# The Refinement Integral

by

*Isidore Fleischer*

*C.R.M., Université de Montréal, C.P.6128*

*Succ. Centre-Ville, Montréal, P.Q. H3C 3J7, Canada*

**Abstract**

The refinement integral already surfaces in implicit form in Darboux's reformulation of the Riemann(-Stieltjes) integral, although the recognition that the upper and lower integrals are refinement depends on the subsequent convergence theory of E.H. Moore, who also observed that the Lebesgue integral could be construed as refinement. For a sub- (or super-) additive integrand, the refinement integral exists as the sup (or inf) over the approximating sums; it can sometimes be evaluated as a standard integral (e.g. for BV functions of quotients of measures by replacing the quotient with the derivative and integrating this point function). Shannon's definitions of information theory notions for continuous distributions can be corrected, and the statisticians' formal integrals made sense of, as refinement integrals. A substantial study of the refinement integral was carried out by Kolmogorov in Math. Ann. 103 (1930) 654–696 — he integrates real-valued integrands of a subset argument on a domain of subsets closed only for intersection. Some extension and application of this basic source will be presented.

Leibniz's conception of integration was as addition of infinitely many infinitely small quantities: To each $x$ he imagined assigned an "infinitesimal" $dy$ whose "sum" for all the $x$'s between $a$ and $b$ was to be the integral $\int_a^b dy$. Eighteenth century mathematicians were unable to make head or tail of this and it was not until Cauchy, who "translated the definition from the language of metaphysics into that of mathematics" that it became usable. Cauchy approximated the "infinitesimal" $f(x)dx$ (for a continuous function $f$) by the finite $f(\xi_i)(x_i - x_{i-1})$ for any $\xi_i$ between $x_{i-1}$ and $x_i$; took the finite sum of these for any partition of the interval by finitely many $x_i$'s; and then defined $\int_a^b f(x)dx$ as the limit of these finite sums as the maximum length of the partitioning subintervals goes to zero. By the (uniform) continuity of $f$ the result is independent of the choice of the $\xi_i$ in the subintervals: indeed, the approach to the limit is uniform in these choices—a uniformity which must be postulated when defining the integral for discontinuous $f$. The approximation to $dy$ on a subinterval $\Delta$ is thus the subset $f(\Delta)\ell(\Delta)$ of products of all values of $f$ on it by its length; this multi-valued function of an interval, the diameter of whose value goes to zero as the interval shrinks to a point, represents the infinitesimal; and the limit of its sum over the partition, the result of the infinite summation.

The multi-valued functions can be avoided by recourse to the Darboux definition: this encompasses the totalities $f(\Delta)$ in terms of their bounds, the sup, $\overline{f(\Delta)}$ and inf, $\underline{f(\Delta)}$ of $f$ on $\Delta$; the upper/lower Darboux sums for a given partition are $\Sigma\overline{f(\Delta_i)}\ell(\Delta_i)/\Sigma(\underline{f\Delta_i})\ell(\Delta_i)$ and the upper/lower integrals, $\inf\Sigma\overline{f(\Delta_i)}\ell(\Delta_i)/\sup\Sigma(\underline{f\Delta_i})\ell(\Delta_i)$ over all partitions. These quantities are actually (refinement) limits: for by adding an additional point to a partition—i.e. by splitting one of the $\Delta_i$ into a pair of contiguous subintervals—the upper sum can at most decrease and the lower at most increase. For any $\epsilon > 0$ there is a partition whose upper/lower sum is within $\epsilon$ of the inf/sup—consequently so is every finer partition. The formulation as a refinement integral permits generalizing the setting from interval functions to set functions defined for subsets of an abstract set; it is also no longer needed to have these of the particular form of products—the additive "length function" $\ell$ can be absorbed by the set function.

Fix a system $\mathcal{M}$ of subsets of some set $X$; a **partition**, $\mathcal{D}E$, of an $E \in \mathcal{M}$ is a representation of $E$ as a finite disjoint union of sets $E_n \in \mathcal{M}$, written $E = \Sigma E_n$. Partition $\mathcal{D}'E$ **refines** $\mathcal{D}E, \mathcal{D}'E > \mathcal{D}E$, if every $E'_n \in \mathcal{D}'E$ is contained in some $E_m \in \mathcal{D}E$—this is transitive: indeed, a partial order on the partitions of $E$.

We'll postulate $\mathcal{M}$ closed for (finite) intersection: Then every $\mathcal{D}E$ induces on every $E' \in \mathcal{M}$ contained in $E$ a $\mathcal{D}E' : E' = \Sigma E_n \cap E'$; every pair of partitions $\mathcal{D}E, \mathcal{D}'E$ have a common (actually a least fine) refinement $(\mathcal{D}\cap\mathcal{D}')E := \Sigma E_m \cap E'_n$. This makes the partition partial order directed, which is what is needed for the usual properties of limit to hold: see Hildebrandt.

Let $F$ be a real-valued function of a set argument (not taking $-\infty$ as value) **differentially** defined i.e. (at least) on the sets making up sufficiently fine partitions: e.g. on all those which refine some given $\mathcal{D}E$. Extend $F$ to a partition function as $F(\mathcal{D}E) = \Sigma F(E_j)$. Thus

$F$ becomes a "net" and if it converges as $\mathcal{D}E$ is refined, $F$ is called (refinement) **integrable**, its limit being denoted $\int_E F(dE)$.

To obtain the standard integral, say of a bounded measurable function $f$ on a finite measure space $X$ with its field $\mathcal{M}$ of subsets measurable for a finitely additive measure $\mu$, take for $F$ the function which assigns $E \in \mathcal{M}$ the product $f(\xi)\mu E$ for some $\xi \in E$. One will require the convergence to $\int f(x)\mu(dx)$ (of $\Sigma f(\xi_i)\mu E_i$) to occur for all choices of the $\xi_i$—and even uniformly in the chosen $\xi_i$: thus convergence of the multi-valued $\Sigma f E_i \cdot \mu E_i$. This can be modeled by an $F$ which assigns each $E \in \mathcal{M}$ a subset of reals rather than a singleton (equivalently, to have uniform convergence for all the single-valued $F$'s sending $E$ into this subset). The convergence of the $F$ extended to partitions is the same as that for the smallest intervals enclosing these image subsets, hence one could restrict to interval-image $F$'s.

Convergence of (a net of) intervals comes to convergence of the left and right endpoints to the same limit. Every interval-valued function $F$ on $\mathcal{M}$ yields two such functions on the partitions $\mathcal{D}'E > \mathcal{D}E$: an upper $\Sigma\overline{F}E'_m$ and a lower $\Sigma\underline{F}E'_m$: the former is $\geq$ the latter on the same partition, but nothing further can be asserted in this general setting. The existence of the integral must thus require explicitly the existence of the refinement limits of the partition functions $\Sigma\overline{F}E'_m$ and $\Sigma\underline{F}E'_m$—these limits could be called the upper and lower integral; (their $\geq$ is still maintained)—as well as their equality, the common value being by definition the integral $\int_E F(dE)$. What this comes to is that the intervals $[\Sigma\underline{F}E'_m,\ \Sigma\overline{F}E'_m]$ refinement converge to this limit.

Corresponding to this three-part requirement is a three-part Cauchy condition: For $\forall\epsilon > 0$ there should exist a $\mathcal{D}E$ such that for all $\mathcal{D}'E > \mathcal{D}E$, $|\Sigma\overline{F}E'_i - \Sigma\overline{F}E_j| < \epsilon$; the same with $\underline{F}$ replacing $\overline{F}$; and finally $\Sigma\overline{F}E'_i - \Sigma\underline{F}E'_i < \epsilon$. (These could be combined into a single condition: $|\Sigma\overline{F}E'_i - \Sigma\underline{F}E''_j| < \epsilon$ for all $\mathcal{D}', \mathcal{D}'' > \mathcal{D}$. Proof: First deduce the last inequality.)

Kolmogorov's take on the Leibniz paradigm is thus to realize the "infinitesimal" at a point as the set function $F$ on the filter of subsets in the domain of $F$ containing the point, the diameter of whose value converges to zero as the subsets shrink to the point, and their "infinite sum" over $E$ as the refinement limit of the values of $F$ as evaluated on the partitions of $E$.

*The integral is additive in the domain of integration*: The subsets over which it exists are closed under disjoint union and its value is the sum of its values on the summands. Proof: Sufficiently fine partitions of the union are obtained by combining partitions of the summands; and the limit of a sum is the sum of the limits (if they exist).

*Integrability is inherited by every $E' \subset E$ which occurs in some $\mathcal{D}E$.* Proof: The Cauchy condition holds also with $\mathcal{D}$ replaced by any refinement, e.g. by $\mathcal{D} \cap \mathcal{D}'$, which induces a partition of $E'$. Complete any refinement of the latter to a partition of $E$ by leaving $\mathcal{D} \cap \mathcal{D}'$ unchanged on the complement of $E'$. The first two Cauchy conditions then follow from their holding for partitions of $E$; the last from the positivity of the difference. By additivity, this may be construed as the integral of the function made zero on the complement.

As a function of $E$, $\int_E F(dE)$ is appropriately designated **indefinite integral** (on the real line this construes $\int f(x)\,dx$ as an additive interval function rather than as a function of its upper limit): it is a single-valued additive set function defined on the subsets of the domain occurring in one of its partitions.

We will now characterize this set function and so obtain an alternate "descriptive" definition of the refinement integral.

To this end, declare two differentially defined functions $f$ and $g$ to be **differentially equivalent** if for every $\epsilon > 0$ there is a $\mathcal{D}E$ such that for every finer $\mathcal{D}'E$, $\Sigma|fE'_n - gE'_n| < \epsilon$. What this comes to is that the set function $|f-g|(E') := |fE' - gE'|$ integrates to 0 over $E$. Since it is non-negative, its indefinite integral is identically zero on $\mathcal{M}E$, the subsets occurring in partitions of $E$ (reducing the integrand to 0 on the complement of its domain results in an integrand dominated by $|f-g|$) a situation which could be called "zero almost everywhere." Since the absolute value of $h := f - g$ integrates identically to zero, it follows (e.g. from $-|h| \leq h \leq |h|$) that so does $h$.

We show the converse: If $\int_G h(dG) = 0$ for all $G \in \mathcal{M}E$ then also $\int_G |h|(dG) = 0$ — specifically, if $|\Sigma\overline{h}(E'_m)|$, $|\Sigma\underline{h}E'_m| < \epsilon$ for all $\mathcal{D}'E > \mathcal{D}E$ then $\Sigma|h(E'_m)| \leq 4\epsilon$ for sufficiently fine $\mathcal{D}'E$ — if not: i.e. if $\lim\limits_{\mathcal{D}'E>\mathcal{D}E} \sup \Sigma|\underline{h}E_n{}'| > 4\epsilon$, one of these $\mathcal{D}'E$ would have the sum of its like-signed terms, say the positive ones, $> 2\epsilon$; the integral over each of the $E'$ sent by $\underline{h}$ to a negative being zero, one can decompose each of these so that the sum of $\underline{h}$ over all these subpartitions is $> -\epsilon$, resulting in a global decomposition of $E$ finer than $\mathcal{D}E$ over which $\underline{h}$ sums to $> \epsilon$.

Thus differential equivalence of $f$ and $g$ comes to $\int_G f(dG) - g(dG) = 0$ for all $G \in \mathcal{M}E$: If one of $f, g$ is integrable, so is the other and the integrals are equal — and conversely. In particular, since the integral is a single-valued additive function on $\mathcal{M}E$, its integral exists and equals its value, so that every integrable function is differentially equivalent to its indefinite integral. Conversely, if $f$ is differentially equivalent to a single-valued additive function, it is integrable with indefinite integral that function. The indefinite integral is thus the unique additive function differentially equivalent to the integrand.

A (single-valued) additive function was seen to be integrable (to its value on the domain $E$); more generally, a **subadditive** function, i.e. $f(E) \leq \Sigma f(E_m)$ on the partitions of every $E$, is integrable to the $\sup \Sigma f(E_m)$ over all partitions of an $E$ (one needs to require the $\Sigma f(E_m)$ to be bounded if one is unwilling to accept infinite values for the integral). If $f$ is additive $|f|$ is subadditive (since $|\ \ |$ is): $\int_E |f(dE)|$ is then the (total) **variation** of $f$ on $E$: e.g. if $F$ is the interval function $F(b) - F(a)$ then this yields the familiar $\sup \Sigma |F(x_i) - F(x_{i-1})|$. A $BV$ additive $F$ can be written as a difference of non-negatives: $\frac{1}{2}\left[\int_G |f(dG)| + f(G)\right] - \frac{1}{2}\left[\int_G |f(dG)| - f(G)\right]$ ($\int \geq \pm f(G)$ by subadditivity)— however without the extremal property of Jordan's.

If $f$ is integrable over $E$, its differential equivalence with its indefinite integral $F$ entails that of $|f|$ with $|F|$. Thus an absolutely integrable $f$ has a $BV$ indefinite integral (in the Lebesgue theory, it is even absolutely continuous).

From every (possibly multi-valued) point function $f(x)$ one can create a set function $f(E) := \bigcup_{x \in E} f(x)$. This is a (complete) $\cup$–morphism. For just a subset–preserving interval–valued set map, define its "measurability" to mean convergence (under refinement) to zero of $\max_i \overline{f} E_i - \underline{f} E_i$ for partitions $E = \Sigma E_i$. Then a product $f\mu$, with $f$ measurable and $\mu$ $BV$, is integrable. If $\mu$ is non-negative and additive, the upper/lower sums are non-increasing/-decreasing under refinement and one obtains the upper/lower integral as an inf / sup as on the real line.

The following convergence theorem can be extracted from [Fl].

Let $f$ be bounded on a BV set $E$ containing an increasing sequence $E_n$ of subsets converging in measure to E; for $f_n\mu$ integrable on $E_n$, let $\sup_{E_n} |fx - f_n x|$ converge to zero. Then $f\mu$ is integrable to $\lim \int_{E_n} f_n \mu$.

Call an interval-valued set function $g$ **differentially contained** (on $E$) in $f$ if on sufficiently fine partitions of $E = \Sigma E_m, \Sigma \underline{f} E_m \leq \Sigma \underline{g} E_m \leq \Sigma \overline{g} E_m \leq \Sigma \overline{f} E_m$. Then if $f$ is integrable over $E$, so is $g$, and to the same value. The condition is certainly satisfied if the values of $g$ on every subset $E'$ (in sufficiently fine partitions of $E$) lie between $\underline{f} E'$ and $\overline{f} E'$—i.e. if values of $g$ are straddled by values of $f$ on every such subset. Values of $g$ are straddled by values of a product $f\mu$ , of which the second factor $\mu$ is single-valued, if and only if values of $g/\mu$ are by those of $f$— this entails that integrability of $f\mu$ yields that of $g$ to the same value.

This justifies the additivity of the integral for integrands which are products: i.e. the integrability of $fg$ and $fg'$ entails that of $f(g+g')$ and its value as the sum of their values: In view of the set identity $F(G+G') \subset FG + FG'$, the values of $f(g+g')$ are contained in those of $fg + fg'$; the latter's integrability to the sum $\int fg + \int fg'$ follows because the extrema of a sum are straddled by the sum of the extrema.

Finally, there is the possibility of evaluating certain refinement integrals as classical Lebesgue or Stieltjes integrals of related integrands. As seen above, the refinement integral of $g$ will exist and equal that of the product$f\mu$ with single-valued $\mu$ if values of $g/\mu$ are straddled by those of $f$ on subsets of sufficiently fine partitions. When $\mu$ is also additive and $f$ derives from a point-function as the image function on subsets, this "Stieltjes-type integral" coincides with the Lebesgue integral when the set-system is that of the measurable sets (due to E.H. Moore)—indeed, a bounded measurable function on a set of finite measure refinement integrates to its Lebesgue integral, since there are partitions on whose sets the oscillation of $f$ is arbitrarily small.

Since the indefinite refinement integral is finitely additive, it can be evaluated as a classical Lebesgue-Stieltjes integral of a given $\sigma$-additive measure $\mu$ for a suitable point integrand, just when it is $\mu$ absolutely continuous. If $\lambda$ is additive and bounded by $\mu$, then $\frac{d\lambda}{d\mu}$ is a bounded measurable function whose values on measurable $E$ straddle $\frac{\lambda(E)}{\mu(E)}$. This is preserved by postcomposition with piecewise monotone (increasing or decreasing) $m$'s—which would serve to reduce also refinement integrals of $m\left[\frac{\lambda(E)}{\mu(E)}\right] \cdot \mu$ (even for BV $m$ by linearity) to ordinary integrals $\int m\left[\frac{d\lambda}{d\mu}\right] d\mu$.

One encounters this form with $m = -\ln$ in Information Theory; the form $m[\frac{\lambda}{\mu}]$ is, even for any convex $m$, subadditive so this refinement integral is also the sup over the approximating sums. Indeed, for any convex function $f$, $vf(\frac{u}{v})$ is a convex function of $u$ and $v > 0$ [V, p. 260]:

$$f\left(\frac{\alpha u + \cdots}{\alpha v + \cdots}\right) = f\left(\frac{av}{\alpha v + \cdots} \cdot \frac{u}{v} + \cdots\right) \leq \frac{\alpha v}{\alpha v + \cdots} f\left(\frac{u}{v}\right) + \cdots;$$

a quotient of additive functions $\lambda/\mu$ sends disjoint unions to convex combinations: $\frac{\lambda+\cdots}{\mu+\cdots} = \frac{\mu}{\mu+\cdots} \cdot \frac{\lambda}{\mu} + \cdots$, and a convex function of these is dominated by the convex combination of its values, hence by the sum of its values—thus a convex function of $\lambda/\mu$ is subadditive.

The basic quantity in Information Theory is the "average self-information" or "uncertainty" of a finite probability distribution

$$H\{p_i\} := \Sigma p_i \ln \frac{1}{p_i} = -\Sigma p_i \ln p_i.$$

In extending this to infinite distributions, Shannon proposed by analogy, for a distribution with density $p$

$$H\{p\} := \int p(x) \ln \frac{1}{p(x)} dx = -\int p(x) \ln p(x) dx.$$

But this is wrong, e.g., it could be negative, which would not be interpretable; more seriously it is not invariant under change of variable.

The correct definition is as a refinement integral

$$\int P(dX) \ln \frac{1}{P(dX)} = -\int P(dX) \ln P(dX);$$

that is, one divides the interval into finitely many disjoint measurable sets $\{E_i\}$ and approximates with finite sums $-\Sigma_i \int_{E_i} pdx \ln \int_{E_i} pdx$. Since $\ln\left(\frac{1}{v}\right)$ is convex, the form is subadditive, whence the integral exists as the sup. It was calculated in [F] and gives $\infty$ except (possibly) for a discrete distribution, for which it refinement integrates to

$$\Sigma p_i \ln \frac{1}{p_i}.$$

Of fundamental significance in communication theory is the average decrease in uncertainty in one marginal (of a joint) distribution due to knowledge of the other. The quantity to be averaged (over the joint distribution) is the difference of the negative ln of the "a priori" (marginal) distribution $\{p_i\}$ and that of the "a posteriori" (conditional—on the other marginal) distribution $p_{i|j}$ :

$$\Sigma p_{ij} \ln \frac{p_{i|j}}{p_i} = \Sigma p_{ij} \ln \frac{p_{ij}}{p_i p_j}.$$

This is seen to be symmetric in the marginals and so is called their "mutual information." It is subadditive under refinement (considering $p_i p_j$ as a product distribution): hence for infinite distributions its refinement integral is the sup over the finite partitions of this expression—also a standard integral of the expression with the quotient of the joint over the product distribution replaced by their derivative [GY, Theorem 1.1].

Renyi's "information gain of order $\alpha \neq 1$" for discrete distributions [R, p. 587] is (except for a constant factor)

$$I_\alpha(Q\|P) = \ln \Sigma \frac{q_k^\alpha}{p_k^{\alpha-1}} = \ln \Sigma \left(\frac{q}{p}\right)^\alpha p.$$

On refinement, this goes over ($x^\alpha$ is convex for $\alpha > 1$, concave for $\alpha < 1$) to

$$\ln \int \left(\frac{dQ}{dP}\right)^\alpha dP.$$

In particular, with $P$ and $Q$ distributions a.c. with respect to Lebesgue measure, the above yields [R, Theorem 2, p. 595].

The paradigm in Statistics is a family $F(x;\theta)$ of cumulative distribution functions indexed by an interval of $\theta$'s; on the basis of an "observed" $n$-tuple one forms an "estimate" $\tilde{\theta}(x_1, \ldots, x_n)$. [W] proposes to differentiate the identity $\int_{-\infty}^{\infty} dF(x;\theta) = 1$ and finds

$$\frac{d}{d\theta} \int_{-\infty}^{\infty} dF(x;\theta) = \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial\theta} \log dF(x;\theta)\right] dF(x;\theta) = 0.$$

The bracket is obviously trying to be

$$[\qquad] = \frac{\frac{\partial}{\partial\theta} dF(x;\theta)}{dF(x;\theta)};$$

and on the next page he explains the bracket as

$$[\qquad] = \lim_{y\uparrow x} \frac{\frac{\partial}{\partial\theta}[F(x;\theta) - F(y;\theta)]}{F(x;\theta) - F(y;\theta)} = \lim_{y\uparrow x} \frac{\frac{\partial}{\partial\theta} P((y,x] \mid \theta)}{P((y,x] \mid \theta)}.$$

However, independent of the existence of this limit or of its expectation, the refinement integral

$$\int d\frac{\partial F(x;\theta)}{\partial\theta} = \int \frac{\partial F}{\partial\theta}(dx;\theta)$$

exists for differentiable $F$ over every bounded interval since $\frac{\partial F}{\partial \theta}$ is, like $F$, additive; only the existence of the limit of $\frac{\partial F}{\partial \theta}$ at $\pm\infty$ is needed for the existence of the integral, and its vanishing to justify the passage of differentiation under $\int_{-\infty}^{\infty}$.

Similarly, his

$$H(\theta, \theta') := \int_{-\infty}^{\infty} [\log dF(x; \theta')] dF(x; \theta)$$

could be interpreted as a refinement integral

$$\int [\log P(dx \mid \theta')] P(dx \mid \theta).$$

An "unbiased estimator" $\tilde{\theta}$ is one for which $\mathcal{E}(\tilde{\theta} \mid \theta) = \theta$, i.e. $\int (\tilde{\theta} - \theta) dF(x_1, \ldots, x_n; \theta) = 0$. Since $\tilde{\theta}$ does not involve $\theta$, differentiation yields $1 = \int (\tilde{\theta} - \theta) \frac{\partial}{\partial \theta} dF = \int (\tilde{\theta} - \theta) d\frac{\partial F}{\partial \theta}$. Let's require $d\frac{\partial F}{\partial \theta}$ to have an $L_2$ quotient with respect to $dF$, i.e. to have $\frac{\partial}{\partial \theta} \log dF$, the "derivative" $d\frac{\partial F}{\partial \theta}/dF, \in L_2(dF)$; then by Schwartz, $1 \le \sigma^2\tilde{\theta} \cdot \mathcal{E}\left(\frac{d\frac{\partial F}{\partial \theta}}{dF}\right)^2$ and we obtain a lower bound for the variance of the estimator

$$\sigma^2\tilde{\theta} \ge \frac{1}{\mathcal{E}\left[\left(\frac{d\frac{\partial F}{\partial \theta}}{dF}\right)^2\right]}.$$

That this "derivative" may be obtained pointwise from left differentiation with respect to $F$ to justify $d\frac{\partial F}{\partial \theta} = \left[\frac{\partial}{\partial \theta} dF/dF\right] dF$, follows because an everywhere left differentiable function is $(dF)$ a.e. differentiable [S, p. 236].

It is worth noting that in the usual settings the two derivatives actually coincide. For discrete distributions $F$ has jumps $p_i$ and $\frac{\partial F}{\partial \theta}$ (if it exists) will have jumps $\frac{dp_i}{d\theta}$ at the same place; in the (absolutely) continuous case, $dF = p(x; \theta)dx$ and, with passage of $\frac{\partial}{\partial \theta}$ into $\int p(x, \theta) dx$ permitted, one has $d\frac{\partial F}{\partial \theta}/dF = \frac{1}{p(x;\theta)} \frac{\partial p(x;\theta)}{\partial \theta}$ by the Chain Rule, both pointwise and for the measures.

An integral $\int \left(\frac{df}{dF}\right)^2 dF$ is the value of a refinement (called a "Hellinger") integral $\int \frac{f^2(dI)}{F(dI)}$, since the value of $\frac{(\Delta f)^2}{(\Delta F)^2}$ ($F$ nondecreasing) at intervals $I$ not including 0 is straddled by values of $\left(\frac{df}{dF}\right)^2$ at points of $I$. Since $x^2$ is convex, this refinement integral is a sup, which is the form conceived by Hellinger in his 1907 Göttingen dissertation.

**References**

[Arb] I.M. Gelfand *et al.*, *Arbeiten zur informationstheorie. II.* Deutscher Verlag der Wissenschaften, Berlin, 1961.

[Cr] H. Cramér, *Mathematical Methods of Statistics*, Princeton Mathematical Series, vol. 9, Princeton University Press, Princeton, 1946.

[F] I. Fleischer, *The central concepts of communication theory for infinite alphabets*, J. Math. Phys. **37** (1958), pp. 223–228.

[Fl] I. Fleischer, *Ein Integrallimesaustauschsatz* Math. Annalen **148** (1962) pp. 367-371.

[G] R.G. Gallager, *Information theory and reliable communication*, Wiley, New York, 1968.

[GKY] I.M. Gelfand, A.N. Kolmogorov and A.M. Yaglom, *On the general definition of the amount of information* (Russian), Doklady Akad. Nauk, SSSR **111** (1956), pp. 745–748. German translation in [Arb].

[GY] I.M. Gelfand and A.M. Yaglom, *Calculation of the amount of information about a random function contained in another such function*, Amer. Math. Soc. Transl. (2) **12** (1959), 199–246.

[H] T.H. Hildebrandt, *Introduction to the Theory of Integration*, Academic Press, New York, 1963.

[Ho] E.W. Hobson, *The theory of functions of a real variable and the theory of Fourier's series*, Vol. I, Dover Publications, Inc., New York, 1958.

[K] A.N. Kolmogoroff, *Untersuchungen über den Integralbegriff*, Math. Ann. **103** (1930), 654–696.

[P] M.S. Pinsker, *Information and information stability of random variables and processes*, Translated and edited by Amiel Feinstein, Holden-Day, Inc., San Francisco, 1964.

[R] A. Rényi, *Probability theory*, North-Holland, 1970.

[S] S. Saks, *Theory of the Integral*, Dover, New York, 1964.

[V] I. Vajda, *Theory of statistical inference and information*, Kluwer Academic Publishers, Dordrecht, 1989.

[W] S.S. Wilks, *Mathematical Statistics*, Princeton University Press, Princeton, N.J., 1943.