# A MEASURE OF ASYMPTOTIC EFFICIENCY FOR TESTS OF A HYPOTHESIS BASED ON THE SUM OF OBSERVATIONS[1]

By Herman Chernoff

*University of Illinois and Stanford University*

**1. Summary.** In many cases an optimum or computationally convenient test of a simple hypothesis $H_0$ against a simple alternative $H_1$ may be given in the following form. Reject $H_0$ if $S_n = \sum_{j=1}^{n} X_j \leqq k$, where $X_1, X_2, \cdots, X_n$ are $n$ independent observations of a chance variable $X$ whose distribution depends on the true hypothesis and where $k$ is some appropriate number. In particular the likelihood ratio test for fixed sample size can be reduced to this form. It is shown that with each test of the above form there is associated an index $\rho$. If $\rho_1$ and $\rho_2$ are the indices corresponding to two alternative tests $e = \log \rho_1 / \log \rho_2$ measures the relative efficiency of these tests in the following sense. For large samples, a sample of size $n$ with the first test will give about the same probabilities of error as a sample of size $en$ with the second test.

To obtain the above result, use is made of the fact that $P(S_n \leqq na)$ behaves roughly like $m^n$ where $m$ is the minimum value assumed by the moment generating function of $X - a$.

It is shown that if $H_0$ and $H_1$ specify probability distributions of $X$ which are very close to each other, one may approximate $\rho$ by assuming that $X$ is normally distributed.

**2. Introduction.** The problem of the efficiency of a test is of relevance to statisticians who are faced with either of the following two problems. The first problem is that of the design of an experiment. The second problem is that of deciding which test combines computational feasibility and efficiency per observation. The measure of efficiency with which we shall deal is especially relevant to problems which involve large samples whose size is determined by the experimenter.

The motivation for the results of this paper may be seen by considering the following simple example. Suppose that under the hypothesis $H_i$,

$$P(X = 1) = p_i, \qquad P(X = 0) = 1 - p_i, \qquad i = 0, 1, \qquad p_1 > p_0. \tag{2.1}$$

Then the likelihood ratio test reduces to that of rejecting $H_0$ if $S_n = \sum_{j=1}^{n} X_j$ exceeds some number $k$. If $n = 400$, $p_0 = .4$, $p_1 = .5$, and $k = 180$, one may reliably proceed to compute the probabilities of error by using the normal approximation to the distribution of $S_n$. On the other hand, if $n$ is very large, (say, 1,000,000) the difference between the means of $S_n$ under $H_0$ and $H_1$ is so large

[1] This paper was prepared with the support of the Office of Naval Research.

compared to the standard deviation of $S_n$ (ratio of 200) that the probabilities to be computed correspond to the extreme tails of the distributions of $S_n$ and the normal approximation is inapplicable. We note that this objection would not be serious for $n = 1{,}000{,}000$ if $p_0$ were very close to $p_1$ (say, $p_0 = .499$) for then

$$(np_1 - np_0)/\sqrt{np_iq_i} = \sqrt{n}(p_1 - p_0)/\sqrt{p_iq_i} \approx 2.$$

This situation immediately gives rise to the question of what is the behaviour of the probability distribution of $S_n$ in the tails of its distribution. This question was treated by H. Cramér [1] and is considered in Section 3 where Theorem 1 states that if $a \leqq E(X)$, $P(S_n \leqq na)$ is roughly like $m^n$ where $m$ is the minimum value assumed by the moment generating function of $X - a$. In Section 4 this result is applied to obtain a theorem which states the following result: If $k$ is selected to minimize $\beta + \lambda\alpha$ where $\lambda$ is some given positive number and $\alpha = P(S_n > k \mid H_0)$ and $\beta = P(S_n \leqq k \mid H_1)$ are the probabilities of error, the minimum value of $\beta + \lambda\alpha$ behaves roughly like $\rho^n$, where $\rho$ does not depend on $\lambda$. Now the notion of efficiency is immediately suggested by the equation

$$\rho_1^{n_1} = \rho_2^{n_2}. \tag{2.2}$$

We may note that in the above example, one may be justified in using the normal approximation to the distribution of $S_n$ for relatively large $n$ if $p_1 - p_0$ is small. This tends to suggest that, if the hypotheses $H_0$ and $H_1$ are very "close" to each other, $\rho$ may be approximated by assuming $X$ to be normally distributed. This conjecture is in fact borne out by the theorems of Section 5.

**3. The distribution of $S_n$ in the tails.** In this section we shall discuss the distribution in the tails of the sum of $n$ independent observations on a chance variable $X$. Excellent results on this problem were obtained by H. Cramér [1] under the conditions that the moment generating function $M(t)$ of $X$ exists (finite) for some interval $-A < t < A$, and that the cumulative distribution function of the chance variables have an absolutely continuous component. This latter condition is not satisfied by discrete distributions. This condition was imposed in order to apply a bound on the error of the normal approximation to the distribution of a sum of chance variables. C. G. Esseen [2] obtained this bound using only the (finite) existence of third order moments. For the case in which we are interested (i.e., $P(S_n \leqq na)$), the former condition may also be relaxed so that $M(t)$ exists (finite) for $-A < t \leqq 0$ if $a < E(X)$.

Since the results of Cramér are extremely more powerful that we require here and the (finite) existence of third order moments is not necessary for the results that we desire, we shall state and briefly outline a proof of Theorem 1. Before doing this we shall first formally state some notation and lemmas which we shall use throughout this paper. These lemmas state known results which are rather obvious, depending mainly on Lebesgue's Theorem on integration of monotone sequences [3].

NOTATION 1. *$S_n$ is the sum of n independent observations $X_1, X_2, \cdots, X_n$ on a chance variable X with moment generating function $M(t) = E(e^{tX})$ and cumulative distribution function $F(x) = P(X \leqq x)$. Let*

$$m(a) = \inf E(e^{t(X-a)}) = \inf e^{-at}M(t) \tag{3.1}$$

*(infimum with respect to real values of t).*

*Unless otherwise specified we shall say that an expectation exists if it is $+\infty$ or if it is $-\infty$. We shall say that $E(g(X))$ fails to exist if both*

$$\int_{g(x)<0} g(x)\, dF(x) = -\infty \quad \text{and} \quad \int_{g(x)>0} g(x)\, dF(x) = +\infty.$$

*We shall denote by $f(\infty)$ the limit of $f(x)$ as x approaches $\infty$.*

LEMMA 1. *$M(t)$ attains its minimum value $m(0)$. This value is attained for finite t unless $P(X > 0) = 0$ or $P(X < 0) = 0$. In that event $m(0) = P(X = 0)$.*

LEMMA 2. *If $P(X \leqq 0) > 0$ and $P(X \geqq 0) > 0$, then $m(0) > 0$.*

LEMMA 3. *For all t in the interior of the interval of finite existence of $M(t)$*

$$\frac{dM}{dt} = \int_{-\infty}^{\infty} xe^{tx}\, dF(x) \tag{3.2}$$

*and*

$$\frac{d^2M}{dt^2} = \int_{-\infty}^{\infty} x^2e^{tx}\, dF(x) \geqq 0. \tag{3.3}$$

*Furthermore, $\frac{d^2M}{dt^2} = 0$ if and only if $P(X = 0) = 1$.*

LEMMA 4. *If $u_1(t), u_2(t), \cdots, u_n(t), \cdots$ is a nondecreasing sequence of functions continuous in the closed interval $[a, b]$,*

$$\lim_{n\to\infty} [\inf_{a \leqq t \leqq b} u_n(t)] = \inf_{a \leqq t \leqq b} [\lim_{n\to\infty} u_n(t)]. \tag{3.4}$$

This statement applies to the extended case where $u_n(t)$ may take on the value $\infty$ and to the case where $a = -\infty$ providing $u_n(-\infty) = \lim_{t\to-\infty} u_n(t)$.

THEOREM 1. *If $E(X) > -\infty$ and $a \leqq E(X)$, then*

$$P(S_n \leqq na) \leqq [m(a)]^n. \tag{3.5}$$

*If $E(X) < +\infty$ and $a \geqq E(X)$, then*

$$P(S_n \geqq na) \leqq [m(a)]^n. \tag{3.6}$$

*If $0 < \epsilon < m(a)$ ($E(X)$ need not exist),*

$$\lim_{n\to\infty} \frac{(m(a) - \epsilon)^n}{P(S_n \leqq na)} = 0 \tag{3.7}$$

*and*

$$\lim_{n\to\infty} \frac{(m(a) - \epsilon)^n}{P(S_n \geqq na)} = 0. \tag{3.8}$$

PROOF. We present here a brief sketch of a proof of Theorem 1. We note first that it suffices to prove (3.5) for $a = 0 \leqq E(X)$, and (3.7) for $a = 0$. Using the extended Tchebycheff inequality [4]

$$E(e^{tS_n}) = [M(t)]^n \geqq P[S_n \leqq 0] \qquad \text{for } t \leqq 0. \tag{3.9}$$

Hence

$$P[S_n \leqq 0] \leqq [\inf_{t \leqq 0} M(t)]^n. \tag{3.10}$$

But $a \leqq E(X)$ implies that

$$\inf_{t \leqq 0} M(t) = \inf M(t) = m(0). \tag{3.11}$$

To establish equation (3.7) we note that it is sufficient to treat the case $a = 0$. Then we see that the cases where $P(X > 0) = 0$ and where $P(X < 0) = 0$ are trivial. Hereafter we shall assume that $P(X > 0) > 0$ and $P(X < 0) > 0$.

We shall now treat the discrete (but not necessarily bounded) case where $P(X = x_i) = p_i > 0$, $i = 1, 2, \cdots$. Given $\epsilon > 0$, one may select an integer $r$ so that

$$\min(x_1, x_2, \cdots, x_r) < 0 < \max(x_1, x_2, \cdots, x_r) \tag{3.12}$$

and

$$\inf\left\{\sum_{i=1}^{r} e^{tx_i} p_i\right\} > \inf\left\{\sum_{i=1}^{\infty} e^{tx_i} p_i\right\} - \frac{\epsilon}{2}. \tag{3.13}$$

In fact, let

$$m^* = \sum_{i=1}^{r} e^{t^* x_i} p_i = \inf\left\{\sum_{i=1}^{r} e^{tx_i} p_i\right\}. \tag{3.14}$$

For this discrete case it now suffices to show that for sufficiently large $n$ there are $r$ positive integers $n_1, n_2, \cdots, n_r$ such that

$$\sum_{i=1}^{r} n_i = n, \tag{3.15}$$

$$\sum_{i=1}^{r} n_i x_i \leqq 0, \tag{3.16}$$

$$P(n_1, n_2, \cdots, n_r) = \frac{n!\, p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}}{n_1!\, n_2! \cdots n_i!} > \left(m^* - \frac{\epsilon}{2}\right)^n. \tag{3.17}$$

For large $n_1, n_2, \cdots, n_r$ (not necessarily integers) Stirling's Formula gives us

$$P(n_1, n_2, \cdots, n_r) \geqq \left\{\prod_{i=1}^{r} \left(\frac{np_i}{n_i}\right)^{n_i}\right\} \cdot \frac{1}{n^{r/2}}. \tag{3.18}$$

Now

$$Q(n_1, n_2, \cdots, n_r) = \prod_{i=1}^{r} \left(\frac{np_i}{n_i}\right)^{n_i} \tag{3.19}$$

can be shown by the method of Lagrange multipliers to attain a maximum of $(m^*)^n$ subject to the restrictions

$$\sum_{i=1}^{r} n_i = n, \tag{3.20}$$

$$\sum_{i=1}^{r} n_i x_i = 0, \tag{3.21}$$

$$n_i > 0, \qquad i = 1, 2, \cdots, r, \tag{3.22}$$

and the maximizing values of $n_1, n_2, \cdots, n_r$ are

$$n_i^{(0)} = np_i e^{t^* x_i}/m^*. \tag{3.23}$$

Assuming that $x_1 \leqq x_i$ for $i \leqq r$, we let

$$n_i^{(1)} = [n_i^{(0)}], \qquad 2 \leqq i \leqq r, \tag{3.24}$$

$$n_1^{(1)} = n - \sum_{i=2}^{r} n_i^{(1)}, \tag{3.25}$$

where $[n_i^{(0)}]$ represents the greatest integer less than or equal to $n_i^{(0)}$. Then for large $n$, the $n_i^{(1)}$ are large positive integers adding up to $n$ for which

$$\textstyle\sum_{i=1}^{r} n_i^1 x_i \leqq 0$$

and

$$Q(n_1^{(1)} n_2^{(1)}, \cdots, n_r^{(1)}) \geqq \left(\frac{p_1}{n}\right)^r (m^*)^n \tag{3.26}$$

and

$$P(n_1^{(1)}, n_2^{(2)}, \cdots, n_r^{(1)}) \geqq \frac{(m^*)^n p_1^r}{n^{3r/2}} \geqq \left(m^* - \frac{\epsilon}{2}\right)^n, \tag{3.27}$$

which was to be shown.

We shall now treat the general case. Let

$$X^{(s)} = \frac{i}{s} \text{ if } \frac{i-1}{s} < X \leqq \frac{i}{s}, \qquad i = \cdots, -1, 0, 1, \cdots, \quad s = 1, 2, \cdots. \tag{3.28}$$

If $S_n^{(s)}$ represents the sum of the $X^{(s)}$ for $n$ independent observations

$$P(S_n \leqq 0) \geqq P(S_n^{(s)} \leqq 0) \tag{3.29}$$

and

$$M^{(s)}(t) = E(e^{tX^{(s)}}) \geqq e^{-|t|/s}M(t). \tag{3.30}$$

Since $P(X > 0) > 0$ and $P(X < 0) > 0$, $M(t)$ attains its minimum for a finite value of $t$ and hence there is an $s$ sufficiently large so that

$$\inf \{M^{(s)}(t)\} \geqq \inf \{M(t)\} - \tfrac{1}{2}\epsilon. \tag{3.31}$$

Our theorem follows from the result for the discrete case and equation (3.29).

**4. The measure of asymptotic efficiency.** In this section some elementary monotonicity and continuity properties of $m(a)$ are obtained. These properties are then used to obtain an index $\rho$ for a test. This index has the property that if $k$ is chosen to minimize

$$\beta + \lambda\alpha = P[S_n \leqq k \mid H_1] + \lambda P[S_n > k \mid H_0], \tag{4.1}$$

the minimum value of $\beta + \lambda\alpha$ is roughly about $\rho^n$. Furthermore, $\rho$ is independent of $\lambda$. From this it is easily seen that if $\rho_1$ and $\rho_2$ are the indices of two tests, $\log \rho_1/\log \rho_2$ is an appropriate measure of the relative efficiency of these tests.

NOTATION 2. *Let $a_e$ be defined by*

$$P(X < a_e) = 0 \tag{4.2}$$

*and*

$$P(X < a_e + \epsilon) > 0 \qquad \text{for every } \epsilon > 0. \tag{4.3}$$

*Let $t(a)$ be given by*

$$m(a) = e^{-at(a)}M[t(a)]. \tag{4.4}$$

Note that Lemma 1 implies that $t(a)$ exists and that Lemma 3 implies that $t(a)$ is unique unless $P(X = a) = 1$.

LEMMA 5. *If $E(X) > -\infty$ and $M(t) = \infty$ for $t < 0$, then $t(a) = 0$ and $m(a) = 1$ for $a \leqq E(X)$.*

PROOF. From the proof of (3.5), it follows that $t(a) \leqq 0$ for $a \leqq E(X)$. Lemma 5 follows immediately.

LEMMA 6. *If $M(t) < \infty$ for some $t < 0$, then $E(x) > -\infty$. Furthermore,*

$$m(a) = 0, \qquad a < a_e, \tag{4.5}$$

$$m(a_e) = P(X = a_e), \tag{4.6}$$

*and*

$$m[E(X)] = 1. \tag{4.7}$$

*Also, $m(a)$ is continuous and strictly monotone increasing for $a_e \leqq a \leqq E(X)$.*

PROOF. That $E(X)$ exists (finite) or is $+\infty$ is apparent. For $a < a_e$ and $t < 0$,

$$\int_{-\infty}^{\infty} e^{t(x-a)}\, dF(x) < e^{t(a_e-a)}. \tag{4.8}$$

Hence $m(a) = 0$. Now we note that if $a_e$ is finite

$$P(X = a) \leqq \int_{-\infty}^{\infty} e^{t(x-a)}\, dF(x), \tag{4.9}$$

$$\lim_{t \to -\infty} \int_{-\infty}^{\infty} e^{t(x-a_e)}\, dF(x) = P(X = a_e) \tag{4.10}$$

and hence $m(a_e) = P(X = a_e)$. If $a_e = -\infty$,

$$\lim_{a \to -\infty} \int_{-\infty}^{\infty} e^{t(x-a)}\, dF(x) = 0, \qquad t < 0, \tag{4.11}$$

so that $\lim_{a \to -\infty} m(a) = 0$. Now we note that

$$\lim_{t \to 0-} \frac{d}{dt}\,[e^{-at} M(t)] = \int_{-\infty}^{\infty} (x - a)\, dF(x). \tag{4.12}$$

Since $(d^2/dt^2)[e^{-at}M(t)] > 0$, unless $P(X = a) = 1$ in which case Lemma 6 is valid, it follows that $t(a) < 0$ for $a < E(X)$ and $t(a) = 0$ for $a = E(X)$. Hence $m[E(X)] = 1$, and $m(a) < 1$ for $a < E(X)$.

We shall now show that for $a_e < a < E(X)$, $t(a)$ is finite and a non-decreasing function of $a$, while $m(a)$ is strictly increasing for $a_e \leqq a < E(X)$. The finiteness of $t(a)$ follows from

$$\int_{-\infty}^{\infty} e^{t(x-a)}\, dF(x) \geqq P(X < a - \epsilon)e^{-\epsilon t} \tag{4.13}$$

for $t < 0$, $\epsilon > 0$. Therefore,

$$m(a - h) \leqq \int_{-\infty}^{\infty} e^{t(a)(x-a+h)}\, dF(x) < m(a), \qquad h > 0, \tag{4.14}$$

and furthermore

$$\int_{-\infty}^{\infty} e^{t'(x-a+h)}\, dF(x) \geqq \int_{-\infty}^{\infty} e^{t(a)(x-a+h)}\, dF(x) \tag{4.15}$$

for $t' > t(a)$, $h > 0$.

It suffices now to show that $m(a)$ is continuous on the right for $a < E(X)$ and continuous on the left for $a_e < a \leqq E(X)$. Given $a < E(X)$ and $\epsilon > 0$, there is a finite $t'$ so that

$$\int_{-\infty}^{\infty} e^{t'(x-a)}\, dF(x) \leqq m(a) + \epsilon, \tag{4.16}$$

$$\lim_{h \to 0+} m(a + h) \leqq \lim_{h \to 0+} \int_{-\infty}^{\infty} e^{t'(x-a-h)} \, dF(x) \tag{4.17}$$
$$\leqq m(a) + \epsilon.$$

Given $a$ so that $a_e < a \leqq E(X)$, there is an $h_1 > 0$ so that $a_e < a - h_1$. For $t(a - h_1) \leqq t \leqq t(a)$, $\int_{-\infty}^{\infty} e^{t(x-a+h)} \, dF(x)$ converges uniformly to $\int_{-\infty}^{\infty} e^{t(x-a)} \, dF(x)$ as $h \to 0+$. Hence $\lim_{h \to 0+} m(a - h) \geqq m(a)$.

NOTATION 3. *$H_0$ and $H_1$ are two hypotheses which specify the distribution of $X$ so that $\mu_0 = E(X \mid H_0) \leqq \mu_1 = E(X \mid H_1)$. For each value of $a$ we consider a test which consists of rejecting $H_0$ if $S_n > na$. Let $\alpha = P(S_n > na \mid H_0)$, $\beta = P(S_n \leqq na \mid H_1)$ and $\lambda$ be any (finite) positive number given in advance. Let*

$$\rho(a) = \max\,[m_0(a), m_1(a)], \tag{4.18}$$

*where*

$$m_i(a) = \inf E(e^{t(X-a)} \mid H_i), \qquad i = 0, 1. \tag{4.19}$$

*Furthermore, let us define the index of the test determined by $X$ by*

$$\rho = \inf_{\mu_0 \leqq a \leqq \mu_1} \rho(a). \tag{4.20}$$

We note that in the event that it is desired to use a test where we reject $H_0$ if $S_n \geqq na$, one may replace $X$ by $-X$ and interchange $H_0$ and $H_1$. The value of $\rho$ is not affected by this transformation.

The customary procedure of minimizing $\beta$ for a fixed value of $\alpha$ does not seem very appropriate when the sample size approaches infinity. We shall instead deal with test which minimize $\beta + \lambda\alpha$ for some fixed value of $\lambda$, $0 < \lambda < \infty$. Such a test is a "Bayes Solution" corresponding to some a priori probability of $H_0$ which depends on $\lambda$. The study of Bayes Solutions may here be justified on grounds not involving any belief in a priori probabilities. In particular, if it is desired to minimize some function $F(\alpha, \beta)$ for large samples and neither $\partial F/\partial\alpha$ nor $\partial F/\partial\beta$ vanish at $\alpha = \beta = 0$, the minimizing test will correspond to a Bayes Solution where $\lambda$ is close to $\dfrac{\partial F(0, 0)}{\partial \beta} \Big/ \dfrac{\partial F(0, 0)}{\partial \alpha}$.

THEOREM 2. *Given $\epsilon$ and $\lambda$, $\epsilon > 0$ and $0 < \lambda < \infty$, then*

$$\lim_{n \to \infty} \left\{ \inf_a \, (\beta + \lambda\alpha)/(\rho + \epsilon)^n \right\} = 0 \tag{4.21}$$

*and if* $0 < \epsilon < \rho$

$$\lim_{n \to \infty} \left\{ \inf_a \, (\beta + \lambda\alpha)/(\rho - \varepsilon)^n \right\} = \infty. \tag{4.22}$$

PROOF. There is a value $a_0$ of $a$ so that $\rho(a_0) \leqq \rho + \epsilon/2$. Applying Theorem 1, equation (4.21) follows immediately. Now we note that

$$\beta = P(S_n \leqq na \mid H_1) \geqq P(S_n \leqq na_1 \mid H_1), \qquad a \geqq a_1, \tag{4.23}$$

$$\inf_{a \geqq a_1} (\beta + \lambda\alpha) \geqq P(S_n \leqq na_1 \mid H_1) \tag{4.24}$$

$$\alpha = P(S_n > na \mid H_0) \geqq P(S_n \geqq na_1 \mid H_0), \qquad a < a_1, \tag{4.25}$$

$$\inf_{a < a_1} (\beta + \lambda\alpha) \geqq \lambda P(S_n \geqq na_1 \mid H_0). \tag{4.26}$$

Theorem 1 gives us our result as soon as we show the existence of an $a_2$ in $[\mu_0, \mu_1]$ so that both $m_0(a_2) \geqq \rho$ and $m_1(a_2) \geqq \rho$. To this end we consider

$$F = \{a : m_1(a) \geqq \rho, \mu_0 \leqq a \leqq \mu_1\}. \tag{4.27}$$

The set $F$ is not empty because $m_1(\mu_1) = 1 \geqq \rho$. Let $a_2 = g.\ l.\ b.\ F$. By continuity on the right $m_1(a_2) \geqq \rho$. Also $m_1(a) < \rho$ for $a < a_2$. Hence $m_0(a) \geqq \rho$ if $\mu_0 \leqq a < a_2$. Since $m_0(a)$ is continuous on the left for $a > \mu_0$, $m_0(a_2) \geqq \rho$ if $a_2 > \mu_0$. Furthermore, if $a_2 = \mu_0$, $m_0(a_2) = 1 \geqq \rho$.

NOTATION 4. *Let $\rho_1$ and $\rho_2$ represent the indices of two tests $T_1$ and $T_2$, respectively. We define the asymptotic relative efficiency of $T_1$ to $T_2$ by*

$$e = \log \rho_1 / \log \rho_2, \tag{4.28}$$

*where e is undefined if $\rho_1 = \rho_2 = 1$. For test $T_i$, $n_i$ is the sample size and*

$$\gamma_i = \inf (\beta + \lambda\alpha) \tag{4.29}$$

*is a function of $n_i$ and $\lambda$.*

The appropriateness of the use of $e$ as a measure of efficiency derives from the following theorem, which is an immediate consequence of Theorem 2.

THEOREM 3. *If* $\lim_{n_1, n_2 \to \infty} \frac{n_2}{n_1} < e (> e)$, *then* $\lim_{n_1, n_2 \to \infty} \frac{\gamma_1}{\gamma_2} = \infty (= 0)$.

Note that $e$ does not depend on $\lambda$.

**5. Some examples.** In this section we shall determine the behaviour of $m(a)$ and $\rho$ for a few simple examples.

EXAMPLE 1. Let $X$ be normally distributed with mean $\mu_i$ and variance $\sigma_i^2$ under hypothesis $H_i$, $i = 0, 1 (\mu_0 < \mu_1)$. Then

$$e^{-at} M_i(t) = e^{(\mu_i - a)t + \frac{1}{2}\sigma_i^2 t^2} \tag{5.1}$$

$$m_i(a) = e^{\frac{1}{2}(\mu_i - a)^2/\sigma_i^2}, \tag{5.2}$$

$$\rho = \rho\left(\frac{\sigma_1 \mu_0 + \sigma_0 \mu_1}{\sigma_1 + \sigma_0}\right) = e^{-\frac{1}{2}((\mu_1 - \mu_0)/(\sigma_1 + \sigma_0))^2}. \tag{5.3}$$

Of course this index applies to a test which is not the likelihood ratio test unless $\sigma_1 = \sigma_0$. The computational problem in obtaining the index of the likelihood ratio test is considerable. However, the results in Section 6 may be easily applied to the likelihood ratio test if $\mu_1 - \mu_0$ and $\sigma_1 - \sigma_0$ are small.

EXAMPLE 2. Let $X/\sigma_i^2$ have a chi-square distribution with $r$ degrees of freedom under hypothesis $H_i$, $i = 0, 1(\sigma_0^2 < \sigma_1^2)$. Then

$$e^{-at} M_i(t) = e^{-at}(1 - 2\sigma_i^2 t)^{-\frac{1}{2}r} \tag{5.4}$$

$$\log m_i(a) = -\tfrac{1}{2}\left[\frac{a}{\sigma_i^2} + r \log \frac{r\sigma_i^2}{a} - r\right], \tag{5.5}$$

$$\log \rho = -\tfrac{1}{2}r[\delta - 1 - \log \delta], \tag{5.6}$$

where

$$\delta = (\log \tau)/(\tau - 1), \tag{5.7}$$

$$\tau = \sigma_0^2/\sigma_1^2 . \tag{5.8}$$

Note that as $\tau$ approaches 1, $\log \rho \approx -r(\tau - 1)^2/16$.

EXAMPLE 3. Let $X$ have the binomial distribution so that

$$P(X = j \mid H_i) = \binom{r}{j} p_i^j q_i^{r-j}, \quad q_i = 1 - p_i, \quad i = 0, 1, \quad j = 0, 1, \cdots, r(p_0 < p_1). \tag{5.9}$$

Then

$$e^{-at}M_i(t) = e^{-at}(p_i e^t + q_i)^r, \tag{5.10}$$

$$\log m_i(a) = (r - a) \log [rq_0/(r - a)] + a \log [rp_0/a], \tag{5.11}$$

$$\log \rho = r\{(1 - c) \log [q_0/(1 - c)] + c \log [p_0/c]\}, \tag{5.12}$$

where

$$c = \frac{\log (q_0/q_1)}{\log (q_0/q_1) + \log (p_1/p_0)}, \quad p_0 < c < p_1. \tag{5.13}$$

Note that as $p_1$ approaches $p_0$, $\log \rho \approx -r(p_1 - p_0)^2/8p_0q_0$.

**6. Normality approximation.** In this section we shall develop some results concerning the conjecture made in the introduction, that if the hypotheses $H_0$ and $H_1$ are very close to one another one may approximate $\rho$ by assuming that $X$ is normally distributed. To this end we shall first investigate more closely the behaviour of $m(a)$ and $t(a)$.

NOTATION 5. *Let* $N(t) = E(Xe^{tX})$ *and* $P(t) = E(X^2e^{tX})$. *Let*

$$t_0 = \text{glb}\{t: M(t) < \infty\}$$

*and if* $t_0 < 0$ *let*

$$a_0 = \inf_{t_0 < t < 0} N(t)/M(t). \tag{6.1}$$

Note that if $E(X) > -\infty$, $a_e < E(X)$ except in the case where $P(X = a_e) = 1$. Furthermore, if $a_e > -\infty$, then $t_0 = -\infty$.

LEMMA 7. *If* $M(t) < \infty$ *for some* $t < 0$, *and* $a_e < E(X)$, *then for* $a_0 < a < E(X)$

$$a = N(t(a))/M(t(a)), \tag{6.2}$$

$$\frac{d[\log m(a)]}{da} = -t(a), \tag{6.3}$$

$$\frac{dt(a)}{da} = \frac{M(t)^2}{M(t)P(t) - N(t)^2}\bigg|_{t=t(a)} > 0. \tag{6.4}$$

*If in addition* $\mu = E(X)$ *and* $\sigma^2 = E[(X - \mu)^2]$ *are finite, then* (6.2), (6.3), and (6.4) *hold for* $a_0 < a \leqq E(X)$, *giving*

$$\frac{d}{da}[\log m(a)]\bigg|_{a=\mu} = 0 \tag{6.5}$$

*and*

$$\frac{dt(a)}{da}\bigg|_{a=\mu} = 1/\sigma^2. \tag{6.6}$$

PROOF. Suppose that $t_0 < t < 0$. Using Lemma 3, there is a unique $a$ so that $t = t(a)$ and this value of $a$ is obtained by

$$\frac{d}{dt}[e^{-at}M(t)] = \int_{-\infty}^{\infty} (x - a)e^{t(x-a)}\, dF(x) = 0 \tag{6.7}$$

and is given by

$$a = N(t)/M(t). \tag{6.8}$$

Considering $a$ as a function of $t$ we may differentiate

$$\frac{da}{dt} = \frac{M(t)P(t) - N(t)^2}{M(t)^2}. \tag{6.9}$$

Applying Schwarz' Inequality, the numerator is at least zero. It can vanish only if $Xe^{tX/2}$ and $e^{tX/2}$ are proportional with probability one. This can occur only if $P[X = a_e] = 1$. This case is excluded by the hypothesis $a_e < E(X)$. Hence $(da/dt) > 0$. Furthermore, as $t \to 0$, $a \to E(X)$. Hence as $t$ varies over $(t_0, 0)$ $a$ ranges continuously (and monotonically) over $(a_0, E(X))$. Equations (6.2) and (6.4) are immediately valid. Equation (6.3) is obtained by differentiating with respect to $a$, $m(a) = E(e^{t(a)(X-a)})$. Equations (6.5) and (6.6) follow from the Lèbesque Convergence Theorem [3] and the fact that if $f(x)$ is continuous at $x = a$, and $f'(x) \to b$ as $x \to a$, then $f'(a) = b$.

If $\nu_0$ and $\nu_1$ are any two probability measures defined on the same Borel Field, we may introduce the measure $\nu = (\nu_0 + \nu_1)/2$. A consequence of the Radon-Nikodym Theorem [3] is the existence of two densities $f_0$ and $f_1$ (unique except possibly on a set of $\nu$ measure zero) so that

$$\nu_i(E) = \int_E f_i(x)\, d\nu(x) \qquad i = 0, 1. \tag{6.10}$$

Hence, except on a set of $\nu$ measure zero, at least one of $f_0(x)$ and $f_1(x)$ are non-zero, and the log of the likelihood may be defined by $\log f_1(x) - \log f_0(x)$.

NOTATION 6. *The outcome of an experiment is denoted by $Y$ and has a probability distribution given by equation (6.10) under hypothesis $H_i$. When an integration sign is unaccompanied by a region of integration it is to be understood that the region is the set of all possible values of $Y$. We shall deal with a chance variable $X$ which is a function of $Y$. In particular the log of the likelihood ratio is defined by* $\log f_1(Y) - \log f_0(Y)$.

$$M_i(t) = E(e^{tX} \mid H_i), \tag{6.11}$$

$$N_i(t) = E(Xe^{tX} \mid H_i), \tag{6.12}$$

$$P_i(t) = E(X^2e^{tX} \mid H_i). \tag{6.13}$$

*We use $m_i(a)$ and $t_i(a)$ to represent the functions $m(a)$ and $t(a)$ under hypothesis $H_i$.*

LEMMA 8. *If $X$ is the log of the likelihood ratio, $X \not\equiv 0$, and $X$ is finite with probability one, then*

$$M_1(t) = M_0(t+1), \qquad N_1(t) = N_0(t+1). \tag{6.14}$$

*As a varies from $\mu_0$ to $\mu_1$, $t_0(a)$ varies continuously from 0 to 1 and*

$$t_0(a) = t_1(a) + 1, \tag{6.15}$$

$$\rho = \inf_{0<t<1} M_0(t). \tag{6.16}$$

PROOF. We note that

$$M_1(t) = \int \left[\frac{f_1(x)}{f_0(x)}\right]^t \frac{f_1(x)}{f_0(x)} f_0(x)\, d\nu(x) = M_0(t+1) \tag{6.17}$$

and

$$N_1(t) = \int \log\left[\frac{f_1(x)}{f_0(x)}\right]\left[\frac{f_1(x)}{f_0(x)}\right]^t \frac{f_1(x)}{f_0(x)} f_0(x)\, d\nu(x) = N_0(t+1). \tag{6.18}$$

It is evident that

$$M_0(0) = M_0(1) = M_1(0) = M_1(-1) = 1. \tag{6.19}$$

It follows that $N_0(t)$ is finite for $0 < t < 1$, that $\mu_1 > 0$, $\mu_0 < 0$, and

$$\lim_{t\to 0-} N_1(t) = N_1(0) = \mu_1, \tag{6.20}$$

$$\lim_{t\to -1+} N_1(t) = N_1(-1) = \mu_0. \tag{6.21}$$

Applying Lemma 7, we find that as $a$ varies from $\mu_0$ to $\mu_1$, $t_1(a)$ varies continuously and (strictly) monotonically from $-1$ to 0. Similarly, $t_0(a)$ varies from 0 to 1. Applying equations (6.2), (6.17), and (6.18)

$$\frac{N_0[t_1(a)+1]}{M_0[t_1(a)+1]} = \frac{N_0[t_0(a)]}{M_0[t_0(a)]} = a \text{ for } \mu_0 < a < \mu_1. \tag{6.22}$$

Equation (6.15) follows.

Since $e^{-at_0(a)}M_0(t_0(a))$ and $e^{-at_1(a)}M_1(t_1(a))$ are both equal and continuous at $a = 0$, the monotonicity properties of Lemma 5 show that

$$\rho = M_0(t_0(0)) = m_0(0) = m_1(0) = \inf_{0<t<1} M_0(t), \tag{6.23}$$

$$\rho = \inf_{0<t<1} \int [f_1(x)]^t [f_0(x)]^{1-t}\, d\nu(x). \tag{6.24}$$

We are interested in likelihood ratio tests for which $\mu_0^2 + \sigma_0^2$ is very small. The following theorem applies to certain classes of tests. In this theorem we are interested in classes of tests where the log of the likelihood ratio has finite means. Hence the restriction of Lemma 8 that $X$ is finite with probability one is automatically satisfied. However the case where $X$ may assume the values $+\infty$ or $-\infty$ with positive probability is of some interest. For this case the above sort of reasoning applies except that all integrals must be taken over the set,

$$G = \{x: -\infty < \log f_1(x) - \log f_0(x) < \infty\}.$$

After the necessary modifications are made, it is seen that (6.24) is valid in general.

THEOREM 4. *If, for a class C of likelihood ratio tests,*

$$\begin{aligned} M_0(t) &= 1 + \mu_0 t + (\mu_0^2 + \sigma_0^2)t^2/2 + o(\mu_0^2 + \sigma_0^2), \qquad 0 < t < 1, \\ M_1(t) &= 1 + \mu_1 t + (\mu_1^2 + \sigma_1^2)t^2/2 + o(\mu_0^2 + \sigma_0^2), \qquad -1 < t < 0, \end{aligned} \tag{6.25}$$

*then*

$$\begin{aligned} \mu_1 &= \sigma_0^2/2 + o(\sigma_0^2), \\ \mu_0 &= -\sigma_0^2/2 + o(\sigma_0^2), \\ \sigma_1^2 &= \sigma_0^2 + o(\sigma_0^2), \end{aligned} \tag{6.26}$$

*and*

$$\begin{aligned} \rho &= e^{-\sigma_0^2/8} + o(\sigma_0^2), \\ \rho &= e^{\mu_0/4} + o(\mu_0), \\ \rho &= e^{-\frac{1}{2}[(\mu_1-\mu_0)/(\sigma_1+\sigma_0)]^2} + o\left[\frac{\mu_1 - \mu_0}{\sigma_1 + \sigma_0}\right]^2, \end{aligned} \tag{6.27}$$

PROOF. *Part 1.*

$$M_1(t-1) = M_0(t), \qquad 0 < t < 1, \tag{6.28}$$

$$\left[\mu_1 - \frac{\mu_1^2 + \sigma_1^2}{2}\right] + t[(\mu_0 - \mu_1) + (\mu_1^2 + \sigma_1^2)] + \frac{t^2}{2}[\mu_0^2 + \sigma_0^2 - \mu_1^2 - \sigma_1^2] = o(\mu_0^2 + \sigma_0^2). \tag{6.29}$$

Hence

$$
\begin{aligned}
\mu_1^2 + \sigma_1^2 &= \mu_0^2 + \sigma_0^2 + o(\mu_0^2 + \sigma_0^2),\\
\mu_1 &= \sigma_1^2/2 + o(\mu_0^2 + \sigma_0^2),\\
\mu_0 &= \sigma_1^2/2 + o(\mu_0^2 + \sigma_0^2),\\
\sigma_1^2 &= \sigma_0^2 + o(\mu_0^2 + \sigma_0^2).
\end{aligned} \tag{6.30}
$$

Equations 6.26 follow immediately.

*Part 2.* By Lemma 8, $\rho = \inf_{0<t<1} M_0(t)$. Minimizing the quadratic approximation we obtain

$$
\rho = 1 - \frac{\mu_0^2}{2(\mu_0^2 + \sigma_0^2)} + o(\mu_0^2 + \sigma_0^2). \tag{6.31}
$$

Applying the results of Part 1, Part 2 follows immediately.

We may also be interested in tests of the form

$$
S_n = \sum_{j=1}^{n} X_j \leq k,
$$

where $X$ is a less efficient statistic than the log of the likelihood ratio. Here again, given a class of tests, we may investigate the behaviour of $\rho$ as the hypotheses get "close" together. For some such classes we state the following theorem.

THEOREM 5. *If, for a class $C^*$ of tests,*

$$
\sigma_i^2 \frac{dt_i(a)}{da} = 1 + o(1) \tag{6.31}
$$

*as $\mu_1 - \mu_0 \to 0$ for $\mu_0 < a < \mu_1$, $i = 0, 1$, then*

$$
\rho = e^{-\frac{1}{2}[(\mu_1-\mu_0)/(\sigma_1+\sigma_0)]^2} + o\left(\frac{\mu_1 - \mu_0}{\sigma_1 + \sigma_0}\right)^2. \tag{6.32}
$$

PROOF.

$$
\begin{aligned}
\log m_0(a) &= \frac{-(a - \mu_0)^2}{2}\left[\frac{1}{\sigma_0^2} + o\left(\frac{1}{\sigma_0^2}\right)\right],\\
\log m_1(a) &= \frac{-(a - \mu_1)^2}{2}\left[\frac{1}{\sigma_1^2} + o\left(\frac{1}{\sigma_1^2}\right)\right].
\end{aligned} \tag{6.33}
$$

Equating the main terms, one obtains

$$
\log \rho = -\frac{1}{2}\left(\frac{\mu_1 - \mu_0}{\sigma_1 + \sigma_0}\right)^2 + o\left(\frac{\mu_1 - \mu_0}{\sigma_1 - \sigma_0}\right)^2. \tag{6.34}
$$

It may be seen that the corresponding value of $a$ satisfies $a \approx (\sigma_1\mu_0 + \sigma_0\mu_1)/(\sigma_1 + \sigma_0)$. Finally we note that equation (6.4) is useful in checking the applicability of Theorem 5.

**7. Measures of information and divergence.** In Section 4 the measure of efficiency $e$, was defined so that $n$ observations for one test is equivalent to $en$ observations for the second test (equivalent from the point of view of the criterion we used). It is evident that it would have been appropriate to use the following equation

$$I(X) = -\log \rho \tag{7.1}$$

to indicate that $-\log \rho$ may be used as a measure of the information per observation for a test based on sums of observations on $X$. (Here $X$ denotes the two specified chance variables associated with $H_0$ and $H_1$, respectively.) In addition we may have written

$$D(Y) = -\log \left[ \inf_{0<t<1} \int [f_1(x)]^t [f_0(x)]^{1-t} \, d\nu(x) \right] \tag{7.2}$$

to indicate that $-\log \rho$ for the likelihood ratio test may be used as a measure of the divergence between the two distributions associated with $Y$. Let $(Y_1, Y_2)$ represent an observation consisting of independent observations on $Y_1$ and $Y_2$ respectively. Then it is easy to see from equation (6.24) that

$$D(Y_1, Y_2) \leqq D(Y_1) + D(Y_2) \tag{7.3}$$

and

$$D(Y, Y) = 2D(Y) \tag{7.4}$$

A measure of divergence used by Kullback and Leibler [5, 6], yields equality in the relation (7.3). The measure (7.2) and that used by Kullback and Leibler are basically two different functionals on the curve relating the type 1 and type 2 errors for likelihood ratio tests.

## REFERENCES

[1] H. Cramér, "Sus un nouveau théorème-limite de la théorie des probabilités," *Actualités Scientifiques et Industrielles*, No. 736, Paris 1938.

[2] C. G. Esseen, "Fourier analysis of distribution functions, *Acta Mathematica*, Vol. 77, (1945) pp. 1–125.

[3] S. Saks, *Theory of the Integral*, 2nd ed., Warsaw, 1937.

[4] A. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Julius Springer, Berlin, 1933.

[5] S. Kullback, and R. A. Leibler, "On information and sufficiency", *Annals of Math. Stat.*, Vol. 22, (1951) pp. 79–86,

[6] H. Jeffreys, *Theory of Probability*, 2nd ed., Oxford University Press, 1948.