The simplest form of Jensen's inequality is that if $\phi(x)$ is a convex function, and $m$ is the arithmetic mean of $x_1, \ldots, x_n$, then the mean of the numbers $\phi(x_n)$ is not less than $\phi(m)$. This inequality can be generalized in several different ways. The function $\phi(x)$ can be replaced by a convex function of several variables, and the arithmetic mean can be replaced by any one of several other means, as has been shown in various proofs. Since the inequality is of considerable utility, it seems worth while to have it established in a form which is general enough to cover a wide assortment of applications.

The proofs will rest on two well known properties of convex sets.† If $K$ is closed and convex and a point $p$ is not in $K$, then $p$ can be separated from $K$ by a hyperplane. If $K$ is closed and convex and $p$ is a boundary point of $K$, there is a hyperplane of support of $K$ passing through $p$.

1. The Inequality in Geometric and in Analytic Form. It will be convenient in the following proofs to use these symbols and definitions:

$R_n$ is $n$-dimensional Euclidean space. Its points will be denoted by $(z_1, \ldots, z_n)$ or by $z$. Linear functions $\sum a_i z_i$ or $R_n$ will be symbolized by $l(z)$.

---

* Presented to the Society, December 31, 1936.
† A set is convex if for every pair $P, Q$ of points of the set the line segment $PQ$ is contained in the set.
$L$ is a linear class of real valued functions $f(x)$ defined on a set $E$. It shall be supposed to have the properties:

(1a) If $f_1, f_2$ are in $L$ and $k_1, k_2$ are real numbers, $k_1f_1 + k_2f_2$ is in $L$.

(1b) The function defined and constantly equal to 1 on $E$ is in $L$.

$Mf$ is a linear mean defined on $L$, with these properties:

(2a) $M1 = 1$.

(2b) If $f_1, f_2$ are in $L$ and $k_1, k_2$ are real numbers, $M(k_1f_1 + k_2f_2) = k_1Mf_1 + k_2Mf_2$.

(2c) If $f(x)$ is in $L$ and $f(x) \geq 0$, then $Mf \geq 0$.

If $f(x)$ is an $n$-tuple of functions $(f_1(x), \cdots, f_n(x))$ of $L$, we denote by $Mf$ the $n$-tuple $(Mf_1, \cdots, Mf_n)$. From (2) we obtain

(3) $Ml(f) = l(Mf)$ for every function $l(z)$ linear on $R_n$.

The geometric formulation of Jensen's inequality is as follows:

**Theorem 1.** Let (1) and (2) be satisfied. Let $K$ be a closed convex point set in $R_n$. Let $f_1(x), \cdots, f_n(x)$ be functions of the class $L$ such that $f = (f_1, \cdots, f_n)$ is in $K$ for all $x$ in $E$. Then $Mf$ is in $K$.

Let $l(z) + c = 0$ be a hyperplane in $R_n$ such that $K$ is entirely to one side, say $l(z) + c \geq 0$ for $z$ in $K$. Then $l(f) + c \geq 0$ for all $x$, and $0 \leq M(l(f) + c) = Ml(f) + Mc = l(Mf) + c$, so that $Mf$ lies on the same side of the hyperplane as $K$. That is, no hyperplane separates $Mf$ from $K$. This is only possible if $Mf$ is in $K$.

The more usual analytical formulation of the inequality is covered by the following theorem:

**Theorem 2.** Let (1) and (2) be satisfied. Let $K$ be a closed convex point set in $R_n$, and let $\phi(x)$ be continuous and convex* on $K$. Let $f_1(x), \cdots, f_n(x)$ be functions of the class $L$ such that $f(x) = (f_1(x), \cdots, f_n(x))$ is in $K$ for all $x$ in $E$, and such, moreover, that $\phi(f(x))$ is in the class $L$. Then $\phi(Mf)$ is defined and

$$(J) \quad \phi(Mf) \leq M\phi(f).$$

* We call $\phi$ convex on $K$ if $\phi(\frac{1}{2}(z_1 + z_2)) \leq \frac{1}{2}(\phi(z_1) + \phi(z_2))$ for all $z_1, z_2$ in $K$. 

Denote the points \((s_1, \ldots, s_{n+1})\) of \(R_{n+1}\) by \(z^1\). These can also be denoted by \((z, s_{n+1})\), where \(z\) is in \(R_n\). Define \(K_1\) to be the set in \(R_{n+1}\) of points \((z, s_{n+1})\) such that \(z\) is in \(K\) and \(s_{n+1} \geq \phi(z)\). It is clear that \(K_1\) is closed and convex, and for all \(x\) in \(E\) the point \((f(x), \phi(f(x)))\) is in \(K_1\). Hence by Theorem 1, the point \((Mf, \phi(f))\) is in \(K_1\); that is, \(Mf\) is in \(K\) and \(M\phi(f) \geq \phi(Mf)\).

Obviously we could weaken our hypotheses somewhat by omitting the requirement that \(\phi\) be continuous and assuming instead that the set \(K_1\) is closed and convex. This would permit \(\phi(z)\) to have infinite discontinuities on the boundary of \(K\).

**Examples:**

1. Let \(\alpha(x_1, \ldots, x_m)\) be a positively monotonic* function of the variables \((x_1, \ldots, x_m)\), and let \(E\) be a set measurable with respect to \(\alpha\) and such that \(0 < m_\alpha E < \infty\). Let \(L\) be the class of all functions \(f(x)\) which are Lebesgue-Stieltjes integrable with respect to \(\alpha\) over \(E\). Let \(Mf = \int_E f d\alpha/\int_E d\alpha\). Then conditions (1) and (2) are satisfied. The conclusion of Theorem 2 assumes the form

\[
\phi\left(\frac{\int f_1 d\alpha}{\int d\alpha}, \ldots, \frac{\int f_n d\alpha}{\int d\alpha}\right) \leq \int \phi(f_1, \ldots, f_n) d\alpha/\int d\alpha.
\]

In the next three examples requirements of convergence or integrability are too obvious to need statement:

2. The range of \(x\) is \((1, \ldots, m)\) or \((1, 2, \ldots)\), so that \(f(x)\) is a (finite or infinite) sequence \((a_1, a_2, \ldots)\), and \(Mf = \sum c_i a_i/\sum c_i\), where \(c_i \geq 0\) and \(0 < \sum c_i < \infty\).

3. The functions \(f\) of \(L\) are continuous, and \(Mf = \int f d\alpha/\int d\alpha\).

4. The functions \(f\) of \(L\) are Lebesgue measurable over the measurable set \(E\), and \(Mf = \int f p d\alpha/\int p d\alpha\), where \(p(x) \geq 0\) and \(0 < \int p d\alpha < \infty\).

5. \(E\) is in the interval \((0,1)\), \(L\) is the class of all bounded functions on \(E\), \(Mf\) is the Banach integral† of \(f\) over \((0,1)\).

6. \(E\) is the set of all real numbers, \(L\) the class of all uniformly almost periodic functions, \(Mf\) is the mean value of \(f\).

* That is, a function whose \(m\)th difference is non-negative.
† Banach, *Théorie des Opérations Linéaires*, p. 31.
(7) More generally, $E$ is any group, $L$ is the class of all functions almost periodic on $E$, $Mf$ is von Neumann’s* mean value of $f$.

(8) In the space $(x_1, x_2, \cdots)$ of infinitely many dimensions, $(0 \leq x_i \leq 1)$, $E$ is a set of finite positive measure,† $L$ is the class of functions summable over $E$, $Mf = \int_E f \, dx/mE$.

2. Conditions for Strict Inequality. In §1 we have not mentioned conditions for strict inequality. To investigate this question it is convenient to define negligible sets. A set $S \subset E$ is negligible (with respect to $L$ and $M$) if there exists a function $f$ in the class $L$ such that

(a) $f(x) \geq 0$ on $E$,  
(b) $f(x) > 0$ on $S$,  
(c) $Mf = 0$.

It follows readily that every subset of a negligible set is negligible, and so is every set which is the sum of a finite number of negligible sets. It is then easy to prove the following theorem:

**Theorem 3.** In Theorem 1, $Mf$ is a boundary point of $K$ only if all points $f(x)$ except those corresponding to a negligible set of $x$ belong to the intersection of $K$ with one of its hyperplanes of support.

If $Mf$ is a boundary point of $K$, through it there passes a hyperplane of support $\pi: l(z) + c = 0$ of $K$; say $l(z) + c \geq 0$ for $z$ in $K$. Let $S$ be the set of $x$ such that $f(x)$ is not in the intersection $\pi K$; then $l(f(x)) + c > 0$ on $S$. Since $l(f(x)) + c \geq 0$ for all $x$ and $M(l(f(x)) + c) = l(Mf) + c = 0$, we see that $S$ is negligible.

I omit the easy proof of the following theorem:

**Theorem 4.** If the set $K$ is strictly convex, and $\phi$ is strictly convex on $K$, then in Theorem 2 equality holds only if $f_i(x) = Mf_i = \text{const.} (i = 1, \cdots, n)$ except on a negligible set.

I am unable to state whether the conditions $f_i = Mf_i$ except on a negligible set are sufficient as well as necessary for equality in Theorem 4. However, by adding a further postulate concern-

ing $L$ and $M$ we can establish this, even when $K$ is not strictly convex. We henceforth restrict our attention to systems $L$, $M$ such that the following condition holds:

(4) If $S$ is any negligible subset of $E$, and $f(x)$ is any (real) function which vanishes on $E - S$, then $f(x)$ is in the class $L$ and $Mf=0$.

In example (1) negligible sets are sets of measure $m_a S = 0$; hence condition (4) is satisfied. In example (2), a set $S$ of integers is negligible if $\sum_{i \in S} i = 0$; in (4), $S$ is negligible if $p(x) = 0$ on almost all of $S$. In (8), $S$ is negligible if $mS = 0$. In (3), (6), and (7), only the empty set is negligible. For all of these (4) is valid. I do not know whether example (5) satisfies (4).

An immediate consequence of conditions (1), (2), and (4) is that if $f(x)$ is any function of class $L$, and $g(x) = f(x)$ except on a negligible set, then $g(x)$ is in $L$ and $Mf = Mg$.

**Theorem 5.** If condition (4) is satisfied, then in inequality (J) equality holds if and only if the following condition holds:

For all $x$ except at most those belonging to a negligible set $S$, the point $(\varphi(x), \cdots, \varphi_n(x))$ belongs to a convex subset $K'$ of $K$ on which $\varphi(x)$ is linear. In particular, if $\varphi(x)$ is strictly convex* equality holds if and only if $\varphi(x) = \text{const.}$ except on a negligible set.

The last statement is an immediate consequence of the first, for if $\varphi$ is strictly convex the only subsets $K'$ on which $\varphi$ is linear consist of single points. Suppose then that the condition of Theorem 5 holds; by redefining $f_i(x)$ on $S$, it will be true that $f$ is in $K'$ for all $x$, without change in $Mf$ or $M(\varphi(f))$. On $K'$ we have $\varphi(x) = l(x) + c$; hence $M\varphi(f) = M(l(f) + c) = M(lf) + c$. But since $K'$ is convex, by Theorem 1 the point $Mf$ is in $K'$, and so $\varphi(Mf) = l(Mf) + c$. Hence equality holds in the inequality (J).

To prove the necessity of our condition we first observe that there may be linear relations $l(f(x)) + c = 0$ holding for all $x$ except those of a negligible set. We choose a maximal set of such linear relations; there is no loss of generality in assuming that these are of the form $f_{i+1}(x) = \cdots = f_n(x) = 0$ except on $S_i$, where $S_i$ is negligible. Then $Mf_{i+1} = \cdots = Mf_n = 0$.

We now change notation. Let $R_n$ be the space of points $(z_1, \cdots, z_n)$; let $H$ be the set of $(z_1, \cdots, z_n)$ such that $(z_1, \cdots, z_n, 0, \cdots, 0)$ is in $K$; let $\psi(z_1, \cdots, z_n) = \varphi(z_1, \cdots, z_n)$.

* We assume only that $K$ is convex, not that it is strictly convex.
0, \cdots, 0) on H; in the space $R_{s+1}$ of points $(z_1, \cdots, z_8, z_{n+1})$
let $H_1$ be the set for which $(z_1, \cdots, z_8)$ is in $H$ and
$z_{n+1} \geq \psi(z_1, \cdots, z_8)$. For $x$ in $E-S_1$ we have $(f_1(x), \cdots, f_s(x))$
in $H$ and $(f_1, \cdots, f_s, \psi(f_1, \cdots, f_s))$ in $H_1$. If equality holds in $(J)$, then
$$M\psi(f_1, \cdots, f_s) = M\phi(f_1, \cdots, f_s, 0, \cdots, 0) = M\phi(f_1, \cdots, f_n)$$
$$= \phi(Mf_1, \cdots, Mf_s, 0, \cdots, 0)$$
$$= \psi(Mf_1, \cdots, Mf_s),$$
so the point $(Mf_1, \cdots, Mf_s, M\psi(f_1, \cdots, f_s))$ is a boundary
point of $H_1$ in $R_{s+1}$.

By Theorem 3, for all points $x$ of $E-S_1$ except a negligible
set $S_2$ the point $(f_1(x), \cdots, f_s(x), \psi(f_1, \cdots, f_s))$ belongs
to the intersection of $H_1$ with a hyperplane of support $\pi$: $a+b_1z_1+\cdots+b_8z_8+cz_{n+1}=0$. That is, except on $S_1+S_2$ the
equation $a+b_1f_1+\cdots+b_sf_s+\psi(f_1, \cdots, f_s)=0$ holds. Here
$c\neq 0$; otherwise we would have a new linear relation between the
$f_i$ independent of the maximal set $f_{s+1}=\cdots=f_n=0$. We
may therefore suppose that $c=1$; hence $\pi$ has the equation
$a+\sum b_i z_i + z_{n+1}=0$. The left member of this equation is positive
for some points $(z_1, \cdots, z_8, z_{n+1})$ of $H_1$, since $z_{n+1}$ is arbitrarily
large. But $\pi$ is a hyperplane of support, so $a+\sum b_i z_i + z_{n+1}$
does not change sign on $H_1$; therefore $a+\sum b_i z_i + z_{n+1} \geq 0$ on $H_1$.

For $(z_1, \cdots, z_8)$ in $H$ the point $(z_1, \cdots, z_8, \psi(z_1, \cdots, z_8))$
is in $H_1$, so that $a+\sum b_i z_i + \psi(z_1, \cdots, z_8) \geq 0$ on $H$. Moreover, if
$z_{n+1} > \psi(z_1, \cdots, z_8)$, then $a+\sum b_i z_i + z_{n+1} > 0$. Hence the points of
$H_1$ which lie in $\pi$ are those of the form $(z_1, \cdots, z_8, \psi(z_1, \cdots, z_8))$
with $(z_1, \cdots, z_8)$ in $H$ and $a+\sum b_i z_i + \psi(z_1, \cdots, z_8)=0$. The
points satisfying these conditions form the intersection $\pi H_1$, which,
being the intersection of convex sets, must be convex, as
is also its projection $H'$ on the space $R_8$. For all $x$ in $E-(S_1+S_2)$
the point $(f_1(x), \cdots, f_s(x), \psi(f_1, \cdots, f_s))$ is in $\pi H_1$, so that
$(f_1(x), \cdots, f_s(x))$ is in $H'$. If we define $K'$ to be the set of all
points $(z_1, \cdots, z_8, 0, \cdots, 0)$ with $(z_1, \cdots, z_8)$ in $H'$, then $K'$ is
convex, and for all $x$ in $E-(S_1+S_2)$ the point $(f_1(x), \cdots, f_s(x),
0, \cdots, 0)$ is in $K'$. If $(z_1, \cdots, z_8)$ is in $K'$ then $(z_1, \cdots, z_8)$
is in $H'$, and
$$\phi(z_1, \cdots, z_8) = \phi(z_1, \cdots, z_8, 0, \cdots, 0)$$
$$= \psi(z_1, \cdots, z_8) = -a - \sum b_i z_i.$$
This establishes the theorem.

3. Extension to Banach Spaces. It is possible to extend Theorems 1, 2, 3, and 4 to functions $f(x)$ assuming values in a Banach space $B$. Suppose that (1) and (2) are satisfied, and that $L$ is a class of functions $f(x)$ defined on $E$ and assuming values in $B$. We shall assume that for every linear function $l(z)$ on $B$ and every $f(x)$ in $L$ the function $l(f(x))$ is in $L$. Further, we shall assume that there is a linear mean $M$ defined on $L$ such that for every linear function $l(z)$ on $B$ and every $f$ in $L$ we have

$$l(Mf) = M(l(f)).$$

We then find that Theorems 1, 2, 3, and 4 extend with only one change. The only properties of convex sets which we used were these: through each boundary point of a convex set there passes a hyperplane of support, and each point which does not belong to a convex set can be separated from it by a hyperplane. These properties have been established for convex bodies* (closed convex sets having interior points). Hence our theorems extend at once, provided that we replace the words "convex set" by "convex body."

University of Virginia