The Annals of Mathematical Statistics
1969, Vol. 40, No. 2, 339–355

GENERALIZED MEANS AND ASSOCIATED FAMILIES OF DISTRIBUTIONS

BY H. K. BRØNS, H. D. BRUNK[1], W. E. FRANCK, AND D. L. HANSON[1]

University of Copenhagen and University of Missouri, Columbia

1. Introduction. The mean of a distribution and the sample mean are key concepts in the theory of statistics. The generalized means studied in this paper share important properties of the expectation, which is seen in this context as a distinguished member of a very large class.

There is an especially close relationship between the sample mean and the theory of estimation associated with one parameter exponential families of distributions. Each of the generalized means determines analogous one parameter families of distributions. Such families are introduced in Section 4. In sampling from such a distribution the maximum likelihood estimate of the generalized mean of the distribution is the generalized mean of the sample. Under appropriate regularity conditions it is a strongly consistent and asymptotically normal estimator.

In Section 2 the generalized means, called ϕ-means, and sample ϕ-means are defined and some of their properties examined. A minimizing property is proved, and they are shown to have the Cauchy mean value property. Also, an extension of Jensen's inequality is observed to be valid for r-means. Asymptotic properties of strong consistency and normality of sample ϕ-means are developed in Section 3. A study is made of conditions under which the sample ϕ-means are infinitely often, or all but finitely often, above or below the distribution ϕ-mean as sample size becomes infinite.

Guenther [6] has recently discussed briefly estimation of λ in sampling from the one parameter family of densities

$$f(x; \lambda) = \lambda x^{\lambda-1}, \qquad 0 < x < 1, \quad \lambda > 0. \tag{1.1}$$

The following observations relative to this example are, on the one hand, irrelevant from the point of view of the discussion in [6]; on the other hand, they do not at all indicate the scope of the present investigation. Nevertheless, they illustrate a partial motivation for considering means other than the usual sample mean. We note first that with changes of variable and of parameter, $Y = -\log X$, $\beta = 1/\lambda$, (1.1) becomes the ordinary exponential distribution. The sample mean is unbiased, sufficient, and efficient (in Cramér's finite sample sense); it is "the natural" estimate of β; the corresponding estimate of λ is then $n/(-\sum_i \log x_i)$. We observe further that (1.1) is a ϕ-family as defined in Section 4, with $\phi(x, \theta) = \log \theta - \log x$, $\theta(u) = \exp(-1/u)$; the ϕ-mean of

Received 26 February 1968.

[1] The work of these authors was partially sponsored by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR Grant No. AF-AFOSR-746-65A.

(1.1) is $\exp(-1/\lambda)$. The maximum likelihood estimate of $\exp(-1/\lambda)$ is the sample ϕ-mean, $\exp(\sum_i \log x_i/n)$ (the geometric mean of $x_1, \cdots, x_n$), yielding again the estimate $n/(-\sum_i \log x_i)$ for λ.

An account of these generalized means containing some of the results of the present paper appears in lecture notes [2] of one of the authors. A subclass of these means was investigated by Huber [7] in connection with a study of robust estimation. A subclass of these, the r-means, has been investigated by Gentleman [4]. Some properties of the r-means are implicit in work by Nikolski [9] and by Ando and Amemiya [1].

2. ϕ-means and some of their properties. Let R denote the set of real numbers, and let $\phi(\cdot, \cdot)$ be an extended real valued function on $R \times R$ such that:

(2.1) $\phi(x, \theta)$ is non-decreasing in θ for each x in R;

(2.2) $\phi(x, \theta) \geqq 0$ for $\theta > x$, and $\phi(x, \theta) \leqq 0$ for $\theta < x$;

(2.3) $\phi(x, \theta)$ is a Borel measurable function of x for each fixed θ in R.

Note that conditions (2.1) and (2.3) are almost enough to make ϕ jointly Borel measurable in x and θ.

If ν is a measure on the Borel subsets of the reals, we define $\lambda(\theta) = \lambda(\theta, \nu, \phi)$ by

$$\lambda(\theta) = \int \phi(x, \theta)\, d\nu(x) = \int \phi^+(x, \theta)\, d\nu(x) - \int \phi^-(x, \theta)\, d\nu(x)$$

if at least one of the summands on the right is finite; here we use the conventions $a^+ = \max\{a, 0\}$ and $a^- = \max\{-a, 0\}$ when a is in $[-\infty, \infty]$. We see that if $\lambda(\theta_0)$ is finite for some θ_0 then $\lambda(\theta)$ is defined (but not necessarily finite) for all θ, and if $\lambda(\theta_1)$ and $\lambda(\theta_2)$ are defined then $\lambda(\theta_1) \leqq \lambda(\theta_2)$ if $\theta_1 \leqq \theta_2$.

Definition 2.1. The real number μ is a ϕ-mean of ν if $\lambda(\theta) \leqq 0$ for all $\theta < \mu$ and $\lambda(\theta) \geqq 0$ for all $\theta > \mu$. The set of ϕ-means of ν will be denoted by $M(\nu)$ or $M(\nu, \phi)$. If ν is the probability measure generated by the distribution function F of the random variable X, then $M(F)$, $M(F, \phi)$, $M(X)$ or $M(X, \phi)$ will also be used to denote the set of ϕ-means. The notation $\bar{\mu} = \sup M(\nu)$ and $\underline{\mu} = \inf M(\nu)$ will be used.

By the support of a measure ν on the Borel subsets of R is meant the intersection of all closed subsets C of R such that $\nu(R - C) = 0$. We define the *interval of support* of ν to be the intersection of all closed intervals I such that $\nu(R - I) = 0$. If the interval of support of ν is a proper subset of R, then $\phi(x, \theta)$ really need not be defined outside the Cartesian product of the interval of support of ν with itself. (Some of our theorems would need minor modifications in this case.) If the interval of support (call it I) of ν is a proper subset of R, and if ϕ is defined on $I \times I$ so as to satisfy (2.1)–(2.3), then ϕ can be extended to $R \times R$ so as to preserve (2.1)–(2.3); all extensions give the same set of ϕ-means of ν. One way of extending ϕ is to let $\phi(x, \theta) = \infty$ if $\theta > x$ and (x, θ) is not in $I \times I$, $\phi(x, x) = 0$ if (x, x) is not in $I \times I$, and $\phi(x, \theta) = -\infty$ if $\theta < x$ and (x, θ) is not in $I \times I$.

If $\phi(x, \theta) = \theta - x$, the ϕ-mean coincides with the usual mean. If $0 < p \leqq$

$p + q = 1$, $\phi(x, \theta) = 1/p$ for $\theta > x$, $\phi(x, x) = 0$, and $\phi(x, \theta) = -1/q$ for $\theta < x$, then the p-mean is the p-fractile. Thus the ϕ-mean is the median if $\phi(x, \theta) = \operatorname{sgn}(\theta - x)$.

DEFINITION 2.2. If $r \geqq 0$ and $\phi(x, \theta) = |\theta - x|^r \operatorname{sgn}(\theta - x)$ when $\theta \neq x$ and $\phi(x, x) = 0$, then the ϕ-mean will be called the r-mean and denoted by $M_r(\nu)$, $M_r(X)$, or $M_r(F)$.

If $g(t)$ is non-decreasing, $g(t) > 0$ for $t > 0$, $g(t) < 0$ for $t < 0$, and $\phi(x, \theta) = g(\theta - x)$, then the interval of ϕ-means of a distribution function F is, in a sense, a location parameter of the distribution. If ψ is increasing then $\phi_1(x, \theta) = \phi[\psi(x), \psi(\theta)]$ satisfied (2.1)–(2.3) as does $\phi_2(x, \theta) = -\phi[\psi(x), \psi(\theta)]$ if ψ is decreasing. Many such means, with $\phi(x, \theta) = \theta - x$ are discussed by Gini [5]. If $h(x)$ is positive and Borel measurable, then $\phi^*(x, \theta) = h(x)\phi(x, \theta)$ again satisfies (2.1)–(2.3). If $x, \theta > 0$ and $\phi(x, \theta) = \theta - x$, one gets the harmonic mean if $h(x) = 1/x$ and the anti-harmonic mean if $h(x) = x$. (Note that $\phi(x) = -1/\theta + 1/x$ also gives the harmonic mean.) One has the "valore divisorio" ([5], page 110) if $x, \theta > 0$, $h(x) = x$, and $\phi(x, \theta) = \operatorname{sgn}(\theta - x)$. The ϕ-mean is the geometric mean if $\phi(x, \theta) = \log \theta - \log x$ for $x, \theta > 0$.

In the following theorem we list some obvious properties of ϕ-means and some properties which follow in a straightforward manner from the definitions of ϕ-mean and $\lambda(\theta)$ and from properties (2.1)–(2.3) of ϕ.

THEOREM 2.1.

(2.4) *If* $\alpha > 0$ *then* $M(\nu) = M(\alpha\nu)$.

(2.5) *If* $\lambda(\mu) = 0$ *then* μ *is in* $M(\nu)$.

(2.6) *If* $M^*(\nu)$ *is the interior of* $M(\nu)$ *and* θ *is in* $M^*(\nu)$ *then* $\lambda(\theta) = 0$.

(2.7) *The set* $M(\nu)$ *is a closed interval (possibly empty).*

(2.8) $M(\nu)$ *is empty if and only if either*

a) $\lambda(\theta)$ *is defined, non-zero, and of constant sign for all* θ

or

b) $\lambda(\theta)$ *is undefined for at least two values of* θ.

(2.9) *Condition* (2.8a) *is impossible if the support of* ν *is bounded.*

(2.10) *If there is exactly one point* μ *such that* $\lambda(\mu)$ *is undefined, then* $M(\nu) = \{\mu\}$.

(2.11) *If for all* θ *and* θ' *with* $\theta < \theta'$ *we have* $\nu\{x: \phi(x, \theta) < \phi(x, \theta')\} > 0$ *then* $M(\nu)$ *consists of at most one point.*

(2.12) *For* θ *in* $M^*(\nu)$, $\phi(x, \theta)$ *is constant in* θ *for almost all* x *(with respect to* ν*).*

(2.13) *If* ν *is* σ*-finite and if the sets* $\{x: \phi(x, \cdot)$ *is discontinuous at* $\theta\}$ *are disjoint for distinct* θ, *then except for a countable collection of* θ*'s,* $\lambda(\theta)$ *is independent of the definition of* $\phi(x, \cdot)$ *at points of discontinuity; thus the set* $M(\nu)$ *is independent of the definition of* $\phi(x, \cdot)$ *at all points of discontinuity. In particular, the last conclusion holds if the only discontinuity (as in the case of the median) is at* $\theta = x$.

(2.14) *If* $\phi^*(x, \theta) = \phi(x, \theta)$ *for* $x \neq \theta$ *and if* $\phi(x, x-) \leqq \phi^*(x, x) \leqq \phi(x, x+)$ *for all* x, *then* $M(\nu, \phi^*) = M(\nu, \phi)$.

COROLLARY 2.1. *If ν is a finite measure with interval of support I, and if ϕ is bounded on $I \times I$, then $M(\nu) \neq 0$.*

DEFINITION 2.3. The real number μ is a ϕ-mean of the finite collection $S = \{x_1, \cdots, x_n\}$ of real numbers (not necessarily distinct) if it is the ϕ-mean of the measure ν such that

$$\nu(B) = \sum_{k=1}^{n} I_B(x_k) \quad \text{where} \quad I_B(x) = 1 \quad \text{if } x \text{ is in } B$$
$$= 0 \quad \text{otherwise.}$$

Equivalently, μ is a ϕ-mean of S if

$$\lambda(\theta) = \sum_{k=1}^{n} \phi(x_k, \theta) \leqq 0 \quad \text{for all } \theta < \mu \tag{2.15}$$

and

$$\lambda(\theta) = \sum_{k=1}^{n} \phi(x_k, \theta) \geqq 0 \quad \text{for all } \theta > \mu. \tag{2.16}$$

We let $M(S)$ denote the collection of ϕ-means of S. If $x_1, \cdots, x_n$ are sample values of a sample of size n from a distribution F, then because of (2.4) μ is also a ϕ-mean of the probability measure corresponding to the sample distribution function F_n (i.e. μ is in $M(F_n)$), and μ will be called a sample ϕ-mean.

We note that

(2.17) if ϕ is finite valued then every finite collection of real numbers has at least one ϕ-mean; by (2.11) it is unique if $\phi(x, \theta)$ is strictly increasing in θ for each x.

If $\mu = EX$ then $E(X - \mu)^2 \leqq E(X - \theta)^2$ for all θ. In the following theorem we extend this optimizing property of the expected value of a random variable so that it applies to most ϕ-means. In the special case where γ is Lebesgue measure and $\phi(x, \theta) = 2(\theta - x)$ the following theorem gives the property of regular means just mentioned.

THEOREM 2.2. *Let γ and ν be sigma-finite measures on the Borel subsets of R, and let $\phi(\cdot, \cdot)$ satisfy (2.1)–(2.3), be Borel measurable on $R \times R$, and satisfy $\phi(x, x) = 0$ for all x. Suppose μ is in $M(\nu)$ and $\lambda(\mu)$ is defined. Define*

(i) $\Phi(x, \theta) = \int_{(x,\theta) \cup [\theta,x)} |\phi(x, u)| \, d\gamma(u)$ *if $\theta > \mu$,*

(ii) $\Phi(x, \theta) = \int_{(x,\theta] \cup (\theta,x)} |\phi(x, u)| \, d\gamma(u)$ *if $\theta < \mu$,* and let $\Phi(x, \mu)$ be given by (i) *or* (ii) *if* $\lambda(\mu) \geqq 0$ or $\lambda(\mu) < 0$ *respectively. Then*

$$\int \Phi(x, \mu) \, d\nu(x) \leqq \int \Phi(x, \theta) \, d\nu(x) \quad \textit{for all } \theta. \tag{2.18}$$

There is equality in (2.18) if and only if either $\int \Phi(x, \mu) \, d\nu(x) = \infty$, or if $\lambda(u) = 0$ almost everywhere (γ) on $(\theta, \mu) \cup [\mu, \theta)$ or $(\theta, \mu] \cup (\mu, \theta)$ depending on whether $\lambda(\mu) \geqq 0$ or $\lambda(\mu) < 0$ respectively.

PROOF. Suppose $\theta > \mu$ and $\lambda(\mu) \geqq 0$. Then

$$\begin{aligned} \int \Phi(x, \mu) \, d\nu(x) &= \int \int_{\{(x,u): x<u<\mu \text{ or } \mu \leqq u < x\}} |\phi| \, d(\nu \times \gamma) \\ &= \int \int_{\{(x,u): x<u<\mu \text{ or } \theta \leqq u < x\}} |\phi| \, d(\nu \times \gamma) \\ &\quad + \int_{[\mu,\theta)} [\int_{(u,\infty)} |\phi(x, u)| \, d\nu(x)] \, d\gamma(u) \end{aligned} \tag{2.19}$$

and

$$\begin{aligned}\int \Phi(x, \theta)\, d\nu(x) &= \int \int_{\{(x,u):x<u<\theta \text{ or } \theta \leqq u<x\}} |\phi|\, d(\nu \times \gamma) \\ &= \int \int_{\{(x,u):x<u<\mu \text{ or } \theta \leqq u<x\}} |\phi|\, d(\nu \times \gamma) \\ &\quad + \int_{[\mu,\theta)} [\int_{(-\infty,u)} |\phi(x, u)|\, d\nu(x)]\, d\gamma(u).\end{aligned}$$

The first terms in the final expressions for $\int \Phi(x, \mu)\, d\nu(x)$ and $\int \Phi(x, \theta)\, d\nu(x)$ are the same; we will compare the second terms. Recall that for $\theta > \mu$ we have $\lambda(\theta) \geqq 0$ so that $\int \phi^+(x, \theta)\, d\nu(x) \geqq \int \phi^-(x, \theta)\, d\nu(x)$. But

$$\begin{aligned}&\int_{[\mu,\theta)} [\int_{(u,\infty)} |\phi(x, u)|\, d\nu(u)]\, d\gamma(u) \\ &\quad = \int_{[\mu,\theta)} [\int \phi^-(x, u)\, d\nu(x)]\, d\gamma(u) \leqq \int_{[\mu,\theta)} [\int \phi^+(x, u)\, d\nu(x)]\, d\gamma(u) \\ &\quad = \int_{[\mu,\theta)} [\int_{(-\infty,u)} |\phi(x, u)|\, d\nu(x)]\, d\gamma(u)\end{aligned}$$

with equality if and only if

$$\int_{[\mu,\theta)} [\int_{(u,\infty)} |\phi(x, u)|\, d\nu(x)]\, d\gamma(u)$$

is infinite or if $\lambda(u) = 0$ a.e. (γ) on $[\mu, \theta)$. Thus in this case (2.18) holds and we have equality if and only if either $\int \Phi(x, \mu)\, d\nu(x) = \infty$ (i.e. one of the two expressions in (2.19) is infinite) or if $\lambda(u) = 0$ a.e. (γ) on $[\mu, \theta) = (\theta, \mu) \cup [\mu, \theta)$.

The proof is essentially the same in each of the other three cases and will be omitted.

Various forms of the preceding theorem can be proved under various assumptions on ϕ, ν, and γ. Different definitions of ϕ and conditions under which there is equality in (2.18) are involved in these various theorems. The theorem given is not necessarily the most general, or even the most useful. It was chosen because it gives a certain amount of generality, is fairly easy to state, and because of certain corollaries to it.

We note that in view of (2.14) the restriction "$\phi(x, x) = 0$ for all x" is not as serious as might first be suspected. It does simplify the theorem statement and the notation involved in its proof.

It can be shown that if μ is in $M(\nu)$ and $\lambda(\mu)$ is undefined, then when $\gamma\{\mu\} = 0$ we can use the definitions and conclusions of the theorem, and when $\gamma\{\mu\} > 0$ we have $\int \Phi(x, \theta)\, d\nu(x) = \infty$ for all θ.

In the special case where γ is non-atomic (i.e. $\gamma\{x\} = 0$ for each x in R) the theorem is much more simply stated since all intervals involved may be assumed to be open. We can define

$$\Phi(x, \theta) = \int_{(x,\theta)\cup(\theta,x)} |\phi(x, u)|\, d\gamma(u) \qquad \text{for all } \theta,$$

and there will be equality in (2.18) if and only if either $\int \Phi(x, \mu)\, d\nu(x) = \infty$ or if $\gamma\{(\theta, \underline{\mu}) \cup (\bar{\mu}, \theta)\} > 0$. If in addition $\gamma(I) > 0$ for every open interval I, then we would have equality in (2.18) if and only if either $\int \Phi(x, \mu) d\nu(x) = \infty$ or if θ as well as μ is in $M(\nu)$. Note that in this case Φ does not depend on μ as it might in Theorem 2.2.

It is perhaps worth noting that if $\phi(x, u)$ is continuous in u for each x, then from (2.12) we see that $\lambda(\mu)$ is the same for all μ in M so that either $M(\nu)$ is a single point or $\lambda(\mu) = 0$ for all μ in $M(\nu)$. In the first case Φ has only a single definition, and in the second case we can use either definition (i) or definition (ii) for all θ without running into trouble.

The following property of the usual mean value of a random variable is a consequence of Theorem 3.2 of [3] and is obtained by setting $M = \{\phi, \Omega\}$ and $Z =$ constant.

THEOREM. *Let X be a random variable, I an interval, and Φ a convex function defined on I, such that $EX^2 < \infty$, $EX = \mu$, $P\{X \varepsilon I\} = 1$, and $E\Phi(X) < \infty$. Then if $\psi^+(x)$ and $\psi^-(x)$ are the right and left derivatives respectively of Φ and $\psi^-(x) \leqq \psi(x) \leqq \psi^+(x)$ for all x, and if $\Delta(x, z) = \Phi(x) - \Phi(z) - (x - z)\psi(z)$, then*

$$\int \Delta(X, \mu)\, dP \geqq \int \Delta(X, \theta)\, dP \qquad \textit{for all } \theta.$$

If in Theorem 2.2 $\phi(x, \theta) = \theta - x$, if γ gives finite measure to finite intervals so that there is a non-decreasing function ψ inducing the measure γ, and if $\nu = P$, then the theorems are essentially equivalent.

The ϕ-means have the Cauchy mean value property ([5], page 57): if S and T are disjoint finite sets, then $M(S \cup T)$ lies between $M(S)$ and $M(T)$. The following theorem may be regarded as a generalization of this observation.

THEOREM 2.3. *Let ν_i with $i = 1, \cdots, N$ or $i = 1, 2, \cdots$ be measures on the Borel subsets of R and let $\nu = \sum_i \nu_i$. Assume $M(\nu_i)$ is non-empty for all i. If $\mu_* = \inf \bigcup_i M(\nu_i)$ and $\mu^* = \sup \bigcup_i M(\nu_i)$ then $M(\nu) \subset [\mu_*, \mu^*]$.*

PROOF. Suppose $\mu_* > -\infty$. If $\theta < \mu_*$ then $\int \phi^+(x, \theta)\, d\nu_i(x) < \int \phi^-(x, \theta)\, d\nu_i(x)$ for all i so $\int \phi^+(x, \theta)\, d\nu(x) \leqq \int \phi^-(x, \theta)\, d\nu(x)$ with equality if and only if both sides of the last expression are infinite. If strict inequality holds for all $\theta < \mu_*$ then $M(\nu) \subset [\mu_*, \infty)$. If equality holds for any $\theta < \mu_*$ then if $\theta \leqq \theta' < \mu_*$ we have $\int \phi^-(x, \theta')\, d\nu(x) \geqq \int \phi^+(x, \theta')\, d\nu(x) \geqq \int \phi^+(x, \theta)\, d\nu(x) = \infty$ so $M(\nu)$ is empty by (2.8b).

Similarly, if $\mu^* < \infty$ then either $M(\nu) \subset (-\infty, \mu^*]$ or $M(\nu)$ is empty.

Putting these together completes the proof of the theorem.

If $\alpha_i > 0$ for all i then since $M(\nu_i) = M(\alpha_i \nu_i)$ the above theorem remains true if ν_i is replaced by $\alpha_i \nu_i$ everywhere. This theorem may be further generalized by assuming that $\nu(A) = \int \nu_\omega(A)\, d\gamma(\omega)$ where ν_ω is a measure for each point ω in Ω from the finite measure space (Ω, Σ, γ), and by asuming that $\nu_\omega(A)$ is a measurable function of ω for each Borel measurable set A. The exact statement of the theorem then becomes a little messy due to the care needed in dealing with sets of γ-measure zero.

One can obtain a little more information about the exact location of $M(\nu)$ in some cases but the additional conditions needed are not particularly nice and the amount of extra information obtained is normally not very large.

In this paper we deal with one fixed ϕ at a time—except in the following theorem which is included for completeness and because it is the analog to Theorem 2.3 when one mixes over the function instead of the measure.

THEOREM 2.4. *Let ϕ_i with $i = 1, \cdots, N$ or $i = 1, 2, \cdots$ be extended real valued functions on $R \times R$ which satisfy (2.1)–(2.3); let $\phi = \sum_i \phi_i$; and let ν be a measure on the Borel subsets of R. Assume $M(\nu, \phi_i)$ is non-empty for all i and that that $\phi(x, x)$ is defined for all values of x. If $\mu_* = \inf \bigcup_i M(\nu, \phi_i)$ and $\mu^* = \sup \bigcup_i M(\nu, \phi_i)$ then $M(\nu, \phi) \subset [\nu_*, \mu^*]$.*

We omit the proof of this theorem since it is essentially the same as the proof of Theorem 2.3. Note that if $\alpha_i > 0$ then $M(\nu, \phi_i) = M(\nu, \alpha_i\phi_i)$ and that the same types of generalizations apply to this theorem as to Theorem 2.3.

We conclude this section by showing that a form of Jensen's inequality is valid for r-means.

LEMMA. *Let X be a random variable, let μ be a real number, and let g be a non-decreasing function on R. The class C of functions f such that $Eg[\theta - f(X)] \leqq 0$ for all $\theta < f(\mu)$ is closed under the operation "$\vee$" where*

$$(f_1 \vee f_2)(x) = \max [f_1(x), f_2(x)].$$

PROOF. Let $f = f_1 \vee f_2$ with f_1 and f_2 in C. If $f(\mu) = f_1(\mu)$ then since g is non-decreasing we have $g[\theta - f(X)] \leqq g[\theta - f_1(X)]$ so that $Eg[\theta - f(X)] \leqq Eg[\theta - f_1(X)]$, and for $\theta < f(\mu) = f_1(\mu)$ we have $Eg[\theta - f_1(X)] \leqq 0$ since f_1 was hypothesized to be in C. Similarly if $f(\mu) = f_2(\mu)$ then $\theta < f(\mu)$ implies $Eg[\theta - f(X)] \leqq 0$. Thus f is in C.

THEOREM 2.5. *Assume that $E|X|^r < \infty$ where $r \geqq 0$, that μ is in $M_r(X)$, that f is convex, and that $M_r[f(X)] \neq 0$. Then $f(\mu) \leqq \sup M_r[f(X)]$. If $r \neq 0$, then $\{\mu\} = M_r(X)$ and $M_r[f(X)]$ each consist of a single point, say $M_r[f(X)] = \{\mu^*\}$; in this case $f(\mu) \leqq \mu^*$.*

PROOF. The fact that $M_r(X)$ and $M_r[f(X)]$ each contain at most one point for $r > 0$ follows from the definition of r-mean (Definition 2.2) and from (2.11). Thus the last part of the theorem is an immediate consequence of $\mu \leqq \sup M_r[f(X)]$; we shall now prove $\mu \leqq \sup M_r[f(X)]$ for $r \geqq 0$.

Set $g(0) = 0$ and $g(x) = |x|^r \operatorname{sgn} x$ for $x \neq 0$. Let $\phi(x, \theta) = g(\theta - x)$. Since $E|X|^r < \infty$ we have $E|g[\theta - f(X)]| < \infty$ for each linear function f. Let $f(x) = ax + b$. If $a = 0$ then $f(\mu) = b$ and $\theta < f(\mu)$ implies $Eg[\theta - f(X)] = g(\theta - b) < 0$. If $a < 0$, then if $\theta < f(\mu)$ we have

$$Eg[\theta - f(X)] = g(a)Eg[(\theta - b)/a - X]$$

and since $\theta < a\mu + b$ we have $(\theta - b)/a > \mu$; thus $Eg[(\theta - b)/a - X] \geqq 0$. However, $g(a) < 0$ so $Eg[\theta - f(X)] \leqq 0$. The same method of proof applies if $a > 0$. We have shown that $\theta < f(\mu)$ implies $Eg[\theta - f(X)] \leqq 0$ when f is linear. It follows from the lemma that the same is true if $f = \bigvee_{i=1}^{n} f_i$ where each f_i is linear. Now suppose f is convex and $\{f_n\}$ is a sequence of polygonal convex functions converging upwards to f and such that $f_n(\mu) = f(\mu)$ for all n. Then $g[\theta - f_n(X)] \downarrow g[\theta - f(X)]$ and for all $\theta < f(\mu)$ we have $0 \geqq Eg[\theta - f_n(X)] \downarrow Eg[\theta - f(X)]$ so that $Eg[\theta - f(X)] \leqq 0$. Thus $\lim_{\theta \uparrow f(\mu)} Eg[\theta - f(X)] \leqq 0$. Since $\sup \{\lambda \mid \lim_{\theta \uparrow \lambda} Eg[\theta - f(X)] \leqq 0\}$ is just $\sup M_r[f(X)]$ the proof of the theorem is complete.

3. Asymptotic properties of sample ϕ-means. Let $X, X_1, X_2, \cdots$, be independent random variables with common distribution function F. For each positive integer n, let F_n be the sample distribution function of $X_1, X_2, \cdots, X_n$: $F_n(x) = \sum_{\{i: i \leqq n, X_i \leqq x\}} 1/n$.

THEOREM 3.1. (1) *If ϕ is finite valued, then for each n $M(F_n) \neq \phi$.*

(2) *If $M(F) \neq \phi$, then for each n, $P[M(F_n) \neq \phi] = 1$.*

PROOF. If ϕ is finite valued, then every finite set has a ϕ-mean, so a sample ϕ-mean always exists. We have $0 \geqq \sum_{i=1}^{n} \phi(X_i, \theta) \geqq -\infty$ for $\theta < \min\{X_1, \cdots, X_n\}$ and $0 \leqq \sum_{i=1}^{n} \phi(X_i, \theta) \leqq \infty$ for $\theta > \max\{X_1, \cdots, X_n\}$. From (2.8) it then follows that $M(F_n) \neq \phi$ unless $\sum_{i=1}^{n} \phi(X_i; \theta)$ is undefined for at least two values of θ. Set

$$b = \inf\{\theta: P[\phi(X, \theta) = \infty] > 0\}, \qquad a = \sup\{\theta: P[\phi(X, \theta) = -\infty] > 0\}.$$

We have $E\phi^+(X, \theta) = \infty$ for $\theta > b$ if $b < \infty$, and $E\phi^-(X, \theta) = \infty$ for $\theta < a$ if $a > -\infty$. Since $M(F) \neq \phi$, it follows from Theorem 2.1 that $a \leqq b$. But if $\theta < b$ or if $\theta > a$ then $\sum_{i=1}^{n} \phi(X_i, \theta)$ is almost surely defined. Thus almost surely $\sum_{i=1}^{n} \phi(X_i, \theta)$ can fail to be defined for at most one value of θ, $\theta = a = b$, and almost surely $M(F_n) \neq \phi$.

The following theorem states the strong consistency of the sample ϕ-mean as an estimator of the population ϕ-mean. A version of this theorem appears in [2] and a version for translation means in [7]. Throughout the remainder of this paper the phrase *almost all n* will be used to mean *all but a finite number of values of n*, and the abbreviations i.o. and f.o. will stand for *infinitely often* and *only finitely often* respectively.

THEOREM 3.2. *If $M(F) \neq \phi$ and if G is an open set containing $M(F)$, then $P\{M(F_n) \subset G$ for almost all $n\} = 1$.*

PROOF. If $a < \mu$ then $\lambda(a) = \int \phi(x, a)\, dF(x) < 0$, for if equality held we should have $a \in M(F)$. By the strong law of large numbers, as $n \to \infty$ we have $(1/n) \sum_{k=1}^{n} \phi(X_k, a) \to \lambda(a)$ almost surely. Similarly, if $b > \bar{\mu}$, as $n \to \infty$ we have $(1/n) \sum_{k=1}^{n} \phi(X_k, b) \to \lambda(b) > 0$ almost surely. The conclusion of the theorem follows.

In Theorem 3.3, results of M. Rosenblatt ([10], Theorem 2.1) on the oscillation of sums of independent random variables are applied to yield information on the oscillation of $\underline{\mu}_n$ and $\bar{\mu}_n$ about $\underline{\mu}$ and $\bar{\mu}$. By $\phi(x, a-)$ we shall understand $\lim_{\theta \uparrow a} \phi(x, \theta)$, and similarly $\phi(x, a+) = \lim_{\theta \downarrow a} \phi(x, \theta)$. Thus, for example, $E\phi(X, a-) = \int \lim_{\theta \uparrow a} \phi(x, \theta)\, dF(x)$, whereas, by contrast, $\lambda(a-) = \lim_{\theta \uparrow a} \int \phi(x, \theta)\, dF(x)$. It follows from the monotone convergence theorem that

$$\text{(3.1)} \qquad \text{if} \quad \lambda(a-) > -\infty \quad \text{then} \quad \lambda(a-) = E\phi(X, a-),$$

$$\text{and if} \quad \lambda(a+) < \infty \quad \text{then} \quad \lambda(a+) = E\phi(X, a+).$$

THEOREM 3.3. *Assume $M(F) \neq \phi$.*

(i) *If either $E\phi(X, a-) = 0$ or $E\phi(X, a+) = 0$, then $P\{\underline{\mu}_n \leqq a$ i.o. and $\bar{\mu}_n \geqq a$ i.o.$\} = 1$.*

(ii) If $E\phi(X, a-) < 0$ *then almost surely $\underline{\mu}_n \geqq a$ for almost all n. If*

$E\phi(X, a+) > 0$ then almost surely $\bar{\mu}_n \leqq a$ for almost all n. If both $E\phi(X, a-) < 0$ and $E\phi(X, a+) > 0$ then almost surely $M(F_n) = \{a\}$ for almost all n.

(iii) *If $E\phi(X, a-) > 0$ then almost surely $\bar{\mu}_n < a$ for almost all n. If $E\phi(X, a+) < 0$ then almost surely $\underline{\mu}_n > a$ for almost all n.*

PROOF. Let $Y, Y_1, Y_2, \cdots$, be independent and identically distributed random variables. Rosenblatt's theorem states that if $E|Y| < \infty$, $P[Y = 0] < 1$, $EY = 0$, and $S_n = \sum_{i=1}^n Y_i$ for $n = 1, 2, \cdots$, then almost surely $S_n > 0$ infinitely often. By Theorem 3.1, the hypothesis $M(F) \neq \phi$ implies $P[M(F_n) \neq \phi] = 1$. In proving (i), we first apply Rosenblatt's theorem with $Y_i = \phi(X_i, a-)$ and again with $Y_i = -\phi(X_i, a-)$, $i = 1, 2, \cdots$. We conclude that almost surely $\sum_{i=1}^n \phi(X_i, a-) \geqq 0$ infinitely often, and almost surely $\sum_{i=1}^n \phi(X_i, a-) \leqq 0$ infinitely often. The first inequality implies that $a \geqq \underline{\mu}_n$ and the second implies that $a \leqq \bar{\mu}_n$. The proof of these conclusions under the hypothesis $E\phi(X, a+) = 0$ is symmetric.

In proving (ii), we note first that if $E\phi(X, a-) < 0$ then the strong law of large numbers assures us that almost surely $\sum_{i=1}^n \phi(X_i, a-) < 0$ for almost all n. This implies $a \leqq \underline{\mu}_n$ for almost all n. Similarly $E\phi(X, a+) > 0$ implies that almost surely $\bar{\mu}_n \leqq a$ for almost all n. Thus if both hold, almost surely $\underline{\mu}_n = \bar{\mu}_n = a$ for almost all n.

Applying the strong law again for (iii), we find that if $E\phi(X, a-) > 0$ then almost surely $\sum_{i=1}^n \phi(X_i, a-) > 0$ for almost all n; i.e., almost surely, $\lim_{t \uparrow a} \sum_{i=1}^n \phi(X_i, t) > 0$ for almost all n. It follows that, almost surely, $\bar{\mu}_n < a$ for almost all n. The proof of the last conclusion is symmetric.

We note that

(3.2) If in part (i) of Theorem 3.3 we assume $P\{\phi(X, a-) = 0\} \neq 1$ in addition to $E\phi(X, a-) = 0$, then $P\{\bar{\mu}_n < a \text{ i.o.}\} = 1$. Similarly, if $P\{\phi(X, a+) = 0\} \neq 1$ as well as $E\phi(X, a+) = 0$, then $P\{\underline{\mu}_n > a \text{ i.o.}\} = 1$.

We will prove the first of the two statements. Under our assumptions Rosenblatt's theorem tells us that $P\{(1/n) \sum_{k=1}^n \phi(X_k, a-) > 0 \text{ i.o.}\} = 1$. The fact that $\phi(x, t) \to \phi(x, a-)$ as $t \uparrow a$ says that if $E_{F_n}\phi(X, a-) > 0$, then $E_{F_n}\phi(X, t) > 0$ for $t < a$ but sufficiently close to a, and consequently $M(F_n) = \phi$ or $\bar{\mu}_n < a$. It follows that $P\{\bar{\mu}_n < a \text{ i.o.}\} = 1$.

COROLLARY 3.1. *If $M(F)$ contains at least two points, then almost surely $\underline{\mu}_n \leqq \underline{\mu}$ infinitely often and $\bar{\mu}_n \geqq \bar{\mu}$ infinitely often.*

PROOF. If $\underline{\mu} < \theta < \bar{\mu}$ then $\lambda(\theta) = 0$. By (3.1) $0 = \lambda(\bar{\mu}-) = E\phi(X, \bar{\mu}-)$ and $0 = \lambda(\underline{\mu}+) = E\phi(X, \underline{\mu}+)$. The conclusion of the corollary follows from (i) of Theorem 3.3.

COROLLARY 3.2. *If there exist real numbers a and b, not necessarily distinct, such that $-\infty < \lambda(a) \leqq 0 \leqq \lambda(b) < \infty$ then almost surely $\underline{\mu}_n \leqq \underline{\mu}$ infinitely often and and $\bar{\mu}_n \geqq \bar{\mu}$ infinitely often.*

PROOF. Since $\lambda(a)$ (also $\lambda(b)$) is finite, $\lambda(\theta)$ is defined for all θ. Thus by (2.8),

$M(F) \neq \phi$. If $M(F)$ contains at least two points then Corollary 3.1 applies. If $M(F) = \{\mu\}$ so that $\mu = \underline{\mu} = \bar{\mu}$, then $a \leqq \mu \leqq b$. We consider three cases: $\lambda(\mu) = 0$, $\lambda(\mu) > 0$, and $\lambda(\mu) < 0$.

CASE 1. If $\lambda(\mu) = 0$ then

$$E\phi(X, \mu-) \leqq E\phi(X, \mu) = 0 \leqq E\phi(X, \mu+).$$

The conclusion now follows from (i) and (ii) of Theorem 3.3.

CASE 2. If $\lambda(\mu) > 0$, then since $\lambda(a) \leqq 0$, $a < \mu$. Since $-\infty < \lambda(a)$, by (3.1) $E\phi(X, \mu-) = \lambda(\mu-) \leqq 0$. Also $E\phi(X, \mu+) \geqq E\phi(X, \mu) = \lambda(\mu) > 0$. Again the conclusion follows from (i) and (ii) of Theorem 3.3.

CASE 3. If $\lambda(\mu) < 0$, we use an argument symmetric to the one of Case 2.

In each of the following examples illustrating Theorem 3.3, X has the standardized normal distribution and $M(X; \phi) = \{0\}$.

EXAMPLE 3.1. $\phi(x, \theta) = 1$ when $\theta > 0$ and $x < \theta$; $\phi(x, \theta) = -1$ when $\theta < 0$ and $x > \theta$; $\phi(x, \theta) = 0$ otherwise. Then $E\phi(X, 0+) = \frac{1}{2}$, $E\phi(X, 0-) = -\frac{1}{2}$, and almost surely $\underline{\mu}_n = \bar{\mu}_n = 0$ for almost all n.

EXAMPLE 3.2. $\phi(x, \theta) = \infty$ for $x \leqq 0, \theta \geqq 0$; $\phi(x, \theta) = 1$ when $x < \theta < 0$; $\phi(x, \theta) = -\infty$ when $x < 0$ and $\theta \leqq x$; $\phi(x, \theta) = 0$ otherwise. Then $E\phi(X, 0-) = \frac{1}{2}$ so almost surely $\bar{\mu}_n < 0$ for almost all n.

EXAMPLE 3.3. $\phi(x, \theta) = 0$ when $0 < \theta < x$, $\phi(x, \theta) = \theta - x$ otherwise. Then $E\phi(X, 0+) > 0$, $E\phi(X, 0-) = 0$. Thus almost surely $\bar{\mu}_n \leqq 0$ for almost all n but also $\bar{\mu}_n \geqq 0$ infinitely often, hence $\bar{\mu}_n = 0$ infinitely often.

Theorem 3.4 below gives conditions for asymptotic normality of the sample ϕ-means. Essentially the same theorem appears in [2]. A version of this theorem for translation means was proved by Huber [7] and his proof extends in a straightforward manner to our situation.

We define $\sigma^2(\theta) =$ variance of $\phi(X, \theta)$ and let $\underline{\sigma}^2 = \sigma^2(\underline{\mu})$ and $\bar{\sigma}^2 = \sigma^2(\bar{\mu})$. Whenever we appear to be performing a Stieltjes integration with respect to some distribution function, the reader should interpret the integral as being performed with respect to the probability measure generated by that distribution function.

LEMMA 3.1. *Let μ be a real number such that $0 < \sigma^2(\mu) < \infty$ and such that $\sigma^2(\theta) \to \sigma^2(\mu)$ as $\theta \to \mu$. Let $\epsilon > 0$ be a real number. Let $\theta_n \to \mu$ as $n \to \infty$ and define $A(n, \epsilon)$ by $A(N, \epsilon) = \{x: |\phi(x, \theta_n) - \lambda(\theta_n)| \geqq \epsilon n^{\frac{1}{2}}\sigma(\theta_n)\}$. Then*

$$(3.3) \qquad \int_{A(n,\epsilon)} [\phi(x, \theta_n) - \lambda(\theta_n)]^2 \, dF(x) \to 0 \qquad as \qquad n \to \infty.$$

PROOF. Fix $h > 0$. Since ϕ is non-decreasing in the second argument, if $\mu - h \leqq \theta \leqq \mu + h$ we have $\phi(x, \mu - h) \leqq \phi(x, \theta) \leqq \phi(x, \mu + h)$ and $\lambda(\mu - h) \leqq \lambda(\theta) \leqq \lambda(\mu + h)$. Set $u(x) = \max\{|\phi(x, \mu + h) - \lambda(\mu - h)|, |\phi(x, \mu - h) - \lambda(\mu + h)|\}$. Then $|\phi(x, \theta) - \lambda(\theta)| \leqq u(x)$ if $\mu - h \leqq \theta \leqq \mu + h$. By hypothesis, $E\phi^2(x, \theta) < \infty$ for θ near μ, so that if h is fixed sufficiently small, we have $\int u^2(x) \, dF(x) \leqq \int \{[\phi(x, \mu + h) - \lambda(u - h)]^2 + [\phi(x, \mu - h) - \lambda(\mu + h)]^2\} \, dF(x) < \infty$. Also, for n sufficiently large, $\mu - h \leqq \theta_n \leqq \mu + h$ and $A(n, \epsilon) \subset \{x: u(x) \geqq \epsilon n^{\frac{1}{2}}\sigma(\theta_n)\}$, so that

$$\int_{A(n,\epsilon)} [\phi(x, \theta_n) - \lambda(\theta_n)]^2 \, dF(x) \leqq \int_{\{x: u(x) \geqq \epsilon n^{\frac{1}{2}}\sigma(\theta_n)\}} u^2(x) \, dF(x) \to 0 \quad \text{as} \quad n \to \infty.$$

THEOREM 3.4. *Suppose* $M(X) = M(F) = [\underline{\mu}, \bar{\mu}]$ *with* $-\infty < \underline{\mu} \leqq \bar{\mu} < \infty$. *Let* $\Phi(z) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{z} e^{-t^2/2}\, dt$ *and let* $\mu_n = \underline{\mu}_n + \alpha_n(\bar{\mu}_n - \underline{\mu}_n)$ *where* $0 \leqq \alpha_n \leqq 1$. *If*

$$\lambda(\theta)/(\theta - \underline{\mu}) \to a \quad as \quad \theta \uparrow \underline{\mu} \quad with \quad 0 < a < \infty, \tag{3.4}$$

$$0 < \lim_{\theta \uparrow \underline{\mu}} \sigma^2(\theta) = \underline{\sigma}^2 < \infty, \tag{3.5}$$

then for all $z < 0$

$$P\{(\mu_n - \underline{\mu})/(\underline{\sigma}/an^{\frac{1}{2}}) < z\} \to \Phi(z) \qquad as \qquad n \to \infty. \tag{3.6}$$

If

$$\lambda(\theta)/(\theta - \mu) \to b \quad as \quad \theta \downarrow \bar{\mu} \quad with \quad 0 < b < \infty, \tag{3.7}$$

$$0 < \lim_{\theta \downarrow \bar{\mu}} \sigma^2(\theta) = \bar{\sigma}^2 < \infty, \tag{3.8}$$

then for all $z > 0$

$$P\{(\mu_n - \bar{\mu})/(\bar{\sigma}/bn^{\frac{1}{2}}) < z\} \to \Phi(z) \quad as \quad n \to \infty. \tag{3.9}$$

PROOF. Let $\theta_n = \underline{\mu} + z\underline{\sigma}/(an^{\frac{1}{2}})$. For arbitrary real z we have

$$[\textstyle\sum_{i=1}^{n} \phi(X_i, \theta_n -) > 0] \subset [\mu_n < \theta_n] \subset [\textstyle\sum_{i=1}^{n} \phi(X_i, \theta_n) \geqq 0].$$

In order to establish (3.6) it suffices to show that for $z < 0$

$$P[\textstyle\sum_{i=1}^{n} \phi(X_i, \theta_n) \geqq 0] \to \Phi(Z) \quad \text{as} \quad n \to \infty$$

and

$$P[\textstyle\sum_{i=1}^{n} \phi(X_i, \theta_n -) > 0] \to \Phi(z) \quad \text{as} \quad n \to \infty.$$

Set

$$Y_{in} = [\phi(X_i, \theta_n) - \lambda(\theta_n)]/\sigma(\theta_n).$$

For fixed n the random variables Y_{in} are independent and identically distributed with $EY_{in} = 0$ and Var $(Y_{in}) = 1$. Also,

$$P[\textstyle\sum_{i=1}^{n} \phi(X_i, \theta_n) \geqq 0] = P[\textstyle\sum_{i=1}^{n} Y_{in}/n^{\frac{1}{2}} \geqq -n^{\frac{1}{2}}\lambda(\theta_n)/\sigma(\theta_n)].$$

From (3.4) and the definition of θ_n we see that $-n^{\frac{1}{2}}\lambda(\theta_n)/\sigma(\theta_n) \to -z$ as $n \to \infty$. Thus it suffices to show that

$$P[\textstyle\sum_{i=1}^{n} Y_{in}/n^{\frac{1}{2}} \geqq -z] \to 1 - \Phi(-z) = \Phi(z) \qquad \text{as} \qquad n \to \infty$$

for every $z < 0$.

We apply the central limit theorem in the form given in ([8], p. 295). The proof that $P\{\sum_{i=1}^{n} \phi(X_i, \theta_n) \geqq 0\} \to \Phi(z)$ will be completed when we have shown that for every $\epsilon > 0$, $E(Y_{in}^2 I\{|Y_{in}| \geqq \epsilon n^{\frac{1}{2}}\}) \to 0$ as $n \to \infty$, where $I\{\cdot\}$ denotes the indicator function of the event described in braces. But $E(Y_{in}^2 I\{|Y_{in}| \geqq \epsilon n^{\frac{1}{2}}\}) = \int_{A(n,\epsilon)} [\phi(x, \theta_n) - \lambda(\theta_n)]^2\, dF(x)/\sigma^2(\theta_n)$ and by (3.3) the last integral approaches zero as n approaches infinity. The proof that $P\{\sum_{i=1}^{n} \phi(X_i, \theta_n -) > 0\} \to \Phi(z)$ as $n \to \infty$ proceeds similarly using $\phi(X_i, \theta_n -)$ instead of $\phi(X_i, \theta_n)$ and a version of Lemma 3.1 for $\phi(X_i, \theta_n -)$. The proof of (3.9) is symmetric to that of (3.6).

COROLLARY 3.3. *Under the hypotheses of Theorem* 3.4, *if* $-\infty < \underline{\mu} < \bar{\mu} < \infty$ *then* $P[\underline{\mu} < \underline{\mu}_n < \bar{\mu}] \to 0$ *as* $n \to \infty$ *and* $P[\underline{\mu} < \bar{\mu}_n < \bar{\mu}] \to 0$ *as* $n \to \infty$.

PROOF. (3.6) implies $\liminf P\{\bar{\mu}_n > \underline{\mu}\} \geqq \frac{1}{2}$ and (3.9) implies $\liminf P\{\underline{\mu}_n > \bar{\mu}\} \geqq \frac{1}{2}$.

COROLLARY 3.4. *If* $M(X) = \{\mu\}$ *and if hypotheses* (3.4) *and* (3.7) *are replaced by the hypothesis that* λ *is differentiable at* μ *with* $\lambda'(\mu) > 0$, *then each of* $\underline{\mu}_n$ *and* $\bar{\mu}_n$ *is asymptotically normal with mean* μ *and standard deviation* $\sigma(\mu)/(n^{\frac{1}{2}}\lambda'(\mu))$.

4. Families of distributions. We consider the class of one parameter families of distributions, the logarithms of whose densities are convex in the parameter. The parameter of such a family may usually be interpreted as a ϕ-mean; and in sampling from a member of such a family the sample ϕ-mean is a maximum likelihood estimator of the parameter. We show how a given function ϕ and a given distribution F determine such a family, and close the section with some examples.

Let κ be a measure on Borel sets of real numbers with support in an interval I (possibly infinite). Let $\Psi(y, \tau)$ be convex in τ for each $y \varepsilon I$. Suppose further that

(4.1)
for each real τ
$f(y, \tau) = \exp\{-\Psi(y, \tau)\}$ is Borel-measurable
for $y \varepsilon I$ and is a probability density
with respect to κ.

$E_\tau h(Y)$ will denote $\int h(y) f(y, \tau)\, d\kappa(y)$. Set $\psi(y, \tau) = d^+\Psi(y, \tau)/d\tau$, adopting the convention that if $\Psi(y, \tau) = \infty$ for $\tau \geqq b(y)$, then for such τ, $\psi(y, \tau) = \infty$; while if $\Psi(y, \tau) = \infty$ for $\tau \leqq a(y)$ then for such τ, $\psi(y, \tau) = -\infty$. We note that if ψ satisfies (2.2) then a point of sign change of $E_\tau\psi(Y, \theta)$ is a ψ-mean of Y. (θ_0 is a point of sign change of a nondecreasing function $f(\theta)$ if $\theta > \theta_0$ implies $f(\theta) \geqq 0$ and $\theta < \theta_0$ implies $f(\theta) \leqq 0$.)

THEOREM 4.1. *The parameter* τ *in* (4.1) *is a point of sign change of* $E_\tau\psi(Y, \theta)$. *It is the unique point of sign change if there exist* τ' *arbitrarily near* τ *above and below such that* $\kappa\{y: \Psi(y, \tau) \neq \Psi(y, \tau')\} > 0$. *If* $y_1, \cdots, y_n$ *are observed sample values in random sampling from* (4.1), *the points of sign change of* $\sum_i \psi(y_i, \theta)$ *are maximum likelihood estimates of* τ.

PROOF. Suppose $\tau' > \tau$, and set $A = \{y: f(y, \tau') \neq f(y, \tau)\}$. If $y \varepsilon A$ then $\exp\{-[\Psi(y, \tau') - \Psi(y, \tau)]\} - 1 > -[\Psi(y, \tau') - \Psi(y, \tau)] \geqq -(\tau' - \tau)\psi(y, \tau')$. (These inequalities are valid also if $\Psi(y, \tau') = \infty$, in which case $\psi(y, \tau') = \infty$ and $\Psi(y, \tau) < \infty$.) Then if $\kappa(A) > 0$,

$$0 = \int_A \{\exp[-\Psi(y, \tau')] - \exp[-\Psi(y, \tau)]\}\, d\kappa(y)$$
$$> -(\tau' - \tau) \int_A \psi(y, \tau') \exp[-\Psi(y, \tau)]\, d\kappa(y).$$

Thus $\int_A \psi(y, \tau') \exp[-\Psi(y, \tau)]\, d\kappa(y) > 0$. If $y \varepsilon A^c$ either $\Psi(y, \cdot)$ assumes its minimum at both τ and τ', or $\tau' > g(y)$, where $g(y)$ is a point of sign change of $\psi(y, \theta)$ as a function of θ. In either case $\psi(y, \tau') \geqq 0$. It follows that

$\int \psi(y, \tau')f(y, \tau)\, d\kappa(y) \geqq 0$, with strict inequality if $\kappa(A) > 0$. The proof of the reverse inequality for $\tau' < \tau$ is symmetric. The proof of the theorem is completed by the observation that the convex function of θ, $\sum_{i=1}^n \Psi(y_i, \theta)$, is minimized by $\theta = \tau$ where τ is any point of sign change of $\sum_{i=1}^n \psi(y_i, \theta)$, the right derivative of $\sum_{i=1}^n \Psi(y_i, \theta)$.

We now give conditions sufficient in order that a given function ϕ satisfying (2.1)–(2.3) and a given distribution function F of a bounded random variable determine a family (4.1).

THEOREM 4.2. *If*

(4.2) F *is the distribution function of a nondegenerate random variable* X *with range in a finite open interval* (a, b);

(4.3) ϕ *restricted to* $[a, b] \times [a, b]$ *satisfies* (2.1)–(2.3);

(4.4) $P[\phi(X, \cdot)$ *is discontinuous at* $\theta] = 0$ *for each* $\theta \,\varepsilon\, [a, b]$;

(4.5) $\phi(\cdot, \cdot)$ *is bounded on* $[a, b] \times [a, b]$;

and if also $\tau_0 \,\varepsilon\, R$, *then there is a non-decreasing function* $\theta(\cdot)$ *defined on* R *with range in* $[a, b]$ *such that*

(4.6) $$\int \exp\{-\textstyle\int_{\tau_0}^{\tau} \phi(x, \theta(u))\, du\}\, dF(x) = 1 \quad \text{for all real} \quad \tau.$$

If in addition

(4.7) $$\phi(x, \theta) < 0 \quad \textit{for} \quad \theta < x, \qquad \phi(x, \theta) > 0 \quad \textit{for} \quad \theta > x,$$

then $\theta(u)$ *is strictly increasing, and* $a < \theta(u) < b$ *for all* u. *If, further,*

(4.8) $$P[\phi(X, \theta_2) > \phi(X, \theta_1)] > 0 \quad \textit{for} \quad a \leqq \theta_1 < \theta_2 \leqq b,$$

then, $\theta(\cdot)$ *is continuous.*

We remark that the family of densities in (4.6) is of form (4.1) with $\Psi(y, \tau) = \int_{\tau_0}^{\tau} \phi(y, \theta(u))\, du$, $\psi(y, \tau) = \phi(y, \theta(\tau+))$. Thus Theorem 4.1 applies. If $\theta(\cdot)$ is continuous, it follows that $\theta(\tau)$ is a ϕ-mean of the distribution (4.6) and that each sample ϕ-mean is a maximum likelihood estimator of $\theta(\tau)$. Theorem 2.2 can also be used here.

PROOF. For $h > 0$, $\exp[-h\phi(x, \theta)] \uparrow \exp[-h\phi(x, a+)]$ as $\theta \downarrow a$. By (4.3), $\phi(x, a+) \leqq 0$ for $x \,\varepsilon\, (a, b)$. Using (4.2) and (4.5) we have

$$\lim_{\theta \downarrow a} \int \exp[-h\phi(x, \theta)]\, dF(x) = \int \exp[-h\phi(x, a+)]\, dF(x) \geqq 1.$$

Similarly $\lim_{\theta \uparrow b} \int \exp[-h\phi(x, \theta)]\, dF(x) = \int \exp[-h\phi(x, b-)]\, dF(x) \leqq 1$. Also (4.4) and (4.5) imply that $\int \exp[-h\phi(x, \theta)]\, dF(x)$ is continuous in θ on $[a, b]$. Hence there exists $\theta_1 \,\varepsilon\, [a, b]$ such that $\int \exp[-h\phi(x, \theta_1)]\, dF(x) = 1$. Using an induction argument, suppose it has been shown that there are $\theta_1, \theta_2, \cdots, \theta_{k-1} (k > 1)$ such that $\theta_1 \leqq \theta_2 \leqq \cdots \leqq \theta_{k-1} \leqq b$ and such that $\int g_j(x)\, dF(x) = 1$, $j = 1, 2, \cdots, k-1$, where $g_j(x) = \exp[-\sum_{\nu=1}^{j} h\phi(x, \theta_\nu)]$, $j = 0, 1, \cdots, k-1$. Set $q(x) = \exp[-h\phi(x, \theta_{k-1})]$. Then $\int g_{k-2}\, dF = \int q g_{k-2}\, dF = 1$, hence by Schwarz inequality $\int q^2 g_{k-2}\, dF \geqq (\int q g_{k-2}\, dF)^2 = 1$.

That is, $\int \exp[-h\phi(x, \theta_{k-1})]g_{k-1}(x)\, dF(x) \geqq 1$. We have also

$\lim_{\theta\uparrow b} \int \exp[-h\phi(x, \theta)]g_{k-1}(x)\, dF(x) = \int \exp[-h\phi(x, b-)]g_{k-1}(x)\, dF(x) \leqq 1.$

Then there exists θ_k such that $\theta_{k-1} \leqq \theta_k \leqq b$ and

$$\int \exp[-\textstyle\sum_{\nu=1}^{k} h\phi(x, \theta_\nu)]\, dF(x) = 1.$$

We conclude that there exists a sequence $\theta_1 \leqq \theta_2 \leqq \cdots$ such that for every positive integer k, $\theta_k \varepsilon [a, b]$ and $\int \exp[-\sum_{\nu=1}^{k} h\phi(x, \theta_\nu)]\, dF(x) = 1$. Thus if $\theta(u) = \theta_j$ for $\tau_0 + (j-1)h \leqq u < \tau_0 + jh, j = 1, 2, \cdots$, we have

$$\int \exp[-\textstyle\int_{\tau_0}^{\tau} \phi(x, \theta(u))\, du]\, dF(x) = 1$$

for $\tau = \tau_0 + jh, j = 0, 1, \cdots$.

Now set $r(x) = \exp[h\phi(x, \theta_1)]$. Then $\int dF(x) = \int [1/r(x)]\, dF(x) = 1$. Hence

$$\int r\, dF = \int r^2(1/r)\, dF \geqq [\textstyle\int r(1/r)\, dF]^2 = 1.$$

As before, we have also that

$$\lim_{\theta\downarrow a} \int \exp[h\phi(x, \theta)]\, dF(x) = \int \exp[h\phi(x, a+)]\, dF(x) \leqq 1$$

so that there exists θ_{-1} such that $a \leqq \theta_{-1} \leqq \theta_1$ and $\int \exp[h\phi(x, \theta_{-1})]\, dF(x) = 1$. Again we show inductively that there are $\theta_{-2}, \theta_{-3}, \cdots$ such that $a \leqq \theta_{-m-1} \leqq \theta_{-m}$, $m = 1, 2, \cdots$, and such that

$$\int \exp[-\textstyle\int_{\tau_0}^{\tau} \phi(x, \theta(u))\, du]\, dF(x) = 1 \qquad \text{for} \qquad \tau = \tau_0 - kh, \quad k = 0, 1, 2, \cdots,$$

where $\theta(u) = \theta_j$ for $\tau_0 + jh \leqq u < \tau_0 + (j+1)h, j = -1, -2, \cdots$. Thus we have for each $h > 0$ a non-decreasing step function $\theta(\cdot)$ on $(-\infty, \infty)$ with range in $[a, b]$ such that $\int \exp[-\int_{\tau_0}^{\tau} \phi(x, \theta(u))\, du]\, dF(x) = 1$ for $\tau = \tau_0 + jh$, $j = \cdots, -1, 0, 1, \cdots$. For fixed $h_0 > 0$, and for an arbitrary positive integer n, let $\theta_n(\cdot)$ denote the step function corresponding to $h = 2^{-n}h_0$. Choose a subsequence $\{\theta_{n_i}(\cdot)\}$ converging at the rationals. Set $\underline{\theta}(u) = \liminf \theta_{n_i}(u)$. Then $\underline{\theta}$ is non-decreasing and $\{\theta_{n_i}(\cdot)\}$ converges also at points of continuity of $\underline{\theta}$. Let $\{\theta_n^*(\cdot)\}$ be a subsequence of $\{\theta_{n_i}(\cdot)\}$ converging also at discontinuity points of $\underline{\theta}$. Then $\{\theta_n^*(\cdot)\}$ converges everywhere to a non-decreasing function $\theta(\cdot)$ with range in $[a, b]$. Set $D = \{y: \theta^{-1}(y)$ is a non-degenerate interval$\}$. D is countable; therefore by (4.4) there exists a set N such that $P[X \varepsilon N] = 0$ and such that if $x \not\varepsilon N$, $\phi(x, \cdot)$ is continuous at y, for all $y \varepsilon D$. We have $\phi(x, \theta_n^*(u)) \to \phi(x, \theta(u))$ unless $\theta(u)$ is a point of discontinuity of $\phi(x, \cdot)$. For fixed $x \not\varepsilon N$, the collection of points, u, such that $\phi(x, \cdot)$ is discontinuous at $\theta(u)$ is countable. Thus for $x \not\varepsilon N$, $\phi(x, \theta_n^*(u)) \to \phi(x, \theta(u))$ for almost all (Lebesgue measure) u, for each τ

$$\textstyle\int_{\tau_0}^{\tau} \phi(x, \theta_n^*(u))\, du \to \int_{\tau_0}^{\tau} \phi(x, \theta(u))\, du.$$

Further,

$$\int \exp[-\textstyle\int_{\tau_0}^{\tau} \phi(x, \theta_n^*(u))\, du]\, dF(x) \to \int \exp[-\int_{\tau_0}^{\tau} \phi(x, \theta(u))\, du]\, dF(x).$$

The last two convergences are justified by the bounded convergence theorem.

If $(\tau - \tau_0)/h_0$ is a binary rational, $\int \exp [-\int_{\tau_0}^{\tau} \phi(x, \theta_n^*(u))\, du]\, dF(x) = 1$ for n sufficiently large. It follows that for such τ,

$$\int \exp [-\int_{\tau_0}^{\tau} \phi(x, \theta(u))\, du]\, dF(x) = 1.$$

Since such τ are dense in the reals and since $\int_{\tau_0}^{\tau} \phi(x, \theta(u))\, du$ is continuous in τ, we have $\int \exp [-\int_{\tau_0}^{\tau} \phi(x, \theta(u))\, du]\, dF(x) = 1$ for all τ.

Now suppose (4.7) holds. Since $a \leqq \theta_n(u) \leqq b$ for all n and all $u \,\varepsilon\, R$, we have $a \leqq \theta(u) \leqq b$ for all u. Also, if $\theta(\tau_1) = b$ for some $\tau_1 > \tau_0$ then $\theta(u) = b$ for $u \geqq \tau_1$. But for $x < b$, $\phi(x, b) > 0$, so that for $\tau > \tau_1$,

$$\begin{aligned}\int \exp [-\textstyle\int_{\tau_0}^{\tau} \phi(x, \theta(u))\, du]\, dF(x) & \\ = \int \exp [-(\tau - \tau_1)\phi(x, b)] \exp [-\textstyle\int_{\tau_0}^{\tau_1} \phi(x, \theta(u))\, du]\, dF(x) & \\ < \int \exp [-\textstyle\int_{\tau_0}^{\tau_1} \phi(x, \theta(u))\, du]\, dF(x) = 1, \qquad \text{a contradiction.} &\end{aligned}$$

Thus $\theta(u) < b$ and similarly $\theta(u) > a$ for all u.

Still assuming (4.7), we now show that $\theta(u)$ is strictly increasing. Suppose the contrary, that there are an interval (c, d) and a number k such that $\theta(u) = k$ for $c < u < d$. Let $\tau_1, \tau \,\varepsilon\, (c, d)$. Then

$$\begin{aligned}\exp \{-\textstyle\int_{\tau_0}^{\tau} \phi(x, \theta(u))\, du\} &= \exp \{-\textstyle\int_{\tau_0}^{\tau_1} \phi(x, \theta(u))\, du - \int_{\tau_1}^{\tau} \phi(x, \theta(u))\, du\} \\ &= \exp \{-(\tau - \tau_1)\phi(x, k)\} \exp \{-\textstyle\int_{\tau_0}^{\tau_1} \phi(x, \theta(u))\, du\}.\end{aligned}$$

Thus if X is a random variable having density $\exp \{-\int_{\tau_0}^{\tau_1} \phi(x, \theta(u))\, du\}$ with respect to dF, the moment generating function of $\phi(X, k)$ is identically 1 in a neighborhood of the origin. Hence $\phi(X, k) = 0$ almost surely, and by (4.7), $X = k$ almost surely, with respect to that distribution. Thus

$$\exp \{-\textstyle\int_{\tau_0}^{\tau_1} \phi(x, \theta(u))\, du\} = 0$$

for almost all $x(dF)$ different from k. But since the exponential is strictly positive for all x, this implies F is degenerate, contradicting (4.2).

We have yet to show that under the additional hypothesis (4.8), $\theta(\cdot)$ is continuous. Set $\Psi(x, \tau) = \int_{\tau_0}^{\tau} \phi(x, \theta(u))\, du$. The derivative from the right with respect to τ is $\psi(x, \tau) = \phi(x, \theta(\tau+))$. Suppose θ is discontinuous at τ_1. Then by (4.8)

$$\begin{aligned}0 &< \int \{\phi(x, \theta(\tau_1+)) - \phi(x, \theta(\tau_1-))\} \exp [-\Psi(x, \tau_1)]\, dF(x) \\ &= \int \phi(x, \theta(\tau_1+))\{\exp [-\Psi(x, \tau_1)] - \exp [-\Psi(x, \tau_1 + h)]\}\, dF(x) \\ &\quad + \int \phi(x, \theta(\tau_1+)) \exp [-\Psi(x, \tau_1 + h)]\, dF(x) \\ &\quad + \int \phi(x, \theta(\tau_1-))\{\exp [-\Psi(x, \tau_1 - h)] - \exp [-\Psi(x, \tau_1)]\}\, dF(x) \\ &\quad - \int \phi(x, \theta(\tau_1-)) \exp [-\Psi(x, \tau_1 - h)]\, dF(x).\end{aligned} \tag{4.9}$$

Since ϕ is non-decreasing in its second argument, we have from the first part of Theorem 4.1

$$\begin{aligned}\int \phi(x, \theta(\tau_1+)) \exp [-\Psi(x, \tau_1 + h)]\, dF(x) & \\ = \int \psi(x, \tau_1) \exp [-\Psi(x, \tau_1 + h)]\, dF(x) \leqq 0 &\end{aligned}$$

and similarly

$$-\int \phi(x, \theta(\tau_1-)) \exp[-\Psi(x, \tau_1 - h)]\, dF(x)$$
$$\leqq -\int \phi[x, \theta[(\tau_1 - h/2)+]] \exp[-\Psi(x, \tau_1 - h)]\, dF(x)$$
$$= -\int \psi(x, \tau_1 - h/2) \exp[-\Psi(x, \tau_1 - h)]\, dF(x) \leqq 0.$$

Thus the second and fourth terms in (4.9) are non-positive. From the boundedness of ϕ we see that $\Psi(x, \tau)$ is for each fixed x continuous at $\tau = \tau_1$, and that this continuity is uniform in x. This and the boundedness of ϕ show that the first and third terms in (4.9) converge to zero as $h \downarrow 0$. Taking the limit in (4.9) as $h \downarrow 0$ thus gives the contradiction

$$0 < \int \{\phi(x, \theta(\tau_1+)) - \phi(x, \theta(\tau_1-))\} \exp[-\Psi(x, \tau_1)]\, dF(x) \leqq 0.$$

Therefore $\theta(\cdot)$ must be continuous and the proof of the theorem is complete.

The reader's attention is called to results on asymptotic efficiency appearing in [2]. It is of interest too to note that a "formal" calculation of the Frechet-Cramér-Rao lower bound for the asymptotic variance of an unbiased estimator of $\theta(\tau)$ in sampling from the distribution (4.6) yields the asymptotic variance of the sample ϕ-mean given by Corollary 3.4.

A first example of ϕ-families of distributions is furnished by the well known exponential families. If $\phi(x, \theta) = \theta - x$, the density generated by ϕ and F (setting $\tau_0 = 0$) is $\exp[x\tau - \Theta(\tau)]$, where $\exp \Theta(\tau) = \int \exp(x\tau)\, dF(x)$. The most general one parameter exponential family for which a sample of arbitrary size has a single sufficient statistic is obtained through change of parameter and change of variable: $\exp\{g(y)\rho(\alpha) - \theta[\rho(\alpha)]\}$.

If $h(x) > 0$ on the support of dF, and if $\phi(x, \theta) = h(x)(\theta - x)$, then the family generated by ϕ and F has densities $f(x, \tau) = \exp h(x)[x\tau - \Theta(\tau)]$, where $\Theta(\tau) = \int_0^\tau \theta(u)\, du$, the existence of $\theta(\cdot)$ being guaranteed by Theorem 4.2. The ϕ-mean is $\theta(\tau)$ which is also the ordinary mean of a distribution having density proportional to $h(x)f(x, \tau)$ with respect to dF. More generally, if $\phi(x, \theta)$ is of the form $\sum_{j=0}^k h_j(x) g_j(\theta)$, the family given by Theorem 4.2 is an ordinary exponential family.

We obtain a translation family if $\phi(x, \theta) = -g(x - \theta)$ where g is non-decreasing, $\operatorname{sgn} g(u) = \operatorname{sgn} u$. Suppose $G'(x) = g(x)$ and $\int \exp[-G(x)]\, dx = 1$ (as may be achieved by addition of a constant to G, if the integral is finite). Setting $\tau_0 = 0$, $\theta(u) = u$, and $F(x) = \int_{-\infty}^x \exp[-G(u)]\, du$ one obtains the densities $\exp[-G(x - \tau) + G(x)]$ with respect to F, or equivalently the densities $\exp[-G(x - \tau)]$ with respect to Lebesgue measure. If $r \geqq 0$ is fixed and $g(x) = |x|^r \operatorname{sgn} x$, then we obtain the family of densities $C_r \exp\{-|x - \tau|^{r+1}/(r + 1)\}$ with respect to Lebesgue measure where C_r is a normalizing constant; τ is the r-mean of a random variable with this density.

We now consider a specific instance of a ϕ-family, not exponential, which might serve as a reasonable model in certain situations. Suppose one is concerned with strictly positive populations, approximately normal, with mean proportional to standard deviation. An appropriate model might appear to be the normal with mean $\lambda > 0$ and variance $\sigma^2\lambda^2$ (σ a known positive constant), truncated at 0. Its density with respect to Lebesgue measure is $\exp\{-(x - \lambda)^2/2\sigma^2\lambda^2\}/(2\pi)^{\frac{1}{2}}\Phi(1/\sigma)\sigma\lambda$, $x \geqq 0$, where Φ is the cumulative standardized normal distribution function. We show that this family of densities is a ϕ-family with $\phi(x, \theta) = (2\theta/c) + (x/\sigma^2) - cx^2/2\theta\sigma^2 = 2(\theta - x)[1 + (c^2x)/(4\sigma^2\theta)]/c$, where $c = (4\sigma^2 + 1)^{\frac{1}{2}} + 1$; $\theta(u) = -c/2u$, $u < 0$; $\tau(\theta) = -c/2\theta$, $\theta > 0$; $dF(x) = [\exp\{-(x - 1/\sigma)^2/2\}/(2\pi)^{\frac{1}{2}}\Phi(1/\sigma)]\,dx$ for $x \geqq 0$. We have

$$\Psi(x, \tau) = \int_{-\sigma}^{\tau} \phi(x, \theta(u))\,du = (x\tau + 1)^2/2\sigma^2 - (x - 1/\sigma)^2/2 - \ln(-\tau) + \ln\sigma, \quad \tau < 0;$$

$$f(x, \tau) = \exp[-\Psi(x, \tau)] = -\tau/\sigma \exp\{-(x\tau + 1)^2/2\sigma^2 + (x - 1/\sigma)^2/2\},$$

$$g(x, \theta) = f(x, \tau(\theta)) = (c/2\sigma\theta) \exp\{-(x - 2\theta/c)^2c^2/8\theta^2\sigma^2 + (x - 1/\sigma)^2/2\}.$$

If we set $\lambda = 2\theta/c$, then

$$g(x, \theta)\,dF(x) = \exp\{-(x - \lambda)^2/2\sigma^2\lambda^2\}\,dx/(2\pi)^{\frac{1}{2}}\Phi(1/\sigma)\sigma\lambda\,dx, \qquad x > 0.$$

(Thus the ϕ-mean is $\theta = c\lambda/2$.)

REFERENCES

[1] Ando, T. and Amemiya, I. (1965). Almost everywhere convergence of prediction sequences in $L_p(1 < p < \infty)$. Zeit. Wahr. **4** 113–120.

[2] Brøns, H. K. (1967). *Forelaesningsnoter i Matematisk Statistik og Sandsynlighedsregning.* Institut for Matematisk Statistik, Københavns Universitet.

[3] Brunk, H. D. (1965). Conditional expectation given a σ-lattice and applications. *Ann. Math. Statist.* **36** 1339–1350.

[4] Gentleman, (1965). Unpublished Ph.D. thesis, Princeton University.

[5] Gini, Corrado (1958). *Le Medie.* Unione Tipografico-Editrice Torinese.

[6] Guenther, W. C. (1967). A best statistic with variance not equal to the Cramér-Rao lower bound. *Amer. Math. Monthly* **74** 993–994.

[7] Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

[8] Loève, Michel (1963). *Probability Theory* (3rd Ed.). Van Nostrand, New York.

[9] Nikol'skiĭ, V. N. (1963). Best approximation by elements of a convex set in a normed linear space. (Russian) *Kalinin Gos. Ped. Inst. Učen. Zap.* **29** 85–119.

[10] Rosenblatt, M. (1952). On the oscillation of sums of random variables. *Trans. Amer. Math. Soc.* **72** 165–178.