PREPRINT

AUG 1971 2

H. K. Brøns

MEAN-VALUES

UNIVERSITY OF COPENHAGEN
INSTITUTE OF
MATHEMATICAL STATISTICS

MEAN-VALUES

H.K.Brøns

Preprint 1971 No. 2
Preprint 1971 No. 2

INSTITUTE OF MATHEMATICAL STATISTICS

UNIVERSITY OF COPENHAGEN

August 1971

Contents.

INTRODUCTION

The computation of averages is a basic tool in statistics, but often the statistician uses other methods of forming "mean-values" e.g. by calculating transformed averages, medians, midranges, etc. In spite of the central role these quantities play in statistical work no attempt has been made to give a comprehensive treatment of their mathematical properties. As a result the ordinary average or mean-value, generalized to the mathematical expectation, dominates the theory in a way that does not look natural from the viewpoint of the practical statistician.

Still there does exist a theory of mean-values. The main contributions to this theory are papers by Kolmogorov [1930]and Nagumo[1930] . A review of the present state of the theory is found in Aczél [1961] . The purpose of the theory is to show that the ordinary average or simple transforms of it are the only functions satisfying axioms of a certain kind, and the effort has been concentrated on weakening these assumptions as much as possible. Because of the emphasis on the average the theory is only of limited interest to the statistician.

It also seems unnatural to reserve the word mean-value for these functions. Intuitively the main quality of a mean-value should be some kind of "inbetweenness"-property, but in Kolmogorov's and Nagumo's axioms a rule of combination (called associativity by de Finetti) is the fundamental assumption.

In a most inspiring paper de Finetti [1931] tried to broaden the point of view, but his intentions have not been followed up by any research.

What is then a natural concept of a mean-value ? First it is a real number that can be computed from any finite set of real numbers. Sometimes one also requires that the mean of a single number is the number itself. Finally the mean-value of the set of numbers must lie inbetween the smallest and the biggest number in the set, i.e. if $x_1,\dots,x_n$, $n = 1,2,\dots$, are real numbers, and m the mean-value then

$$\min(x_1,\dots,x_n) \leq m(x_1,\dots,x_n) \leq \max(x_1,\dots,x_n).$$

Evidently this property is much too weak to give rise to any interesting mathematical theory.

The following treatment of mean-values will be based on the definition

$$(1)\quad \min(m(x_1,\dots,x_k),m(x_{k+1},\dots,x_n)) \leqq m(x_1,\dots,x_n) \leqq \max(m(x_1,\dots,x_k),m(x_{k+1},\dots,x_n)),$$

$k = 1,\dots,n$. These inequalities state that if two samples of real numbers are combined into one sample then the mean-value of the total sample will lie between the mean-values of the subsamples.

If a denotes ordinary average:

$$a(x_1,\dots,x_n) = \frac{x_1+\dots+x_n}{n}$$

then

$$a(x_1,\dots,x_n) = \frac{k}{n}a(x_1,\dots,x_k) + \frac{n-k}{n}a(x_{k+1},\dots,x_n),$$

which shows that a satisfies (1).

The following is a preliminary report on basic properties of mean-values. The statistical applications will be treated later.

Part of the work was carried out during the summer 1962 while the author was visiting University of California, Riverside, on a grant from NATO Science Fellowship Programme.

The present paper has appeared in two preliminary versions in 1963 and 1964. This final version is identical with the 1964 version except for a few necessary changes and the addition of proofs for the theorems in sections II and III.

I. Sample mean-values.

1. The definition and simple examples of sample means.

Let X be an abstract set. The *sample space* of X is the set of all ordered finite subsets of X, i.e. the set

$$X^* = \bigcup_{k=1}^{\infty} X^k$$

where X^k, $k = 1,2,\dots,$ denotes the k'th cartesian power of X. The points in X^* are *samples* from X. If

$$x^* = (x_1,\dots,x_n) \in X^*$$

and

$$y^* = (y_1,\dots,y_n) \in X^*$$

then x^*y^*denotes the *combined sample*

$$(x_1,\dots,x_n, y_1,\dots,y_n).$$

A *sample function on X* is a numerical function on X^*, i.e. a mapping from X^* into the extended real line R'.

A *mean-value (mean) on X* is a sample function m on X that satisfies the conditions

$$\min(mx^*,my^*) \leqq mx^*y^* \leqq \max(mx^*,my^*)$$

for all $x^* \in X^*$ and $y^* \in X^*$.

It follows from this definition that if $x_i^* \in X^*$, $i = 1,\dots,k$, $k = 1,2,\dots,$ and

$$mx_1^* = \dots = mx_k^* ,$$

then

$$mx_1^* \, x_2^* \dots x_k^* = mx_1^* .$$

Also if f is a monotone numerical function on R' then fm is a mean-value on X.

If $Y \subset X$ then the restriction of m to Y^* is a mean-value on Y.

Example 1. <u>The average.</u> The ordinary average a is a mean-value on R' defined by

$$a(x^*) = \begin{cases} \frac{1}{n}(x_1 + \dots + x_n) & \text{when meaningful} \\ 0 & \text{elsewhere,} \end{cases}$$

$x^* = (x_1,\dots,x_n) \in R'^*$. Instead of 0 one could have used any fixed real number.

Example 2. <u>Max and min.</u> The function max defined on R'^* by

$$\max(x^*) = \max(x_1,\dots,x_n)$$

$x^* = (x_1,\dots,x_n) \in R'^*$ is clearly a mean-value on R'. Similarly for min.

For $x^* = (x_1,\dots,x_n) \in R'^*$ let $x^*_{(k)}$, $k = 1,\dots,n$, denote the k'th smallest among $x_1,\dots,x_n$.

Example 3. <u>The medians.</u> The sample functions m_0 and m_1 defined by

$$m_0 x^* = \begin{cases} x^*_{\left(\left[\frac{n}{2}\right]+1\right)} & \text{for n uneven} \\ x^*_{\left(\frac{n}{2}\right)} & \text{for n even} \end{cases}$$

and

$$m_1 x^* = x^*_{\left(\left[\frac{n}{2}\right]+1\right)},$$

$x^* = (x_1,\dots,x_n) \in R'^*$, are mean-values on R'. $[x]$ is the integral part of x.

Since

$$m_0 x^* = -m_1(-x^*), \qquad x^* \in R'^*,$$

it is sufficient to prove that m_1 is a mean-value.

Put

$$x^* = (x_1,\dots,x_n) \in R'^*$$

and

$$y^* = (y_1,\dots,y_n) \in R'^*.$$

To the left of the point $\min(m_1 x^*, m_1 y^*)$ there are at most

$$\left[\frac{n}{2}\right] + \left[\frac{s}{2}\right] < \left[\frac{n+s}{2}\right] + 1$$

points of the combined sample x^*y^*, i.e.

$$\min(m_1 x^*, m_1 y^*) \leqq m_1 x^*y^* .$$

To the left of or in the point $\max(m_1 x^*, m_1 y^*)$ there are at least

$$\left[\frac{n}{2}\right] + 1 + \left[\frac{s}{2}\right] + 1 \geqq \left[\frac{n+s}{2}\right] + 1$$

points of the combined sample, i.e.

$$m_1 x^*y^* \leqq \max(m_1 x^*, m_1 y^*).$$

By the <u>ordered sample</u> corresponding to

$$x^* = (x_1,\dots,x_n) \in R'^*$$

is meant the sample

$$(x^*_{(1)},\dots,x^*_{(n)}).$$

All the mentioned examples of mean-values are symmetric, i.e. depend only on the ordered sample. In a general set a mean-value is symmetric if it takes the same value in all samples which are permutations of the sample.

The R'-mean-values

$$m'(x_1,\dots,x_n) = x_1$$

and

$$m''(x_1,\dots,x_n) = x_n$$

are not symmetric.

Example 4. The linear mean-values on R.

We want to determine all mean-values m on R that are linear i.e. of the form

$$m(x_1,\dots,x_n) = \sum_{i=1}^{n} a_{ni}x_i, \quad (x_1,\dots,x_n) \in R^n ,$$

$a_{ni} \in R$, $i = 1,\dots,n$, $n = 1,2,\dots$.

The inequalities

$$\min(m1,0) \leqq m(\overbrace{0,\dots,0}^{i-1},1,\overbrace{0,\dots,0}^{n-i}) \leqq \max(m1,0)$$

or

$$\min(a_{11},0) \leqq a_{ni} \leqq \max(a_{11},0)$$

show that if $a = a_{11}$ is zero then all the coefficients are zero, and if $a \neq 0$ then all have the same sign.

For $a \neq 0$ the general case is reduced to the case $a = 1$ by multiplication by a^{-1}. Put

$$x = m(\overbrace{0,\dots,0}^{i-2},1,\overbrace{0,\dots,0}^{n-i}) = a_{n-1,i-1} ,$$

$i = 2,\ldots,n$, $n = 2,3,\ldots$. The equation

$$x = m(x,\overbrace{0,\ldots,0}^{i-2},1,\overbrace{0,\ldots,0}^{n-i})$$

is equivalent to

$$a_{n-1,i-1} = a_{n1}a_{n-1,i-1} + a_{ni}$$

or

$$(1) \qquad a_{ni} = a_{n-1,i-1}(1-a_{n1}),\ i = 2,\ldots,n.$$

Similarly the equation

$$x = m(\overbrace{0,\ldots,0}^{i-2},1,\overbrace{0,\ldots,0}^{n-i},x)$$

gives

$$a_{n-1,i-1} = a_{n,i-1} + a_{nn}a_{n-1,i-1}$$

or

$$a_{n,i-1} = a_{n-1,i-1}(1-a_{nn}),\ i = 2,\ldots,n.$$

For $i = 2$ here and $i = n$ in (1) one gets

$$a_{nn} = a_{n-1,n-1}(1-a_{n1})$$

and

$$a_{n1} = a_{n-1,1}(1-a_{nn})$$

so that

$$(2) \qquad a_{n1}(1-a_{n-1,1}a_{n-1,n-1}) = a_{n-1,1}(1-a_{n-1,n-1}).$$

The equation

$$m(1,1) = m1$$

is equivalent to

$$a_{21} + a_{22} = 1.$$

In the case $a_{21} > 0$ put

$$a_{21} = \frac{1}{1+b}\ , \quad b \geqq 0$$

so that

$$a_{22} = \frac{1}{1+b}\ .$$

It follows by induction from (1) and (2) that

$$a_{ni} = \frac{b^{i-1}}{1+b+\ldots+b^{n-1}}$$

and for general a:

$$a_{ni} = \frac{ab^{i-1}}{1+b+\ldots+b^{n-1}}\ , \quad i = 1,\ldots,n,\ n = 1,2,\ldots\ .$$

With these coefficients m satisfies the equality

$$m(x_1,\ldots,x_n) = \frac{1+b+\ldots+b^{k-1}}{1+b+\ldots+b^{n-1}}\, m(x_1,\ldots,x_k) + \frac{b^k+\ldots+b^{n-1}}{1+b+\ldots+b^{n-1}}\, m(x_{k+1},\ldots,x_n),$$

which shows that m is a mean.

For $a_{21} = 0$ $(b = +\infty)$ one finds

$$a_{nn} = a,\ a_{ni} = 0,\ i = 1,\ldots,n-1, \quad n = 1,2,\ldots,$$

which corresponds to the mean

$$m(x_1,\ldots,x_n) = ax_n\ .$$

Example 5. The quantiles. We want to determine all mean-values on R' of the form

$$mx^* = x^*_{(r_n)}, \quad x^* = (x_1,\ldots,x_n) \in R'^*,$$

$r_n = 1,2,\dots,n$. For $a < b$, $a \in R$, $b \in R$, consider the number

$$m(\overbrace{a,\dots,a}^{s},\overbrace{b,\dots,b}^{n-s})$$

$s = 0,1,\dots,n$, $n = 1,2,\dots$. This number depends only on $\frac{s}{n}$ since

$$m(\overbrace{a,\dots,a}^{ks},\overbrace{b,\dots,b}^{k(n-s)}) = m(\overbrace{a,\dots,a}^{s},\overbrace{b,\dots,b}^{n-s})$$

for $k = 1,2,\dots,$ because m is symmetric. The function g on the rational numbers in $[0,1]$ defined by

$$g(\tfrac{s}{n}) = m(\overbrace{a,\dots,a}^{s},\overbrace{b,\dots,b}^{n-s}),$$

$s = 0,\dots,n$, $n = 1,2,\dots$, is decreasing.

First

$$a = ma \leqq g(\tfrac{s}{n}) \leqq mb = b.$$

Next

$$\min(ma, g(\tfrac{s}{n-1})) \leqq g(\tfrac{s+1}{n}) \leqq \max(ma, g(\tfrac{s}{n-1}))$$

or

$$ma \leqq g(\tfrac{s+1}{n}) \leqq g(\tfrac{s}{n-1}) \;,$$

$s = 0,\dots,n-1$, $n = 2,3,\dots$. Also

$$\min(mb, g(\tfrac{s}{n-1})) \leqq g(\tfrac{s}{n}) \leqq \max(mb, g(\tfrac{s}{n-1}))$$

or

$$g(\tfrac{s}{n-1}) \leqq g(\tfrac{s}{n}) \leqq mb.$$

This shows that

$$g(\tfrac{s+1}{n}) \leqq g(\tfrac{s}{n}) \;,$$

$s = 0,\dots,n-1$, $n = 2,3,\dots$, which is sufficient.

Now g is given by

$$g(\tfrac{s}{n}) = \begin{cases} b & \text{for } s < r_n \\ a & \text{for } s \geqq \dfrac{r_n-1}{n} \end{cases}$$

or equivalently

$$g(r) \begin{cases} b & \text{for } r \leqq \dfrac{r_n-1}{n} \\ a & \text{for } r \geqq \dfrac{r_n}{n} \end{cases}$$

$r \in [0,1]$, r rational. Necessary and sufficient for this to define an decreasing function of r is that

(3) $$\frac{r_s}{s} > \frac{r_n-1}{n}$$

for $s = 1,2,\dots$, $n = 1,2,\dots$.

From (3) follows

$$\left|\frac{r_s}{s} - \frac{r_n}{n}\right| \leqq \frac{1}{n} \quad \text{for } s \geqq n,$$

so that $\dfrac{r_n}{n}$ for $n \to \infty$ has a limit $p \in [0,1]$.

From (3) one gets

(4) $$\frac{r_n-1}{n} \leqq p \leqq \frac{r_n}{n}$$

by letting $s \to \infty$ for fixed n and vice versa.

(4) can be written

$$np \leqq r_n \leqq np + 1, \quad n = 1,2,\dots .$$

For p irrational $r_n = [np] + 1, \quad n = 1,2,\dots$.

For $p = 0$ $\quad r_n = 1, \quad n = 1,2,\dots$.

For $p = 1$ $\quad r_n = n, \quad n = 1,2,\dots$.

For $p \in]0,1[$, p rational, there are two solutions

$$r_n = \begin{cases} np \text{ for } [np] = np \\ [np] + 1 \text{ for } [np] < np \end{cases}$$

and

$$r_n = [np]+ 1,$$

since $r_s = sp$ and $r_n = np + 1$ gives

$$\frac{r_s}{s} = \frac{r_n - 1}{n},$$

in contradiction with (3). That all these sample functions are mean-values is proved as for the medians.

2. Construction of mean-values by minimalization.

Let f be a numerical function on a subset A of a linear space. f is said to be <u>quasi-convex</u> on A if

$$f((1-\lambda)x_0 + \lambda x_1) \leqq \max(fx_0), f(x_1))$$

for $x_0 \in A$, $x_1 \in A$, $(1-\lambda)x_0 + \lambda x_1 \in A$, $\lambda \in [0,1]$. f is <u>quasi-concave</u> on A if under the same conditions

$$f((1-\lambda)x_0 + \lambda x_1) \geqq \min(f(x_0), f(x_1)).$$

If f is both quasi-convex and quasi-concave on A, it is <u>monotone</u> on A.

A numerical function defined on a convex subset A of a linear space is quasi-convex if and only if the set of all $x \in A$ for which

$$f(x) \leqq \tau$$

is convex for all $\tau \in R'$.

Theorem 1. Let F be an arbitrary family of quasi-convex functions defined on the convex subset $\mathcal{A}$ of a linear space. Then the supremum

$$g(x) = \sup_{f \in F} f(x), \quad x \in \mathcal{A}$$

is quasi-convex.

Proof. For $\tau \in R'$ is

$$\mathcal{B} = \{x | g(x) \leqq \tau\} = \bigcap_{f \in F} \{x | f(x) \leqq \tau\}$$

which shows that $\mathcal{B}$ is convex.

A numerical function on a convex subset $\mathcal{A}$ of a linear space is <u>convex</u> if

$$f((1-\lambda)x_0 + \lambda x_1) \leqq (1-\lambda)f(x_0) + \lambda f(x_1)$$

for all $x_0 \in \mathcal{A}$, $x_1 \in \mathcal{A}$, $\lambda \in]0,1[$, such that the right hand side is defined. f is <u>concave</u> if $-f$ is convex. If f is both convex and concave it is <u>affine</u>.

A convex function is quasi-convex.

Let now f be a quasi-convex function on an interval $J \subset R$.

A point $\theta \in J$ is a <u>point of decrease</u> for f if $f(\theta') \geqq f(\theta)$ for all $\theta' \leqq \theta$, which belongs to J. J's left hand endpoint is a point of decrease if it belongs to J.

Let θ be a point of decrease for f and let $\theta_1 \leqq \theta$ be a point in J. If $\theta' \leqq \theta_1$ is a point in J, then

$$f(\theta_1) \leqq \max(f(\theta'), f(\theta))$$

implies $f(\theta_1) \leqq f(\theta')$, i.e. θ_1 is a point of decrease for f.

Hence the set S_f of all points of decrease for f is an interval with the same left hand endpoint as J. S_f is possibly empty.

A point $\theta \in J$ is a <u>point of increase</u> for f if $f(\theta') \geqq f(\theta)$ for all $\theta' \geqq \theta$, which belong to J. J's right hand endpoint is a point of increase if it belongs to J. Analogously to the argument above it is proved that the set D_f is empty or an interval with same right hand endpoint as J.

We have that

$$S_f \cup D_f = J, \tag{2}$$

for if $\theta \in J \setminus (S_f \cup D_f)$ there exist points θ' and θ'' in J such that

$$\theta' < \theta < \theta'', \; f(\theta') < f(\theta), f(\theta'') < f(\theta)$$

which is in contradiction with (1).

Now define

$$m_0 f = \inf D_f, \quad m_1 f = \sup S_f,$$

$m_0 f$ is J's right hand endpoint when D_f is empty, and $m_1 f$ is J's left hand endpoint if S_f is empty.

(2) shows that $m_0 f \leqq m_1 f$.

If $m_0 f$ and $m_1 f$ are equal we denote their common value by mf.

The set $S_f \cap D_f$ is a (possibly empty) interval with endpoints $m_0 f$ and $m_1 f$. It consists of all points in which f assumes its minimum value.

f is decreasing on S_f and increasing on D_f. It is therefore immediate that a necessary and sufficient condition for a numerical function f on an interval $J \subset R$ to be quasi-convex is that J is the union of two disjoint (possibly empty) intervals such that f is decreasing on the left interval and increasing on the right.

A family F of numerical functions on an interval $J \subset R$ is said to be <u>quasi-convex under addition</u> if for all $(f_1, \ldots, f_n) \in F^*$ the function

$$f_1 + \ldots + f_n$$

is defined, quasi-convex, and not identically equal to $+\infty$.

A family of finite convex functions on J has this property.

Theorem 2. Let to every $x \in X$ correspond a numerical function f_x on the interval $J \subset R$ such that the family $\{f_x\}$ is quasi-convex under addition. The sample functions

$$m_0(f_{x_1}+\dots+f_{x_n}) \ , \ m_1(f_{x_1}+\dots+f_{x_n}), \ (x_1,\dots,x_n) \in X^*,$$

are then symmetric mean-values on X.

Proof. Let g_0 and g_1 be quasi-convex functions on J such that $g = g_0 + g_1$ is defined and quasi-convex on J. It is assumed that neither g_0, g_1, nor g are identically equal to $+\infty$.

Let

$$\theta \in S_{g_0} \cap S_{g_1} \ .$$

If $\theta' \in J$ and $\theta' < \theta$ then

$$g_0(\theta') \geqq g_0(\theta)$$

$$g_1(\theta') \geqq g_1(\theta).$$

By addition one gets

$$g(\theta') \geqq g(\theta),$$

i.e. θ belongs to S_g, so

$$S_{g_0} \cap S_{g_1} \subset S_g . \qquad (3)$$

Let now

$$\theta \in (J \setminus S_{g_0}) \cap (J \setminus S_{g_1}).$$

There exist points θ' and θ'' in J such that

$$g_0(\theta') < g_0(\theta) \qquad \theta' < \theta \ , \qquad (4)$$

$$g_1(\theta'') < g_1(\theta), \qquad \theta'' < \theta \ . \qquad (5)$$

Assume $\theta' \geqq \theta''$. From the inequality $g_1(\theta') \leqq \max(g_1(\theta''), g_1(\theta))$ and (5) it follows that

$$(6) \qquad g_1(\theta') \leqq g_1(\theta).$$

If $g_1(\theta)$ is finite, addition of (4) and (6) gives $g(\theta') < g(\theta)$, the condition for θ to belong to $J \setminus S_g$.

Since $\theta \in J \setminus S_{g_1}$, it is impossible that $g_1(\theta) = -\infty$.

If $g_1(\theta) = +\infty$, then $g_1(\theta_1) = +\infty$, $\theta_1 \geqq \theta$, $\theta_1 \in J$, since $\theta \in D_{g_1}$. It follows that

$$g(\theta_1) = +\infty$$

for all $\theta_1 \geqq \theta$. g is not identically equal to $+\infty$. There must therefore exist a point $\theta_2 \in J$, such that $g(\theta_2) < g(\theta)$, $\theta_2 < \theta$, i.e. $\theta \in J \setminus S_g$.

We have thus shown that

$$J \setminus S_g \subset (J \setminus S_{g_0}) \cap (J \setminus S_{g_1})$$

or

$$(7) \qquad S_g \subset S_{g_0} \cup S_{g_1}.$$

It follows from (3) and (7) that $\min(m_1 g_0, m_1 g_1) \leqq m_1 g \leqq \max(m_1 g_0, m_1 g_1)$. Similarly one finds that $\min(m_0 g_0, m_0 g_1) \leqq m_0 g \leqq \max(m_0 g_0, m_0 g_1)$.

The theorem is proved by applying these results to the functions

$$g_0 = f_{x_1} + \ldots + f_{x_k}, \quad g_1 = f_{x_{k+1}} + \ldots + f_{x_n}.$$

Example 6. The average a on R is the m-function of the family of finite convex functions on R defined by

$$f_x(\theta) = (x-\theta)^2, \quad x \in R, \quad \theta \in R.$$

Example 7. The smallest and the largest p-quantile on R, $p \in]0,1[$ respectively, are m_0- and m_1-functions of the family of finite convex functions on R defined by

$$f_x(\theta) = \begin{cases} \frac{1}{p}(\theta - x) & x \leqq \theta \\ \frac{1}{q}(x-\theta) & x \geqq \theta, \end{cases}$$

$x \in R$, $\theta \in R$.

max is the m-function of the family

$$f_x(\theta) = \begin{cases} \theta - x & x \leqq \theta \\ +\infty & x > \theta, \end{cases}$$

$x \in R$, $\theta \in R$, and min is the m-function of the family

$$f_x(\theta) = \begin{cases} +\infty & x \leqq \theta \\ x - \theta & x > \theta . \end{cases}$$

Corollary. Let to every $x \in X$ correspond a finite numerical function f_x on the interval $J \subset R$ such that the family $\{f_x\}$ is quasi-convex under addition. For $b \in]0,+\infty[$ the function

$$f_{x_1} + bf_{x_2} + b^2 f_{x_3} + \ldots + b^{n-1} f_{x_n}$$

$x^* = (x_1,\ldots,x_n) \in X^*$, is quasi-convex on J, and the sample functions

$$m_0(f_{x_1} + \ldots + b^{n-1} f_{x_n}),\ m_1(f_{x_1} + \ldots + b^{n-1} f_{x_n})$$

are mean-values on X.

Proof. When b is rational it is evident that

$$f_{x_1} + bf_{x_2} + \ldots + b^{n-1} f_{x_n}$$

is quasi-convex and for b irrational it follows by passing to the limit through rational values.

The second assertion is proved by remarking that if f is a quasi-convex function on J then $b^k f$, $k = 0,1,2,\ldots$, is quasi-convex and has the same m_0- and m_1- values as f, and then applying the result from the proof of theorem 2 to the functions

$$g_0 = f_{x_1} + bf_{x_2} + \ldots + b^{k-1} f_{x_k}$$

$$g_1 = b^k (f_{x_{k+1}} + bf_{x_{k+2}} + \ldots + b^{n-k-1} f_{x_n})$$

$$g = f_{x_1} + bf_{x_2} + \ldots + b^{n-1} f_{x_n} = g_0 + g_1 .$$

The corollary shows how to construct unsymmetric mean-values.

Example 8. The R-mean-value

$$m(x^*) = \frac{x_1 + bx_2 + \ldots + b^{n-1} x_n}{1 + b + \ldots + b^{n-1}}$$

$x^* = (x_1,\ldots,x_n) \in R^*$, $b \in]0,+\infty[$, is the m-function computed from the function

$$(x_1-\theta)^2 + b(x_2-\theta)^2 + \ldots + b^{n-1}(x_n-\theta)^2,$$

$\theta \in R$.

A numerical function f on an interval $J \subset R$ is <u>strictly quasi-convex</u> if $f((1-\lambda)\theta_0 + \lambda\theta_1) < \max(f(\theta_0),f(\theta_1))$ for all x_0 and x_1 in J, $x_0 \neq x_1$, $\lambda \in]0,1[$.

f is strictly quasi-convex if and only if it is quasi-convex and not constant on any non-degenerate subinterval of J.

If f is strictly quasi-convex then $m_0 f = m_1 f$.

The maximum of a finite family of strictly quasi-convex functions is strictly quasi-convex.

Theorem 3. Let to every $x \in X$ correspond a strictly quasi-convex function f_x defined on the interval $J \subset R$. The sample function

$$m(\max(f_{x_1},\ldots,f_{x_n}))$$

is then a mean-value on X.

Proof. Let g_0 and g_1 be strictly quasi-convex functions on J, and put $g = \max(g_0,g_1)$. g is then strictly quasi-convex.

Let $\theta \in S_{g_0} \cap S_{g_1}$. If $\theta' \in J$ and $\theta' < \theta$ then

$$g_0(\theta') \geqq g_0(\theta)$$

$$g_1(\theta') \geqq g_1(\theta).$$

It follows that $g'(\theta') \geqq g(\theta)$, i.e. θ belongs to S_g, so

$$S_{g_0} \cap S_{g_1} \subset S_g. \tag{7}$$

By applying this result to the functions

$$h_0(\theta) = g_0(-\theta), \quad h_1(\theta) = g_1(-\theta),$$

$$h(\theta) = \max(h_0(\theta),h_1(\theta)) = g(-\theta), \qquad -\theta \in J,$$

we get

$$D_{g_0} \cap D_{g_1} \subset D_g$$

or

$$S_g \subset S_{g_0} \cup S_{g_1} \tag{8}$$

since $D_g = J \setminus S_g$ etc.

From (7) and (8) it follows that $\min(mg_0,mg_1) \leqq mg \leqq \max(mg_0,mg_1)$, and the theorem is proved by putting

$$g_0 = \max(f_{x_1},\ldots,f_{x_k}), \quad g_1 = \max(f_{x_{k+1}},\ldots,f_{x_n}).$$

$$g = \max(g_0,g_1) = \max(f_{x_1},\ldots,f_{x_n}).$$

$$g_0 = \max(f_{x_1},\ldots,f_{x_k}), \quad g_1 = \max(f_{x_{k+1}},\ldots,f_n),$$

$$g = \max(g_0,g_1) = \max(f_{x_1},\ldots,f_{x_n}).$$

Unsymmetric mean-values can be constructed by considering the m-function for

$$\max(f_{x_1},bf_{x_2},\ldots,b^{n-1}f_{x_n}),$$

where $b \in [0,+\infty[$.

3. Miscellaneous remarks.

Fisher [1925] has noted that if a sample has independent identically distributed components, a sample function which is (minimal) sufficient for some parameter for all sample sizes has the property that the value in a combined sample is a function of the values in the subsample. It is easy to see that this property implies the associativity assumed in Kolmogorov [1930] and Nagumo [1930] . Further all mean-values have the necessary monotonicity such that the Kolmogorov-Nagumo theorem can be restated in the following way:

Under certain (not very strong) regularity conditions a mean-value that is sufficient for some parameter in the case of independent identically distributed observations $x_1,\ldots,x_n$, $n = 1,2,\ldots$, has the form

$$\frac{f(x_1)+\ldots+f(x_n)}{n} .$$

From this it follows easily that the distribution must be of the Darmois-Koopman type.

These problems will be treated at a later occasion.

Spitzer [1956] proved an important combinatorial lemma on sums or ordinary averages. The lemma was used to simplify the proofs of certain theorems due to Sparre Andersen.

It is not difficult to see that the lemma is true for all mean-values. Brunk [1961] published a proof of this and used it to give an easy interpretation of the theorems of Sparre Andersen and to derive some tests for trend.

The following theorem shows that mean-values are "limitierungsprozessen".

Theorem 2. Let m be a mean-value on X. If for $x_n \in X$, $n = 1,2,\dots$, mx_n has a finite limit for $n \to \infty$, then $m(x_1,\dots,x_n)$ is also convergent with a finite limit.

Proof. Put $\lim\limits_{n\to\infty} mx_n = a$ and $m(x_1,\dots,x_n) = y_n$, $n = 1,2,\dots$. $\{y_n\}$ is bounded, since

$$\inf_k mx_k \leqq y_n \leqq \sup_k mx_k$$

for all n. Assume first $\limsup y_n = b > a$. Let $c \in \,]a,b[$. Then there exists an N such that $mx_n < c$ for $n > N$.

To every $n > N$ there corresponds a $k > n$ such that $y_k > c$. Now

$$c < y_k \leqq \max(y_n, mx_{m+1},\dots,mx_k)$$

i.e. $y_n > c$ for $n > N$.

From

$$c < y_{n+1} \leqq \max(y_n, mx_{n+1})$$

it follows that $y_{n+1} \leqq y_n$, i.e. the sequence y_n is decreasing for $n > N$ therefore has a limit.

The case $\liminf y_n < a$ is treated similarly.

II. Extension of mean-values.

1. Symmetric mean-values and monotone functions.

Let again X be an abstract set. The system of all subsets of X will be denoted by $\mathcal{S}$. The indikatorfunction I is a function on X x $\mathcal{S}$ defined by

$$I_x A = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \in A^c, \end{cases}$$

$x \in X$, $A \in \mathcal{S}$. For every $x \in X$ the restriction I_x of I is a probability measure on $\mathcal{S}$ which we shall call the point measure in x. The set of all point-measures is called $\mathcal{I}$.

More general: to every sample x^* from X there corresponds a probability measure E_{x^*} on $\mathcal{S}$ defined by

$$E_{x^*}A = \frac{1}{n}\sum_{i=1}^{n} I_{x_i} A \ ,$$

$x^* = (x_1,\ldots,x_n) \in X^*$, $A \in \mathcal{S}$. This measure is the sample distribution corresponding to x^* or the empirical probability measure. $\mathcal{E}$ is the set of all such measures.

If

$$x^* = (x_1,\ldots,x_n) \in X^*$$

$$y^* = (y_1,\ldots,y_s) \in X^*$$

then

$$E_{x^*y^*} = \frac{n}{n+s} E_{x^*} + \frac{s}{n+s} E_{y^*}.$$

Theorem 1. A sample function is a symmetric mean-value on X if and only if it depends on the sample through a monotone function of the sample distribution.

Proof. Let m be a symmetric mean-value on X, and let x^* and y^* be samples from X of size n and s, respectively, and with a common empirical distribution.

Put

$$z^* = \overbrace{x^*x^* \ldots x^*}^{s}$$

and

$$w^* = \overbrace{y^*y^* \ldots y^*}^{n}.$$

Now $E_{z^*} = E_{x^*}$ and $E_{w^*} = E_{y^*}$. Therefore the samples z^* and w^* have the same empirical distributions, and since they are of the same size w^* is a permutation of z^*. It follows that $mx^* = mz^* = mw^* = my^*$, i.e. m depends on the empirical distribution. Put

$$mx^* = m'E_{x^*},\ x^* \in X^* .$$

Assume

$$v^* = (v_1,\ldots,v_n) \in X^* ,$$

$$u^* = (u_1,\ldots,u_s) \in X^* ,$$

$$E_{x^*} = (1-\lambda)E_{v^*} + \lambda E_{u^*} ,$$

$\lambda \in [0,1]$, then λ must be rational or $E_{v^*} = E_{u^*}$. Put $= \frac{p}{q}$, $p = 0,\ldots,q$, $q = 1,2,\ldots$. The equation

$$E_{x^*} = \frac{q-p}{q}\frac{1}{n}\sum_{i=1}^{n} I_{v_i} + \frac{p}{q}\frac{1}{s}\sum_{j=1}^{s} I_{u_j}$$

$$= \frac{1}{qns}\left(\sum_{i=1}^{n}(q-p)sI_{v_i} + \sum_{j=1}^{s} pnI_{u_j}\right)$$

shows that E_{x^*} is the empirical distribution of the sample

$$w^* = \overbrace{v^*v^*\ldots v^*}^{(q-p)s}\overbrace{u^*u^*\ldots u^*}^{pn} .$$

We thus have

$$\min(mv^*,mu^*) \leqq mw^* \leqq \max(mv^*,mu^*)$$

or

$$\min(m'E_{v^*},m'E_{u^*}) \leqq m'E_{x^*} \leqq \max(m'E_{v^*},m'E_{u^*}).$$

The sufficiency follows from the fact that the empirical distribution of a combined sample is a convex combination of the empirical distribution of the subsamples.

Theorem 1 was proved in a special case in example 5 of chapter I.

In the following we shall consider symmetric mean-values only, and shall therefore omit the adjective symmetric, and use the words mean-value and monotone function synonymously.

A function F on S defined by

$$FA = \sum_{i=1}^{n} \lambda_i I_{x_i} A, \qquad A \in S,$$

$$(x_1,\dots,x_n) \in X^*, \ \lambda_i \geqq 0, \ \sum_{i=1}^{n} \lambda_i = 1,$$

is a probability measure on $\mathcal{S}$. Such a measure is called a <u>finite atomic probability measure</u>, and $\mathcal{F}$ is the set of all these.

To be able to consider more general probability measures one must restrict the domain of definition. Let therefore $\mathcal{A}$ be a sub-σ-algebra of $\mathcal{S}$, i.e. a σ-field of subsets of X containing X, and let $\mathcal{P}_{\mathcal{A}}$ be the set of all probability measures on X with domain of definition including $\mathcal{A}$. If no confusion is possible we shall write $\mathcal{P}$ instead of $\mathcal{P}_{\mathcal{A}}$. We have the relations

$$\mathcal{I} \subset \mathcal{E} \subset \mathcal{F} \subset \mathcal{P}.$$

Two elements in $\mathcal{P}$ are equal if they have the same domain of definition and coincide there. Let $P_0 \in \mathcal{P}$ be defined on $\mathcal{A}_0$ and $P_1 \in \mathcal{P}$ on $\mathcal{A}_1$. For $\lambda \in]0,1[$

$$(1-\lambda)P_0 + \lambda P_1$$

is the element P in $\mathcal{P}$ defined by

$$PA = (1-\lambda)P_0A + \lambda P_1A, \quad A \in \mathcal{A}_0 \cap \mathcal{A}_1 .$$

For $\lambda = 0$ we put

$$P = (1-\lambda)P_0 + \lambda P_1 = P_0,$$

and for $\lambda = 1$

$$P = (1-\lambda)P_0 + \lambda P_1 = P_1.$$

Finite convex combinations

$$\sum_{i=1}^{n} \lambda_i P_i,$$

where $P_1,\dots,P_n$ belong to $\mathcal{P}$ and $\lambda_i \geqq 0$, $\sum_{i=1}^{n} \lambda_i = 1$ are defined by induction in n.

The concepts: convex subset of $\mathcal{P}$, quasi-convex, monotone etc. functions on subsets of $\mathcal{P}$ are now well defined, even if we shall not have occasion to consider $\mathcal{P}$ as a subset of a linear space.

$\mathcal{P}$ is a convex set, and

$$\text{co } \mathcal{I} = \text{co } \mathcal{E} = \mathcal{F},$$

where co stands for convex hull.

If a set $\mathcal{K} \subset \mathcal{P}$ with P_0 and P_1 contains

$$(1-r)P_0 + rP_1$$

for all rational r in $[0,1]$, $\mathcal{K}$ is said to be rationally convex. The rationally convex hull of a set $\mathcal{K}$ is denoted rco $\mathcal{K}$. We have

$$\text{rco } \mathcal{I} = \mathcal{E}.$$

Example. The quantiles. Let f be an $\mathcal{A}$-measurable numerical function. For $P \in \mathcal{P}$ let m_0P denote the smallest p-quantile of f with respect to P, $p \in [0,1]$. m_0 is defined by $m_0P \in \text{cl } f(X)$ (cl = closure)

$$P\{f \leqq a\} < p \quad \text{for } a < m_0 P$$

$$P\{f \leqq a\} \geqq p \quad \text{for } a \geqq m_0 P.$$

m_0 is monotone on $\mathcal{P}$. Let namely $P_0 \in \mathcal{P}, P_1 \in \mathcal{P}, \lambda \in [0,1]$, and put $p = (1-\lambda)P_0 + \lambda P_1$. For every $a < \min(m_0P_0, m_0P_1)$ is

$$P\{f \leqq a\} = (1-\lambda)P_0\{f \leqq a\} + \lambda P_1\{f \leqq a\} < p.$$

It follows that $m_0P > a$, and consequently $m_0P \geqq \min(m_0P_0, m_0P_1)$.

For $b = \max(m_0P_0, m_0P_1)$ is in the same way $P\{f \leqq b\} \geqq p$, and so $m_0P \leqq b$.

The largest p-quantile m_1P, defined by $m_1P \in \mathrm{cl}\, f(X)$, $P\{f < a\} \leqq p$ for $a \leqq m_1P$, $P\{f < a\} > p$ for $a > m_1P$, is also a mean-value on $\mathcal{P}$.

Later we shall discuss the possibility of extending sample mean-values to mean-values on probability measures.

From the statistical point of view this is an investigation into the possibility of using Fisher's first definition of a consistent estimator. According to this definition an estimator is consistent for some parameter if it is the "same" function of the sample as the parameter is of the population.

2. Simple convergence of probability measures.

In the sequel we shall consider $\mathcal{P}$ as a topological space provided with the <u>simple topology</u>, i.e. the topology, for which the convex sets

$$\mathcal{N}(P) = \{Q \in \mathcal{P} \mid |QA_j - PA_j| < \varepsilon_j,\ j = 1,\dots,k\},$$

$A_1,\dots,A_k \in \mathcal{A}$, $\varepsilon_1,\dots,\varepsilon_k > 0$, $k = 1,2,\dots$, form a neighbourhoodbasis of open sets for $P \in \mathcal{P}$.

A generalized sequence in $\acute{P}$ converges to $P \in \acute{P}$ if and only if the values on every set $A \in \acute{A}$ converge to PA.

The limit is, therefore, only identified by its values on $\acute{A}$ which shows that $\acute{P}$ in general is not a Hausdorff space.

For $B \in \acute{A}$, $P \in \acute{P}$, $B > 0$, the symbol P^B denotes the element in $\acute{P}$ defined by

$$P^B A = \frac{P(B \cap A)}{PB}, \quad A \in \acute{A},$$

the conditional distribution given B corresponding to P. Note that if $B_k \in \acute{A}$, $k = 1,2,\ldots$, $B_k \to F$, $k \to \infty$, $PB > 0$, then P^{B_k} is defined from some number on and

$$P^{B_k} \to P^B$$

for $k \to \infty$. $(1-\lambda)P_0 + \lambda P_1$ is a continuous function of (P_0, P_1, λ), which shows that

$$\text{cl rco } \acute{K} = \text{cl co } \acute{K}$$

for all $\acute{K} \subset \acute{P}$.

Theorem 2. $\acute{E}$ is dense in $\acute{P}$.

Proof. We must prove that each $\acute{N}(P)$ contains an element of $\acute{E}$. Since $\acute{E}$ is the set of all convex combinations with rational coefficient of elements in $\acute{I}$, $\acute{E}$ is dense in $\acute{F}$, and it is sufficient to prove that $\acute{N}(P)$ contains an element of $\acute{F}$. If

$$C = \{C \mid C = D_1 \cap \ldots \cap D_k,\ D_j = A_j \text{ or } A_j^c,\ j = 1,\ldots,k\}$$

then

$$X = \bigcup_{C \in \acute{C}} C, \quad \acute{C} \subset \acute{A}.$$

Choose in each of the finitely many C's a point x_C (if C is empty choose any point in X) and form

$$F = \sum_{C \in \mathcal{C}} PC \; I_{x_C} \quad .$$

Then $F \in \mathcal{F}$ and $FA_j = PA_j$, $j = 1,\ldots,k$.

We note that in general one needs generalized sequences to reach from $\mathcal{E}$ to an element in $\mathcal{P}$, because countable sequences in $\mathcal{E}$ can only create atomic measures with finitely or countably many atoms.

It follows from theorem 2 that $\mathrm{cl}\ \mathrm{co}\ \mathcal{I} = \mathrm{cl}\ \mathrm{co}\ \mathcal{E} = \mathrm{cl}\ \mathcal{E} = \mathrm{cl}\ \mathcal{F} = \mathcal{P}$.

3. Extension of mean-values from $\mathcal{E}$ to $\mathcal{P}$.

Theorem 2 suggests the possibility of making this extension by continuity.

Consider again the two extreme p-quantiles, but now for $p \in \,]0,1[$ only. The continuity properties of m_0 and m_1 are expressed in the equations

$$(1) \qquad \begin{aligned} m_1P &= \limsup_{\mathcal{P} \ni Q \to P} m_0Q, \qquad && P \in \mathcal{P} \\ m_0P &= \liminf_{\mathcal{P} \ni Q \to P} m_1Q, \qquad && P \in \mathcal{P}. \end{aligned}$$

For the definition and simple properties of limit superior and limit inferior, see Bourbaki [1960] .

We prove the second equation. First m_0 is lower semicontinuous, because the set

$$\{P \mid m_0P > a\} \;\; = \{P \mid P\ f \leqq a\} < p\}$$

is an open set. Since $m_0 \leqq m_1$ we have

$$m_0P \leqq \liminf m_0Q \leqq \liminf m_1Q \;.$$

If now $m_0P = m_1P$ the proof is finished.

Assume then $m_0P < m_1P$. For every $y \in]m_0P, m_1P[$ is

$$P\{f \leqq y\} = p > 0.$$

Choose x in $\{f \leqq y\}$, and put

$$Q_n = (1-\frac{1}{n})P + \frac{1}{n} I_x, \quad n = 1,2,\ldots .$$

Since

$$Q_n\{f \leqq y\} = p + \frac{1}{n}(1-p) > p$$

is $m_1Q_n \leqq y$. But $Q_n \to P$ for $n \to \infty$ so

$$\liminf m_1Q \leqq y,$$

and consequently

$$\liminf m_1Q \leqq m_0P,$$

Q.E.D.

From (1) follows that m_0 is lower and m_1 is upper semicontinuous, and that m_0 and m_1 are continuous in P if and only if the p-quantile for f with respect to P is unique.

Let now m be any mean-value on $\not{E}$. The example leads to consideration of lower and upper semicontinuous regularizations m_0 and m_1 of m:

$$m_0P = \liminf_{\not{E} \ni E \to P} mE, \quad P \in \not{P},$$

$$m_1P = \limsup_{\not{E} \ni E \to P} mE, \quad P \in \not{P}.$$

If m_0 and m_1 take the same value in a point $E \in \not{E}$ then $mE = m_0E = m_1E$.

The following theorem is of basic importance.

Theorem 3. Let f be a quasi-convex function on $\acute{E}$. The lower semi-continuous regularization f_0 of f is a quasi-convex function on $\acute{P}$.

Proof. It is clear that

$$\{f_0 \leqq a\} = \bigcap_{a < b \in R'} \mathrm{cl}\, \{P \in E \mid f(P) \leqq b\}$$

for all a $\in$ R'. Since the sets

$$\{P \in E \mid f(P) \leqq b\}$$

are rationally convex for all b $\in$ R' their closures are convex. Hence $\{f_0 \leqq a\}$ is convex, i.e. f_0 is quasi-convex.

The similar theorem for the upper semi-continuous regularization of a quasi-convex function is not true in general. It is here necessary to introduce a stronger topology on $\acute{P}$. This is defined as the simpel topology by considering the unit interval [0,1] as]0,1[in the ordinary topology plus two isolated points 0 and 1.

4. The expectation.

Let f be a numerical function on X which does assume at most one of the values $-\infty$ and $+\infty$. The average of f is the mean-value.

$$m(x_1,\ldots,x_n) = \frac{f(x_1)+\ldots+f(x_n)}{n}, \quad (x_1,\ldots,x_n) \in X^*.$$

The corresponding monotone function on $\acute{E}$ is denoted af and is defined by

$$(af)(E) = \sum_{x \in X} f(x)E\{x\}, \quad E \in \acute{E}.$$

The following theorems describe the behavious of the lower and upper semi-continuous regularizations of af, denoted by $\int_0 f(x)P(dx)$ and $\int_1 f(x)P(dx)$, respectively.

Theorem 4a. If f is unbounded from below and does not assume the value $+\infty$ then $\int_0 f(x)p(dx) = -\infty$.

Proof. Let $\acute{N}$ be any neighbourhood of P and pick $E \in \acute{N} \cap \acute{E}$. Take $x_1, x_2, \ldots$ to be a sequence of elements in X for which $f(x_n) \leqq -n^2$ for all n. If

$$Q_n = \frac{n-1}{n} E + \frac{1}{n} I_x,$$

then $Q_n \in \acute{E} \cap \acute{N}$ for all sufficiently large n. Now

$$(af)(Q_n) = \frac{n-1}{n}(af)(E) + \frac{1}{n} f(x_n) \leqq \frac{n-1}{n}(af)(E) - n,$$

which tends to $-\infty$ for $n \to \infty$.

Theorem 4b. If f is a finite simple function given by $f(x) = \sum_{j=1}^{k} a_j IA_j$, $x \in X$, $A_j \in \acute{A}$, $j = 1,\ldots,k$, then

$$\int_0 f(x)P(dx) = \int_1 f(x)P(dx) = \sum_{j=1}^{k} a_j PA_j \ .$$

The theorem is proved by taking limits in the evident equation

$$(af)(E) = \sum_{j=1}^{k} a_j EA_j, \ E \in \acute{E}.$$

Theorem 4c. If f is bounded and $\acute{A}$-measurable then $\int_0 f(x)P(dx) = \int_1 f(x)P(dx)$.

Proof. Let b_0 be a lower and b_1 an upper bound for f and define

$$A_{kj} = \left\{b_0 + (b_1-b_0)\frac{j}{k} < f \leqq b_0 + (b_1-b_0)\frac{j+1}{k}\right\}, \ k=1,2,\ldots, \ j = 0,1,\ldots k-1,$$

$$g_k = \sum_{j=0}^{k-1} \left(b_0 + (b_1-b_0)\frac{j}{k}\right) I_x A_{kj} + b_0 \ I\{f = b_0\} \ ,$$

$$h_k = \sum_{j=0}^{k-1} \left(b_0 + (b_1-b_0)\frac{j+1}{k}\right) I_x A_{kj} + b_0 I\{f = b_0\} \ .$$

Then g_k and h_k are finite simple functions and $a_k \leqq f \leqq h_k$. Therefore $ag_k(E) \leqq af(E) \leqq ah_k E$ and so

$$\int_0 g_k(x)P(dx) \leqq \int_0 f(x)P(dx) \leqq \int_0 h_k(x)P(dx)$$

for all $P \in \acute{P}$. According to theorem 4b

$$\int_0 h_k(x)P(dx) - \int_0 g_k(x)P(dx) = \sum_{j=0}^{k-1} (b_1-b_0)\frac{1}{k} PA_{kj} \leqq \frac{b_1-b_0}{k} .$$

It follows that

$$\int_0 f(x)P(dx) = \inf_k \int_0 h_k(x)P(dx)$$

for all $P \in \acute{P}$, which shows that $\int_0 f(x)P(dx)$ as a function of P is the lower bound of a family of continuous functions and therefore itself upper semi-continuous.

Theorem 4d. If $f_1, f_2, \ldots$ is an increasing sequence of measurable functions, bounded from below and tending to f, then $\int_0 f_k(x)P(dx) \uparrow \int_0 f(x)P(dx)$ for $k \to \infty$.

Proof. Since $(af_k)(E) \uparrow (af)(E)$ for all $E \in \acute{E}$, the sequence $\int_0 f_k(x)P(dx)$ is increasing and

$$\lim_{k \to \infty} \int_0 f_k(x)P(dx) \leqq \int_0 f(x)P(dx).$$

To prove the opposite inequality let b be a constant lower bound for f_1, and put $B = \{f < +\infty\}$. First consider the case $PB = 1$. Let ε be a positive number and define $B_k = \{f_k > f-\varepsilon\}$. Then $B_k \uparrow B$ for $k \to \infty$, and for $E \in \acute{E}$

$$(af_k)(E) \geqq a((f-\varepsilon)IB_k)(E) + bEB_k^c ,$$

and so for $EB_k > 0$

$$(af_k)(E) \geqq EB_k(af)(E^{B_k}) - \varepsilon EB_k + bEB_k^c.$$

By taking lower limits for $E \to P$ it follows that

$$\int_0 f_k(x)P(dx) \geqq PB_k \int_0 f(x)P^{B_k}(dx) - \varepsilon PB_k + bPB_k^c ,$$

because $(af)(E^{B_k})$ is bounded from below and $E^{B_k} \to P^{B_k}$. By computing lower limits for $k \to \infty$ and observing that

$$\int_0 f(x)P^{B_k}(dx)$$

is bounded from below and $P^{B_k} \to P^B = P$, it is proved that

$$\lim_{k \to \infty} \int_0 f_k(x)P(dx) \geqq \int_0 f(x)P(dx) - \varepsilon.$$

This holds for all $\varepsilon > 0$ and so

$$\lim_{k \to \infty} \int_0 f_k(x)P(dx) \geqq \int_0 f(x)P(dx).$$

Next consider the case $PB < 1$. Let K be a finite number and define $C_k = \{f_k > K\} \cap B^c$. Then $C_k \uparrow B^c$ for $k \to \infty$, and from the inequality

$$(af_k)(E) \geqq KEC_k + bEC_k^c$$

it follows that

$$\int_0 f_k(x)P(dx) \geqq KPC_k + bPC_k^c ,$$

which for $k \to \infty$ yields

$$\lim_{k \to \infty} \int_0 f_k(x)P(dx) \geqq KPB^c + bPB .$$

Since this is true for all K it follows that

$$\lim_{k \to \infty} \int_0 f_k(x)P(dx) = +\infty .$$

Theorem 4c justifies the following definition for f $\mathcal{A}$-measurable:

$$\int f(x)P(dx) = \begin{cases} \int_0 f(x)P(dx) & \text{for f bounded from below} \\ \int_1 f(x)P(dx) & \text{for f bounded from above,} \end{cases}$$

and theorem 4b and 4d show that the definition gives the ordinary Lebesgue integral.

The definition of the integral can be extended by the usual trick to functions that are neither bounded from below nor from above, but we shall not consider such integrals.

5. Complete mean-values.

Let $(Y,\mathcal{B},L)$ be a probability space, and let to each $y \in Y$ correspond an element P_y of $\mathcal{P}$ such that $P.A, A \in \mathcal{A}$, is a $\mathcal{B}$-measurable function. It follows from theorem 4 that $Q = \int P_y L(dy)$ is in $\mathcal{P}$.

A numerical function on $\mathcal{P}$ is a <u>complete mean-value</u> on $\mathcal{P}$ if $mP_y \leqq a$ for all $y \in Y$ implies $mQ \leqq a$, and $mP_y \geqq a$ for all $y \in Y$ implies $mQ \geqq a$ for $a \in R$.

A complete mean-value on $\mathcal{P}$ is a mean-value on $\mathcal{P}$.

Theorem 5. A continuous mean-value on $\mathcal{P}$ is a complete mean-value.

If f is an $\mathcal{A}$-measurable numerical function bounded from below then $\int f(x)P(dx)$ is a complete mean-value on $\mathcal{P}$.

Proof. The first statement is an immediate consequence of lemma 4 and the second statement follows from lemma 3 below.

Lemma 1. If f is a bounded $\mathcal{A}$-measurable function and $a \in R$ then $\int af(x)P(dx) = a\int f(x)P(dx)$.

III. Conditional mean-values and limit theorems.

1. Conditional mean-values.

Let $(Y,\mathcal{B},L)$ be a probability space, and let there to every $B \in \mathcal{B}$ with $LB > 0$ correspond an element P^B of P such that $P^B = P^{B'}$ if B and B' differ by an L-null set, and if $B = \cup B_i$ is a countable disjoint union then PB is a countable convex combination of $P^{B_1}, P^{B_2}, \ldots$. We shall call such a mapping a <u>decomposition</u> of $P = P^Y$.

Example 1. If to every $y \in Y$ there corresponds an element P_y in P such that $P.A$, $A \in \mathcal{A}$, is a $\mathcal{B}$-measurable then the mapping

$$Q^B = \frac{1}{LB} \int_B P_y L(dy), \quad B \in \mathcal{B},\ LB > 0,$$

is a decomposition of $Q = Q^Y$. If a decomposition can be defined in this way it is <u>regular.</u>

Example 2. Let $\mathcal{B}$ be a σ-algebra contained in $\mathcal{A}$, and let $P \in \mathcal{P}$. The mapping from $\mathcal{B}$ into $\mathcal{P}$ defined by

$$P^B A = \frac{P(B \cap A)}{PB} \quad , A \in \mathcal{A}$$

$B \in \mathcal{B}$, $PB > 0$ is a decomposition of P.

If there exists a regular conditional probability given $\mathcal{B}$ then this decomposition is regular.

Theorem 6. Let P be decomposed as described above. Then to every complete mean-value m there exists a $\mathcal{B}$-measurable function f such that for every $a \in R$

$$mP^B \geqq a \text{ for } B \in B,\ LB > 0,\ B \subset \{f \geqq a\}$$

$$mP^B \leqq a \text{ for } B \in B,\ LB > 0 \quad B \subset \{f \leqq a\}.$$

A function like f is called the <u>conditional m-mean-value given the decomposition</u>.

It is determined up to an L-equivalence.

If the decomposition is regular and mP_y is $\mathcal{B}$-measurable $f(y) = mP_y$ for L-almost all $y \in Y$.

We begin the proof of theorem 6 with a

Lemma (Decomposition). Let m be a complete mean-value on $\mathcal{A}$. To all real a there exists a set $B \in \mathcal{B}$ such that $mP^F \leqq a$ for $F \subset B$, $F \in \mathcal{B}$, $LF > 0$, and $mP^F \geqq a$ for $F \subset B^c$, $F \in \mathcal{B}$, $LF > 0$.

Proof. A $\mathcal{B}$-measurable set A shall be called negative if $mP^F \leqq a$ for $F \subset A$, $F \in A$, $PF > 0$.

The empty set is negative, and the difference between two negative sets is negative, because $F \subset A_1 \cup A_2^c$ implies $F \subset A_1$. The union of a sequence of pairwise disjoint negative sets is again negative. If namely $A_1, A_2, \ldots$ is such a sequence with union A then for $F \subset A$, $F \in \mathcal{A}$, $PF > 0$

$$P^F = \sum_{i=1}^{\infty} \frac{P(F \cap A_i)}{PF} \, P^{F \cap A_i}$$

where the summation is extended over all i such that $P(F \cap A_i) > 0$, but $mP^{F \cap A_i} \leqq a$ for all i and so $mP^F \leqq a$.

It follows that countable unions of netative sets are negative.

Let β denote the supremum of PA over all negative sets A, and let $B_1, B_2, \ldots$ be a sequence of negative sets such that $PB_n \to \beta$. If

$$B = \bigcup_{n=1}^{\infty} B_n$$

then $PB = \beta$ and B is negative. We shall show that B also satisfies the second condition in the lemma.

Assume that on the contrary there exists a measurable set $E_0 \subset B^c$ such that $PE_0 > 0$ and $mP^{E_0} < a$. E_0 can not be negative because in that case $E_0 \cup B$ is negative, but $P(B \cup E_0) = PB + PE_0 > \beta$. Let k_1 be the smallest positive integer such that there is a measurable $E_1 \subset E_0$ with $PE_1 \geqq \frac{1}{k_n}$

and $mP^{E_1} > a$. $P(E_0 \cap E_1^c)$ cannot be zero since this implies $a > mP^{E_0} = mP^{E_1} > a$. Therefore $P^{E_0 \cap E_1^c}$ is well-defined, and from $\min(mP^{E_0 \cap E_1^c}, mP^{E_1}) \leqq mP^{E_0} < a$, it follows that $mP^{E_0 \cap E_1^c} < a$. $E_0 \cap E_1^c$ is not negative. Let k_2 be the smallest positive integer for which there exists a measurable $E_2 \subset E_0 \cap E_1^c$ with $PE_2 \geqq \frac{1}{k_2}$ and $mP^{E_2} > a$.

Continuing in this way a sequence $E_1, E_2, \ldots$ of sets and a sequence $k_1, k_2, \ldots$ of positive integers are found. Put

$$E = \bigcup_{n=1}^{\infty} E_n .$$

Then

$$PE_0 \geqq PE = \sum_{n=1}^{\infty} PE_n \geqq \sum_{n=1}^{\infty} \frac{1}{k_n}$$

and so $k_n \to \infty$ for $n \to \infty$. Define $F_0 = E_0 \cap E^c$. PF_0 cannot be zero because then $a > mP^{E_0} = mP^{E} \geqq a$. If $F \subset F_0$, $F \in \mathcal{A}$, $PF > 0$, and $mP^F > a$, then $PF < \frac{1}{k_n - 1}$ for all n and therefore $PF = 0$. It follows that F_0 is negative, but then $B \cup F$ is negative, and $P(B \cup F_0) > \beta$.

Proof of theorem 6. According to the decomposition lemma there corresponds to every rational r a set B_r^+ and a set B_r^- such that B_r^+ and B_r^- are $\mathcal{B}$-measurable, $Y = B_r^+ \cup B_r^-$, $B_r^+ \cap B_r^- = \emptyset$ and for which

$$mP^B \geqq r \text{ for } B \in \mathcal{B},\ LB > 0,\ B \subset B_r^+$$

$$mP^B \leqq r \text{ for } B \in \mathcal{B},\ LB > 0,\ B \subset B_r^- .$$

For $y \in \cup B_r^+$ define $f(y) = \sup_{x \in B_r^+} r$ and for $y \in (\cup B_r^+)^c = \cap B_r^-$ put $f(y) = -\infty$.

For all real a is $\{f > a\} = \bigcup_{r > a} B_r^+$, so f is a $\mathcal{B}$-measurable function, and $\{f \leqq a\} = \bigcap_{r > a} B_r^-$, which shows that for $B \subset \{f \leqq a\}$, $B \in \mathcal{B}$, and $LB > 0$, is $mP^B \leqq r$ for all rational $r > a$, and therefore $mP^B \leqq a$. If $B \subset \{f > a\}$ $B \in \mathcal{B}$, and $LB > 0$ then $B \subset \{f > b\}$ for all $b < a$. Put $B = \bigcup_{r > b} B_r$

where $B_r \subset B_r^+$ for all $r > b$ and the B_r's are pairwise disjoint. It follows that $mP^B \geqq b$ for all $b < a$ and therefore $mP^B \geqq a$. This shows that f is a conditional m-meanvalue given the decomposition.

Assume now that g is a $\mathcal{B}$-measurable numerical function such that $L\{f \neq g\} = 1$. If $B \in \mathcal{B}$ and $LB > 0$ then $B \subset \{g \geqq a\}$ implies $mP^B = mP^{B\cap\{f \geqq a\}} \geqq a$ and $B \subset \{g \leqq a\}$ implies $mP^B = mP^{B\cap\{f \leqq a\}} \leqq a$ and so g is also a conditional mean-value.

Next, let f and g be two conditional mean-values. Then $\{f < g\} = \cup(\{f \leqq a\}\cap\{g \geqq b\})$ where the union is taken over all rational a and b such that $a < b$. If $P(\{f \leqq a\}\cap\{g \geqq b\}) > 0$ then $a \geqq mP^{(\{f\leqq a\}\cap\{g\geqq b\})} \geqq b$ which is impossible. Therefore $P\{f < g\} = 0$.

The last statement of the theorem follows immediately from the definitions.

2. Limit Theorems.

Let $P \in \mathcal{P}$ be decomposed as described in 1.

Definition. Let $\mathcal{C}$ be a σ-algebra contained in $\mathcal{A}$. A mean-value m on $\mathcal{P}$ is said to have the monotone convergence property with respect to the decomposition restricted to $\mathcal{C}$ if for n = 1,2,...

$$B_n \in \mathcal{B},\ LB_n > 0,\ LC > 0$$

$$B_n \uparrow C \in \mathcal{C} \text{ or } B_n \downarrow C \in \mathcal{C}$$

implies $mP^{B_n} \to mP^C$.

If $P^{B_n} \to P^C$, and m is continuous in all P^C, $C \in C$, then it has this property. For integrals it is implied by theorem 4d.

Theorem 7 (Martingale). Let $\mathcal{C}_1, \mathcal{C}_2, \ldots$ be an increasing sequence of σ-algebras contained in $\mathcal{B}$, and let $\mathcal{C}$ be the smallest σ-algebra over their union. Let m be a complete mean-value with the monotone convergence property with respect to the decomposition determined by $\mathcal{C}$. If $f, f_1, f_2, \ldots$ are conditional m-mean-values of P with respect to the decomposition determined by $\mathcal{C}, \mathcal{C}_1, \mathcal{C}_2, \ldots$. then $f_n \to f$ for $n \to \infty$ L-almost surely.

Proof (Comp. Andersen and Jessen [1948]). The functions $\underline{f} = \liminf f_n$ and $\overline{f} = \limsup f_n$ are $\mathcal{C}$-measurable, since

$$\{\underline{f} > a\} = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \{f_n > a\}$$

$$\{\overline{f} < a\} = \bigcup_{k=1} \bigcap_{n=k} \{f_n < a\}$$

for all real a.

Now assume $a < +\infty$, and let $C \in \cup\mathcal{C}_n$, $C \subseteqq \{\underline{f} \leqq a\} = H$, and $LC > 0$. Put

$$H_n = \{\inf_p f_{n+p} < a_n\} = \bigcup_{p=1}^{\infty} \{f_{n+p} < a_n\}$$

and

$$H_{np} = \begin{cases} \{f_{n+1} < a_n\} & \text{for } p = 1 \\ \bigcap_{q=1}^{p-1} (\{f_{n+q} \geqq a_n\} \cap \{f_{n+p} < a_n\}) & \text{for } p > 1, \end{cases}$$

where $a_1 > a_2 > \ldots > a_n > \ldots$ is a strictly decreasing sequence with limit a. It is clear that $H_{np} \in \mathcal{C}_{n+p}$, that $H_{np} \subset \{f_{n+p} < a_n\}$, that for a fixed n all H_{np} are mutually disjoint, and that

$H_n = \bigcup_{p=1}^{\infty} H_{np}$. Further is $H_1 \supset H_2 \supset \ldots$ and $H = \bigcap_{n=1}^{\infty} H_n$.

There exists an n_0 such that $C \in \mathcal{C}_n$ for $n \geqq n_0$, and therefore $C \cap H_{np} \in \mathcal{C}_{n+p}$ for $n \geqq n_0$ and all p. It follows that

$$mP^{C \cap H_{np}} \leqq a_n$$

if $L(C \cap H_{np}) > 0$, and so $mP^C = mP^{(C \cap H_n)} \leqq a_n$ for $n \geqq n_0$, which shows that $mP^C \leqq a$.

Consider next the set $\mathcal{C}_0$ of all $C \in \mathcal{C}$ such that $C \subseteqq \{\underline{f} \leqq a\}$ and either $LC = 0$ or $mP^C \leqq a$. Because of the monotone convergence property this class is a monotone class. The preceding argument shows that $\mathcal{C}_0$ contains the field of all $C \in \cup\mathcal{C}_n$ such that $C \subset \{\underline{f} \leqq a\}$. It follows that $\mathcal{C}_0$ contains the smallest σ-field containing this field, i.e. the set of all $C \in \mathcal{C}$ such that $C \subset \{\underline{f} \leqq a\}$. We have thus shown that $C \subset \{\underline{f} \leqq a\}$, $C \in \mathcal{C}$, $LC > 0$ implies $mP^{C} \leqq a$.

In a similar way it is proved that $mP^C \geqq a$ for $C \subset \{\overline{f} \geqq a\}$, $C \in \mathcal{C}$, and $LC > 0$. It follows that $\underline{f}$ and $\overline{f}$ both are conditional m-mean-values given the decomposition restricted to $\mathcal{C}$, and therefore $\underline{f} = \overline{f}$ almost surely with respect to L.

Theorem 8 (Reversed martingale). Let $\mathcal{C}_1, \mathcal{C}_2, \ldots$ be a decreasing sequence of σ-algebras contained in $\mathcal{B}$, and let $\mathcal{C}$ be their intersection. Let m be a complete mean-value with the monotone convergence property with respect to the decomposition determined by $\mathcal{C}$, then $f_n \to f$ for $n \to \infty$ L-almost surely.

Proof (Comp. Andersen and Jessen [1948]). $\underline{f} = \liminf f_n$ and $\overline{f} = \limsup f_n$ are clearly $\mathcal{C}$-measurable. Now assume $C \subset \{\sup f_n > a\}$, $C \in \mathcal{C}$, $LC > 0$. Put

$$H_n = \{\max_{p \leqq n} f_p > a\}$$

and

$$H_{np} = \begin{cases} \{f_p > a\} \cap \bigcap_{q=p+1}^{n} \{f_q \leqq a\} & \text{for } p < n \\ \{f_n > a\} \text{ for } p = n. \end{cases}$$

Then $H_{np} \in \mathcal{C}_p$, $H_{np} \subset \{f_p > a\}$ and

$$H_n = \bigcup_{p \leqq n} H_{np}.$$

It follows that $mP^{C \cap H_n} \geqq a$. Since $H_n \uparrow \{\sup f_n > a\}$, the monotone convergence property implies $mP^C \geqq a$.

One proves in a similar way that $mP^C \leqq a$ for $C \subset \{\inf f_n < a\}$, $C \in \mathcal{C}$, and $LC > 0$. This shows that both $\underline{f}$ and $\overline{f}$ are conditional mean-values given the decomposition restricted to $\mathcal{C}$ and therefore $\underline{f} = \overline{f}$ a.s. with respect to L.

Theorem 9 (Ergodic theorem). Let T be an invertible $\mathcal{A}$-measurable transformation of X into itself, and let $\mathcal{C}$ be the σ-algebra of T-invariant sets. Let m be a complete mean-value on $\mathcal{P}$ that has the monotone convergence property with respect to the decomposition determined by $\mathcal{C}$. Let $P \in \mathcal{P}$ be T-invariant and put

$$f_n(x) = m\left(\frac{1}{n}\sum_{i=0}^{n-1} I_{T^i x}\right), \qquad x \in X.$$

If f_n is $\mathcal{A}$-measurable, and $\limsup_n f_n$ and $\liminf_n f_n$ are $\mathcal{C}$-measurable then f_n for $n \to \infty$ tends to the conditional m-meanvalue of P given $\mathcal{C}$, almost surely with respect to P's restriction to $\mathcal{C}$.

Our proof is a generalization of a proof of the ergodic theorem due to A.N. Kolmogorov, see Khinchine [1949] , p. 19, seq.

We start with some combinatorial lemmas. Let $\ldots, x_{-1}, x_0, x_1, \ldots$ be a fixed sequence of elements of X and a a fixed real number. For integers $j < k$ let $[j,k)$ denote the interval $\{j' \mid j \leqq j' < k\}$. An interval $[j,k)$ is <u>proper</u> if $m(x_j, \ldots, x_{k-1}) > a$ and $m(x_j, \ldots, x_{j'-1}) \leqq a$ for $j < j' < k$.

Lemma 1. If two proper intervals overlap each other, then one is contained in the other.

Proof. Let $[j,k)$ and $[j_1,k_1)$ be proper intervals such that $j \leqq j_1 < k \leqq k_1$. Since

$$m(x_j,\ldots,x_{j_1-1}) \leqq a < m(x_j,\ldots,x_{k-1}) \leqq \max(m(x_j,\ldots,x_{j_1-1}), m(x_{j_1},\ldots,x_{k-1}))$$

it follows that $m(x_{j_1},\ldots,x_{k-1}) > a$ and therefore $k = k_1$.

For $n = 1,2,\ldots$ an interval $[j,k)$ is called <u>n-proper</u> if it is n-proper and $k-j \leqq n$. $[j,k)$ is <u>maximal n-proper</u> if it is n-proper and not contained in a bigger n-proper interval.

Lemma 2. Every n-proper interval is contained in a maximal n-proper interval.

Proof. If two n-proper intervals contain the given one, they overlap and according to lemma 1 one is contained in the other. The set of n-proper intervals containing the given one is finite and so must contain a maximal element. This interval must be maximal n-proper.

Lemma 3. Two different maximal n-proper intervals are disjoint.

Proof. Lemma 1 shows that if they are not disjoint one must contain the other, which is impossible according to the definition.

Lemma 4. It is necessary and sufficient for

$$\max_{1 \leqq k \leqq n} m(x_0,\dots,x_{k-1}) > a$$

that there exists a maximal n-proper interval containing 0.

Proof. Necessity: Let k be the smallest integer $\geqq 1$ such that $m(x_0,\dots,x_{k-1}) > a$. Then $[0,k)$ is n-proper and therefore according to lemma 2 contained in a maximal n-proper interval. Sufficiency: Let $[j,k)$ be a maximal n-proper interval containing 0. Then $1 \leqq k \leqq n$, and the inequalities $m(x_j,\dots,x_{-1}) \leqq a < m(x_j,\dots,x_{k-1}) \leqq \max(m(x_j,\dots,x_{-1}), m(x_0,\dots,x_k))$ show that $m(x_0,\dots,x_k) > a$.

Turning to the proof of the theorem define

$$E^n_x = \frac{1}{n} \sum_{k=0}^{n-1} I_{T^k x}$$

for $n = 1,2,\dots$, $x \in X$, and put $\underline{f} = \liminf f$ and $\overline{f} = \limsup f_n$ where $f_n = mE^n$. It is sufficient to prove that $B \in \mathcal{C}$, $PB > 0$, and $B \subset \{\sup f_n > a\}$ implies $mP^B \geqq a$, because $\{\inf f_n < a\}$ can be treated analogously.

Let $H_n = \left\{ \max_{p \leqq n} f_p > a \right\}$ and $F_{jk} = \{[-j,-j+k) \text{ maximal n-proper}\}$ for $k = 1,\dots,n$, $j = 0,1,\dots,k-1$ or $-j \leqq 0 < -j+k$. Lemma 3 proves that different F's are disjoint, and lemma 4 that

$$H_n = \bigcup_{k=1}^{n} \bigcup_{j=0}^{k-1} F_{jk} .$$

If A^n stands for $T^{-n}A$ for $A \subset X$ it is seen that $F^j_{jk} = F_{0k}$ and therefore $P(B \cap F_{jk}) = P(B \cap F_{0k})$. Similarly

$$P^{F_{jk}} = \frac{P(F_{jk} \cap A)}{PF_{jk}} = \frac{P(F_{0k} \cap A^j)}{PF_{0k}} = P^{F_{0k}}{}_{A^j}$$

for $A \in \mathcal{A}$ and $PF_{jk} > 0$ and so

$$P^{B\cap H_n}{}_A = \sum_{k=1}^{n} \sum_{j=0}^{k-1} \frac{P(B\cap F_{jk})}{P(B\cap H_n)} P^{B\cap F_{jk}}{}_A = \sum_{k=1}^{n} \sum_{j=0}^{k-1} \frac{P(B\cap F_{0k})}{P(B\cap H_n)} P^{B\cap F_{0k}}{}_{A^j}$$

$$= \sum_{k=1}^{n} \frac{kP(B\cap F_{0k})}{P(B\cap H_n)} \frac{1}{k} \sum_{j=0}^{k-1} P^{B\cap F_{0k}}{}_{A^j} = \sum_{k=1}^{n} \frac{kP(B\cap F_{0k})}{P(B\cap H_n)} \int E_x^k(A) P^{B\cap F_{0k}}(dx)$$

Since $mE_x^k > a$ for $x \in B\cap F_{0k}$ it follows that $mP^{B\cap H_n} \geqq a$ which implies $mP^B \geqq a$ because $H_n \uparrow \{\sup f_n > a\}$.

The ergodic theorem can be used to prove that consistency of estimators in the sense mentioned on p.25 implies consistency in the usual sense.

References.

J. Aczél [1961]: Vorlesungen über Funktionalgleichungen und ihre Anvendungen. Basel und Stuttgart 1961.

E.S. Andersen and B. Jessen [1948]: Some limit theorems on set functions, Danske Vid. Selsk. Mat-Fys. Medd. 25 (1948).

N. Bourbaki [1960]: Eléments de mathématique Première partie. Livre III. Topologie générale, Chapitre 4, Troisième édition. Paris 1960.

H.D. Brunk [1961]: A theorem of E. Sparre Andersen and its application to tests against trend, Math. Scand. 8 (1961), 305-326.

B. de Finetti [1931]: Sul concetto di media, Giorn. Inst. Ital. Attuarii 2 (1931), 369-396.

R.A. Fisher [1925]: Theory of statistical estimation, Proc. Cambridge Phil. Soc. 12 (1925), Pt.5, 700-725.

A.I. Khinchin [1949]: Mathematical foundations of statistical mechanics. New York 1949.

A.N. Kolmogorov [1930]: Sur la notion de la moyenne, R.C. Accad.Lincei (6) 12. (1930), 388-391.

M. Nagumo [1930]: Über eine Klasse Mittelwerten, Japanese J. Math. 7 (1930), 520-526.