# ASYMPTOTIC SOLUTIONS OF COMPOUND DECISION PROBLEMS

by

JAMES FRANCIS HANNAN

Special report to THE UNITED STATES AIR FORCE
under Contract AF 18(600)-83 monitored by the
Office of Scientific Research.

Institute of Statistics
Mimeograph Series No. 68

April 23, 1953

ACKNOWLEDGEMENT

I wish to express my sincere appreciation to Professor Herbert Robbins for his guidance, interest and valuable suggestions in this research. Thanks are also due to Professor Wassily Hoeffding for his helpful suggestions and wise criticisms.

I wish also to thank the Institute of Statistics, the Office of Naval Research and the U. S. Air Force for financial assistance during a substantial part of this investigation.

# TABLE OF CONTENTS

Table of Contents(Continued)

# INTRODUCTION

Robbins [12][1] introduced the idea of a compound decision problem and showed that in the case where the component decisions were between $N(-1,1)$ and $N(1,1)$ there exists a decision function $\underline{t}^*$ whose risk approaches the risk of the best "simple" decision function based on a knowledge of the proportion of component problems in which the true distribution is $N(1,1)$. In Chapter I we generalize this result to two (arbitrary) fully specified populations and strengthen the conclusion by the replacement of "simple" by "invariant".

In Chapter II we prove a theorem which made this strengthening possible, obtaining its proof by a density Central Limit Theorem (DCLT) for $\ell n$-concave densities, which we apply to the generalized binomial (GB).

In Chapter III we find the exact distribution of the statistic $D_n^+$ and use this to obtain an elementary derivation of the asymptotic distribution (originally obtained by Kolmogoroff [11]). These results were announced in [8]. Birnbaum and Tingey [2] have published a similar derivation of the exact distribution and have tabulated certain percentage points of the distribution.

---

1. Numbers in square brackets refer to bibliography.

CHAPTER I.

ASYMPTOTIC SOLUTIONS OF THE COMPOUND DECISION PROBLEM

FOR TWO COMPLETELY SPECIFIED POPULATIONS

## 1. Summary of Prior Results.

In this section we will give an outline of the results of [12] in the notation of the present chapter and an indication of the degree of our indebtedness to an earlier unpublished version of [9] due to Robbins.

The problem and results of [12] may be outlined as follows: let $\underline{x} = (x_1, \ldots, x_n)$ be a random vector such that the components $x_i$ are independent and $x_i$ has the distribution function $\Phi(x+1-2\theta_i)$ with $\theta_i$ equal to 0 or 1 and $\Phi(v)$ the distribution function of a normal variable with mean 0 and variance 1. It is required to decide for each i whether $\theta_i$ is 0 or 1 when the loss of the decision $\underline{d} = (d_1, \ldots, d_n)$, if $\underline{\theta} = (\theta_1, \ldots, \theta_n)$ is the true parameter vector, is taken to be

$$W(\underline{d},\underline{\theta}) = \frac{1}{n} \sum_1^n [\theta_i(1-d_i) + (1-\theta_i)\, d_i] .$$

The decision functions of an essentially complete class T are expressible in the form $\underline{t}(\underline{x}) = (t_1(\underline{x}), \ldots, t_n(\underline{x}))$ where $t_i(\underline{x}) = \Pr[d_i = 1 \mid \underline{x}]$ and the risk of a $\underline{t}$ in T is

$$R(\underline{t},\underline{\theta}) = E[W(\underline{t}(\underline{x}), \underline{\theta})] .$$

If S denotes the class of simple (cf. Section 3) decision functions and $\underline{t} = t$ is in S, $R(\underline{t},\underline{\theta})$ is representable by

$$R(t,\underline{\theta}) = p(\underline{\theta}) \int (1-t(x))d\Phi(x-1) + (1-p(\underline{\theta})) \int t(x)d\Phi(x+1)$$

with $p(\underline{\theta}) = n^{-1}\Sigma\theta_j$. Minimization of a natural extension of this function introduces a complete (relative to S) class of simple decision functions $t_p$ defined $0 \leq p \leq 1$ by

$$t_p(x_i) = \begin{cases} 1 & \text{if} \quad x_i > \frac{1}{2} \ln \frac{1-p}{p} \\ 0 & \text{if} \quad x_i < \frac{1}{2} \ln \frac{1-p}{p} \end{cases},$$

and having the property that

$$\inf_{t \text{ in } S} R(t,\underline{\theta}) = R(t_{p(\underline{\theta})}, \underline{\theta}) \qquad \text{for all } \underline{\theta} \ .$$

The existence of estimators $p_n(\underline{x})$ such that $p_n(\underline{x}) - p(\underline{\theta})$ approaches 0 in probability was shown and for the choice, $p_n^*(\underline{x})$ = truncation to $[0,1]$ of the unbiased estimator $n^{-1}\Sigma\,[(x_i+1)/2]$, it was noted that the result

$$\lim_{n\to\infty} \sup_{\underline{\theta}} \left[R(t_{p_n^*}, \underline{\theta}) - R(t_{p(\underline{\theta})}, \underline{\theta})\right] = 0$$

was both plausible and capable of rigorous proof. Thus, the non-simple decision function $\underline{t}^* = t_{p_n^*}$ does about as well for large n as the "best" simple decision function based on a knowledge of $p(\underline{\theta})$. Finally, the Bayes solutions corresponding to symmetric weights were explicitly displayed and it was shown that the simple decision function $t_{\frac{1}{2}}$ was the unique (and, hence, admissible) minimax decision function for the compound problem. Since $R(t_{p(\underline{\theta})}, \underline{\theta}) = \phi(p(\underline{\theta}))$, a function of $p(\underline{\theta})$ such that $\phi(0) = \phi(1) = 0$ and

$$\phi(p(\underline{\theta})) \leq R(t_{\frac{1}{2}}, \underline{\theta}) \equiv \underline{\Phi}(-1)$$

with equality holding only for $\underline{\theta}$ with $p(\underline{\theta}) = \frac{1}{2}$, the attractiveness for large n (at least by comparison with the minimax decision function) of non-simple decision functions of the type $\underline{t}^*$ is apparent.

In an earlier version of [9] Robbins treated the generalization which constitutes the subject matter of the present chapter. The statement of the problem in Section 2, treatment of simple decision functions in Section 3, and of Bayes and minimax solutions in Section 7, are, with the exception of simplifications attendant on the reduction of the class of decision functions effected in 2.2 to 2.12, an inheritance from this earlier version. Robbins had there shown (cf. Theorem 2) the existence of an estimator $p_n(\underline{x})$ such that $p_n(\underline{x}) - p(\underline{\theta})$ is a uniformly consistent estimate of 0 and had proved a theorem similar to Theorem 4 under the assumption (See

Section 2 for notation and definitions)

$$\mu_0 \left[ x \,\middle|\, \frac{f(x,1)}{f(x,0)} = \lambda \right] = 0 \qquad \text{for every } 0 < \lambda < \infty .$$

The material of Section 2 to Section 7, with some additions, will form the body of [9] and is presented here in its totality because it was thought preferable to do this rather than refer to an unpublished (and, consequently, not generally available) paper.

## 2. Statement and Reduction of the Problem.

The problem of testing a simple statistical hypothesis against a simple alternative can be formulated as follows. Let $x$ be a random variable (of arbitrary dimensionality) which is known to have one of the two distinct distribution functions $F(x,\theta)$, $\theta = 0$ or 1. On the basis of a single observation on $x$ (we consider only the non-sequential case) it is required to decide whether the true value of the unknown parameter $\theta$ is 0 or 1.

Statistical decision problems of the same formal structure often occur, or can be considered, in large groups. We shall, therefore, take as a single entity the following compound decision problem. Let $n$ be a fixed positive integer and let $x_1, \ldots, x_n$ be independent random variables each of which has the distribution function $F(x,\theta)$ with respective parameter values $\theta_1, \ldots, \theta_n$, $\theta_i = 0$ or 1. Let $\underline{x} = (x_1, \ldots, x_n)$ denote the vector of observations and $\underline{\theta} = (\theta_1, \ldots, \theta_n)$ the unknown vector of parameters; $\underline{\theta}$ is known to belong to the set $\underline{(\,\bar{}\,)}$ consisting of all $2^n$ possible vectors of n components, each 0 or 1. On the basis of $\underline{x}$ it is required to decide the true value of $\underline{\theta}$, which amounts to deciding for every $i = 1, \ldots, n$ whether $\theta_i = 0$ or 1.

Any vector of n functions $\underline{t} = (t_1(\underline{x}), \ldots, t_n(\underline{x})$ is a (randomized) decision function for the compound decision problem if for $i = 1, \ldots,$ n, $0 \leq t_i(\underline{x}) \leq 1$, and if the conditional probabilities, given $\underline{x}$, of deciding that $\theta_i = 0$ or 1 are respectively $1 - t_i(\underline{x})$, $t_i(\underline{x})$. If for some function $t(x)$, $t_i(\underline{x}) = t(x_i)$, $i = 1, \ldots, n$, then $\underline{t}$ will be called simple and will be denoted by t.

We assume given by the practical background of the problem two positive constants

(2.1)

a = loss incurred in deciding "$\theta_i = 0$" when true value of $\theta_i = 1$,

b = loss incurred in deciding "$\theta_i = 1$" when true value of $\theta_i = 0$.

For any Borel set B we define

$$\mu_\theta(B) = \int_B dF(x,\theta) \qquad \theta = 0, 1, \tag{2.2}$$

$$\mu(B) = \mu_0(B) + \mu_1(B) \ . \tag{2.3}$$

Since $\mu_0$ and $\mu_1$ are absolutely continuous with respect to the finite measure $\mu$, generalized probability density functions $f(x,0)$ and $f(x,1)$ exist such that for any Borel set B

$$(2.4) \qquad \mu_\theta(B) = \int_B f(x,\theta)d\mu \qquad \theta = 0,\ 1.$$

We note that the relation

$$(2.5) \qquad f(x,0) + f(x,1) = 1 \qquad \text{a.e. } (\mu)$$

is obtainable from the identity in Borel B,

$$\int_B 1\ d\mu = \mu(B) = \mu_0(B) + \mu_1(B) = \int_B f(x,0) + f(x,1)\ d\mu\ .$$

The joint generalized probability density function of $\underline{x}$ with respect to the product measure $\mu^n$, when the parameter vector is $\underline{\theta} = (\theta_1, \ldots \theta_n)$, is $f(\underline{x},\underline{\theta}) = \prod_1^n f(x_i,\theta_i)$, and the expected loss on the i-th decision in using a decision function $\underline{t} = (t_i(\underline{x}), \ldots, t_n(\underline{x}))$ is for $i = 1, \ldots, n$,

$$(2.6) \quad R_i(\underline{t},\underline{\theta}) = \int [a\theta_i(1-t_i(\underline{x})) + b(1-\theta_i)t_i(\underline{x})]f(\underline{x},\underline{\theta})\ d\mu^n\ .$$

The average expected loss on all n decisions, or as we shall say, the risk, is therefore

$$R(\underline{t},\underline{\theta}) = \frac{1}{n}\sum_1^n R_i(\underline{t},\underline{\theta})$$

(2.7)

$$= \frac{a}{n}\sum_1^n \theta_i \int (1-t_i(\underline{x}))f(\underline{x},\underline{\theta})d\mu^n + \frac{b}{n}\sum_1^n(1-\theta_i)\int t_i(\underline{x})f(\underline{x},\underline{\theta})d\mu^n.$$

By the remark (2.5) $f(\underline{x},\underline{\theta})$ is expressible

$$(2.8)\quad f(\underline{x},\underline{\theta}) = \prod_1^n f(x_i,\theta_i) = \prod_1^n[\theta_i f(x_i,1) + (1-\theta_i)(1-f(x_i,1))]$$

and the Halmos-Savage [5] form of the theorem linking sufficient statistics and density factorization shows that $(f(x_1,1), \ldots, f(x_n,1))$ is a sufficient statistic for $\underline{\theta}$ in $\underline{\Theta}$. Let

$$(2.9)\qquad \underline{z} = (z_1, \ldots, z_n) \qquad \text{where } z_i = f(x_i,1)\ ,$$

$$(2.10)\qquad \nu_\theta(I) = \mu_\theta(x \mid f(x,1) \text{ in } I) \quad \text{for all Borel sets } I,\ \theta = 0,1,$$

$$(2.11)\qquad \nu(I) = \nu_0(I) + \nu_1(I)\ .$$

It is easily verified (See Sec. 32 [4]) that the sufficient statistic $\underline{z}$ (and, consequently, the measures $\nu_0$, $\nu_1$, and $\nu$) is independent of the choice (2.3) of an underlying measure $\mu$ relative to which $\mu_0$ and $\mu_1$ are absolutely continuous. We note that $\underline{z}$ has, with respect to the product measure $\nu^n$, the generalized probability density

$$(2.12) \qquad d(\underline{z},\underline{\theta}) = \prod_{1}^{n} [\theta_i z_i + (1-\theta_i)(1-z_i)] .$$

Returning to (2.7) we see that $E[\underline{t} \mid \underline{z}] = (E[t_1(\underline{x}) \mid \underline{z}], \ldots, E[t_n(\underline{x}) \mid \underline{z}])$ is a decision rule having the same risk as $\underline{t}$. We denote this rule by $\underline{t}(\underline{z}) = (t_1(\underline{z}), \ldots t_n(\underline{z}))$ and can express (2.7) as

$$(2.13) \quad R(\underline{t},\underline{\theta}) = \frac{a}{n} \sum_{1}^{n} \theta_i \int (1-t_i(\underline{z}))d(\underline{z},\underline{\theta})d\nu^n + \frac{b}{n} \sum_{1}^{n} (1-\theta_i) \int t_i(\underline{z})d(\underline{z},\underline{\theta})d\nu^n.$$

3. Simple Decision Functions.

If $\underline{t} = t$ is a simple decision function, then (2.13) simplifies to

$$(3.1) \quad R(t,\underline{\theta}) = \frac{a}{n} \sum_{1}^{n} \theta_i \int (1-t(z))z d\nu + \frac{b}{n} \sum_{1}^{n} (1-\theta_i) \int t(z)(1-z)d\nu .$$

Setting

$$(3.2) \quad p(\underline{\theta}) = \frac{1}{n} \sum_{1}^{n} \theta_i = \text{proportion of 1's among the n components of } \underline{\theta},$$

we can write (3.1) in the form

$$(3.3) \qquad R(t,\underline{\theta}) = \int ap(\underline{\theta})(1-t(z))z + b(1-p(\underline{\theta}))t(z)(1-z)\, d\nu .$$

The value $p(\underline{\theta})$ is necessarily rational, of the form $k/n$, $k = 0, \ldots, n$. We now define for any real number $p$, $0 \leq p \leq 1$, and any $\nu$-measurable function $t(z)$ such that $0 \leq t(z) \leq 1$, the expression

$$B(t,p) = \int ap(1-t(z))z + b(1-p)t(z)(1-z)\,d\nu \tag{3.4}$$

so that (3.3) becomes

$$R(t,\underline{\theta}) = B(t,p(\underline{\theta})) \; . \tag{3.5}$$

From (3.4) it is clear that for any fixed $p$, $B(t,p)$ is a minimum if and only if for almost every ($\nu$) $z$, $t(z)$ is of the form

$$t_p(z) = \begin{cases} 1 & \text{if } apz > b(1-p)(1-z) \\ 0 & \text{if } apz < b(1-p)(1-z) \\ \text{arbitrary} & \text{if } apz = b(1-p)(1-z) \end{cases} \; . \tag{3.6}$$

The minimum value of $B(t,p)$ is

$$\begin{aligned} \phi(p) = B(t_p,p) &= \int \min\,[apz, b(1-p)(1-z)]\,d\nu \\ &= ap\int_0^{C(p)} d\nu_1 + b(1-p)\left(1-\int_0^{C(p)} d\nu_0\right) \end{aligned} \tag{3.7}$$

where

$$C(p) = \frac{b(1-p)}{ap+b(1-p)} \qquad 0 \leq p \leq 1 \; . \tag{3.8}$$

It follows from (3.5) that, for any simple decision function t and any $\underline{\theta}$ in $\underline{\Omega}$,

$$(3.9) \qquad R(t,\underline{\theta}) \geq \phi(p(\underline{\theta})) ,$$

equality holding if and only if t(z) is of the form (3.6) with $p = p(\underline{\theta})$.

We shall now establish some properties of the function $\phi(p)$ defined for $0 \leq p \leq 1$ by (3.7). The identity, $\min(u,v) = (u+v)/2 - |u-v|/2$, clearly holds for all real numbers u,v, and using this in (3.7) we obtain

$$(3.10) \qquad \phi(p) = \frac{ap+b(1-p)}{2} - \frac{1}{2}\int |apz - b(1-p)(1-z)|\, d\nu .$$

For any two numbers $0 \leq p, p' \leq 1$, we therefore have

$$\phi(p)-\phi(p') = \frac{(a-b)(p-p')}{2} - \frac{1}{2}\int |apz-b(1-p)(1-z)| - |ap'z-b(1-p')(1-z)|\, d\nu$$

whence setting

$$(3.11) \qquad c = \max(a,b)$$

we obtain

$$(3.12) \qquad \begin{aligned} |\phi(p) - \phi(p')| &\leq \frac{|a-b|}{2}|p-p'| + \frac{1}{2}\int |a(p-p')z + b(p-p')(1-z)|\, d\nu \\ &\leq \frac{|a-b|}{2}|p-p'| + \frac{a+b}{2}|p-p'| = c\,|p-p'| , \end{aligned}$$

which proves that $\phi$ is continuous.

Using (3.7) we have for $0 \leq p < p' \leq 1$, $0 < s < 1$

$$s\phi(p)+(1-s)\phi(p') = \int \min \left[ aspz, bs(1-p)(1-z) \right]$$

$$+ \min\left[ a(1-s)p'z, b(1-s)(1-p')(1-z) \right] d\nu$$

$$\leq \int \min \left\{ a \left[ sp + (1-s)p' \right] z \, , \right. \tag{3.13}$$

$$\left. b(1 - \left[ sp + (1-s)p' \right])(1-z) \right\} d\nu$$

$$= \phi(sp + (1-s)p')$$

which proves $\phi(p)$ is a concave function of p. Further, it is easy to see that the inequality is strict unless

$$\nu \left[ z \mid apz < b(1-p)(1-z), \; ap'z > b(1-p')(1-z) \right] = 0 \, .$$

This condition can be expressed

$$\nu \left[ C(p') < z < C(p) \right] = 0 \, . \tag{3.14}$$

Since $\phi(0) = \phi(1) = 0$, (3.13) and (3.14) imply

$$\phi(p) > 0 \qquad 0 < p < 1, \text{ unless } \nu \left[ 0 < z < 1 \right] = 0 \, . \tag{3.15}$$

The exception in (3.15) is the trivial case where $\nu_0$ and $\nu_1$ (and hence $\mu_0$ and $\mu_1$) are disjoint measures.

For any $0 \leq p,\ p' \leq 1$ we have from (3.4)

$$(3.16)\qquad B(t_{p'},p) = \int ap(1-t_{p'}(z))z + b(1-p)t_{p'}(z)(1-z)\ d\nu\ .$$

For fixed $p'$ this is a linear function of $p$ which for $p = p'$ takes the value $\phi(p')$. Hence,

$$(3.17)\qquad \begin{aligned} \left|B(t_{p'},p) - \phi(p')\right| &= |p'-p| \cdot \left|a \int (1-t_{p'}(z))d\nu_1 - b \int t_{p'}(z)d\nu_0\right| \\ &\leq c\ |p'-p| \end{aligned}$$

From (3.17) and (3.12) it follows that

$$(3.18)\qquad |B(t_{p'},p)-\phi(p)| \leq |B(t_{p'},p)-\phi(p')| + |\phi(p')-\phi(p)| \leq 2c|p'-p|\ ,$$

and hence that

$$(3.19)\qquad 0 \leq B(t_{p'},p) - \phi(p) \leq 2c\ |p'-p|\ .$$

From (3.5) it follows that for any $0 \leq p' \leq 1$ and any $\underline{\theta}$ in $\underline{(\,)}$ ,

$$(3.20)\qquad 0 \leq R(t_{p'},\underline{\theta}) - \phi(p(\underline{\theta})) \leq 2c\ |p'-p(\underline{\theta})|\ .$$

Thus, we have proved

Theorem 1.

Suppose that in some manner an approximate value $p'$ of the true proportion $p(\underline{\theta})$ is known. Then the statistician who uses the simple decision function $\underline{t} = t_{p'}$ will achieve a risk $R(t_{p'}, \underline{\theta})$ which is within $2c\, |p'-p(\underline{\theta})|$ of the minimum attainable risk $\phi(p(\underline{\theta}))$ in the class of all simple decision functions.

## 4. "Consistent" Estimation of a Proportion.

We use (without loss of generality) the canonical form of the problem (see (2.2) to (2.12)), assuming only that the original measures, $\mu_0$ and $\mu_1$ are not identical. We are, thus, concerned with the random vector $\underline{z} = (z_1, \ldots, z_n)$ having the density $d(\underline{z}, \underline{\theta})$ defined by (2.12) and consider the problem of finding an estimator $p_n(\underline{z})$ for the proportion $p(\underline{\theta})$ of 1's in the first n coordinates of $\underline{\theta}$. Since our principal interest is in the asymptotic estimation problem, we will consider the sequence of problems defined for $n = 1, 2, \ldots$ as embedded in the probability space of infinite sequences $\underline{z} = (z_1, z_2, \ldots)$ with p-measure induced by the density $d(\underline{z}, \underline{\theta})$ defined on the first n coordinates of $\underline{z}$ and $\underline{\theta}$, respectively. We emphasize this aspect in this and the following section by referring to $\underline{\theta}$ in $\Omega_\infty$.

Avoiding a discussion of "optimum" estimation, we devote this section to the consideration of a subclass H of the class U of all unbiased estimators of $p(\underline{\theta})$. H is to be the class of all estimators of the form

$$(4.1) \qquad h_n(\underline{z}) = \frac{1}{n} \sum_1^n h(z_i) \quad ,$$

where $h(z)$ is an unbiased estimator of $\theta$.

As a measure of the risk of an estimator in H we use its variance

$$(4.2) \qquad \operatorname{Var} h_n(\underline{z}) = p(\underline{\theta})\, V_1(h) + (1-p(\underline{\theta}))\, V_0(h)$$

where $V_1(h) = \int (h(z)-1)^2 \, d\nu_1$ and $V_0(h) = \int (h(z))^2 \, d\nu_0$. We single out an interesting subclass of H by investigating the existence and representation of elements h minimizing for fixed p, with $0 < p < 1$,

$$(4.3) \qquad pV_1(h) + (1-p)\, V_0(h) \; .$$

Define for any pair of real numbers $\lambda = (\lambda_0, \lambda_1)$ an extension of (4.3) to the set of all $g(z)$ such that $\int g(z)\, d\nu < \infty$ ;

$$(4.4) \quad F_\lambda(g) = p\left[\int g^2 d\nu_1 - 1\right] + (1-p)\int g^2 d\nu_0 - 2\lambda_1\left[\int g\, d\nu_1 - 1\right] - 2\lambda_0 \int g\, d\nu_0 \; .$$

Using the density representation of the integrals of (4.4) and the restriction on the domain of g, we have

$$(4.5) \quad F_\lambda(g) = \int g^2(z)\left[pz + (1-p)(1-z)\right] - 2g(z)\left[\lambda_1 z + \lambda_0(1-z)\right] d\nu - p + 2\lambda_1 \; .$$

For fixed $\lambda_0$, $\lambda_1$ the integrand in (4.5) is a minimum for each fixed z, if and only if

$$(4.6) \qquad g(z) = g_\lambda(z|p) = \frac{\lambda_1 z+\lambda_0(1-z)}{pz+(1-p)(1-z)} \qquad \text{a.e. } (\nu) .$$

Since $|g_\lambda(z|p)| \leq \max [|\lambda_1|, |\lambda_0|] / \min [p, 1-p]$, $g_\lambda(z|p)$ is a unique ($\nu$) minimum of $F_\lambda(g)$ over its domain, and if there exists a determination of $\lambda_0$ and $\lambda_1$ such that $g_\lambda(z|p)$ is an unbiased estimator of $\theta$, we will denote it by $h(z|p)$ and have for all unbiased h

$$(4.7) \qquad \begin{aligned} pV_1(h) + (1-p)V_0(h) = F_\lambda(h) \geq \\ F_\lambda(h(\ |p)) = pV_1(h(\ |p)+(1-p)V_0(h(\ |p)) \end{aligned}$$

where equality holds only if $h(z) \equiv h(z|p)(\nu)$.

The estimator $g_\lambda(z|p)$ will be unbiased if $\lambda_0$, $\lambda_1$ satisfy

$$(4.8) \qquad \begin{aligned} \lambda_1 \int \frac{z}{pz+(1-p)(1-z)}\, d\nu_0 + \lambda_0 \int \frac{1-z}{pz+(1-p)(1-z)}\, d\nu_0 = 0 \\ \lambda_1 \int \frac{z}{pz+(1-p)(1-z)}\, d\nu_1 + \lambda_0 \int \frac{1-z}{pz+(1-p)(1-z)}\, d\nu_1 = 1 . \end{aligned}$$

The determinant of these equations is

$$(4.9) \quad \left[ \int \frac{z(1-z)}{pz+(1-p)(1-z)}\, d\nu \right]^2 - \int \frac{(1-z)^2}{pz+(1-p)(1-z)}\, d\nu \cdot \int \frac{z^2}{pz+(1-z)1-z)}\, d\nu$$

and is non-positive by the Schwarz inequality and if equal to zero, z and 1-z are linearly dependent a.e. ($\nu$). Since z and 1-z are densities, linear dependence would imply that $\nu_0$ and $\nu_1$ are identical and this possibility has been excluded by us. Thus, an explicit representation of estimators in H, minimizing (4.2) for $\underline{\theta}$ such that $p(\underline{\theta}) = p$, is obtained

$$(4.10) \qquad h_n(\underline{z}|p) = \frac{1}{n}\sum_1^n h(z_i|p) \qquad 0 < p < 1$$

where $h(z|p) = g_\lambda(z|p)$ and $\lambda$ is a solution of (4.8). This class of estimators merits our interest since each is admissible relative to H and each satisfies the following theorem.

<u>Theorem</u> 2.

Let $h_n(\underline{z})$ be any estimator in H such that $|h(z)| + 1 < M < \infty$. Define

$$(4.11) \qquad p_n(\underline{z}) = \begin{cases} 0 & \text{if} \quad h_n(\underline{z}) < 0 \\ h_n(\underline{z}) & \text{if} \quad 0 \leq h_n(\underline{z}) \leq 1 \\ 1 & \text{if} \quad 1 < h_n(\underline{z}) \ . \end{cases}$$

Then

for any $\varepsilon > 0$ there exists an $N(\varepsilon)$ such that for any $\underline{\theta}$ in $\underline{\bigcirc}_\infty$

(4.12)

$$\Pr \left[\, |p_n(\underline{z}) - p(\underline{\theta})| > \varepsilon \qquad \text{for any } n \geq N(\varepsilon) \right] \leq \varepsilon \ ,$$

$$0 \leq p_n(\underline{z}) \leq 1 \quad . \tag{4.13}$$

Proof:

It will clearly be sufficient to show that $h_n(\underline{z})$ satisfies (4.12) and for this purpose we introduce

$$\bar{h}_i(k) \qquad i = 0, 1 \qquad k = 1, 2, \ldots\ldots \tag{4.14}$$

to denote the arithmetic mean of k independent random variables each having the probability measure $\nu_i h^{-1}$. We express

$$h_n(\underline{z}) = \frac{1}{n} \sum_1^n [\theta_j h(z_j) + (1-\theta_j)\, h(z_j)]$$

$$= p(\underline{\theta}) \frac{\sum_1^n \theta_j h(z_j)}{\sum_1^n \theta_j} + (1-p(\underline{\theta})) \frac{\sum_1^n (1-\theta_j) h(z_j)}{\sum_1^n (1-\theta_j)} \tag{4.15}$$

$$= p(\underline{\theta})\, \bar{h}_1(np(\underline{\theta})) + (1-p(\underline{\theta}))\, \bar{h}_0(n(1-p(\underline{\theta})))$$

$$h_n(\underline{z}) - p(\underline{\theta}) = p(\underline{\theta})\, [\bar{h}_1(np(\underline{\theta}))-1] + (1-p(\underline{\theta}))\, \bar{h}_0(n(1-p(\underline{\theta}))) . \tag{4.16}$$

The strong law of large numbers yields the existence of functions $N_1(\eta)$, $N_0(\eta)$, defined for $\eta > 0$ and such that

$$\Pr [|\bar{h}_i(k)-i| > \eta \qquad \text{for any } k \geq N_i(\eta)] \leq \eta \qquad i = 0, 1 . \tag{4.17}$$

Fix $\varepsilon > 0$ and consider the term $p(\underline{\theta})\,[\bar{h}_1(np(\underline{\theta}))-1]$ of (4.16). Since $p(\underline{\theta}) \leq 1$, $|\bar{h}_1(np(\underline{\theta}))-1| \leq M$, we have

$$(4.18)\quad |p(\underline{\theta})\,[\bar{h}_1(np(\underline{\theta}))-1]| \leq \min [p(\underline{\theta})\,M,\ |\bar{h}_1(np(\underline{\theta}))-1|] .$$

Hence, if N is any fixed integer

$$\Pr[p(\underline{\theta})(\bar{h}_1(np(\underline{\theta}))-1) > \varepsilon/2 \text{ for any } n \geq N]$$

$$(4.19)\qquad \leq \Pr[|\bar{h}_1(np(\underline{\theta}))-1| > \varepsilon/2 \text{ and } \sum_1^n \theta_j > nM^{-1}\varepsilon/2, \text{ for any } n \geq N]$$

$$\leq \Pr[|\bar{h}_1(k)-1| > \varepsilon/2 \text{ for any } k > NM^{-1}\varepsilon/2] .$$

Thus, if $N \geq 2M\varepsilon^{-1}N_1(\varepsilon/2)$, it follows from (4.17) for $i = 1$ that the right-hand side of (4.19) $< \varepsilon/2$ uniformly for all $\underline{\theta}$ in $\Omega_\infty$ .

We can deal in a similar fashion with the term

$$(1-p(\underline{\theta}))\,\bar{h}_0(n(1-p(\underline{\theta})))$$

of (4.16) and thus obtain that

$$(4.20)\qquad N(\varepsilon) = 2M\varepsilon^{-1} \max [N_1(\varepsilon/2),\ N_0(\varepsilon/2)]$$

satisfies (4.12).

## 5. Non-Simple Decision Functions $\underline{t}^*$.

We have seen in Section 3 that if p' is a good approximation to $p(\underline{\theta})$ then the simple decision function $\underline{t} = t_{p'}$ (see (3.6)) does about as well as is possible for any simple decision function. Although a good approximation $p'$ to $p(\underline{\theta})$ is not generally available to the statistician, we have seen in Section 4 that for large n a good estimator $p_n(\underline{z})$ of $p(\underline{\theta})$ is always available. It is natural to combine these two results by using the decision function $t_{p'}$ with the constant $p'$ replaced by the random variable $p_n(\underline{z})$; this amounts to using the non-simple decision function $\underline{t}^* = (t_1^*(\underline{z}), \ldots, t_n^*(\underline{z}))$ such that for $i = 1, \ldots, n$,

$$(5.1) \qquad t_i^*(\underline{z}) = \begin{cases} 1 \text{ if } z_i > C(p_n(\underline{z})) \\ \\ 0 \text{ otherwise} \end{cases}$$

(We have chosen, arbitrarily, one way of resolving the ambiguity in the definition (3.6) when $z = C(p')$.)

It should be remarked that $\underline{t}^*$ can be used in practice only when all the values $z_1, \ldots, z_n$ are known before the individual decisions on the values of $\theta_1, \ldots, \theta_n$ have to be made.

We shall now investigate the behavior, for large n, of the risk function $R(\underline{t}^*, \underline{\theta})$. We begin by considering the loss of the decision rule $t_{p'}$ where $\underline{z}$ is observed and $\underline{\theta}$ is the true parameter point. We denote this random variable by $W(\underline{z}, p', \underline{\theta})$ and have

$$W(\underline{z},p',\underline{\theta}) = \frac{1}{n}\sum_1^n \left[a\theta_i(1-t_{p'}(z_i))+b(1-\theta_i)\ t_{p'}(z_i)\right]$$

(5.2)

$$= ap(\underline{\theta})\frac{\sum_1^n \theta_i(1-t_{p'}(z_i))}{\sum_1^n \theta_i} + b(1-p(\underline{\theta}))\frac{\sum_1^n (1-\theta_i)t_{p'}(z_i)}{\sum_1^n (1-\theta_i)} .$$

Let $S_i(v \mid k)$ be the sample distribution function of k independent random variables, each distributed with probability measure $\nu_i$, where $i = 0$ or $1$, $k = 1, 2, \ldots\ldots$ . For future use we define

$$D_1^+(k) = \sup_{0 \le v \le 1} \left[ S_1(v \mid k) - \int_0^v d\ \nu_1 \right]$$

(5.3)

$$D_0^-(k) = \sup_{0 \le v \le 1} \left[ \int_0^v d\ \nu_0 - S_0(v \mid k) \right] .$$

From (5.1) we obtain that (5.2) has the representation

(5.4) $W(\underline{z}, p', \underline{\theta}) =$

$$ap(\underline{\theta})S_1(C(p') \mid np(\underline{\theta}))+b(1-p(\underline{\theta})) \left[ 1-S_0(C(p') \mid n(1-p(\underline{\theta}))) \right] .$$

We rewrite (5.4) in the form

$$W(\underline{z},p',\underline{\theta}) = ap(\underline{\theta}) \left[ S_1(C(p') \mid np(\underline{\theta})) - \int_0^{C(p')} d\ \nu_1 \right]$$

$$+ b(1-p(\underline{\theta}))\left[ \int_0^{C(p')} d\ \nu_0 - S_0(C(p') \mid n(1-p(\underline{\theta}))) \right]$$

(5.5)

$$+ ap(\underline{\theta}) \int_0^{C(p(\underline{\theta}))} d\ \nu_1 + b(1-p(\underline{\theta}))\left[ 1 - \int_0^{C(p(\underline{\theta}))} d\ \nu_0 \right]$$

$$+ ap(\underline{\theta}) \int_{C(p(\underline{\theta}))}^{C(p')} d\, \nu_1 - b(1-p(\underline{\theta})) \int_{C(p(\underline{\theta}))}^{C(p')} d\, \nu_0$$

and from (3.7) and (5.3) have

$$W(\underline{z},p',\underline{\theta}) \leq ap(\underline{\theta})\; D_1^+(np(\underline{\theta})) + b(1-p(\underline{\theta}))\; D_0^-(n(1-p(\underline{\theta})))$$

(5.6)

$$+ \phi(p(\underline{\theta})) + \int_{C(p(\underline{\theta}))}^{C(p')} [\, ap(\underline{\theta})z - b(1-p(\underline{\theta}))(1-z)\,]\, d\nu \quad .$$

The integrand in (5.6) is a linear function which vanishes at $C(p(\underline{\theta}))$; hence, is absolutely bounded by

(5.7) $$\left| ap(\underline{\theta})C(p') - b(1-p(\underline{\theta}))(1-C(p')) \right| = \frac{ab\; |p'-p(\theta)|}{ap'+b(1-p')} \leq c\; |p'-p(\underline{\theta})| \; .$$

From (5.6) and (5.7) follows

$$W(\underline{z},p',\underline{\theta}) \leq \phi(p(\underline{\theta}))$$

(5.8)

$$+ ap(\underline{\theta})\; D_1^+(np(\underline{\theta})) + b(1-p(\underline{\theta}))\; D_0^-(n(1-p(\underline{\theta})))$$

$$+ 2c\; |p' - p(\underline{\theta})| \; .$$

Since (5.8) holds for each $0 \leq p' \leq 1$ we have, for any estimate $p_n(\underline{z})$ satisfying (4.13),

$$W(\underline{z},p_n(\underline{z}),\underline{\theta}) - \phi(p(\underline{\theta})) \leq ap(\underline{\theta}) D_1^+(np(\underline{\theta})) + b(1-p(\underline{\theta}) D_0^-(n(1-p(\underline{\theta})))$$

(5.9)

$$+ 2c\; |p_n(\underline{z}) - p(\underline{\theta})| \quad .$$

By the Glivenko-Cantelli Theorem [3]

(5.10) $$\Pr[\lim_{k=\infty} D_1^+(k) = 0] = \Pr[\lim_{k=\infty} D_0^-(k) = 0] = 1 \quad .$$

This is equivalent to the existence of functions $N_1^*(\eta)$, $N_0^*(\eta)$, defined $\eta > 0$, such that

$$(5.11)\qquad \begin{aligned} &\Pr[D_1^+(k) > \eta \quad \text{for any } k \geq N_1^*(\eta)] \leq \eta \\ &\Pr[D_0^-(k) > \eta \quad \text{for any } k \geq N_0^*(\eta)] \leq \eta . \end{aligned}$$

As in Section 4 (4.17) to (4.20), we have that if

$$(5.12)\qquad N^*(\varepsilon) = 2\varepsilon^{-1} \max[aN_1^*(\varepsilon/2),\ bN_0^*(\varepsilon/2)] \qquad \text{for } \varepsilon > 0,$$

then, for any $\underline{\theta}$ in $\Omega_{oo}$ ,

(5.13)

$$\Pr[ap(\underline{\theta})D_1^+(np(\underline{\theta}))+b(1-p(\underline{\theta}))D_0^-(n(1-p(\underline{\theta}))) > \varepsilon \text{ for any } n \geq N^*(\varepsilon)] \leq \varepsilon .$$

Returning to (5.9) we see that (5.13) and Theorem 2 combine to furnish for any $\varepsilon > 0$

$$(5.14)\qquad n_0(\varepsilon) = \max[N^*(\varepsilon/2),\ N(\varepsilon/4c)]$$

such that, for any $\underline{\theta}$ in $\Omega_{oo}$,

$$(5.15)\qquad \Pr[W(\underline{z}, p_n(\underline{z}), \underline{\theta}) - \phi(p(\underline{\theta})) > \varepsilon \quad \text{for any } n \geq n_0(\varepsilon)] < \varepsilon .$$

The argument from (5.2) to (5.15) proves

Theorem 3.

If $p_n(\underline{z})$ satisfies the conclusions (4.12), (4.13) of Theorem 2, then the decision function $\underline{t}^*$ defined by (5.1) is such that to any

$\varepsilon > 0$ there corresponds an $n_0(\varepsilon)$ such that for any $\underline{\theta}$ in $\underline{\Omega}_\infty$

(5.16) $\Pr[W(\underline{z},p_n(\underline{z}),\underline{\theta}) < \phi(p(\underline{\theta})) + \varepsilon \text{ for all } n \geq n_0(\varepsilon)] > 1-\varepsilon$ .

Since $W(\underline{z},p_n(\underline{z}),\underline{\theta}) \leq c$, Theorem 3 implies

Theorem 4.

Under the assumptions of Theorem 3, $\underline{t}^*$ is such that to any $\varepsilon > 0$ there corresponds an $n_1(\varepsilon)$ such that for any $n \geq n_1(\varepsilon)$ and any $\underline{\theta}$ in $\underline{\Omega}$,

$$R(\underline{t}^*, \underline{\theta}) < \phi(p(\underline{\theta})) + \varepsilon \tag{5.17}$$

Thus, $\underline{t}^*$ is a non-simple decision function which for large n does about as well as could be done by any simple decision function even if $p(\underline{\theta})$ were known exactly.

## 6. Invariant and R-Invariant Decision Functions.

Letting

(6.1) $W(\underline{t},\underline{\theta}) = \frac{1}{n}\sum_1^n [a\theta_i(1-t_i)+b(1-\theta_i)t_i]$ on $[0,1]^n \times \underline{\Omega}$

we have from (2.13)

$$R(\underline{t},\underline{\theta}) = \int W(\underline{t}(\underline{z}),\underline{\theta})\, d(\underline{z},\underline{\theta})\, d\nu^n \tag{6.2}$$

Let $(P(1), \ldots, P(n))$ be an arbitrary permutation of $(1, \ldots, n)$. Define, for any real vector $\xi = (\xi_1, \ldots, \xi_n)$,

$$(6.3) \qquad P\xi = (\xi_{P(1)}, \ldots, \xi_{P(n)}) .$$

We note that

$$(6.4) \qquad W(P\underline{t},P\underline{\theta}) = W(\underline{t},\underline{\theta})$$

$$(6.5) \qquad d(P\underline{z},P\underline{\theta}) = d(\underline{z},\underline{\theta})$$

and obtain for any decision function $\underline{t}$, by a change of variable of integration and the use of (6.4) and (6.5),

$$\begin{aligned} R(\underline{t},P\underline{\theta}) &= \int W(\underline{t}(\underline{z}),P\underline{\theta})\, d(\underline{z},P\underline{\theta})\, d\nu^n \\ &= \int W(\underline{t}(P\underline{z}),P\underline{\theta})\, d(P\underline{z},P\underline{\theta})\, d\nu^n \\ &= R(P^{-1}\underline{t}P,\underline{\theta}) \quad . \end{aligned}$$

We call a decision function $\underline{t}$ invariant if

$$(6.7) \qquad P^{-1}\underline{t}P = \underline{t} \qquad \text{for all } P,$$

and denote by R-invariant a decision function $\underline{t}$ such that

$$(6.8) \qquad R(P^{-1}\underline{t}P,\underline{\theta}) = R(\underline{t},\underline{\theta}) \qquad \text{for all } P.$$

The risk of an R-invariant decision function can be expressed as an explicit function of $P(\underline{\theta})$ by formally averaging the representation (6.2) over all P. From (6.2), (6.6) and (6.8)

$$(6.9) \qquad \begin{aligned} R(\underline{t},\underline{\theta}) &= (n!)^{-1}\sum_P R(\underline{t},P\underline{\theta}) \\ &= (n!)^{-1}\sum_P \int W(\underline{t}(\underline{z}),P\underline{\theta})d(\underline{z},P\underline{\theta})d\nu^n \end{aligned}$$

and changing the order of summations and integration, we obtain that if $\underline{t}$ is R-invariant its risk, $R(\underline{t},\underline{\theta})$, is expressible in the form

$$(6.10) \qquad \begin{aligned} R(\underline{t},\underline{\theta}) = \frac{1}{n}\sum_1^n \int \Big[ & a(1-t_i(\underline{z}))(n!)^{-1}\sum_P \theta_{P(i)}d(\underline{z},P\underline{\theta}) \\ & + bt_i(\underline{z})(n!)^{-1}\sum_P(1-\theta_{P(i)})d(\underline{z},P\underline{\theta})\Big]d\nu^n . \end{aligned}$$

For any real vector $\underline{r} = (r_1, \ldots, r_n)$ we define for every integer k

$$(6.11)\quad L(k,n,\underline{r}) = \begin{cases} (n!)^{-1}\sum\limits_{P}\prod\limits_{1}^{k} r_{P(j)} \prod\limits_{k+1}^{n}(1-r_{P(j)}) & k = 0, 1, \ldots, n, \\ \\ 0 & \text{otherwise,} \end{cases}$$

and

$$(6.12)\quad L(k,n-1,\check{r}_i) = L(k,n-1,(r_1,\ldots,r_{i-1},r_{i+1},\ldots,r_n)) \qquad i=1,2,\ldots n.$$

Then

$$\sum_P \theta_{P(i)} d(\underline{z},P\underline{\theta}) = z_i \sum_{P \ni \theta_{P(i)}=1} \prod_{j\neq i} [z_j^{\theta_{P(j)}}(1-z_j)^{1-\theta_{P(j)}}]$$

$$(6.13)\qquad = z_i(\sum_1^n \theta_k)(n-1)!\ L(\sum_1^n \theta_k - 1, n-1, \check{z}_i)$$

and similarly

$$(6.14)\quad \sum_P (1-\theta_{P(i)}) d(\underline{z},P\underline{\theta}) = (1-z_i)\sum_1^n(1-\theta_k)(n-1)!\ L(\sum_1^n \theta_k, n-1, \check{z}_i)\ .$$

It follows from (6.10) that for R-invariant $\underline{t}$

$$R(\underline{t},\underline{\theta}) = \frac{1}{n}\sum_1^n \int [a(1-t_i(\underline{z}))z_i p(\underline{\theta})\ L(np(\underline{\theta})-1,n-1,\check{z}_i)$$

$$(6.15)\qquad + bt_i(\underline{z})(1-z_i)(1-p(\underline{\theta}))\ L(np(\underline{\theta}),n-1,\check{z}_i)]dv^n$$

$$= \frac{1}{n}\sum_1^n R_i(\underline{t},\underline{\theta})\ .$$

For later use we note that if $\underline{t}$ is actually invariant

$$(6.16) \qquad R_i(\underline{t},\underline{\theta}) = R_1(\underline{t},\underline{\theta}) = R(\underline{t},\underline{\theta}) \qquad i = 1, \ldots, n,$$

follows from that property and a permutation of the variables of integration in the representation of $R_i(\underline{t},\underline{\theta})$ implicit in (6.15). It further follows from this representation that, for any fixed $\underline{\theta}$, the integrand of $R_i(\underline{t},\underline{\theta})$ is, for each fixed $\underline{z}$, at a minimum with respect to $t_i$ if and only if $t_i$ is of the form defined for $p = 0, 1/n, \ldots, n-1/n, 1$, by

$$(6.17) \quad t_{ip}(\underline{z}) = \begin{cases} 1 & az_i \, pL(np-1,n-1,\check{z}_i) > b(1-z_i)(1-p)L(np,n-1,\check{z}_i) \\ 0 & < \\ \text{arbitrary} & = \end{cases}$$

with p equal to $p(\underline{\theta})$. We denote $(t_{1p}(\underline{z}), \ldots, t_{np}(\underline{z}))$ by $\underline{t}_p$. It is easily verified that, with a proper determination of the arbitrary part, any decision function of the form (6.17) is invariant. Hence, the representation of its risk in the form (6.15) is valid and $R(\underline{t},\underline{\theta})$ is for each fixed $\underline{\theta}$ minimized over the class of R-invariant decision functions by an invariant decision function of the form (6.17).

Let

$$(6.18) \qquad \phi_n(p(\underline{\theta})) = \inf_{\text{R-invariant } \underline{t}} R(\underline{t},\underline{\theta}) \, .$$

From (6.15), (6.16) and (6.17) we obtain

$$\phi_n(p(\underline{\theta})) = R_1(\underline{t}_{p(\underline{\theta})}, \underline{\theta})$$

$$(6.19) \qquad = \int \min\left[ az_1 p(\underline{\theta}) L(np(\underline{\theta})-1, n-1, \check{z}_1) , \right.$$

$$\left. b(1-z_1)(1-p(\underline{\theta})) L(np(\underline{\theta}), n-1, \check{z}_1) \right] d\nu^n .$$

Since $L(k,n-1,\check{z}_1)$ is a symmetric function of $z_2, \ldots, z_n$, we have for any symmetric measurable set S

$$(6.20) \qquad \int_S L(k,n-1,\check{z}_1)\, d\nu^{n-1} = \nu_1^k \times \nu_0^{n-1-k}(S) .$$

It follows from (6.20) that the integral with respect to $z_2, \ldots, z_n$ of a $z_1$-section of the integrand of (6.19) is

$$az_1 p(\underline{\theta})\, \nu_1^{np(\underline{\theta})-1} \times \nu_0^{n-np(\underline{\theta})} (\overline{S}_{z_1})$$

(6.21)

$$+ b(1-z_1)(1-p(\underline{\theta}))\, \nu_1^{np(\underline{\theta})} \times \nu_0^{n-1-np(\underline{\theta})} (S_{z_1})$$

where $S_{z_1}$ is the symmetric set

$$S_{z_1} = \left[ (z_2, \ldots, z_n) \,\middle|\, az_1 p(\underline{\theta})\, L(np(\underline{\theta})-1, n-1, \check{z}_1) \right.$$

(6.22)

$$\left. > b(1-z_1)(1-p(\underline{\theta})) L(np(\underline{\theta}), n-1, \check{z}_1) \right] .$$

The developments from (6.9) to (6.22) set the stage for the following theorem.

Theorem 5.

If $\phi(p(\underline{\theta}))$ and $\phi_n(p(\underline{\theta}))$ are defined by (3.7) and (6.19), respectively, then for any $\varepsilon > 0$ there exists $N(\varepsilon)$ such that for any $n \geq N(\varepsilon)$ and any $\underline{\theta}$

$$(6.23) \qquad \phi(p(\underline{\theta})) - \varepsilon < \phi_n(p(\underline{\theta})) \leq \phi(p(\underline{\theta})) \ .$$

Proof.

$\phi_n \leq \phi$ for all $\underline{\theta}$ follows from the fact that every simple decision function is invariant.

For the non-trivial part of the proof fix $\varepsilon > 0$. From the continuity of $\phi(p)$ ((3.12)) we obtain a $\delta(\varepsilon)$ such that

$$(6.24) \qquad \phi(p) - \varepsilon < 0 \text{ for } p < \delta \text{ and } p > 1-\delta \ .$$

Thus, it will be sufficient to show the existence of an $N(\varepsilon)$ which suffices for the theorem when $0 < \delta \leq p(\underline{\theta}) \leq 1-\delta$. We will obtain this as a corollary from the following measure theoretic theorem which is itself a corollary to the principal theorem of Chapter II.

(T) If $m_0$, $m_1$ are non disjoint p-measures on a $\sigma$-algebra of subsets $X$ of a set $X$, then for every $\varepsilon$, $\delta > 0$ there exists $N(\varepsilon,\delta)$ such that for any pair of positive integers

$r,s$ with $r+s \geq N(\varepsilon,\delta)$ and $\delta \leq r/(r+s) \leq 1-\delta$

$$\left| m_0^r \times m_1^s(S) - m_0^{r-1} \times m_1^{s+1}(S) \right| < \varepsilon$$

uniformly for all symmetric $S$ in $\chi^{r+s}$.

From (T) we obtain that if $n-1 \geq N(\eta,\delta^*)$ and $\delta^* \leq \dfrac{np(\underline{\theta})}{n-1} \leq 1-\delta^*$

$$(6.25)\quad [1-\nu_1^{np(\underline{\theta})-1} \times \nu_0^{n-np(\underline{\theta})}(S_{z_1})] + \nu_1^{np(\underline{\theta})} \times \nu_0^{n-1-np(\underline{\theta})}(S_{z_1}) > 1-\eta$$

uniformly for all $z_1$ and from this it follows that $\phi_n(p(\underline{\theta})) > (1-\eta)\phi(p(\underline{\theta})) = \phi(p(\underline{\theta})) - \eta\phi(p(\underline{\theta}))$. Since the $\max\limits_p \phi(p) \leq c$ we see that in view of (6.24) the choice $N(\varepsilon) = N(c^{-1}\varepsilon,\ 2^{-1}\delta(\varepsilon))$ will complete the proof of Theorem 5.

We note that Theorem 5 can be combined with Theorem 3 or Theorem 4 of the previous section, leading to the strengthened endorsement of $\underline{t}^*$ : for large $\underline{n}$, $\underline{t}^*$ does about as well as could be done by any R-invariant decision function even if $p(\underline{\theta})$ were known exactly.

## 7. Bayes and Minimax Solutions.

Let $\underline{t} = (t_1(\underline{z}), \ldots, t_n(\underline{z}))$ be any decision function in the compound decision problem. For any $\underline{\Theta}$ in $\Omega$ we can write the risk (2.13) in the form

$$(7.1)\quad R(\underline{t},\underline{\Theta}) = \int \frac{1}{n}\sum_1^n \left[ a\Theta_i(1-t_i(\underline{z}))d(\underline{z},\underline{\Theta}) + b(1-\Theta_i)t_i(\underline{z})d(\underline{z},\underline{\Theta}) \right] d\,\nu^n \quad .$$

By a weight function we mean any function $\beta(\underline{\Theta}) \geq 0$ defined on $\Omega$ and not identically 0. For any weight function $\beta(\underline{\Theta})$ and any decision function $\underline{t}$ we define the weighted risk of $\underline{t}$ relative to $\beta(\underline{\Theta})$ as

$$B(\underline{t},\beta) = \sum_{\underline{\omega}} \beta(\underline{\omega}) \cdot R(\underline{t},\underline{\omega})$$

$$(7.2)\qquad = \int \frac{1}{n}\sum_1^n \left[ (1-t_i(\underline{z}))\, a \sum_{\underline{\omega}} \beta(\underline{\omega})\omega_i\, d(\underline{z},\underline{\omega}) \right.$$

$$\left. + t_i(\underline{z}) b \sum_{\underline{\omega}} \beta(\underline{\omega})(1-\omega_i)d(\underline{z},\underline{\omega}) \right] d\nu^n \quad .$$

For fixed $\beta(\underline{\Theta})$ this will be a minimum if and only if for almost every $(\nu^n)$ $\underline{z}$ and every $i = 1, \ldots, n$, $t_i(\underline{z})$ is of the form

$$(7.3)\quad t_i(\underline{z}\mid\beta) = \begin{cases} 1 & a\sum_{\underline{\omega}} \beta(\underline{\omega})\omega_i d(\underline{z},\underline{\omega}) > b\sum_{\underline{\omega}} \beta(\underline{\omega})(1-\omega_i)d(\underline{z},\underline{\omega}) \\ 0 & \qquad\qquad < \\ \text{arbitrary} & \qquad\qquad = \end{cases} \quad .$$

Any decision function $\underline{t}$ of the form (7.3) is called a Bayes solution relative to the weight function $\beta(\underline{\Theta})$.

For the remainder of the section we restrict our attention to symmetric $\beta(\underline{\theta})$, i.e.,

(7.4) $\quad \beta(\underline{\theta}) = \beta_k$ for all $\underline{\theta}$ with $\sum_1^n \theta_j = k$, $k = 0, \ldots, n$.

Using (6.12) we have

$$a \sum_{\underline{\omega}} \beta(\underline{\omega})\omega_i d(\underline{z},\underline{\omega}) = a \sum_0^n \left[ \beta_k \sum_{\underline{\omega} \ni \Sigma\omega_j = k} \omega_i \prod_1^n \left[ z_j^{\omega_j}(1-z_j)^{1-\omega_j} \right] \right]$$

(7.5)

$$= az_i \sum_0^n \left[ \beta_k \binom{n-1}{k-1} L(k-1,n-1,\check{z}_i) \right] ,$$

and similarly obtain

(7.6) $$b \sum_{\underline{\omega}} \beta(\underline{\omega})(1-\omega_i)d(\underline{z},\underline{\omega}) = b(1-z_i) \sum_0^n \left[ \beta_k \binom{n-1}{k} L(k,n-1,\check{z}_i) \right] .$$

Abbreviating $\binom{n-1}{k} L(k,n-1,\check{z}_i)$ by $b(k,n-1,\check{z}_i)$, it follows from (7.3) that for fixed $\beta(\underline{\theta})$ satisfying (7.4) $B(\underline{t},\beta)$ is a minimum if and only if $\underline{t}$ is of the form

(7.7) $$t_i(\underline{z}|\beta) = \begin{cases} 1 & az_i \sum_0^n \beta_k b(k-1,n-1,\check{z}_i) > b(1-z_i)\sum_0^n \beta_k b(k,n-1,\check{z}_i) \\ 0 & < \\ \text{arbitrary} & = \end{cases}$$

Several particular cases of (7.4) hold special interest for us. If for some integer $k_0$ with $0 \leq k_0 \leq n$ (7.4) takes the form

$$(7.8) \qquad \beta_{k_0} = 1, \ \beta_k = 0 \text{ for } k = 0, \ldots k_0-1, k_0+1, \ldots, n$$

then the corresponding Bayes solution (7.7) is a decision function of the form (6.17).

If for some constant p with $0 \leq p \leq 1$ (7.4) takes the form

$$(7.9) \qquad \beta_k = p^k(1-p)^{n-k} \quad ,$$

then the corresponding Bayes solution (7.7) is a decision function of the form (3.6) provided the Bayes solution is required to be simple on the arbitrary part of its definition.

Referring back to Section 3, we have by the previous paragraph that each of the simple decision functions $\underline{t} = t_p$ , $0 \leq p \leq 1$, defined by (3.6) is a Bayes solution for the compound decision problem. It can be shown (we omit the simple proof) that there always exists a value $0 < p' < 1$, and a determination of $t_{p'}$ for which the coefficient of p in (3.16) vanishes; letting r be the constant value of $B(t_{p'},p)$ it follows that

$$(7.10) \qquad R(t_{p'},\underline{\theta}) \equiv r \qquad \text{for every } \underline{\theta} \text{ in } \underline{\Theta} \quad .$$

The decision function $\underline{t}' = t_{p'}$ has the following properties:

(a) it is simple,

(7.11) (b) it is a Bayes solution, relative to a weight function which is positive for every $\underline{\theta}$ in $\underline{\Theta}$,

(c) it has a constant risk r for every $\underline{\theta}$ in $\underline{\Theta}$ ,

whence it follows that it is also an admissible minimax; that is,

$$\sup_{\underline{\theta} \text{ in } \underline{\Theta}} \left[ R(\underline{t},\underline{\theta}) \right] = \text{minimum for } \underline{t} = \underline{t}' \ , \tag{7.12}$$

and if $\underline{t}$ is any other decision function minimizing

$$\sup_{\underline{\theta} \text{ in } \underline{\Theta}} \left[ R(\underline{t},\underline{\theta}) \right] \tag{7.13}$$

then $R(\underline{t},\underline{\theta}) \equiv R(t_{p'},\underline{\theta})$.

## 8. Conclusion.

Since the minimax decision function, $\underline{t}' = t_{p'}$, is simple and has constant risk, it follows that

$$\sup_{\underline{\theta} \text{ in } \underline{\Theta}} R(t_{p'},\underline{\theta}) = R(t_{p'},\underline{\theta}) \geq \inf_{t} R(t,\underline{\theta}) = \phi(p(\underline{\theta})).$$

Since $\phi(0) = \phi(1) = 0$, proper inequality actually holds for at least two points in the parameter space as long as $\mu_0$ and $\mu_1$ are non-disjoint. From Theorem 4 we can conclude that there exists $\varepsilon > 0$ such that for sufficiently large n the minimax decision function is $\varepsilon$-inadmissible, since decision functions of the type $\underline{t}^*$ are $\varepsilon$-better. The present paper has failed to exhibit a $\underline{t}^*$ which is admissible (or even to show the existence of such), and it is hoped that this result can be obtained from a more intensive investigation of the Bayes solutions (7.3).

In closing we would like to emphasize the fact that the practical significance of the performance of decision functions of the type $\underline{t}^*$ is greater than that of their use in presently existing situations. Their performance is distinctly more satisfactory than that of any decision function in the component problem (unless the component problem is the product of a completely known probability mechanism) and it may, consequently, be worthwhile to design experiments so as to utilize this advantage of the compound approach.

CHAPTER II.

ASYMPTOTIC THEOREMS.

## 1. Introductory Remarks.

The principal theorem of this chapter is Theorem 1. Our interest in it was occasioned by the need for the theorem (T) (used in obtaining Theorem 5 of Chapter I), since (T) is an obvious corollary of Theorem L. The density Central Limit (DCLT) for the generalized binomial (GB) could probably be obtained by a refinement of Von Mises' [13] use of the Fourier inversion theorem. The use of Theorem 3 for this task seemed to have the advantages of novelty of approach and less uniformity difficulty. The work of Hammersley [6],[7] on DCLT was called to our attention after the completion of the present chapter.

## 2. Statement of the Theorem.

Let $X$ be an arbitrary set, $\mathcal{X}$ be a $\sigma$-field of subsets of $X$, $\mathfrak{M}$ be the family of all probability measures defined on $\mathcal{X}$. For $\mu, \nu$ any two elements of $\mathfrak{M}$ we define and use the following notation for the relations of $\nu$-absolute continuity, equivalence and orthogonality, respectively:

$$\mu \ll \nu \equiv \text{If } T \text{ in } \mathcal{X} \text{ and } \nu(T) = 0, \text{ then } \mu(T) = 0.$$

$$\mu \sim \nu \equiv \mu \ll \nu \text{ and } \nu \ll \mu .$$

$$\mu \perp \nu \equiv \text{There exists } T \text{ in } \mathcal{X} \text{ with } \mu(T) = \nu(X-T) = 0.$$

Let $X^N$ be the cartesian product of X with itself to N factors, $R_N$ be the family of subsets of $X^N$ which are product sets with components in X, and $E_N$ be the field of subsets of $X^N$ which have a representation as a finite union of sets in $R_N$. Let $X^N$ be the extension of $R_N$ to be a σ-field and for $\nu_j$ in $\mathfrak{M}$, $j = 1, \ldots, N$, let $\nu_1 \ldots \nu_N \equiv \overset{N}{\underset{1}{X}} \nu_j$ be the unique extension to a measure on $X^N$ of the measure defined on $E_N$ which assigns to a set $T = T_1 \times \ldots \times T_N$ in $R_N$ the number $\prod_1^N \nu_j(T_j)$. If $\nu_1 = \ldots = \nu_k = \mu, \nu_{k+1} = \ldots = \nu_N = \nu$, we abbreviate $\overset{N}{\underset{1}{X}} \nu_j$ by $\mu^k \nu^{N-k}$. We use the letter P for the permutation $(P(1), \ldots, P(N))$ of $(1, \ldots, N)$ and as an operator on the product space, $P(x_1, \ldots, x_N) = (x_{P(1)}, \ldots, x_{P(N)})$. We refer to a subset S of $X^N$ as symmetric if $P(S) = S$ for all P.

Theorem 1.

If $\mu_0$ and $\mu_1$ are any two non-orthogonal elements of $\mathfrak{M}$ and $k(n,m)$ is any integral valued function on pairs of positive integers such that $k^2(n,m) = o(\frac{m \cdot n}{m+n})$, then there exist functions $N(\epsilon)$, $M(\epsilon)$, defined $\epsilon > 0$, such that for any pair of positive integers $(n,m)$ with $n \geq N(\epsilon)$, $m \geq M(\epsilon)$

$$\left| \mu_0^n \mu_1^m (S) - \mu_0^{n-k(n,m)} \mu_1^{m+k(n,m)} (S) \right| < \epsilon$$

uniformly for all symmetric S in $X^{n+m}$.

We remark that if $R_{n+m}$ were substituted for $X^{n+m}$ in the statement of the theorem, the proof would follow readily from the fact that $a^n b^m - a^{n-k(n,m)} b^{m+k(n,m)} = o(1)$ uniformly for all $(a,b)$ on the unit square, since S symmetric in $R_{n+m}$ implies $S = T^{n+m}$ with T in X.

The necessity of the non-orthogonality of $\mu_0$ and $\mu_1$ can be shown as follows. Let T in X be such that $\mu_0(T) = \mu_1(X-T) = 0$ and let $K = (X-T)^n T^m$. Then K is in $R_{n+m}$, $\mu_0^n \mu_1^m(K) = 1$ and $\mu_0^{n-1} \mu_1^{m+1}(K) = 0$. With the obvious meaning attached to P(K), we readily verify that $S = \bigcup_P P(K)$ is symmetric and $\mu_0^n \mu_1^m(S) - \mu_0^{n-1} \mu_1^{m+1}(S) = 1$.

## 3. Reduction to a DCLT for GB.

For the proof of Theorem 1 we introduce a class of probability measures on X. For $0 < p < 1$ let $\mu_p$ be the probability measure defined by

$$\mu_p(T) = p\mu_1(T) + (1-p)\mu_0(T) \text{ for all T in X.} \tag{3.1}$$

Clearly $\mu_0 \ll \mu_p$, $\mu_1 \ll \mu_p$ whence $\mu_{p'} \ll \mu_p$ and $\mu_{p'} \sim \mu_p$ for $0 < p$, $p' < 1$, and we consequently use the notation, a.e. ($\mu$), to refer to a set in X having $\mu_p$-measure 1 for any $0 < p < 1$. Let $f_0(t,p), f_1(t,p)$ be the Radon-Nikodym derivatives $\frac{d\mu_0}{d\mu_p}$, $\frac{d\mu_1}{d\mu_p}$, defined a.e. ($\mu$) by

$$\mu_i(T) = \int_T f_i(t,p)d\mu_p \text{ for all T in X,} \qquad i = 0, 1. \tag{3.2}$$

We digress to develop some of the properties of the functions $f_i(t,p)$ which will be needed later in the proof. For any T in $\mathcal{X}$

$$\int_T pf_1(t,p) + (1-p)f_0(t,p)d\mu_p = p\mu_1(T) + (1-p)\mu_0(T) = \int_T 1 d\mu_p,$$

from which we conclude

$$(3.3) \qquad pf_1(t,p) + (1-p)f_0(t,p) = 1 \quad \text{a.e. } (\mu),\ 0 < p < 1.$$

To make explicit the degree of dependence on p of the functions $f_i(t,p)$ we let $f_i(t) = f_i(t, \frac{1}{2})$ i = 0, 1 and obtain from

$$\int_T \frac{f_i(t)}{pf_1(t)+(1-p)f_0(t)}\, d\mu_p = \int_T \frac{f_i(t)}{pf_1(t)+(1-p)f_0(t)} \frac{d\mu_p}{d\mu_{\frac{1}{2}}}\, d\mu_{\frac{1}{2}}$$

$$= \int_T f_i(t)d\mu_{\frac{1}{2}} = \mu_i(T) = \int_T f_i(t,p)d\mu_p \quad i=0,1$$

that

$$(3.4) \qquad f_i(t,p) = \frac{f_i(t)}{pf_1(t)+(1-p)f_0(t)} \qquad \text{a.e. } (\mu).$$

From (3.4) we obtain

$$\int f_1(t,p)f_0(t,p)d\mu_p = \int f_1(t,p)d\mu_0 = \int \frac{f_1(t)}{pf_1(t)+(1-p)f_0(t)}\, d\mu_0$$

$$\geq \int_{f_1(t)\leq 1} \frac{f_1(t)}{f_0(t)}\, d\mu_0 + \int_{f_1(t)>1} 1\, d\mu_0$$

and, since the right-hand side is independent of p and is equal to zero only if $f_1(t) = 0$ a.e. $(\mu_0)$, we conclude that for any $\mu_0 \not\perp \mu_1$, there exists $h > 0$ such that

$$\int f_0(t,p)f_1(t,p)d\mu_p \geq h \text{ uniformly } 0 < p < 1. \tag{3.5}$$

From (3.3) we obtain

$$\int pf_1(t,p)(1-p)f_0(t,p)d\mu_p = p \int (1-p)f_0(t,p)d\mu_1 \leq p$$

and

$$\int pf_1(t,p)\cdot(1-p)f_0(t,p)d\mu_p = (1-p) \int pf_1(t,p)d\mu_0 \leq (1-p)$$

whence

$$\int pf_1(t,p)\cdot(1-p)f_0(t,p)d\mu_p \leq \min(p,1-p). \tag{3.6}$$

This is an obviously improvable bound which turns out to be sufficient for our purposes.

Returning to our main development at (3.2) we obtain that

$$\frac{d\mu_0^n\mu_1^m}{d\mu_p^{n+m}} = \prod_1^n f_0(x_j,p)\cdot \prod_{n+1}^{n+m} f_1(x_j,p) \qquad \text{a.e. } (\mu^{n+m}) \tag{3.7}$$

from the fact that the right-hand side is a solution of the defining equation for S in $R_{n+m}$, hence by the extension theorem, for S in $\chi^{n+m}$. Letting $\psi_S(x) = \psi_S(x_1, \ldots, x_{n+m})$ be the characteristic set function of the subset S of $X^{n+m}$ it follows from (3.7) that for any S in $X^{n+m}$

$$\mu_0^n\mu_1^m(S) = \int_S \prod_1^n f_0(x_j,p) \prod_{n+1}^{n+m} f_1(x_j,p)\,d\mu_p^{n+m}$$

(3.8)

$$= \int \psi_S(x) \prod_1^n f_0(x_j,p) \prod_{n+1}^{n+m} f_1(x_j,p)\,d\mu_p^{n+m} \quad .$$

Since $\mu_p^{n+m}(P(S)) = \mu_p^{n+m}(S)$ for S in $X^{n+m}$ and any P, the right-hand side of (3.8) is invariant under permutations of the integrand variables and we consequently obtain a representation of $\mu_0^n\mu_1^m(S)$ as a formal average over all P

$$\mu_0^n\mu_1^m(S) = \frac{1}{(n+m)!} \sum_P \int \psi_S(P(x))\prod_1^n f_0(x_{P(j)},p)\prod_{n+1}^{n+m} f_1(x_{P(j)},p)\,d\mu_p^{n+m}$$

(3.9)

$$= \int \frac{1}{(n+m)!} \sum_P \psi_S(P(x))\prod_1^n f_0(x_{P(j)},p) \prod_{n+1}^{n+m} f_1(x_{P(j)},p)\,d\mu_p^{n+m}$$

If P(S) = S for every P, $\psi_S(P(x)) = \psi_S(x)$ for every P and the right-hand side of (3.9) reduces to

$$\text{(3.10)} \qquad \int \psi_S(x) \frac{1}{(n+m)!} \sum_P \prod_1^n f_0(x_{P(j)},p) \prod_{n+1}^{n+m} f_1(x_{P(j)},p)\,d\mu_p^{n+m}$$

which thus furnishes us with a class of density representation of the function defined for all S in $X^{n+m}$ by

$$\text{(3.11)} \qquad \mathrm{sm}\ \mu_0^n\mu_1^m(S) = \frac{1}{(n+m)!} \sum_P \mu_0^n\mu_1^m(P(S)) .$$

The function sm $\mu_0^n\mu_1^m$ is the unique extension to $X^{n+m}$ of $\mu_0^n\mu_1^m$ on the symmetric sets of $X^{n+m}$ such that sm $\mu_0^n\mu_1^m(P(S))$ = sm $\mu_0^n\mu_1^m(S)$.

If $N$ is any integer, $z = (z_1, \ldots, z_N)$ any point in $[0,1]^N$ let $b_N(k,z)$ be defined for $k = 0, 1, \ldots, N$ by

$$\text{(3.12)} \qquad k!(N-k)!b_N(k,z) = \sum_P \prod_1^k z_{P(j)} \prod_{k+1}^N (1-z_{P(j)}) .$$

For any $0 < p < 1$ let

$$\text{(3.13)} \qquad b(k;\ N,p) = \binom{N}{k} p^k (1-p)^{N-k} \qquad k = 0, 1, \ldots, N.$$

Using the abbreviation

$$\text{(3.14)} \qquad pf_1(x_j,p) = z_j(p)$$

we obtain from (3.3) and (3.10) through (3.14)

$$\text{(3.15)} \qquad \mathrm{sm}\ \mu_0^n\mu_1^m(S) = \int \psi_S(x) \frac{b_{n+m}(m,z(p))}{b(m;\ n+m,p)}\ d\mu_p^{n+m} .$$

From (3.15) we obtain for any integer $k$ such that $0 < n-k \leq n+m$

$$\text{(3.16)} \qquad \mathrm{sm}\ \mu_0^n\mu_1^m(S) - \mathrm{sm}\ \mu_0^{n-k}\mu_1^{m+k}(S) =$$

$$\int \psi_S(x)\left[\frac{b_{n+m}(m,z(p))}{b(m;\ n+m,p)} - \frac{b_{n+m}(m+k,z(p))}{b(m+k;n+m,p)}\right] d\mu_p^{n+m} .$$

Since the integrand of the representation (3.16) is maximal for each fixed x if S is taken to be the symmetric set

$$(3.17) \qquad S_{max} = \left[ x \;\middle|\; \frac{b_{n+m}(m,z(p))}{b(m;n+m,p)} > \frac{b_{n+m}(m+k,z(p))}{b(m+k;n+m,p)} \right]$$

we conclude that $\mu_0^n\mu_1^m(S) - \mu_0^{n-k}\mu_1^{m+k}(S)$ is maximized over the class of symmetric sets by $S_{max}$. It follows similarly that $\mu_0^n\mu_1^m(S)-\mu_0^{n-k}\mu_1^{m+k}(S)$ is minimized over the class of symmetric sets by c $S_{max}$(by c we denote complimentation) and since $\mu_0^n\mu_1^m(X^{n+m}) = \mu_0^{n-k}\mu_1^{m+k}(X^{n+m}) = 1$ we obtain that $S_{max}$ maximizes $|\mu_0^n\mu_1^m(S) - \mu_0^{n-k}\mu_1^{m+k}(S)|$ in the class of symmetric sets of $X^{n+m}$. Letting

$$(3.18) \qquad M(n,m,k) = 2\left|\mu_0^n\mu_1^m(S_{max}) - \mu_0^{n-k}\mu_1^{m+k}(S_{max})\right|$$

it follows from (3.16) and (3.17) that M(n,m,k) has the class of representations

$$(3.19) \quad M(n,m,k) = \int \left| \frac{b_{n+m}(m,z(p))}{b(m;\ n+m,p)} - \frac{b_{n+m}(m+k,z(p))}{b(m+k;n+m,p)} \right| d\mu_p^{n+m} \quad .$$

The remainder of the proof of Theorem 1 will be devoted to showing

$$(3.20) \qquad \lim_{\substack{n=\infty \\ m=\infty}} M(n,m,k(n,m)) = 0.$$

This is clearly a necessary and sufficient condition for Theorem 1.

The sequence M(n,m,k(n,m)) is independent of p and the advantage of the class of representations lies in the fact that we may, for each (n,m), choose a p(n,m) in such a way as to reduce the analytic problems involved in the approximation of the integrand of (3.19) over a portion of the range of integration. We make that choice $p(n,m) = m(n+m)^{-1}$ and writing

$$p' = m(n+m)^{-1}$$

(3.21)

$$k' = k(n,m)$$

express $M(n,m,k')$ in the form

$$M(n,m,k') = M_1(n,m,k',p') + M_2(n,m,k',p') \tag{3.22}$$

where $M_1$ and $M_2$ are defined by (3.22) and

$$M_1(n,m,k',p') = \int_{cR_{n,m}} \left| \frac{b_{n+m}(m,z(p'))}{b(m;n+m,p')} - \frac{b_{n+m}(m+k',z(p'))}{b(m+k';n+m,p')} \right| d\mu_{p'}^{n+m} ,$$

(3.23)

$$cR_{n,m} = [x \mid s_{n+m}^2(z(p')) \leq \frac{h}{2} \; \frac{m \cdot n}{m+n} ]$$

with $s_{n+m}^2(z)$ defined for every z in $[0,1]^{n+m}$ by

$$s_{n+m}^2(z) = \sum_1^{n+m} z_j(1-z_j)$$

and the constant, h, taken to be the $h(\mu_0,\mu_1)$ possessing the property (3.5). We will prove (3.20) by showing

$$\lim_{\substack{n=\infty \\ m=\infty}} M_1(n,m,k',p') = 0 \tag{3.24}$$

(3.25) $$\lim_{\substack{n=\infty \\ m=\infty}} M_2(n,m,k',p') = 0 \ .$$

For the proof of (3.24) we introduce several lemmas.

Lemma 1.

$$\mu_{p'}^{n+m} [x \mid s^2_{n+m}(z(p')) \leq \frac{h}{2} \ \frac{m\cdot n}{n+m}] \leq \frac{4}{h^2} \ \frac{(n+m)^2}{n\cdot m\cdot \max(n,m)}$$

Proof.

For any integrable $(\mu_{p'}^{n+m})$ function $g(x)$ defined on $X^{n+m}$ let $Eg$ denote $\int g(x)d\mu_{p'}^{n+m}$ . We have

(3.26) $$E\ s^2_{n+m}(z(p')) = (n+m)\ E\ z_1(p')(1-z_1(p'))$$

and consequently by (3.5)

(3.27) $$E\ s^2_{n+m}(z(p')) \geq (n+m)p'(1-p')h = \frac{n\cdot m}{n+m}\ h \ .$$

Omitting unessential arguments it follows from (3.27) that

(3.28) $$[s^2 \leq \frac{h}{2}\ \frac{m\cdot n}{n+m}] \subset [s^2 - Es^2 \leq -\frac{h}{2}\ \frac{m\cdot n}{m+n}] \subset [|s^2 - Es^2| \geq \frac{h}{2}\ \frac{m\cdot n}{m+n}]$$

Tchebycheff's inequality yields that

(3.29) $$\mu_{p'}^{n+m} [|s^2 - Es^2| \geq \frac{h}{2}\ \frac{m\cdot n}{m+n}] \leq \frac{4}{h^2}\ \frac{(n+m)^2}{n^2 m^2}\ E(s^2 - Es^2)^2 \ .$$

Further

$$E(s^2-Es^2)^2 = (n+m)\ E\left[z_1(1-z_1) - E\ z_1(1-z_1)\right]^2$$

(3.30) $$= (n+m)\ E\left[z_1(1-z_1)\right]^2 - \left[E\ z_1(1-z_1)\right]^2$$

$$\leq (n+m)\ E\left[z_1(1-z_1)\right]^2 \quad .$$

From (3.6) and $|z_1(1-z_1)| \leq 1$ it follows that

(3.31) $$E\left[z_1(1-z_1)\right]^2 \leq E\ z_1(1-z_1) \leq \min(p',1-p') \quad .$$

Since $\frac{(n+m)\min(p',1-p')}{n\cdot m} = \frac{1}{\max(n,m)}$ the proof is completed by the chain (3.28) through (3.31).

Lemma 2.

$$\lim_{\substack{n=\infty \\ m=\infty}} \left[2\pi\ \frac{m\cdot n}{m+n}\right]^{\frac{1}{2}} b(m+k';n+m,p') = 1 \quad .$$

Proof.

By the Stirling approximation we obtain

$$\left[2\pi\ \frac{m\cdot n}{m+n}\right]^{\frac{1}{2}} b(m+k';n+m,p') \sim$$

(3.32)

$$\exp - \left[(m+k')\ln(1+\frac{k'}{m}) + (n-k')\ln(1-\frac{k'}{n})\right]$$

Since $k' = o(\sqrt{\frac{m\cdot n}{m+n}})$, $k' = o(m^{\frac{1}{2}})$ and $o(n^{\frac{1}{2}})$, and the lemma follows with the aid of the easily established bound

$$(3.33) \qquad |\ln(1+\eta) - \eta| < \eta^2 \qquad |\eta| < \frac{1}{2} \quad .$$

Since $b_{n+m}(j,z) \geq 0$ and

$$(3.34) \qquad \sum_{0}^{n+m} b_{n+m}(j,z) = \prod_{1}^{n+m} [z_j + (1-z_j)] = 1$$

it follows that the sup over $X^{n+m}$ of the integrand of (3.23) $\leq \max( b(m;n+m,p')^{-1}, b(m+k';n+m,p')^{-1} )$, while from Lemma 2 this max is $\sim [2\pi \frac{m\cdot n}{m+n}]^{\frac{1}{2}}$ . This remark and Lemma 1 imply

$$M_1(n,m,k',p') \lesssim \frac{4(2\pi)^{\frac{1}{2}}}{h^2} \frac{(n+m)^{3/2}}{n^{\frac{1}{2}}m^{\frac{1}{2}}\max(n,m)} \leq$$

$$(3.35) \qquad \frac{4(2\pi)^{\frac{1}{2}}}{h^2} \frac{2^{3/2}}{[\min(n,m)]^{\frac{1}{2}}}$$

which completes the proof of (3.24).

For the proof of (3.25) we introduce Theorem 2 which will itself be proved by Lemmas 3 to 7 and Theorem 3, following its use in the completion of Theorem 1. For the normal density and distribution functions we employ the notation

$$(3.36) \qquad \phi(t) = (2\pi)^{-\frac{1}{2}} e^{-\frac{t^2}{2}}, \quad \Phi(t) = \int_{-\infty}^{t} \phi(v)\, dv.$$

Theorem 2.

For any $z$ in $[0,1]^N$, $N = 1, 2, \ldots$, let

$$(3.37) \qquad b_N(k,z) = \Sigma' \prod_{1}^{k} z_{P(j)} \prod_{k+1}^{N} (1-z_{P(j)}) \qquad k = 0, \ldots, N$$

where $\Sigma'$ indicates summation over all $\binom{N}{K}$ permutations $P$ with $P(1) < \cdots < P(k)$ and $P(k+1) < \cdots < P(N)$. Let

$$(3.38) \qquad s_N^2(z) = \sum_{1}^{N} z_j(1-z_j)$$

$$(3.39) \qquad k^* = (k - \sum_{1}^{N} z_j) s_N^{-1}(z) \qquad k = 0, \ldots, N.$$

Then for any $\epsilon > 0$ there exists $M(\epsilon)$ such that

$$(3.40) \qquad \max_{k} \left| s_N(z) b_N(k,z) - \phi(k^*) \right| < \epsilon$$

for all $N, z$ such that $s_N(z) > M(\epsilon)$.

Proof.

We will obtain that an $M(\epsilon)$ of the form $(\phi(0)+3c)/\delta(\epsilon)$, where $c$ will be obtained from Lemma 3 and $\delta(\epsilon)$ from Theorem 3, will suffice for Theorem 2. For the moment we defer the proof and show that Theorem 2 will enable us to complete the proof of Theorem 1.

Returning to the proof of Theorem 1, let

$$R_{n,m} = [\, x \mid s^2_{n+m}(z(p')) > \frac{h}{2}\, n\cdot m(n+m)^{-1} \,]$$

$$I_{n,m}(x) =$$

(3.41)

$$\left| \alpha' s_{n+m}(z(p'))b_{n+m}(m,z(p')) - \beta' s_{n+m}(z(p'))b_{n+m}(m+k',z(p)) \right|$$

where

$$(\alpha')^{-1} = [\, 2\pi\ m\cdot n(n+m)^{-1} \,]^{\frac{1}{2}}\ b(m;n+m,p')$$

$$(\beta')^{-1} = [\, 2\pi\ m\cdot n(n+m)^{-1} \,]^{\frac{1}{2}} b(m+k';n+m,p').$$

From (3.22) and (3.23) we express

$$M_2(n,m,k',p') = \int_{R_{n,m}} \left[ \frac{2\pi\ m\cdot n(n+m)^{-1}}{s^2_{n+m}(z(p'))} \right] I_{n,m}(x)\, d\mu^{n+m}_{p'}$$

(3.42)

$$\leq \left[ \frac{4\pi}{h} \right]^{\frac{1}{2}} \sup_{R_{n,m}} I_{n,m}(x).$$

By the triangle inequality

$$\sup_{R_{n,m}} I_{n,m}(x) \le \alpha' \sup_{R_{n,m}} \left| s_{n+m}(z(p'))b_{n+m}(m,z(p')) - \phi(m^*) \right|$$

$$(3.43) \qquad +\beta' \sup_{R_{n,m}} \left|\phi((m+k')^*) - s_{n+m}(z(p'))b_{n+m}(m+k',z(p'))\right|$$

$$+ \sup_{R_{n,m}} \left|\alpha'\phi(m^*) - \beta' \phi((m+k')^*)\right| .$$

Since $\lim\limits_{\substack{n=\infty \\ m=\infty}} \frac{n\cdot m}{n+m} = \infty$, Theorem 2 implies

$$(3.44) \quad \lim_{\substack{n=\infty \\ m=\infty}} \sup_{R_{n,m}} \left| s_{n+m}(z(p'))b_{n+m}(m,z(p')) - \phi(m^*) \right| = 0,$$

$$(3.45) \quad \lim_{\substack{n=\infty \\ m=\infty}} \sup_{R_{n,m}} \left| \phi((m+k')^*) - s_{n+m}(z(p'))b_{n+m}(m+k',z(p')\right| = 0.$$

Since $\phi(t) \le \phi(0)$ and $\left|\phi(t_1) - \phi(t_2)\right| \le |\phi'(1)|\,|t_1-t_2|$

$$\left|\phi(m^*) - \phi((m+k')^*)\right| \le |\phi'(1)|\,|k'| \,/s_{n+m}(z(p'))$$

$$(3.46) \qquad \le |\phi'(1)| \sqrt{\frac{2}{h}} \left| \frac{k'}{\sqrt{\frac{m\cdot n}{m+n}}} \right|$$

for x in $R_{n,m}$. From (3.44), (3.45), (3.46) and Lemma 2 we obtain

$$(3.47) \qquad \lim_{\substack{n=\infty \\ m=\infty}} \sup_{R_{n,m}} I_{n,m}(x) = 0.$$

(3.25) follows from (3.42) and (3.47) and the proof of Theorem 1 is complete.

## 4. A CLT for Ln-Concave Extensions of GB.

In order to prove Theorem 2 we will prove Lemmas 3, 4, 5, 6, 7 and finally obtain its proof as a corollary to Theorem 3.

We will make considerable use of the abbreviations

(4.1) $[t]$ = greatest integer less than or equal to t $\quad -\infty < t < \infty$,

(4.2) $$t^{-*} = s_N(z)\cdot t + \sum_1^N z_j \qquad z \text{ in } [0,1]^N, \quad -\infty < t < \infty,$$

as well as of those definitions and abbreviations previously introduced.

Lemma 3.

There exists a constant c such that

(4.3) $$\sup_t \left| \sum_0^{[t^{-*}]} b_N(k,z) - \Phi(t) \right| \leq c s_N^{-1}(z).$$

Proof.

$b_N(k,z)$ $k = 0, \ldots, N$ has a realization as the distribution of the sum of N independent random variables $\xi_1, \ldots, \xi_N$ where $\xi_j$ assumes the values 1 and 0 with probability $z_j$ and $1-z_j$, respectively. Berry's Theorem [1] can be applied and asserts that there exists a constant c such that

$$(4.4) \qquad \sup_t \left| \sum_{0}^{[t^{-*}]} b_N(k,z) - \Phi(t) \right| \leq c\Gamma_N^3(z) s_N^{-3}(z).$$

Since $\Gamma_N^3(z) = \sum_1^N [z_j^3(1-z_j)+(1-z_j)^3 z_j] \leq s_N^2(z)$, Lemma 3 follows from (4.4).

If $a_0, a_1, \ldots, a_N$ is any finite sequence of non-negative numbers we shall say that it is a $\ell n$-concave finite sequence if

$$a_k^2 \geq a_{k-1}a_{k+1} \qquad k = 1, 2, \ldots, N-1.$$

An equivalent definition is

$$a_{k_2}^{k_3-k_1} \geq a_{k_1}^{k_3-k_2} \cdot a_{k_3}^{k_2-k_1} \qquad 0 \leq k_1 \leq k_2 \leq k_3 \leq N.$$

Lemma 4.

For each positive integer N and each z in $[0,1]^N$, $b_N(k,z)$ $k = 0, 1, \ldots, N$ is a $\ell n$-concave finite sequence.

Proof.

A stronger result is an easy generalization of Theorem 144 [10], namely that $L_k = \binom{N}{k}^{-1} b_N(k,z)$ $k = 0, 1, \ldots, N$ is a $\ell n$-concave finite sequence. Assuming this result, it follows that for $1 \leq k < N$

$$b_N(k-1,z)b_N(k+1,z) = \binom{N}{k-1}\binom{N}{k+1}L_{k-1}L_{k+1} \leq \binom{N}{k-1}\binom{N}{k+1}L_k^2 = \tag{4.5}$$

$$\left(1+\frac{1}{k}\right)^{-1}\left(1+\frac{1}{N-k}\right)^{-1} b_N^2(k,z) \leq b_N^2(k,z)$$

with equality if and only if $b_N^2(k,z) = 0$.

<u>Lemma</u> 5.

For each positive integer N and each z in $[0,1]^N$, $b_N(k,z)$ has a continuous extension $\overset{\circ}{b}_N(k,z)$ which is $\ell n$-concave on $-\infty < k < \infty$ and such that

$$\sup_{-\infty < r < \infty} \left| \int_{-\infty}^{r} \overset{\circ}{b}_N(k,z)dk - \sum_{0}^{[r]} b_N(k,z) \right| < \max_k b_N(k,z) \ . \tag{4.6}$$

<u>Proof</u>.

Fix N,z and abbreviate $b_N(k,z)$ by $b_k$. Let m be any k such that $b_k = \max$. If $k < k' \leq m$ it follows from Lemma 4 that

$$b_{k'} \geq b_k^{\frac{m-k'}{m-k}} \cdot b_m^{\frac{k'-k}{m-k}}$$

whence either $b_k = 0$ or

$$\ell n\ b_{k'} \geq \frac{m-k'}{m-k}\ \ell n\ b_k + \frac{k'-k}{m-k}\ \ell n\ b_m$$

and in either case it follows that

(4.7) $$b_k \leq b_{k'} \qquad k < k' \leq m .$$

By a similar argument we obtain

(4.8) $$b_k \geq b_{k'} \qquad m \leq k < k'$$

and conclude that the set of $k$ such that $b_k \neq 0$ is an interval, $I = [L,R]$. We define a class of functions $g(t,M)$ on $-\infty < t < \infty$, by letting $g(t,M) = 0$ for $t$ outside the interval $(L-1, R+1)$ and determining $g(t,M)$ for $L-1 < t < R+1$ by

(4.9) $$\ln g(t,M) = \begin{cases} \ln b_L + M \ln(t-(L-1)) & L-1 < t < L \\ \ln b_{[t]} + (\ln b_{[t]+1} - \ln b_{[t]})(t-[t]) & L \leq t \leq R \\ \ln b_R + M \ln(R+1-t) & R < t < R+1. \end{cases}$$

It is easily verified that if

(4.10) $$M > \max(\ln b_{L+1} - \ln b_L,\ \ln b_{R-1} - \ln b_R)$$

$g(t,M)$ is a continuous $\ln$-concave extension of $b_k$, $k = 0, 1, \ldots, N$ and

(4.11) $$\int_{L-1}^{L} g(t,M)\,dt = \frac{b_L}{M+1}, \qquad \int_{R}^{R+1} g(t,M)\,dt = \frac{b_R}{M+1} .$$

It follows from (4.7) and (4.8) that if $m$ is any integer such that $b_m = \max b_k$ and $j$ is an integer

$$b_{j-1} \leq \int_{j-1}^{j} g(t,M)\,dt \leq b_j \qquad L < j \leq m$$

(4.12)

$$b_j \leq \int_{j-1}^{j} g(t,M)\,dt \leq b_{j-1} \qquad m < j \leq R \quad .$$

From (4.11), (4.12) and the obvious inequality

$$(4.13) \quad \int_{-\infty}^{[r]} g(t,M)\,dt \leq \int_{-\infty}^{r} g(t,M)\,dt \leq \int_{-\infty}^{[r]+1} g(t,M)\,dt$$

it is easily verified that for $M \geq 0$

$$(4.14) \qquad \sup_{-\infty < r < \infty} \left| \int_{-\infty}^{r} g(t,M)\,dt - \sum_{0}^{[r]} b_j \right| < b_m \ .$$

Since any $M$ satisfying (4.10) is necessarily positive, a $\overset{\circ}{b}_N(k,z)$ which satisfies the conclusion of the lemma is obtainable from (4.9) with $M = \max\,(\ln b_{L+1} - \ln b_L,\ \ln b_{R-1} - \ln b_R)$.

Lemma 6.

If $\overset{\circ}{b}_N(k,z)$ is any function on $-\infty < k < \infty$ which satisfies the conclusion of Lemma 5, then for every $N,z$

$$s_N(z)\ \overset{\circ}{b}_N(t^{-*},z) \qquad -\infty < t < \infty$$

is continuous, non-negative, and $\ln$-concave.

Proof.

The only non-obvious part of the conclusion is the $\ln$-concavity. For any $0 < c < 1$, $-\infty < t_1, t_2 < \infty$

$$\ln s\cdot b^{\circ}(s[ct_1+(1-c)t_2]+\Sigma) = \ln s+\ln b^{\circ}(c[st_1+\Sigma]+(1-c)[st_2+\Sigma])$$

and by the concavity of $\ln b^{\circ}(k)$ it follows that

$$\ln b^{\circ}(c[st_1+\Sigma]+(1-c)[st_2+\Sigma]) \geq c \ln b^{\circ}(st_1+\Sigma)+(1-c)\ln b^{\circ}(st_2+\Sigma).$$

From these relations we obtain the desired result:

$$\ln s\cdot b^{\circ}(s[ct_1+(1-c)t_2]+\Sigma) \geq c \ln sb^{\circ}(st_1+\Sigma)+(1-c) \ln b^{\circ}(st_2+\Sigma) .$$

Lemma 7.

If $b^{\circ}_N(k,z)$ is any function on $-\infty < k < \infty$ which satisfies the conclusion of Lemma 5

$$\sup_t \left| \int_{-\infty}^{t} s_N(z)\ b^{\circ}_N(v^{-*},z)dv - \underline{\Phi}(t) \right| < \frac{\phi(0)+3c}{s_N(z)}$$

where the constant c is the c of Lemma 3.

Proof.

By the triangle inequality and a change of variable of integration we obtain that

$$\left| \int_{-\infty}^{t} s_N(z)b^{\circ}_N(v^{-*},z)dv - \underline{\Phi}(t) \right| \leq$$

$$\left| \int_{-\infty}^{t^{-*}} \overset{\circ}{b}_N(k,z)dk - \sum_{0}^{[t^{-*}]} b_N(k,z) \right| + \left| \sum_{0}^{[t^{-*}]} b_N(k,z) - \Phi(t) \right| .$$

By Lemma 5 the first term on the right $< \max_k b_N(k,z)$ and it follows from Lemma 3 that the sum on the right is less than

$$\max_k b_N(k,z) + \frac{c}{s_N(z)} < \sup_t [\Phi(t^*) - \Phi((t-1)^*)] + \frac{3c}{s_N(z)} < \frac{\phi(0)+3c}{s_N(z)}$$

and this completes the proof.

## 5. A DCLT for Ln-Concave Functions.

**Theorem 3.**

Let C be the class of all functions f(t) such that f(t) is defined on $-\infty < t < \infty$ and is continuous, non-negative and $\ln$-concave on that domain. For any f in C let D(f) =

$\sup_t \left| \int_{-\infty}^{t} f(v)dv - \Phi(t) \right|$ . For any $\epsilon > 0$, there exists $\delta = \delta(\epsilon)$ such that

$$D(f) < \delta(\epsilon) \text{ implies } \sup_t |f(t) - \phi(t)| < \epsilon$$

uniformly for all f in the class C.

Proof.

We shall restrict our attention here and throughout the proof to the subclass of C for which $D(f) < \frac{1}{2}$.

It follows from $\int_{-\infty}^{\infty} f(v)dv > \frac{1}{2}$ that $\sup f(t) > 0$. From the $\ln$-concavity of f we obtain that $[t \mid f(t) \geq \frac{1}{2} \sup f(t)]$ is an interval, while from the continuity of f and $\int_{-\infty}^{\infty} f(v)dv < 3/2$ it follows that this interval is closed and bounded. We deduce from this result that $M(f) = [t \mid f(t) = \max]$ is non-empty and, by the $\ln$-concavity and continuity of f, is a closed interval, say $[a(f), b(f)]$.

If $a(f) \leq t_1 < t_2$ we have from $\ln$-concavity

$$\ln f(t_1) \geq \frac{t_2-t_1}{t_2-a(f)} \ln f(a(f)) + \frac{t_1-a(f)}{t_2-a(f)} \ln f(t_2)$$

and since $\ln f(a(f))$ is maximal it follows that

$$(5.1) \qquad f(t_1) \geq f(t_2) \qquad a(f) \leq t_1 < t_2.$$

We similarly obtain that

$$(5.2) \qquad f(t_1) \leq f(t_2) \qquad t_1 < t_2 \leq b(f).$$

If $2D(f) < \delta$ and $b(f) \geq 3\eta$ for any choice of positive $\delta$ and positive $\eta$, it follows from (5.1), (5.2), and the observation that $\phi$ is also in C with $M(\phi) = [0,0]$, that

$$\eta\, \phi(\eta)-\delta < \int_0^{\eta} \phi(v)\,dv-\delta < \int_0^{\eta} f(v)\,dv \le \eta f(\eta) \le$$

(5.3)

$$\eta f(2\eta) \le \int_{2\eta}^{3\eta} f(v)\,dv < \int_{2\eta}^{3\eta} \phi(v)\,dv + \delta < \eta\, \phi(2\eta) + \delta.$$

Since the choice of $\delta = \eta[\phi(\eta) - \phi(2\eta)]/2$ contradicts the above, we conclude that $4D(f) < \eta\,[\phi(\eta) - \phi(2(\eta))]$ implies $b(f) < 3\eta$. A similar argument leads to the conclusion that if $4D(f) < \eta[\phi(-\eta) - \phi(-2\eta)]$ then $a(f) > -3\eta$. Letting

$$(5.4) \qquad \delta_M(\epsilon) = \begin{cases} \frac{\epsilon}{12}\,[\phi(\frac{\epsilon}{3}) - \phi(\frac{2\epsilon}{3})] & \epsilon \le 1 \\ \frac{1}{12}\,[\phi(\frac{1}{3}) - \phi(\frac{2}{3})] & \epsilon > 1, \end{cases}$$

it follows that

$$-\epsilon < a(f) \le b(f) < \epsilon$$

uniformly for all f in C such that $D(f) < \delta_M(\epsilon)$.

It is easily seen that $\phi(t)$ satisfies a Lipschitz condition on $-\infty < t < \infty$. Let L be such that $|\phi(t_1) - \phi(t_2)| < L\,|t_1-t_2|$. Let $\epsilon$,t be arbitrary positive numbers and let $\delta = \min[\epsilon/2L, t/2]$.

For any f in C such that $D(f) < \min[\delta\epsilon/4, \delta_M(t/2)]$ we obtain

$$\delta\phi(t)-\delta\epsilon < \delta\phi(t+\delta)-\frac{\delta\epsilon}{2} < \int_t^{t+\delta}\phi(v)dv-\frac{\delta\epsilon}{2} < \int_t^{t+\delta} f(v)dv \leq \delta f(t)$$

and

$$\delta f(t-\delta) \leq \int_{t-\delta}^{t} f(v)dv < \int_{t-\delta}^{t}\phi(v)dv+\frac{\delta\epsilon}{2} < \delta\phi(t-\delta)+\frac{\delta\epsilon}{2} < \delta\phi(t)+\delta\epsilon$$

whence

$$\phi(t)-\epsilon < f(t) \leq f(t-\delta) < \phi(t)+\epsilon.$$

Since $t/2$ and $\delta_M(t/2)$ are non-decreasing functions of t on $(0,\infty)$ we obtain for any $\epsilon > 0$, $h > 0$ that if $D(f) < \min[\frac{\epsilon^2}{8L}, \frac{\epsilon h}{8}, \delta_M(h/2)]$ then $|f(t)-\phi(t)| < \epsilon$ uniformly for all $t \geq h$. As is probably clear from the symmetry of $\phi(t)$ we can by a similar argument obtain the same conclusion for $t \leq -h$. Letting

$$(5.5) \qquad \delta_h(\epsilon) = \min[\frac{\epsilon^2}{8L}, \frac{\epsilon h}{8}, \delta_M(h/2)] \qquad \epsilon > 0,\ h > 0$$

it follows that

$$(5.6) \qquad \sup_{|t|\geq h} |f(t)-\phi(t)| < \epsilon$$

uniformly for all f in C such that $D(f) < \delta_h(\epsilon)$.

If $|t| < h$ and f is in C

$$\ln f(t) \geq \frac{h-t}{2h}\ln f(-h) + \frac{t+h}{2h}\ln f(h)$$

and if $D(f) < \delta_h(\varepsilon/2)$

$$\ln f(t) \geq \frac{h-t}{2h} \ln[\phi(-h) - \frac{\varepsilon}{2}] + \frac{t-h}{2h} \ln[\phi(h) - \frac{\varepsilon}{2}] = \ln[\phi(h) - \frac{\varepsilon}{2}]$$

whence $\qquad f(t) \geq \phi(h) - \frac{\varepsilon}{2}$

and $\qquad f(t) - \phi(t) \geq \phi(h) - \phi(t) - \frac{\varepsilon}{2} > -\varepsilon$ uniformly $|t| < h$

provided $h < \varepsilon/4L$.

If $0 \leq t \leq h$, $0 < \eta$ and $D(f) < \delta_h(\eta)$

$$\ln f(h) \geq \frac{h}{2h-t} \ln f(t) + \frac{h-t}{2h-t} \ln f(2h)$$

and hence

$$\begin{aligned}
\ln f(t) &\leq \frac{2h-t}{h} \ln f(h) - \frac{h-t}{h} \ln f(2h) \\
&\leq \frac{2h-t}{h} \ln [\phi(h)+\eta] - \frac{h-t}{h} \ln [\phi(2h)-\eta] \\
&\leq \ln[\phi(h)+\eta] + \frac{h-t}{h} \left\{ \ln[\phi(h)+\eta] - \ln[\phi(2h)-\eta] \right\} \\
&\leq \ln[\phi(h)+\eta] + \left\{ \ln[\phi(h)+\eta] - \ln[\phi(2h)-\eta] \right\}
\end{aligned}$$

uniformly $0 \leq t \leq h$. If $\eta + 2Lh < \phi(0)$ we have

$$\frac{\phi(h)+\eta}{\phi(2h)-\eta} - 1 = \frac{\phi(h)-\phi(2h)+2\eta}{\phi(2h)-\eta} \leq \frac{Lh+2\eta}{\phi(0)-2Lh-\eta}$$

and the inequality $\ln(1+x) < x \quad x \neq 0$ yields

$$\ln f(t) < \ln [\phi(h)+\eta] + \frac{Lh+2\eta}{\phi(0)-2Lh-\eta} \qquad 0 \leq t \leq h$$

or equivalently

$$f(t) < [\phi(h)+\eta] \cdot \exp\left[\frac{Lh+2\eta}{\phi(0)-2Lh-\eta}\right] \qquad 0 \leq t \leq h$$

whence, if $3Lh + 3\eta < \phi(0)$ we obtain from the inequality $e^x < \frac{1}{1-x}$, $x < 1$

$$f(t) < [\phi(h)+\eta] \cdot \left[\frac{\phi(0)-2Lh-\eta}{\phi(0)-3Lh-3\eta}\right] \qquad 0 \leq t \leq h$$

and

$$\begin{aligned} f(t)-\phi(t) &< \left[\phi(h) \frac{\phi(0)-2Lh-\eta}{\phi(0)-3Lh-3\eta} - \phi(t)\right] + \eta \frac{\phi(0)}{\phi(0)-3Lh-3\eta} \\ &< \phi(h) \frac{Lh+2\eta}{\phi(0)-3Lh-3\eta} + \eta \frac{\phi(0)}{\phi(0)-3Lh-3\eta} \\ &< \frac{\phi(0)(Lh+3\eta)}{\phi(0)-3Lh-3\eta} \qquad 0 \leq t \leq h . \end{aligned}$$

Replacing $\eta$ by $\varepsilon/12$ we have that if $6Lh + \varepsilon/2 \leq \phi(0)$ our previous conditions of this type will be satisfied, and for any h such $2Lh < \varepsilon/2$

$$(5.7) \qquad f(t) - \phi(t) < \varepsilon \qquad 0 \leq t \leq h$$

uniformly for all f in C such that $D(f) < \delta_h(\varepsilon/12)$.

A similar analysis could be applied to yield the same result for the t-interval $[-h,0]$ and the combination of this with the results of the two preceding paragraphs yields for any $\varepsilon > 0$, $h > 0$ such that $4Lh < \varepsilon$ and $6Lh + \varepsilon/2 < \phi(0)$

$$(5.8) \qquad \sup_{|t|\leq h} |f(t) - \phi(t)| < \varepsilon$$

uniformly for all f in C such that $D(f) < \delta_h(\varepsilon/12)$.

With $h'(\varepsilon) = \varepsilon/5L$ we define

$$(5.9) \qquad \delta(\varepsilon) = \begin{cases} \delta_{h'(\varepsilon)}(\varepsilon/12) & 0 < \varepsilon \leq 4\phi(0)/7 \\ \delta(4\phi(0)/7) & 4\phi(0)/7 < \varepsilon \end{cases}$$

and since $\delta(\varepsilon) \leq \delta_{h'(\varepsilon)}(\varepsilon/2)$ the proof of Theorem 3 is complete with this choice of $\delta(\varepsilon)$.

Examination of $\delta_{h'(\varepsilon)}(\varepsilon/12)$ reveals that $\delta(\varepsilon)$ is expressible

$$(5.10) \quad \delta(\varepsilon) = \begin{cases} \frac{\varepsilon}{120L}\left[\phi(\frac{\varepsilon}{30L})-\phi(\frac{2\varepsilon}{30L})\right] & 0 < \varepsilon \leq 4\phi(0)/7 \\ \delta(4\phi(0)/7 & 4\phi(0)/7 \leq \varepsilon \ . \end{cases}$$

There would be considerable interest in securing a larger $\delta(\varepsilon)$ which would satisfy Theorem 3 and this would seem readily possible since our construction had little concern with the attainment of arithmetically optimum results.

## 6. A DCLT for GB: Completion of Proof.

We complete the proof of Theorem 2 by noting that Lemmas 5, 6, and 7 imply there exists $\mathring{b}_N(k,z)$ such that

$$s_N(z)\mathring{b}_N(t^{-*},z) \text{ is in } C$$

and

$$D(s_N(z)\mathring{b}_N(t^{-*},z)) < (\phi(0)+3c)/s_N(z).$$

Thus, if $s_N(z) > (\phi(0)+3c)/\delta(\varepsilon)$ Theorem 3 implies

$$\sup_{-\infty<t<\infty} \left| s_N(z)\mathring{b}_N(t,z)-\phi(t^*) \right| = \sup_{-\infty<t<\infty} \left| s_N(z)\mathring{b}_N(t^{-*},z)-\phi(t) \right| < \varepsilon$$

and hence

$$\max_k \left| s_N(z)b_N(k,z) - \phi(k^*) \right| < \varepsilon.$$

This finishes the proof of Theorem 2 and displays an $M(\varepsilon) = (\phi(0)+3c)/\delta(\varepsilon)$.

CHAPTER III.

EXACT AND ASYMPTOTIC DISTRIBUTIONS

ASSOCIATED WITH A KOLMOGOROFF STATISTIC.

1. Definitions.

Let $X_1, X_2, \ldots$ be a sequence of independent rectangular random variables on (0, 1),

$S_n(x)$, the empirical d. f. defined by the set $X_1, X_2, \ldots, X_n$,

$D_n^+ = \sup_x [ S_n(x) - F(x) ]$,

$P_n(c) = \Pr [ D_n^+ \leq \frac{c}{n} ]$,

$X^{(k)}$, the k-th order statistic of the set $X_1, \ldots, X_n$,

2. Expressions for $P_n(c)$, c an Integer Between 0 and n.

$$
\begin{aligned}
P_n(c) &= \Pr [ S_n(x) \leq x + \tfrac{c}{n} \text{ for all } x \text{ in } (0, 1) ] \\
&= \Pr [ S_n(X^{(k)}) \leq X^{(k)} + \tfrac{c}{n} \quad k = 1, 2, \ldots, n ]
\end{aligned}
\tag{2.1}
$$

since $S_n(x) - x$ is monotone decreasing in x except for the points $X^{(k)}$. Equivalently,

$$P_n(c) = Pr\left[ X^{(k)} > \frac{k-c}{n} \qquad k = 1, 2, \ldots, n \right]$$

$$= n! \int_{\frac{n-c}{n}}^{1} \int_{\frac{n-c-1}{n}}^{x_n} \cdots \int_{\frac{1}{n}}^{x_{c+2}} \int_{0}^{x_{c+1}} \cdots \int_{0}^{x_2}$$

(2.2)

$$dx_1 \ldots dx_c \, dx_{c+1} \ldots dx_{n-1} \, dx_n$$

$$= n! \int_{\frac{n-c}{n}}^{1} \int_{\frac{n-c-1}{n}}^{x_n} \cdots \int_{\frac{1}{n}}^{x_{c+2}} \frac{(x_{c+1})^c}{c!} \, dx_{c+1} \ldots dx_{n-1} dx_n .$$

Transforming by

$$x_{c+k} - \frac{k}{n} = y_k \qquad k = 1, 2, \ldots, n-c,$$

we can express $P_n(c)$ as

$$(2.3) \frac{n!}{c!} \int_0^{\frac{c}{n}} \int_0^{y_{n-c}+\frac{1}{n}} \cdots \int_0^{y_3+\frac{1}{n}} \int_0^{y_2+\frac{1}{n}} (y_1 + \frac{1}{n})^c dy_1 dy_2 \ldots dy_{n-c-1} dy_{n-c} .$$

Define

$$R(k,a,t) = k! \int_0^t \int_0^{z_k+a} \cdots \int_0^{z_2+a} dz_1 \ldots dz_{k-1} dz_k \qquad k = 1, 2, \ldots$$

$$R(0,a,t) = 1.$$

We obtain from (2.3)

$$P_n(c) = \frac{n!}{(c+1)!} \int_0^{\frac{c}{n}} \cdots \int_0^{y_3+\frac{1}{n}} (y_2+\frac{2}{n})^{c+1} dy_2 \cdots dy_{n-c}$$

$$- \binom{n}{c+1}(\frac{1}{n})^{c+1} R(n-c-1,\frac{1}{n},\frac{c}{n}),$$

$$= \frac{n!}{(c+2)!} \int_0^{\frac{c}{n}} \cdots \int_0^{y_4+\frac{1}{n}} (y_3+\frac{3}{n})^{c+2} dy_3 \cdots dy_{n-c}$$

$$- \sum_{j=1}^{2} \binom{n}{c+j}(\frac{j}{n})^{c+j} R(n-c-j,\frac{1}{n},\frac{c}{n}),$$

and by an obvious induction

$$(2.4) \quad P_n(c) = 1 - \sum_{j=1}^{n-c} \binom{n}{c+j}(\frac{j}{n})^{c+j} R(n-c-j,\frac{1}{n},\frac{c}{n}) .$$

In order to evaluate the R-function involved in (2.4) we introduce the following lemma.

Lemma.

$$R(k,a,t) = t(t + ka)^{k-1} \qquad k = 0, 1, \ldots$$

Proof:

Obviously true for $k = 0, 1$ and assuming for $k = r-1$

$$R(r,a,t) = r \int_0^t (z_r+a)(z_r+ra)^{r-2} dz_r$$

$$= \left[ \frac{r(z+a)(z+ra)^{r-1}}{r-1} - \frac{(z+ra)^r}{r-1} \right]_0^t = t(t+ra)^{r-1} .$$

Using this result in(2.4) we obtain an expression for $P_n(c) =$ $P\left[D_n^+ \leq \frac{c}{n}\right]$ (with c integral),

$$P_n(c) = 1 - \sum_{k=1}^{n-c} \frac{c}{n} \binom{n}{k+c} \left(\frac{k}{n}\right)^{k+c} \left(1 - \frac{k}{n}\right)^{n-k-c-1} ,$$

(2.5)

$$= 1 - \sum_{k=1}^{n-c} \frac{c}{n-k} B(n, k+c; \frac{k}{n}),$$

where $B(n,r;\, p) = \binom{n}{r} p^r (1-p)^{n-r}$.

We can obtain an alternative expression for $P_n(c)$ by first expressing $(y_1 + \frac{1}{n})^c$ in terms of R-functions, then evaluating (2.3) directly. For it can easily be shown that

$$(y_1 + \frac{1}{n})^c = \sum_{j=1}^{c} \binom{c}{c-j} \left(\frac{-j}{n}\right)^{c-j} R(k, \frac{1}{n}, y_1 + \frac{1}{n}),$$

whence, inserting in (2.3) we obtain

$$P_n(c) = \sum_{k=1}^{c} \binom{n}{c-k} \left(\frac{-k}{n}\right)^{c-k} R(n-c+k, \frac{1}{n}, \frac{c}{n})$$

(2.6)

$$= \sum_{k=1}^{c} \frac{c}{n+k} B(n, -k+c; \frac{-k}{n}).$$

Comparison of the expressions for $P_n(c)$, (2.5) and (2.6), leads to the identity

$$\sum_{k=-c}^{k=n-c} \frac{c}{n-k} B(n, k+c; \frac{k}{n}) = 1 \qquad c \neq 0.$$

3. Expression for $P_n(c)$, c Any Number Between 0 and n.

Let $\bar{c}$ be the greatest integer less than or equal to c. By a slight modification of the derivation of (2.5) we obtain

$$(3.1) \qquad P_n(c) = 1 - \sum_{k=1}^{n-\bar{c}} \frac{c}{n-k+(c-\bar{c})} B(n, k+\bar{c}; \frac{k-(c-\bar{c})}{n}).$$

4. The Asymptotic Distribution of $n^{\frac{1}{2}} D_n^+$.

Theorem 4.1.

Given any sequence $z_n$ such that

(a) $\lim_{n=\infty} z_n = z > 0$ and $|z_n - z| < Kn^{-\frac{1}{2}}$

(b) $n^{\frac{1}{2}} z_n$ is integral for each n

then

$$\lim_{n = \infty} P_n(n^{\frac{1}{2}} z_n) = 1 - e^{-2z^2} .$$

<u>Theorem</u> 4.2.

For any $z \geq 0$ $\lim_{n=\infty} P_n(n^{\frac{1}{2}} z) = 1 - e^{-2z^2}$ .

<u>Proof of Theorem</u> 4.2.

Result is trivial $z = 0$, and $z > 0$ the sequences with respective n-th terms

$$n^{-\frac{1}{2}}(\overline{n^{\frac{1}{2}} z}), \ n^{-\frac{1}{2}}(\overline{n^{\frac{1}{2}} z + 1})$$

satisfy the hypothesis of Theorem 4.1 and

$$P_n(\overline{n^{\frac{1}{2}} z}) \leq P_n(n^{\frac{1}{2}} z) < P_n(\overline{n^{\frac{1}{2}} z + 1}).$$

<u>Proof of Theorem</u> 4.1.

Throughout the remainder of Section 4 we will consider $c = n^{\frac{1}{2}} z_n$. From (2.5) we obtain

$$P_n(n^{\frac{1}{2}} z_n) = 1 - \sum_{k=1}^{n-c} \frac{nc}{n-k} B(n,k+c; \frac{k}{n}) \frac{1}{n} . \tag{4.1}$$

It can easily be shown via Stirling relation and the usual $\ell n$ expansion that

$$\lim_{n=\infty} f_n(t) =$$

(4.2) $$[2\pi t(1-t)^3]^{-\frac{1}{2}} z \exp[-z^2/2t(1-t)] \text{ uniformly } 0 < a \leq t \leq b < 1$$

where

$$f_n(\tfrac{k}{n}) = \frac{nc}{n-k} B(n,k+c;\tfrac{k}{n}) \ .$$

Hence

$$\lim_{\substack{a=0\\b=1}} \lim_{n=\infty} \sum_{k=na}^{k=nb} f_n(\tfrac{k}{n}) \frac{1}{n} =$$

(4.3) $$\int_0^1 [2\pi t(1-t)^3]^{-\frac{1}{2}} z \exp[-z^2/2t(1-t)] dt = L(z).$$

Transforming $L(z)$ by $t = \sin^2 \frac{\omega}{2}$ and performing some elementary reductions we obtain

$$L(z) = e^{-2z^2} \ .$$

It will be sufficient to complete the proof to show

(4.4) $$\lim_{a=0} \overline{\lim_{n=\infty}} \sum_{k=1}^{na} f_n(\tfrac{k}{n})\frac{1}{n} = \lim_{b=1} \overline{\lim_{n=\infty}} \sum_{k=nb}^{n-c} f_n(\tfrac{k}{n})\frac{1}{n} = 0.$$

For this purpose consider $f_n(\frac{k}{n})$ in the form

$$\frac{nc}{n-k} B(n,k;\ \frac{k}{n}) \cdot \prod_{j=1}^{c} \frac{(1 - \frac{j-1}{n-k})}{(1 + \frac{j}{k})} \quad .$$

From the Stirling relation we obtain an upper bound for the first factor

$$(4.5) \qquad \frac{nc}{n-k}\ B(n,k;\ \frac{k}{n}) < \left[2\pi(\frac{k}{n})(1-\frac{k}{n})^3\right]^{-\frac{1}{2}} z_n e^{1/12},$$

while using the inequality

$$\frac{r}{s+r} < \ln(1+\frac{r}{s}) < \frac{r}{s} \qquad \frac{r}{s} > -1, \neq 0,$$

we obtain

$$\ln \prod_{j=1}^{c} \frac{(1 - \frac{j-1}{n-k})}{(1 + \frac{j}{k})} < - \sum_{j=1}^{c} \left[\frac{j-1}{n-k} + \frac{j}{k+j}\right] \leq$$

$$- \sum_{j=1}^{c} \left[\frac{j-1}{n-k} + \frac{j}{k+c}\right] = - \frac{c(c-1)}{2(n-k)} - \frac{c(c+1)}{2(k+c)} \quad .$$

But, since $1 \leq k \leq n-c$ and c is (as throughout this section) $n^{\frac{1}{2}} z_n$

$$(4.6) \qquad \ln \prod_{j=1}^{c} \frac{(1 - \frac{j-1}{n-k})}{(1 + \frac{j}{k})} < \frac{1}{2} - \frac{c^2}{2(n-k)} - \frac{c^2}{2(k+c)} \leq \frac{1}{2} - \frac{c^2}{2(n-k)} - \frac{z_n^2}{2} \quad .$$

(4.5) and (4.6) yield

$$\sum_{k=p_1}^{k=p_2} f_n\left(\frac{k}{n}\right) \frac{1}{n} < (2\pi)^{-\frac{1}{2}} z_n \exp\left[ \frac{1}{12} + \frac{1}{2} - \frac{z_n^2}{2} \right]$$

$$\cdot \sum_{k=p_1}^{k=p_2} \left(\frac{k}{n}\right)^{-\frac{1}{2}} \left(1- \frac{k}{n}\right)^{-\frac{3}{2}} \exp\left[-z_n^2/2\left(1-\frac{k}{n}\right) \right] \cdot \frac{1}{n} \quad .$$

Thus

$$\sum_{k=1}^{na} f_n\left(\frac{k}{n}\right)\frac{1}{n} < M(z_n) \sum_{t=1/n}^{t=a} t^{-\frac{1}{2}} \exp\left[-z_n^2/2(1-t)\right] \cdot \frac{1}{n} \, (1-t)^{-\frac{3}{2}} \, .$$

To use (a) of the hypothesis we note that $|z_n - z| < Kn^{-\frac{1}{2}}$ implies

that $z_n > z - Kn^{-\frac{1}{2}}$ and therefore

$$-z_n^2 < -z^2 + 2Kn^{-\frac{1}{2}} \, .$$

Hence

$$\sum_{k=1}^{na} f_n\left(\frac{k}{n}\right)\frac{1}{n} < M(z_n) e^{-\frac{Kn^{-\frac{1}{2}} z}{a}} \sum_{t=\frac{1}{n}}^{t=a} t^{-\frac{1}{2}} (1-t)^{-\frac{3}{2}} e^{-\frac{z^2}{2(1-t)}} \cdot \frac{1}{n}$$

and

$$\lim_{a=0} \overline{\lim_{n=\infty}} \sum_{k=1}^{na} f_n\left(\frac{k}{n}\right)\frac{1}{n}$$

$$\leq M(z) \lim_{a=0} \int_0^a t^{-1/2} (1-t)^{-3/2} e^{-\frac{z^2}{2(1-t)}} \, dt = 0.$$

Similarly

$$\sum_{k=nb}^{n-c} f_n(\frac{k}{n})\frac{1}{n} < M(z_n)\ e^{\frac{Kz}{z_n}} \sum_{t=b}^{t=1-\frac{1}{n}} t^{-1/2}(1-t)^{-3/2}\ e^{-\frac{z^2}{2(1-t)}}\ \frac{1}{n}$$

and

$$\lim_{b=1}\ \overline{\lim_{n=\infty}} \sum_{k=nb}^{n-c} f_n(\frac{k}{n})\frac{1}{n}$$

$$\leq M(z)e^K \lim_{b=1} \int_b^1 t^{-1/2}(1-t)^{-3/2}\ e^{-\frac{z^2}{2(1-t)}}\ dt = 0,$$

thus completing the proof of Theorem 4.1 by proving (4.4).

## 5. Some Associated Probabilities.

We can use the machinery of Section 2 with some slight modifications to evaluate some similar probabilities. Thus we find

$$\Pr[S_n(x) \leq \lambda x \quad \text{for all } x \text{ in } (0,\ 1)] \qquad \lambda \geq 1$$

$$= \Pr[S_n(X^{(k)}) \leq \lambda X^{(k)}\ \ k = 1,\ \ldots,\ n]$$

$$(5.1) \qquad = \Pr[X^{(k)} > \frac{k}{n\lambda}\ \ k = 1,\ \ldots,\ n]$$

$$= \Pr[X^{(k)} - \frac{k}{n\lambda} > 0\ \ k = 1,\ \ldots,\ n]$$

$$= R(n, \frac{1}{n\lambda},\ 1 - \frac{1}{\lambda}) = 1 - \frac{1}{\lambda}\ .$$

If $\lambda \geq 1$ and $c$ is integral such that $0 < c < n$,

$$\Pr\left[ S_n(x) \leq \lambda x + \frac{c}{n} \text{ for all } x \text{ in } (0, 1) \right]$$

(5.2)

$$= 1 - \sum_{k=1}^{n-c} \frac{n(\lambda-1)+c}{n-k} B(n, k+c; \frac{k}{n\lambda}) \quad .$$

For $\lambda = 1$ this reduces to (2.5).

BIBLIOGRAPHY


1. Berry, A. C., "The Accuracy of the Gaussian Approximation to the Sum of Independent Variates," Transactions of the American Mathematical Society, 49 (1941), 122-136.

2. Birnbaum, Z. W. and Tingey, F. H., "One-Sided Confidence Contours for Probability Distribution Functions," Annals of Mathematical Statistics, 22 (1951), 592-596.

3. Frechet, M., Generalites sur les probabilites; variables aleatoires, Paris, Gauthier-Villars, 1937.

4. Halmos, P. R., Measure Theory, New York, D. Van Nostrand Company, 1950.

5. Halmos, P. R. and Savage, L. J., "Application of the Radon-Nikodym Theorem to the Theory of Sufficient Statistics," Annals of Mathematical Statistics, 20 (1949), 225-242.

6. Hammersley, J. M., "On Estimating Restricted Parameters," Journal Royal Statistical Society, B, 12 (1950), 192-240.

7. Hammersley, J. M., "Tauberian Theory for the Asymptotic Forms of Statistical Frequency Functions," Proceedings of the Cambridge Philosophical Society, 48 (1952), 592-599.

8. Hannan, J. F., "Some Tests Based on the Empirical Distribution Function. (Preliminary Report).," (abstract), Annals of Mathematical Statistics, 21 (1950), 312.

9. Hannan, J. F. and Robbins, H. E., "Asymptotic Solutions of the Compound Decision Problem for Two Completely Specified Populations," (unpublished paper), (1953).

10. Hardy, G. H., Littlewood, J. E. and Polya, G., Inequalities, Cambridge, Cambridge University Press, 1934.

11. Kolmogorov, A. N., "Sulla determinatione empirica di una leggi di distributione," Giornale dell' Istituto Italiano degli Attuari, 4 (1933), 83-91.

12. Robbins, H. E., "Asymptotically Subminimax Solutions of Compound Statistical Decision Problems," Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1951.

13. von Mises, R., Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und Theoretischen Physik, Leipzig, Deuticke, 1931.