# ON DISJOINT SETS OF DISTRIBUTION FUNCTIONS

AGNES BERGER

1. **Introduction.** Let us call a function $F(x)$ of the real variable $x$ a distribution function, *d. f.*, if

$$(C_1) \qquad \begin{array}{l} F(x) \text{ is non-decreasing, continuous to the right,} \\ F(-\infty) = 0, \quad F(+\infty) = 1. \end{array}$$

Let further $\{F_i(x)\}$ and $\{G_j(x)\}$ be two sets of d. f.'s with $i$ ranging over the set of values $I=\{i\}$ and $j$ over $J=\{j\}$. We shall assume

$$(C_2) \qquad \begin{array}{l} F_i(x) \not\equiv F_j(x), \quad G_i(x) \not\equiv G_j(x), \\ F_i(x) \not\equiv G_j(x) \end{array} \qquad \text{for all } i \text{ and } j,$$

and attempt to answer the following question proposed by Neyman:

What conditions must be satisfied by sets $\{F_i(x)\}$ and $\{G_j(x)\}$ so that there exists at least one Borel set $W$ such that

$$(A) \qquad \int_W dF_i(x) \neq \int_W dG_j(x) \qquad \text{for all } i, j?$$

This question is of interest in the theory of statistical hypotheses. To indicate this briefly: Let $\xi$ be a random variable of which it is known only that its distribution belongs to the union of the two sets $\{F_i(x)\}$ and $\{G_j(x)\}$. Let $H$ be the hypothesis asserting that the distribution of $\xi$ belongs to $\{F_i(x)\}$. Should there be no Borel set $W$ of the property (A) it would cause difficulty in testing the hypothesis H.

2. **The simplest cases.** To begin with we examine the simplest cases $I=1$, $J=1$ and $I=1$, $J=\{1, 2\}$. It is easy to see that in these cases one can always find an interval solution for (A):

$I=1$, $J=1$. Since by $(C_2)$, there exists an $x_0$ such that $F_1(x_0) \neq G_1(x_0)$, but by $(C_1)$, $F_1(-\infty)=G_1(-\infty)$, the interval $(-\infty, x_0)$ satisfies (A).

$I=1$, $J=\{1, 2\}$. By $(C_2)$, either of the inequalities $F_1(x) \neq G_j(x)$, $j=1, 2$, has at least one solution. If they have a simultaneous one, say $x_0$, then $(-\infty, x_0)$ satisfies (A), or else, if whenever $F_1(x_1) \neq G_1(x_1)$, we have $F_1(x_1)=G_2(x_1)$ and whenever $F_1(x_2) \neq G_2(x_2)$, then $F_1(x_2) = G_1(x_2)$ so that $(x_1, x_2)$ satisfies (A).

Received by the editors August 23, 1948 and, in revised form, October 27, 1948.

3. **Discontinuous d. f.'s.** So far no restrictions have been imposed upon the d. f.'s. Now we show that as long as the d. f.'s are allowed to have jumps, (A) does not always admit a solution when $I=1$, $J=\{1, 2, 3\}$. Namely we shall prove the following theorem:

THEOREM I. *Whatever be the distribution function $F(x)$, as long as it has at least two discontinuities, it is always possible, and in a continuum of ways, to determine three different distribution functions $G_j, j=1, 2, 3$, each different from $F(x)$ and such that for every Borel set $W$*

$$\int_W dF = \int_W dG_j \qquad \textit{for some } j.$$

PROOF. Let $a_1 < a_2$ be two discontinuity points of $F(x)$ and $a_3 > a_2$ an arbitrary point on the $x$-axis. Let the set of three points $a_1, a_2, a_3$ be denoted by $\sigma$. Let the saltus of $F(x)$ at $a_1, a_2, a_3$ be respectively $p_1, p_2, p_3$, where $p_1 > 0$, $p_2 > 0$, but $p_3 \geqq 0$ and $\sum_{i=1}^{3} p_i = s \leqq 1$. Then every Borel set $W$ can be represented as a sum $W' + W''$, where $W' = \sigma W$ and $W'' = W - W'$ and

$$\int_W dF = \int_{W'} dF + \int_{W''} dF.$$

Let $a_0 = -\infty$, $a_4 = +\infty$. Let $G_j(x), j=1, 2, 3$, be three d. f.'s satisfying the relations $G_j(x) - G_j(a_k) = F(x) - F(a_k)$ for $a_k \leqq x < a_{k+1}$ for $k = 0, 1, 2, 3$. Then, for every $W''$ disjoint with $\sigma$

$$(*) \qquad \int_{W''} dG_j(x) = \int_{W''} dF(x) \qquad \text{for every } j.$$

In order to complete the definition of the functions $G_j(x)$ it remains to specify their saltus at the points $a_1, a_2, a_3$. Thus let us denote the saltus of $G_j$ at $a_i$ by $q_{ji}$, $j, i=1, 2, 3$, and put

$$q_{jj} = p_j, \qquad j = 1, 2, 3,$$

while we choose arbitrarily any $q_{ji}$ such that

$$0 \leqq q_{ji} \neq p_i \qquad \text{for } i \neq j;\ i, j = 1, 2, 3,$$

and satisfying $\sum_{i=1}^{3} q_{ji} = s$, $j=1, 2, 3$. Then $G_j \not\equiv F$, $j=1, 2, 3$, but for $W' = a_i$, $i=1, 2, 3$,

$$\int_{W'} dF = p_j = \int_{W'} dG_j \qquad \text{for some } j.$$

For $W' = \sigma - a_i$, $i=1, 2, 3$,

$$\int_{W'} dF = s - p_j = \int_{W'} dG_j \qquad \text{for some } j,$$

and since

$$\int_{\sigma} dF = \int_{\sigma} dG_j \qquad \text{for all } j,$$

we have by (*), for arbitrary $W$,

$$\int_{W} dF = \int_{W} dG_j \qquad \text{for some } j.$$

Using Theorem I, we can show now that also in the case $I=\{1, 2\}$, $J=\{1, 2\}$, there exist sets of discontinuous d. f.'s for which (A) can not be satisfied. Namely let $F$, $G_1$, $G_2$, $G_3$ be a set of d. f.'s satisfying Theorem I and suppose further that $p_2<q_{32}<2p_2$. (This is obviously allowed by the construction.) Then $p_i+q_{2i}-q_{3i}\geqq 0$ for $i=1, 2, 3$, so that the function $F(x)+G_2(x)-G_3(x)$ is a d. f.

Let now

$$F_1^* = F, \qquad F_2^* = F + G_2 - G_3, \qquad G_1^* = G_1, \qquad G_2^* = G_2.$$

The sets $\{F_i^*\}$, $\{G_j^*\}$, $i=1, 2$, $j=1, 2$, obviously satisfy ($C_1$) and ($C_2$). Let $F-G_j=D_j$, $j=1, 2, 3$, and $F_i^*-G_j^*=D_{ij}^*$, $i=1, 2, j=1, 2$. Then

$$D_{11}^* = D_1, \qquad D_{12}^* = D_2, \qquad D_{21}^* = F + G_2 - G_3 - G_1, \qquad D_{22}^* = D_3,$$

so that the set $\{D_{ij}^*\}$, $i=1, 2, j=1, 2$, contains the set $\{D_j\}, j=1, 2, 3$. Since, by assumption, whatever be $W$, $\int_W dF-\int_W dG_j=\int_W d(F-G_j)=\int_W dD_j=0$ for at least one of the values $j=1, 2, 3$; also,

$$\int_{W} dD_{ij}^* = 0 \qquad \text{for at least one pair } i, j.$$

Therefore we can state the following theorem:

THEOREM II. *Whenever the number of distribution functions in the set $\{F_i\}$ multiplied by the number of distribution functions in the set $\{G_j\}$ is not less than 3 and the distribution functions are allowed to have jumps, it is not true that to every such pair of disjoint sets of distribution functions a Borel set $W$ exists for which*

$$\int_{W} dF_i - \int_{W} dG_j \neq 0 \qquad \textit{for all } i, j.$$

**4. Continuous distribution functions.** On the other hand we shall show that if the d. f.'s are continuous, the following theorem holds:

THEOREM III. *Let* $\{F_i\}$ *and* $\{G_j\}$ *be two denumerable sequences of distribution functions, each satisfying* ($C_1$) *and* ($C_2$) *and each continuous in* $x$. *Then there always exists a Borel set* $W$ *of arbitrarily small measure and such that*

$$\int_W dF_i - \int_W dG_j \neq 0 \qquad \textit{for all } i, j.$$

In the proof we shall make repeated use of the terms "point of change of a function" and "interval of change of a function," defined as follows:

Let $f$ be any function defined on a closed interval $[X, Y]$ and such that

$$f(X) \neq f(Y).$$

Let $Z$ be the least upper bound of those points $z$ of the interval $[X, Y]$ for which $f(z) \neq f(Y)$. $Z$ will be called the point of change of $f$ in $[X, Y]$. Obviously $Z$ has the property that given any $\delta > 0$, there exists a point $z$ for which

$$|z - Z| < \delta$$

and

$$f(z) \neq f(Z).$$

We shall say that $z$ and $Z$ determine an interval of change of $f$ in the $\delta$-neighborhood of $Z$.

PROOF OF THEOREM III. Let $F_i(x) - G_j(x) = D_{ij}(x)$. Since

$$\int_W dF_i - \int_W dG_j = \int_W d(F_i - G_j),$$

we have to find a solution for

$$\text{(A}') \qquad \int_W dD_{ij}(x) \neq 0 \qquad \text{for all } i, j.$$

Since the sequences $\{F_i\}$ and $\{G_j\}$ are enumerable, so is the set $\{D_{ij}\}$ and its elements may be enumerated in a single indexed sequence

$$D_1(x), D_2(x), \cdots, D_i(x), \cdots.$$

Thus (A′) becomes

$$\text{(A}''\text{)}\qquad \int_W dD_i(x) \neq 0 \qquad \text{for all } i$$

or, introducing the notation

$$\text{(A}'''\text{)}\qquad \begin{aligned} &\int_W dD_i(x) = V_i(W), \\ &V_i(W) \neq 0 \qquad \text{for all } i. \end{aligned}$$

By $(C_1)$, $(C_2)$, the functions $D_i(x)$ satisfy the following conditions:

$$(C_{12})\qquad \begin{aligned} &D_i(-\infty) = D_i(+\infty) = 0, \\ &D_i(X_i) \neq 0 \qquad \text{for at least one finite point } X_i. \end{aligned}$$

But since $D_i(x)$ is continuous, for each $D_i(x)$ there exists at least one other finite point $Y_i$ such that $D_i(Y_i)\neq 0$ and

$$(C_3)\qquad D_i(X_i) \neq D_i(Y_i).$$

Therefore we can assign to every $D_i(x)$ a point of change $Z_i$ and to every given $\delta$ an interval of change in the $\delta$-neighborhood of $Z_i$.

Let now $\epsilon>0$ be given in advance. First we shall show that it is possible to construct a sequence of Borel sets $W_1, W_2, \cdots, W_r, \cdots$ with the following properties:

(i) $W_r$ consists of a finite number of disjoint intervals;

(ii) The measure of $W_r$ is smaller than $\epsilon/2^r$;

(iii) $V_1(W_1)\neq 0$;

(iv) If $V_i(W_1)\neq 0$ for $i\leqq i_1$ but $V_{i_1+1}(W_1)=0$, then there exists an $i_2>i_1$ such that $V_i(W_2)\neq 0$ for $i\leqq i_2$ and, in general, if $V_i(W_r)\neq 0$ for $i\leqq i_r$, but $V_{i_r+1}(W_r)=0$, then $V_i(W_{r+1})\neq 0$ for $i\leqq i_{r+1}>i_r$.

If in the process of construction we arrive at a $W_r$ such that $V_i(W_r)\neq 0$ for all $i$, the theorem is proved. If, however, we arrive at an infinite sequence, we shall show that this sequence has a non-empty limit set $W$ which satisfies the requirements of the theorem.

The construction of the sequence $W_r$ is done by induction: Let $Z_1$ be a point of change of $D_1$ and $W_1$ an interval of change of $D_1$ in the $\epsilon/2$-neighborhood of $Z_1$. Then (iii) holds. Assume now that we succeeded in finding $r$ terms of the sequence, $W_1, W_2, \cdots, W_s, W_{s+1}, \cdots, W_r$ such as requested. Then we know that

$$\begin{aligned} &(1)\qquad V_i(W_s) \neq 0, \quad i \leqq i_s, \\ &(2)\qquad V_{i_s+1}(W_s) = 0. \end{aligned} \qquad \text{for } s = 1, 2, \cdots, r,$$

Let us for brevity put $i_r+1=\nu$ and let $Z_\nu$ be a point of change of $D_\nu$.

Since by assumption $W_r$ consists of a finite number of disjoint intervals, we can and should choose a $\delta_r$ such that $0<\delta_r<\epsilon/2^r$ and so small that the $\delta_r$-neighborhood of $Z_\nu$, and therefore every interval of change in it, should not contain any boundary point of $W_r$ except perhaps $Z_\nu$.

By (1) for every $s \leqq r$, there exists an $\eta_s>0$ such that

$$(3) \qquad |V_i(W_s)| > \eta_s, \qquad i \leqq i_s,\ s \leqq r.$$

Let now $\sum_1^\infty \alpha_i = \alpha < 1$, $\alpha_i > \alpha_{i+1} > 0$, be any given series and $[A_\nu, B_\nu]$ any closed interval $A_\nu < Z_\nu < B_\nu$. Since $D_1, D_2, \cdots, D_{i_r}$ are continuous in $[A_\nu, B_\nu]$, $\delta_r$ can and should be further adjusted to be so small, that for every subinterval $I(\delta_r)$ of length smaller than $\delta_r$ in $[A_\nu, B_\nu]$

$$(4) \qquad |V_i[I(\delta_r)]| < \eta_s \alpha_r, \qquad s \leqq r,\ i \leqq i_r,$$

should hold.

Having $\delta_r$ thus fixed, let $\Delta_r$ be an interval of change of $D_\nu$ in the $\delta_r$-neighborhood of $Z_\nu$, so that

$$(5) \qquad |\Delta_r| < \delta_r < \epsilon/2^r$$

and

$$(6) \qquad |V_\nu(\Delta_r)| = |V_{i_r+1}(\Delta_r)| > 0.$$

Let now

$$W_{r+1} = W_r + (-1)^{p_r}\Delta_r,$$

where $p_r = 2$ if $\Delta_r$ contains no inner points of $W_r$, and $p_r = 1$ if $\Delta_r$ contains inner but no outer points of $W_r$. Then by (3), (4), and (5)

$$|V_i(W_{r+1})| \geqq |V_i(W_r)| - |V_i(\Delta_r)| > \eta_r - \eta_r\alpha_r > 0 \qquad \text{for } i \leqq i_r,$$

and by (2) and (6)

$$|V_{i_r+1}(W_{r+1})| = |V_{i_r+1}(\Delta_r)| > 0.$$

Let us now put

$$W = W_1 + \sum_1^\infty (-1)^{p_r}\Delta_r.$$

By (5), $W$ is not empty and its measure is less than $\epsilon$. By (4) we have

$$(7) \quad |V_i(\Delta_r)| < |V_i(W_s)|\alpha_r, \quad i \leqq i_s,\ s = 1, 2, \cdots, r;\ r = 1, 2, \cdots.$$

Let $i$ be a fixed arbitrary index. Find in the sequence $i_1, i_2, \cdots$ the indices $i_{k-1}, i_k$ for which

$$i_{k-1} < i \leqq i_k.$$

By (7)

$$|V_i(W)| \geqq |V_i(W_k)| - \sum_{j=0}^{\infty} |V_i(\Delta_{k+j})|,$$

while for any $j$

$$|V_i(\Delta_{k+j})| < |V_i(W_k)| \alpha_{k+j},$$

thus

$$\sum_{j=0}^{\infty} |V_i(\Delta_{k+j})| < |V_i(W_k)| \sum_{0}^{\infty} \alpha_{k+j}$$

and

$$|V_i(W)| > |V_i(W_k)| \left(1 - \sum_{0}^{\infty} \alpha_{k+j}\right) \geqq 0,$$

$$|V_i(W)| > 0,$$

Q.E.D.

Further results concerning the case when $I$ and $J$ range over a continuum of indices are discussed in a joint paper with Wald in one of the recent issues of the Annals of Mathematical Statistics.

While preparing this paper I had opportunities to discuss it with Professor Neyman and I wish to acknowledge my indebtedness for his generous advice and many helpful suggestions.

NEW YORK CITY