

5. There is an analogous image for the theory of the theta functions of r variables. In it the lattice space is of $2r(r+1)$ dimensions, and the appropriate configurations lie on a system of r four-dimensional spheres immersed in the higher space. For $r > 1$ the image is not expressible in terms of reflexions alone.

THE UNIVERSITY OF WASHINGTON.

NOTE ON THE MEDIAN OF A SET OF NUMBERS.

BY PROFESSOR DUNHAM JACKSON.

(Read before the American Mathematical Society September 7, 1920.)

LET a_1, a_2, \dots, a_n be a set of real numbers, which may or may not be all distinct. Let

$$S_2(x) = \sum_{i=1}^n (x - a_i)^2.$$

The value of x which reduces $S_2(x)$ to a minimum is the arithmetical mean of the numbers a_1, \dots, a_n . If the condition that $S_2(x)$ be a minimum is replaced by the condition that

$$S_1(x) = \sum_{i=1}^n |x - a_i|$$

be reduced to a minimum, the median of the a 's is obtained. It is uniquely defined whenever n is odd; if the numbers a_i are arranged in order of magnitude, so that

$$a_1 \leq a_2 \leq \dots \leq a_n,$$

and if $n = 2k - 1$, the median is simply a_k , the middle one of the a 's. The median is uniquely defined also when n is even, $n = 2k$, if it happens that $a_k = a_{k+1}$, being then equal to this common value. Otherwise, the definition is satisfied by any number x belonging to the interval

$$a_k \leq x \leq a_{k+1},$$

and the median is to this extent indeterminate.

The purpose of the following paragraphs is to show that

for each value of $p > 1$ there is a definite number $x = x_p$ which minimizes the sum

$$S_p(x) = \sum_{i=1}^n |x - a_i|^p,$$

and that x_p approaches a definite limit X as p approaches 1. The value of X coincides with the median as already defined, in the cases where that definition is determinate; and when $n = 2k$ and $a_k \neq a_{k+1}$, X is a definite number between a_k and a_{k+1} . It serves thus to supplement the former definition, on the theoretical side at any rate.

It is not unlikely that the same average has been discussed, as a curiosity at least, by writers on statistics. The accurate proof of the statements involved, however, seems to be an exercise in pure mathematics. It will have to be admitted that the number does not lend itself readily to direct computation, except in the simplest cases.

In the first place, for any particular value of $p > 1$, $S_p(x)$ is a continuous function of x which is always positive or zero, and which becomes infinite as x goes out to infinity in either direction. Consequently it attains a minimum for some one value of x at least.

Furthermore, each term $|x - a_i|^p$ has a continuous derivative for all values of x , the value $x = a_i$ not excepted. Hence $S_p(x)$ also has a continuous derivative everywhere, and this derivative must vanish at any point where a minimum is reached. But the derivative of $|x - a_i|^p$ always increases when x increases; hence the same thing is true of the derivative of $S_p(x)$, and $S_p'(x)$ can vanish only once. That is, the minimizing value x_p is indeed uniquely determined, and is moreover characterized by the vanishing of $S_p'(x)$.

To be more precise with regard to the formulas involved,

$$\frac{d}{dx} |x - a_i|^p = \frac{d}{dx} (x - a_i)^p = p(x - a_i)^{p-1}, \quad x > a_i;$$

$$\frac{d}{dx} |x - a_i|^p = \frac{d}{dx} (a_i - x)^p = -p(a_i - x)^{p-1}, \quad x < a_i;$$

while for $x = a_i$ the derivative is correctly represented by either formula, being equal to zero. If i is any one of the indices 1, 2, \dots , $n - 1$, and x any number in the interval

$$a_i \leq x \leq a_{i+1},$$

$$S_p(x) = (x - a_1)^p + \cdots + (x - a_i)^p \\ + (a_{i+1} - x)^p + \cdots + (a_n - x)^p,$$

$$(1) \quad \frac{1}{p} S_p'(x) = (x - a_1)^{p-1} + \cdots + (x - a_i)^{p-1} \\ - (a_{i+1} - x)^{p-1} - \cdots - (a_n - x)^{p-1}.$$

Now suppose n is odd, $n = 2k - 1$. Let ϵ be any positive quantity, and let r be an index such that

$$a_r < a_k + \epsilon \leq a_{r+1}.$$

Then $r \geq k$. By the formula just written down,

$$\frac{1}{p} S_p'(a_k + \epsilon) = (a_k + \epsilon - a_1)^{p-1} + \cdots + (a_k + \epsilon - a_r)^{p-1} \\ - (a_{r+1} - a_k - \epsilon)^{p-1} - \cdots - (a_n - a_k - \epsilon)^{p-1}.$$

When p approaches 1, each of the first k terms on the right approaches 1 as a limit. Any one of the remaining $k - 1$ terms, sign included, has the limit 1, 0, or -1 , as the case may be. At any rate,

$$\lim_{p=1} S_p'(a_k + \epsilon) > 0.$$

There exists a positive δ_1 such that

$$S_p'(a_k + \epsilon) > 0$$

whenever $1 < p \leq 1 + \delta_1$.

Similarly, there is a positive δ_2 such that

$$S_p'(a_k - \epsilon) < 0$$

if $1 < p \leq 1 + \delta_2$. If δ is the smaller of δ_1 , δ_2 , and $1 < p \leq 1 + \delta$, it is certain that $S_p'(x)$ vanishes between $a_k - \epsilon$ and $a_k + \epsilon$. But the point where $S_p'(x)$ vanishes is the point x_p ; consequently

$$a_k - \epsilon < x_p < a_k + \epsilon$$

for $1 < p \leq 1 + \delta$, that is,

$$\lim_{p=1} x_p = a_k.$$

Analogous reasoning shows that when $n = 2k$ and $a_k = a_{k+1}$,

$$\lim_{p=1} x_p = a_k = a_{k+1}.$$

It remains to consider the case of principal interest, the case that $n = 2k$ and $a_k < a_{k+1}$. By reasoning similar to that already employed, it is seen that $\lim_{p=1} S_p'(a_k) < 0$. For at least k terms in the expression for $S_p'(a_k)$ approach the limit -1 , the term $(x - a_k)^{p-1}$ is zero, and not more than $k - 1$ terms are positive. In the same way, $\lim_{p=1} S_p'(a_{k+1}) > 0$, and therefore

$$a_k < x_p < a_{k+1}$$

when p is sufficiently near to 1.

When $x > a_i$, $(x - a_i)^{p-1}$ can be represented by the series

$$\begin{aligned} (x - a_i)^{p-1} &= e^{(p-1)\log(x-a_i)} \\ &= 1 + (p-1)\log(x-a_i) + \frac{1}{2}(p-1)^2\log^2(x-a_i) + \dots \\ &= 1 + (p-1)\log(x-a_i) + (p-1)^2\varphi_i(x, p), \end{aligned}$$

where $\varphi_i(x, p)$ is a function which remains finite if x is fixed and p approaches 1. Similarly,

$$(a_i - x)^{p-1} = 1 + (p-1)\log(a_i - x) + (p-1)^2\varphi_i(x, p)$$

when $x < a_i$. The work so far has been based on the use of the first term as an approximation to the value of the series; it will be convenient now to make use of the explicit form of the first two terms.

If the expressions just indicated are substituted in the right-hand member of (1), for a value of x between a_k and a_{k+1} , there will be k terms each equal to $+1$ and k terms each equal to -1 , which will cancel, and a relation will be obtained which can be written in the form

$$\begin{aligned} \frac{1}{p(p-1)} S_p'(x) &= \log(x - a_1) + \dots + \log(x - a_k) \\ &\quad - \log(a_{k+1} - x) - \dots - \log(a_n - x) \\ &\quad + (p-1)\varphi(x, p) \\ &= \log \frac{(x - a_1) \dots (x - a_k)}{(a_{k+1} - x) \dots (a_n - x)} + (p-1)\varphi(x, p), \end{aligned}$$

where $\varphi(x, p)$ is a function which remains finite for fixed x as p approaches 1.

The fraction on the right in the last relation is equal to 0 for $x = a_k$; it increases steadily when x increases from a_k

toward a_{k+1} , each factor of the numerator increasing while each factor of the denominator decreases; and it becomes infinite as x approaches a_{k+1} . Hence there is just one intermediate point, $x = X$, at which the fraction is equal to 1 and its logarithm to 0.

For $x = X - \epsilon$, where ϵ is an arbitrarily small positive quantity, the logarithm has a negative value, independent of p . Hence $S_p'(X - \epsilon)$ is negative if $p - 1$ is sufficiently small. Similarly, $S_p'(X + \epsilon)$ is positive for small values of $p - 1$. This means that x_p is between $X - \epsilon$ and $X + \epsilon$ if p is sufficiently near to 1, that is,

$$\lim_{p=1} x_p = X.$$

So the original assertion is proved.

The number X is characterized by the equation

$$(X - a_1) \cdots (X - a_k) = (a_{k+1} - X) \cdots (a_n - X).$$

If $n = 2$,

$$X = \frac{a_1 + a_2}{2}.$$

If $n = 4$,

$$X = \frac{a_4 a_3 - a_2 a_1}{(a_4 + a_3) - (a_2 + a_1)}.$$

When $n > 4$, the explicit determination of X will involve the solution of an algebraic equation of more or less high degree.

The discussion has been concerned primarily with values of p not much greater than 1. The proof of the existence and uniqueness of x_p , however, applies equally well when p is arbitrarily large. It is readily seen that

$$\lim_{p=\infty} x_p = Y = \frac{1}{2}(a_1 + a_n),$$

the intermediate a 's having no effect on the result. For if ϵ is fixed and p is very large, the single term $(x - a_1)^p$ will be greater than the whole sum $S_p(Y)$ when $x \geq Y + \epsilon$, with a corresponding inequality on the other side, so that x_p must be between $Y - \epsilon$ and $Y + \epsilon$.

THE UNIVERSITY OF MINNESOTA,
MINNEAPOLIS, MINN.