our development in Sec. 4 of the paper can be brought into play.

Fig. 2 shows several frequency distributions of auditory r.t.'s obtained from a single listener who was responding to different intensities. The stimulus was a 1000 c.p.s. tone rising in roughly linear form to full intensity over approximately 100 milliseconds. The different distributions signify changes in intensity from 50 db to 100 db sound pressure level. Although we do not see much of their tails, the distributions appear to have the characteristic candy-cane appearance demanded by the conditional density function of a general-gamma process. Many latency distributions do not, particularly those obtained in free operant behavior; see, for example, Hearst (1958).

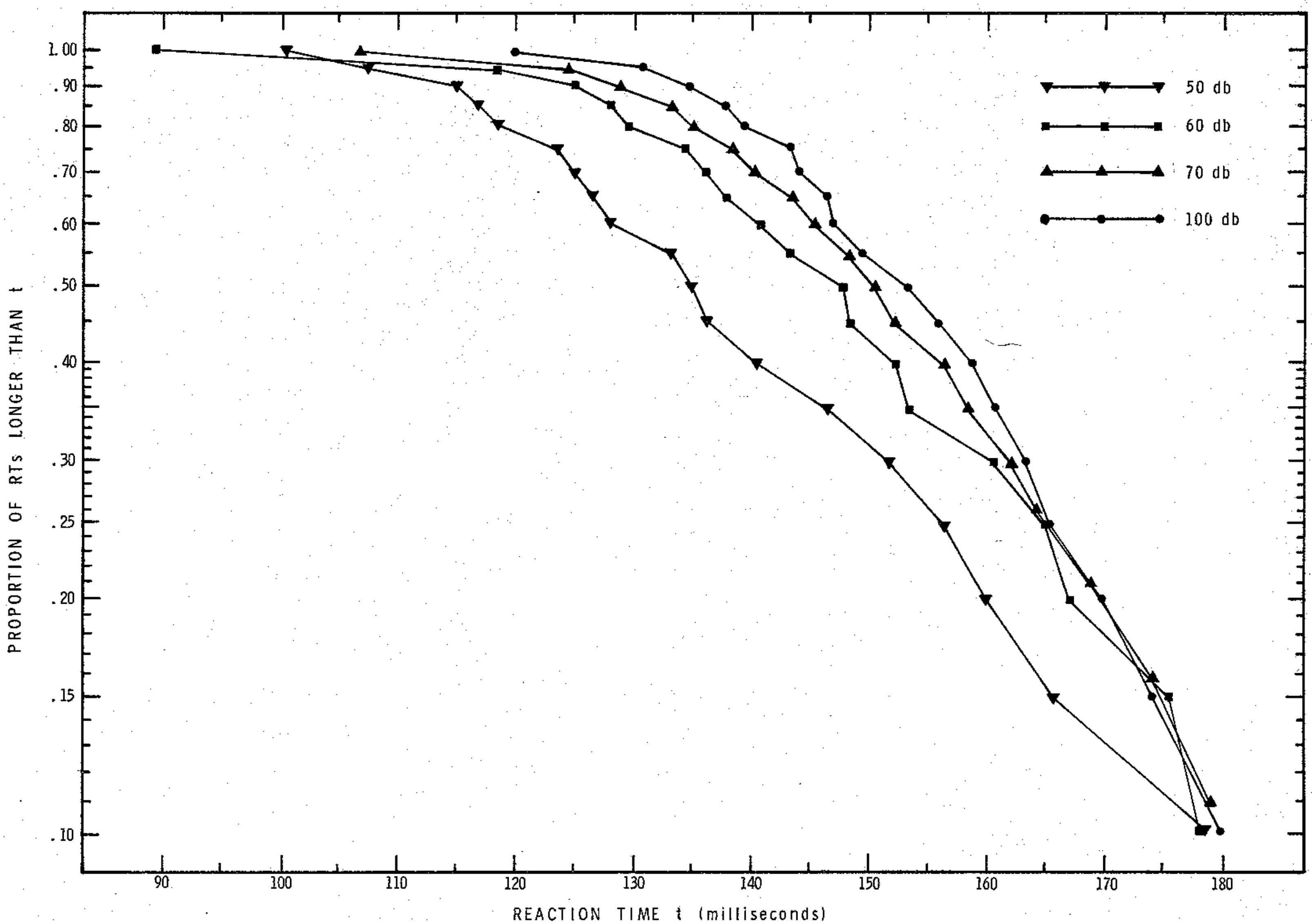


Fig. 2. Distributions of reaction times to brief (.5 sec.) tones (1000 cps) at intensities (sound pressure levels) shown. Each distribution based on 100 responses of one listener. Data at 50, 60, and 70 db are shifted in time to allow comparison with distribution at 100 db (correctly placed). Actual locations are as follows: 50 db -- add 70 milliseconds to time scale; 60 db -- add 45 milliseconds; 70 db -- add 31 milliseconds.

In any event these reaction-time distributions seem to meet our initial test. Reaction times gathered under comparable conditions also exhibit conformity with the restriction on the conditional density function; see Greenbaum (1963).

Now how do we find the vestige of the response in the distribution of reaction times using the general-gamma model? The easiest and most obvious approach is to bypass detection problems by going to a very intense stimulus. If transmission times exhibt low-order random variation, the tail of the reaction time distribution will be heavily weighted by the response component. Moreover the prominence of this linear tail will depend on the prominence of the stage that generates it, i.e., on the size of its time constant relative to the other time constants of the process. As intensity decreases, detection should require more and more time. This should then not only spread the distributions corresponding to different intensities apart (in time), it should also make the approach to the linear tail slower. Finally, if it happens that the detection stage takes longer than the response, the tail-asymptote should change.

The distributions in Fig. 2 have been shifted together in order to allow easy comparison of their tails. There appears to be a common linear asymptote between 60 db and 100 db, and the approach is slower for the weaker stimuli. The slope of the tail is approximately .05, yielding 20 milliseconds as the average random delay of the response stage. This is not an unreasonable figure and in fact it is in fairly good agreement with the physiological data (see Sashin & Allen, 1960) but it does present certain difficulties.

The r.t. distribution generated by a 60 db tone is located approximately 45 milliseconds beyond the distribution at 100 db. All of this difference might be due to increased detection time. To show this, suppose that the impulse rate produced by a tone at 100 db is 300 per second and suppose further that 5 impulses must be counted to meet the detection criterion. Detection times would then be of the order of 16-17 milliseconds. If the rate dropped to 80 per second at 60 db, the detection time would increase to an average of 62-63 milliseconds, and the separation of the distributions would be approximately what we have found. The impulse rates would stand in a ratio of nearly 4 to 1 and this turns out to be not far below the loudness ratio of 100 db to 60 db set by the same listener whose r.t. distributions provide the basis of our arguments. Hence we might easily expect a shift of 45 milliseconds or more (depending on impulse rates and detection criteria) just from differences in detectability between 60 db and 100 db. How can this large detection delay be squared with a response delay of only 20 milliseconds and still yield the response latency in the asymptote of the r.t. distribution?

The fact is that the problem cannot be surmounted unless most of the delay introduced by decreasing the stimulus intensity is fixed, i.e., nonrandom, or unless the detection process consists of substages arranged in the form of a gamma process. In that event the proper comparison would be between 20 milliseconds and the delay expected of a single count in the detection stage. Since impulse rates averaging

10 per second imply 100 millisecond delays, it is apparent that even single counts might take appreciable time.

Our principal conclusion is that the detection time for medium intensities is probably quite long, and our general-gamma model will certainly fail unless one of the possibilities suggested in the previous paragraph presents itself. Thus the question of whether we really see the response stage in the tail of the r.t. distribution should be pushed. For instance, we must account for the linear relation between the average and standard deviation of empirical r.t. distributions; see Chocholle (1940); Restle (1961); and Greenbaum (1963). In the general-gamma process the relation is a consequence of putting varying impulse rates into a detector that must reach a certain count before it detects. The independence of the response stage can only weaken this relation at high intesities (short reaction times). Whether the "purified distributions" arrived at by peeling the final stage away via Eq. 14 can be made to yield an improved linear relation is a resolvable question, but better tail data are required in order to answer it. If the relation fails to improve with this operation, perhaps the response stage is not in the tail of the reaction time distribution at all and is really located among the fixed delays at the beginning of the distribution.

An interesting way to try to settle this last question is to measure r.t.'s from the same subject instructing him (on cue) to make one or the other of two responses (right index finger and left foot, for example). Two different distributions are then built up around the two responses which are deliberately chosen so as to introduce different response delays into the reaction times. We can then check our reasoning by using Eq. 14 to peel off the response stage. The purified distributions should match one another since detection and input conditions are identical. On the other hand, if the response appears among the initial delays, Eq. 14 will fail, but the distributions may be matched by sliding them along the time scale, i.e., by subtracting out two different delay constants.

References

Audley, R. J. A stochastic model for individual choice behavior. Psychol. Rev., 1960, 67, 1-15.

Bartlett, M. S. An introduction to stochastic processes. Cambridge: Cambridge Univer. Press, 1955.

Bateman, H. A solution of a system of differential equations occurring in the theory of radioactive transformations. Proc. Cambridge Philosophical Soc., 1910, 15, 423-427.

Bharucha-Reid, A. T. Elements of the theory of Markov processes and their applications. New York: McGraw Hill, 1960.

Bush, R. R. & Mosteller, C. F. Stochastic models for learning. New York: Wiley, 1955.

Chocholle, R. Variation des temps de réaction auditifs en fonction de l'intensité a diverse fréquences. Année Psychol., 1940, 41, 5-124.

Cox, D. R., & Smith, W. L. Queues. London: Methuen & Co., 1961.

Christie, L. S. & Luce, R. D. Decision structure and time relations in simple choice behavior. Bull. math. Biophys., 1956, 18, 89-112.

Davis, D. J. An analysis of some failure data. J. Amer. stat. Assn., 1952, 47, 113-150.

Erlang, A. K. The application of the theory of probability in telephone administration. Proc. Scandinavian H. C. Örsted Congress, Copenhagen, 1920. (Reprinted in The life and works of A. K. Erlang by E. Brockmeyer, H. Halstrøm, & A. Jensen. Trans. Danish Academy of Technical Sciences. Copenhagen, Denmark, 1948, Pp. 172-200, Whole No. 2.)

Feller, W. On the theory of stochastic processes with particular reference to applications. In J. Neyman (Ed.), *Proceedings of the Berkeley symposium on mathematical statistics and probability*. Vol. 1, Berkeley and Los Angeles: Univer. California Press, 1949, Pp. 403-432.

Feller, W. *An introduction to the theory of probability and its applications*. (2nd ed.) New York: Wiley, 1957.

Green, D. M. Psychoacoustics and detection theory. *J. acoust. Soc. Amer.*, 1960, *32*, 1189-1203.

Greenbaum, H. B. Simple reaction time: a case in signal detection. Unpublished doctoral dissertation, Columbia University, 1963.

Hearst, E. The behavioral effects of some temporally defined schedules of reinforcement. *J. exp. anal. Behav.*, 1958, *1*, 45-55.

Jeffreys, H. & Jeffreys, B. *Methods of mathematical physics*. (3rd ed.) Cambridge, England: University Press, 1956.

Jensen, A. An elucidation of Erlang's statistical works through the theory of stochastic processes. In E. Brockmeyer, H. L. Halstrøm, & A. Jensen, *The life and works of A. K. Erlang*. Transactions of the Danish Academy of Technical Sciences, Copenhagen, Denma Pp. 23-100 (Whole No. 2).

LaBerge, D. A recruitment theory of simple behavior. *Psychometrika*, 1962, *27*, 375-396.

Lundberg, O. *On random processes and their application to sickness and accident statistics*. Uppsala, Sweden: University of Uppsala, 1940.

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, Pp. 309-360.

Miller, G. A. & McGill, W. J. A statistical description of verbal learning. _Psychometrika_, 1952, _17_, 369-396.

Morse, P. M. _Queues, inventories and maintenance_. New York: Wiley, 1958.

Palm, C. Reference is to a paper in Swedish, published in _Teckn. Medd. Fran. Kungl. Tele._, 1946. Cited by Jensen, A., An elucidation of Erlang's statistical works through the theory of stochastic processes. In E. Brockmeyer, H. L. Halstrøm, & A. Jensen, _The life and works of A. K. Erlang_. Transactions of the Danish Academy of Technical Sciences, Copenhagen, Denmark, Pp. 23-100 (Whole No. 2).

Parzen, E. _Stochastic process_. San Francisco: Holden Day, 1962.

Restle, F. _Psychology of judgment and choice_. New York: Wiley, 1961.

Saaty, T. L. _Elements of queuing theory_. New York: McGraw Hill, 1961.

Sashin, P. & Allen, J. The latency of the human electromyogram in relation to reaction time. _Quarterly Progress Report of the Research Laboratory of Electronics_. Massachusetts Institute of Technology, July 1960.

Smith, W. L. On the distribution of queuing times. _Proc. Cambridge Philosophical Soc._ 1953, _49_, 449-461.

Starkey, B. J. _Laplace transforms for electrical engineers_. New York: Philosophical Library, 1955.

Stone, M. Models for choice reaction time. _Psychometrika_, 1960, _25_, 251-260.

Swets, J. A., Tanner, W. P., & Birdsall, T. G. Decision processes in perception. Psychol. Rev., 1961, 68, 301-340.

Woodbury, M. A. On a probability distribution. Ann. math. Statist., 1949, 20, 311-313.

Van Liew, H. D. Semilogarithmic plots of data which reflect a continuum of exponential processes. Science, 1962, 138, 682-683.