PSYCHOMETRIKA—VOL. 25, NO. 3
SEPTEMBER, 1960

# MODELS FOR CHOICE-REACTION TIME

MERVYN STONE

MEDICAL RESEARCH COUNCIL*

In the two-choice situation, the Wald sequential probability ratio decision procedure is applied to relate the mean and variance of the decision times, for each alternative separately, to the error rates and the ratio of the frequencies of presentation of the alternatives. For situations involving more than two choices, a fixed sample decision procedure (selection of the alternative with highest likelihood) is examined, and the relation is found between the decision time (or size of sample), the error rate, and the number of alternatives.

This paper develops to the point of usefulness several mathematical models for choice-reaction time. The working details are confined to appendices and only definitions and results appear in the text. It is hoped that this method of presentation will assist the reader in making a quick "calculated-observed" analysis of the data he may have. The choice of models is made mainly by analogy with statistical decision procedures, but no model is presented which is psychologically unreasonable. Also no comparisons are made with experimental data for several reasons: (i) the paucity of available data means that the field should be kept open to avoid premature rejections; (ii) published data are often summarized in directions orthogonal to our interests; (iii) for the most powerful discrimination, experiments will need to be designed with specific models in mind.

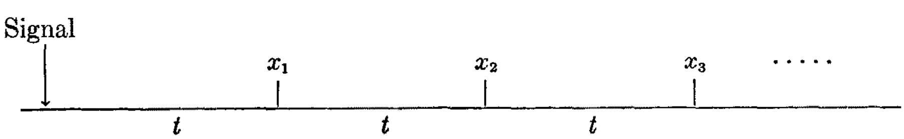
The models are envisaged as applying to the situation in which the subject ($S$) is given a time-stationary stimulus or signal and is required to identify some attribute of the signal and make an appropriate reaction. The signal remains present until the reaction is made. $S$ is presented with signal after signal and the successive attributes form a random sequence; that is, for a given run of signals, the attributes of different signals are mutually independent and their probabilities of presentation do not change with time. The models assume that $S$ has a settled mode of response. They will be hydrodynamic in the following sense. At the onset of each signal, a stream of information about the signal flows at a uniform rate into $S$. After a certain time, the input time, the front of this stream reaches $S$'s decision taking mechanism or "computer." After a further time, the decision time, $S$ makes a response. The time taken for the response to be recorded will be called the motor time. Thus the choice-reaction time is made up of three components:

*Applied Psychology Research Unit, 15 Chaucer Road, Cambridge, England.

the input time, $T_i$ ; the decision time, $T_d$ ; the motor time, $T_m$ . The models apply to $T_d$ , which will be related to the environmental variables (the number of signals and their frequencies of presentation) and the rate at which $S$ makes incorrect responses. By concentrating on $T_d$ in this way, it is not implied that $T_i$ and $T_m$ are necessarily independent of these factors.

### *Likelihood Ratio Models for the Two-Choice Situation*

It is assumed that the subject knows when the signal (either $s_0$ or $s_1$ , say) commences; that is, he knows when to start examining the stream of information arriving at the computer. (This stream is "noisy" until the stream from the signal is added to it.) This assumption holds in the self-paced condition and also when some preparatory warning signal is given. It is supposed that there is some overlap in the information; that is, some patterns of information may arise from either $s_0$ or $s_1$ . If there is no uncertainty in this sense, there is no need for a statistical computer. The uncertainty may arise from the external situation, from noise added at the input stage, or from both sources. We will suppose that the information on which $S$'s computer operates is equivalent to a series of independent random variables at short time intervals $t$ and that each random variable has the (stationary) distribution of a random variable $x$ (dependent on which signal has occurred) until the response is made.




Let $p_0(x)$ and $p_1(x)$ be the probabilities of $x$ when the signal is $s_0$ and $s_1$ , respectively. If the $x$'s are instantaneous samples of an almost continuous stream of information then the assumption of independence implies zero auto-correlation between parts of the stream not less than time $t$ apart. If the $x$'s are integrals of the stream over the successive intervals, then the assumption requires zero auto-correlation for all time lags (or at least for those not small compared with $t$). Suppose the computer transforms each $x$ to a quantity $c(x)$ which is then stored in an adder.

#### *Sequential Case*

The computer makes a running total of $c(x_1)$, $c(x_2)$, $\cdots$ . Constant log $A$ and log $B$ with $A > B$ are preselected so that $S$ decides for $s_0$ (and makes the appropriate motor action) as soon as the total falls below log $B$, provided the total has not previously exceeded log $A$ when the decision would have been made for $s_1$ . (The odd way of expressing the constants facilitates later references.) If the decision is made at the $n$th sample $T_d = nt$. The theory of the sequential probability ratio test [1] shows that the optimum choice of the function $c(x)$ is

$$c(x) = \log p_1(x) - \log p_0(x). \tag{1}$$

Such a function implies that $S$ is familiar with the probability distributions $p_0(x)$ and $p_1(x)$. Such familiarity may be the result of a process of learning, provided $S$ has performed many trials of the discrimination task and is given knowledge of results. $S$'s computer may be thought of as exploratory, trying out different $c(x)$'s until the optimal one is found. However it is conceivable that the distributions can be deduced by $S$ from the structure of the situation and then imposed on his computer. The optimality of (1) is stated by Wald [1] in the following terms: let $\bar{n}_0$, $\bar{n}_1$ be the averages of the number of samples necessary for decision when the signals presented are $s_0$, $s_1$, respectively. If $\bar{n}_0^*$, $\bar{n}_1^*$ are the averages for any other decision procedure based on $x_1$, $x_2$, etc., with smaller probabilities of incorrect response to $s_0$ and $s_1$, then $\bar{n}_0^* \geq \bar{n}_0$ and $\bar{n}_1^* \geq \bar{n}_1$. It is possible that this form of optimality does not appeal to $S$, who may have to be trained to use it by suitable reward.

Before testing the model, it must be remembered that it is $T$ which is measured and not $T_d$. Even so, a test is available which requires only the following assumption. Consider trials leading to a decision for $s_0$. The assumption is, given the value of $T_d$, that the distribution of $T_i + T_m$ is the same whether the decision is right or wrong. (The same assumption is made for decisions for $s_1$.) This does not exclude the possibility that $T_i + T_m$ and $T_d$ be correlated. The length of time, $T_i$, may affect the uncertainty in the information presented to the computer and therefore may affect $T_d$; alternatively, if $T_d$ is long, $T_m$ may be deliberately shortened. However, it does assume that $T_m$ cannot be influenced by information processed since the initiation of the motor action. In Appendix 1 it is shown that, with mild restrictions on $p_0(x)$ and $p_1(x)$, the distribution of the $n$'s, and therefore of the $T_d$'s, leading to a decision for $s_0$ (or of those leading to $s_1$) is the same whether the decisions are correct or incorrect. With the above assumption, this implies that the same result should hold for a comparison of the correct and incorrect $T$'s leading to $s_0$ (and for a comparison of those leading to $s_1$). This provides the basis of a reasonable test of the model. However, a fair proportion of errors would be needed to give a powerful test.

Without making assumptions about $p_0(x)$ and $p_1(x)$, it is difficult to think of more ways of examining the validity of the model. Since $x$ is an intervening variable without operational definition, it would clearly be unwise to assume much about $p_0(x)$ and $p_1(x)$. However, there is one assumption, called the "condition of symmetry," which in some discrimination tasks may be reasonable. This is that the distribution of $p_1(x)/p_0(x)$, when $x$ is distributed according to $p_0(x)$, is identical with that of $p_0(x)/p_1(x)$, when $x$ is distributed according to $p_1(x)$. It is shown in Appendix 2 that, if this condition holds,

$$\bar{n}_1/\bar{n}_0 = J(\beta, \alpha)/J(\alpha, \beta); \tag{2}$$

$$\begin{aligned} J(\alpha, \beta)v_1 - J(\beta, \alpha)v_0 \\ = 4[J(\beta, \alpha)\alpha(1 - \alpha)\bar{n}_1^2 - J(\alpha, \beta)\beta(1 - \beta)\bar{n}_0^2]/(1 - \alpha - \beta)^2, \end{aligned} \tag{3}$$

where $\alpha$ and $\beta$ are the probabilities of incorrect response to a single $s_0$ and $s_1$, respectively, $v_i$ is the variance of the sample sizes when $s_i$ is presented, and

$$J(\alpha, \beta) = \alpha \log [\alpha/(1 - \beta)] + (1 - \alpha) \log [(1 - \alpha)/\beta].$$

If it is feasible to estimate $T_d$ directly for each trial by eliminating $T_i + T_m$ from $T$, then (2) and (3) imply

$$\bar{T}_{d1}/\bar{T}_{d0} = J(\beta, \alpha)/J(\alpha, \beta), \tag{4}$$

$$\begin{aligned} J(\alpha, \beta) \text{ var } T_{d1} - J(\beta, \alpha) \text{ var } T_{d0} \\ = 4[J(\beta, \alpha)\alpha(1 - \alpha)\bar{T}_{d1}^2 - J(\alpha, \beta)\beta(1 - \beta)\bar{T}_{d0}^2]/(1 - \alpha - \beta)^2. \end{aligned} \tag{5}$$

Equations (4) and (5) are most relevant if $S$ can be persuaded to achieve different $(\alpha, \beta)$ combinations without changing the distributions $p_0(x)$ and $p_1(x)$. When $\alpha = \beta$, then $\bar{n}_0 = \bar{n}_1$ and $v_0 = v_1$; with the assumptions that $T_i + T_m$ is (i) uncorrelated with $T_d$ and (ii) independent of the signal presented, this implies equality of means and variances of reaction times to the signals. So, for the latter special case, it is not necessary to measure $T_d$.

For the "condition of symmetry" it is sufficient that, with $x$ represented as a number, $p_0(x) = p_1(x - d)$ for some number $d$ with $p_0(x)$ symmetrical about its mean. This might occur when $s_0$, $s_1$ are signals which are close together on some scale and the error added to the signals to make $x$ has the same distribution for each signal. Symmetry would not be expected in absolute threshold discriminations or in the discrimination of widely different colors in a color-noisy background. Another sufficient condition is that $x$ be bivariate, $[x(1), x(2)]$, the probabilities under $s_0$ obtained from those under $s_1$ by interchanging $x(1)$ and $x(2)$. For instance, $x(1)$ and $x(2)$ may be the inputs on two noisy channels and $s_0$ consists of stimulation of the first while $s_1$ consists of stimulation of the second.

A further prediction of the model for the symmetrical case can be made when $S$ is persuaded by a suitable reward to give equal weight to errors to $s_0$ and $s_1$, that is to minimize his unconditional error probability, by adjustment of the constants $A$ and $B$ in his computer. If $p_0$ is the frequency of presentation of $s_0$ then the error probability is $p_0 \alpha + (1 - p_0)\beta$ or $e$, say, and the average decision time is $p_0\bar{T}_{d0} + (1 - p_0)\bar{T}_{d1}$ or $\bar{T}_d$, say. It is shown in Appendix 3 that, provided $10e < p_0 < 1 - 10e$, the minimization results in the following relation between $\bar{T}_d$, $e$ and $p_0$ :

$$\bar{T}_d \propto [J(e, 1 - e) - J(p_0, 1 - p_0)].$$

*The Non-Sequential Fixed-Sample Case*

If $S$ has an incentive to react quickly and correctly, then the advantage of the sequential decision procedure is that those discriminations which by chance happen to be easy are made quickly and time is saved. However it is possible that $S$ may adopt a different, less efficient strategy—which is to fix $T_d$ for all trials at a value which will give a certain accepted error rate. Let the sample size corresponding to this decision time be $n$. The likelihood ratio procedures are as follows: decide for $s_0$ if $c(x_1) + \cdots + c(x_n) < \log C$; decide for $s_1$ if $c(x_1) + \cdots + c(x_n) \geq \log C$; $c(x) = \log p_1(x) - \log p_0(x)$ and $C > 0$. These procedures are optimal in the sense that, if any other procedure based on $x_1, \cdots, x_n$ is used, there exists one of the likelihood ratio procedures with smaller error probabilities. It was remarkable that in the sequential case useful predictions were obtainable under mild restrictions on $p_0(x)$ and $p_1(x)$. Unfortunately this does not hold for the fixed-sample case, making more difficult the problem of testing whether such a model holds.

If there is no input storage, it is possible that the results of the self-imposed strategy just outlined are equivalent to those obtainable when the experimenter himself cuts off the signals after an exposure time $T_d$. But this is the type of situation considered by Peterson and Birdsall [2]. The emphasis of these authors is mainly on the external parameters (such as energy) rather than on any supposed intervening variable. They define a set of physical situations for auditory discrimination in terms of a parameter $d$, which is equivalent to the difference between the means of two normal populations with unit variance. (For, in the cases considered, it happens that the logarithm of the likelihood ratio of the actual physical random variables for the two alternatives is normally distributed with equality of variance under the two alternatives.) This parameter sets a limit to the various performances (error probabilities to $s_0$ and $s_1$) of any discriminator using the whole of the physical information. It therefore sets an upper bound on the performance of $S$ who can only use less than the whole. In [2] the authors make the assumption that the information on the basis of which $S$ makes his discrimination nevertheless gives normality of logarithm of the likelihood ratio. They examine data to see whether $S$ is producing error frequencies that lie on a curve defined by a $d$ greater than that in the external situation.

## *More than Two Alternatives*

For $m$ alternatives there are $m$ probability distributions for the intervening variable $x$ (which may be multivariate); that is, signal $s_i$ induces an

$x$ with the probability distribution $p_i(x)$ for $i = 1, \cdots, m$. We will consider the consequences of a fixed-sample decision procedure based on $x_1, \cdots, x_n$, where $n$ is fixed.

If the signals are presented independently with probabilities $p_1, \cdots, p_m$ (adding to unity) and if $\alpha_i(\mathfrak{D})$ is the probability of error to signal $s_i$ when the decision procedure $\mathfrak{D}$ (based on $x_1, \cdots, x_n$) is used, then the probability of error to a single presentation is

$$e = \sum_1^m p_i \alpha_i(\mathfrak{D}).$$

It is shown in Appendix 4 that the $\mathfrak{D}$ minimizing $e$ is that which effectively selects the signal with maximum posterior probability. In this section, this minimum $e$ will be related to $n$ (or $T_d/t$) and $m$ when distributions are normal. However in the validation of the model it might be necessary to supplement $T_d$ with a time $T_{\bar{d}}$, representing the time the computer requires to examine the $m$ posterior probabilities to decide which is the largest. For, although it might be reasonable to suppose that $T_i + T_m$ is independent of $m$, one would expect $T_{\bar{d}}$ to vary with $m$. The simplest model for $T_{\bar{d}}$ would be to suppose that $T_{\bar{d}} = (m - 1)t'$, where $t'$ is the time necessary to compare any two of the probabilities and decide which is the larger.

We will state the relation between $n$ and $m$ when $e$ is constant in the following special case (treated by Peterson and Birdsall [3], who stated the relation between $e$ and $m$ when $n$ is held constant by the experimenter): we take $p_1 = p_2 = \cdots = p_m = 1/m$ and $x$ a multivariate random variable $x(1), \cdots, x(m)$. Under $s_i$, suppose that $x(1), \cdots, x(m)$ are independent and that $x(i)$ is normally distributed with mean $\mu > 0$ and unit variance, while the other components of $x$ are normal with zero means and unit variances. Thus there is all-round symmetry. $x(1), \cdots, x(m)$ can be regarded as the inputs on $m$ similar channels. The $i$th channel is stimulated under $s_i$. It is readily seen that the optimal procedure is to choose the signal corresponding to the channel with the largest total. It is shown in Appendix 5 that, with this procedure,

$$n\mu^2 = \{1 + [0.64(m - 1)^{-1/2} + 0.45]^2\}[\Phi^{-1}(1 - e) - \Phi^{-1}(1/m)]^2$$

for those $m$ for which $e < 1 - (1/m)$. $\Phi^{-1}$ is the inverse of the normal standardized distribution function. The values of $n\mu^2$ for certain values of $e$ and $m$ have been calculated. If $\mu$ is independent of $m$, then $T_d$ is proportional to $n\mu^2$ and the results are plotted in Figure 1. It can be seen that $T_d$ is very nearly linear against log $m$, which agrees with some experimental findings in this field.

The question may be raised whether any $m$-choice task can obey the symmetry condition of the model. Peterson and Birdsall apply the model to the case where an auditory signal is presented in one of four equal periods

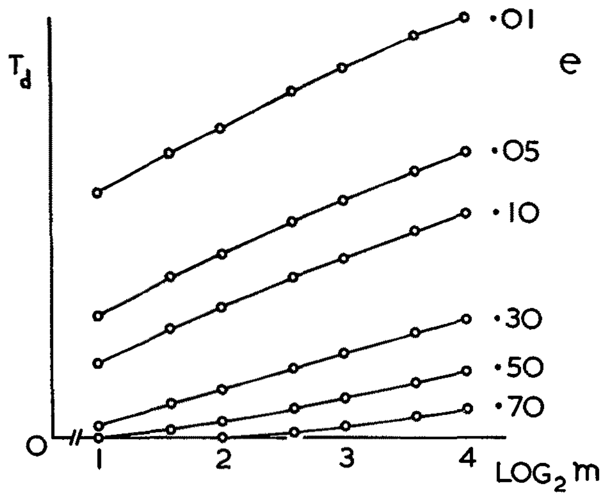



FIGURE 1
The Decision Time ($T_d$) for Error Rate ($e$) and Number of Equally Likely Alternatives ($m$)

of an exposure of $S$ to "white" noise. In this case symmetry is superficially present, but any memory difficulties of $S$ would upset it. We would not expect the model to apply to the case of response to one of $m$ fairly easily discriminable lights arranged in some display, for the noise would be highly positional. However, in the case where the lights are patches of white noise on one of which a low intensity visual signal is superimposed so that response is difficult, the positional effect may not be important and there may be symmetry.

*Appendix* 1

Let $n_{ij}$ be the sample size for a decision in favor of $s_i$ when $s_j$ is presented. The distribution of $n_{ij}$ is completely determined by its moment generating function, $\psi_{ij}$. From $A5.1$ of [1], if

$$\phi_i(t) = \sum_x p_i(x)[p_1(x)/p_0(x)]^t,$$

then

$$(1 - \alpha)B^t\psi_{00}[-\log \phi_0(t)] + \alpha A^t\psi_{10}[-\log \phi_0(t)] \equiv 1, \tag{6}$$

$$\beta B^t\psi_{01}[-\log \phi_1(t)] + (1 - \beta)A^t\psi_{11}[-\log \phi_1(t)] \equiv 1, \tag{7}$$

provided the quantities $E_i$, $V_i$ defined in Appendix 2 are small. If $\alpha < 0.1$ and $\beta < 0.1$ then to a good approximation $A = (1 - \beta)/\alpha$ and $B = \beta/(1 - \alpha)$. Now $\phi_0(1 + u) = \phi_1(u)$; so, putting $t = 1 + u$ in (6) and (7),

$$\beta B^u\psi_{00}[-\log \phi_1(u)] + (1 - \beta)A^u\psi_{10}[-\log \phi_1(u)] \equiv 1,$$

$$(1 - \alpha)B^u\psi_{01}[-\log \phi_0(u)] + \alpha A^u\psi_{11}[-\log \phi_0(u)] \equiv 1.$$

By comparing these equations with (6) and (7), it is found that $\psi_{00} = \psi_{01}$ and $\psi_{10} = \psi_{11}$. Therefore the distributions of $n_{00}$ and $n_{01}$ (and similarly those of $n_{10}$ and $n_{11}$) are identical.

*Appendix* 2

In the case of symmetry,

$$\sum_x p_0(x) \log [p_0(x)/p_1(x)] = \sum_x p_1(x) \log [p_1(x)/p_0(x)] = E,$$

and

var log $[p_0(x)/p_1(x)]$ under $p_0(x)$ = var log $[p_1(x)/p_0(x)]$ under $p_1(x) = V$.

From $A$:72 of [1], if $E$ and $V$ are small,

$$(8) \qquad \bar{n}_0 = J(\alpha, \beta)/E; \qquad \bar{n}_1 = J(\beta, \alpha)/E.$$

Therefore

$$\bar{n}_1/\bar{n}_0 = J(\beta, \alpha)/J(\alpha, \beta).$$

By differentiating (6) twice with respect to $t$ and substituting $t = 0$, using (8) and the fact that $\psi_{ij}$ is the moment generating function of $n_{ij}$,

$$v_0 = [VJ(\alpha, \beta)/E^3] - 4[\alpha(1 - \alpha)\bar{n}_1^2/(1 - \alpha - \beta)^2].$$

By symmetry

$$v_1 = [VJ(\beta, \alpha)/E^3] - 4[\beta(1 - \beta)\bar{n}_0^2/(1 - \alpha - \beta)^2].$$

Hence

$$J(\alpha, \beta)v_1 - J(\beta, \alpha)v_0 = 4[J(\beta, \alpha)\alpha(1 - \alpha)\bar{n}_1^2 - J(\alpha, \beta)\beta(1 - \beta)\bar{n}_0^2]/(1 - \alpha - \beta)^2.$$

*Appendix* 3

If $\alpha < 0.1$ and $\beta < 0.1$ then, by (8), $\bar{T}_d \propto p_0 J(\alpha, \beta) + (1 - p_0)J(\beta, \alpha)$. Keeping $e$ [or $p_0\,\alpha + (1 - p_0)\beta$] constant at a value in the range given by $10e < p_0 < 1 - 10e$, the condition on $\alpha$ and $\beta$ will be satisfied. It is found by the usual methods that the minimum $\bar{T}_d$ is proportional to $J(e, 1 - e) - J(p_0, 1 - p_0)$.

*Appendix* 4

Let $X$ be the set of all possible values of $x = (x_1, \cdots, x_n)$ and $X_i$ the set of $x$ for which a decision is made for $s_i$. Then

$$e = \sum_{i=1}^{m} p_i \sum_{x \varepsilon X - X_i} p_i(x).$$

Suppose $X_i$ and $X_j$ have a common boundary; then, for $e$ to be a minimum, it will not be changed by small displacements in this boundary. Hence, on the boundary, $p_i p_i(x) = p_j p_j(x)$; that is, the posterior probability of $s_i$ equals that of $s_j$. Considering all possible boundaries, the solution is that $X_i$ is the set of $x$'s for which $s_i$ has greater posterior probability than the other signals.

*Appendix* 5

Write

$$\bar{x}(i) = \sum_{s=1}^{n} x_s(i)/n.$$

Then, under $s_1$, $\sqrt{n}\bar{x}(1)$ is $N(\sqrt{n}\mu, 1)$ and $\sqrt{n}\bar{x}(i)$ is $N(0, 1)$ for $i \neq 1$. Therefore,

$$\alpha_1(\mathfrak{D}) = \cdots = \alpha_m(\mathfrak{D})$$

$$= 1 - (2\pi)^{-1/2} \int_{-\infty}^{\infty} [\Phi(u)]^{m-1} \exp [-\tfrac{1}{2}(u - \sqrt{n}\,\mu)^2]\, du.$$

On integration by parts,

$$(9) \qquad e = \sum p_i \alpha_i(\mathfrak{D})$$

$$= (m - 1)(2\pi)^{-1/2} \int_{-\infty}^{\infty} \Phi(u)^{m-2} \Phi(u - \sqrt{n}\,\mu) \exp(-\tfrac{1}{2}u^2)\, du$$

$$= e_m(\theta),$$

say, where $\theta = \sqrt{n}\mu$. Peterson and Birdsall [3] use this form as the basis of their tabulation. However $e_m(\theta) \to 0$ as $\theta \to \infty$ and $e_m(\theta) \to 1$ as $\theta \to -\infty$; while $e_m'(\theta) \leq 0$. Therefore $|e_m'(\theta)|$ is a "probability density function" for $\theta$. The characteristic function and hence the distribution of $\theta$ turns out to be the same as that of $v + w$, where $w = \max(v_1, \cdots, v_{m-1})$ and $v, v_1, \cdots, v_{m-1}$ are $m$ independent standard normal variables. Referring to Graph 4.2.2(7) of [4], it can be seen that, for $m < 20$, the first and second moment quotients of $w$ are not very different from those of a normal distribution. Also the addition of $v$ to $w$ will improve normality. Hence $\theta$ is approximately normal, agreeing with the calculations of Peterson and Birdsall. If $\theta$ is $N(\nu, \sigma^2)$, we determine $\nu$ and $\sigma^2$ as follows. From (9), $e_m(0) = 1 - (1/m)$. Also $e_m(0) = 1 - \Phi(-\nu/\sigma)$. Therefore

$$\nu/\sigma = -\Phi^{-1}(1/m).$$

Also $\sigma^2 = \operatorname{var} v + \operatorname{var} w$ and from Graph 4.2.2(6) of [4], $\operatorname{var} w = [0.64\,(m - 1)^{-\frac{1}{2}} + 0.45]^2$ for $m < 20$, which determines $\sigma^2$. Putting $e_m(\theta) = e$, the constant error rate,

$$n\mu^2 = \{1 + [0.64(m - 1)^{-1/2} + 0.45]^2\}[\Phi^{-1}(1 - e) - \Phi^{-1}(1/m)]^2.$$

## REFERENCES

[1] Wald, A. *Sequential analysis*, New York: Wiley, 1947.
[2] Peterson, W. W. and Birdsall, T. G. The theory of signal detectability. Tech. Rep. No. 13, Electronic Defense Group, Univ. Michigan, 1953.
[3] Peterson, W. W. and Birdsall, T. G. The probability of a correct decision in a forced choice among $M$ alternatives. Quarterly Prog. Rep. No. 10, Electronic Defense Group, Univ. Michigan, 1954.
[4] Gumbel, E. J. *Statistics of extremes*. New York: Columbia Univ. Press, 1958.

*Manuscript received 10/26/59*

*Revised manuscript received 1/4/60*