Dmytro Taranovsky
December 18, 2001; Last Modified: September 18, 2003; Formatting Changes: February 17, 2004

Measure and Integration

Measure

Definition 1.1: A σ-ring $\Re$ a collection of sets such that

- Union of a countable collection of sets from $\Re$ is in $\Re$, and
- If $A\in\Re$ and $B\in\Re$, then $A\text{-}B=\{x|x\in A \text{ and } x\notin B\}\in\Re$.

Corollary 1.1.1: Intersection of a countable collection of sets from $\Re$ belongs to $\Re$.
Interpretation 1.1.2: Elementary operations on sets from $\Re$ only lead to sets in $\Re$.
Definition 1.2: A function *f* (defined on sets) is called σ-additive if $f\left(\bigcup_{i\in I} A_i\right)=\sum_{i\in I} f(A_i)$ for every countable collection of pairwise-disjoint sets $\{A_i\}$ when the right side is defined.
Interpretation 1.2.1: Partition X into a countable collection $\{X_i\}$. $f(X)$ is the sum of $f(X_i)$.
Definition 1.3: Measure on a set X is a σ-additive function μ that assigns a nonnegative real number or ∞ for certain subsets of X, called measurable sets, such that the collection of measurable sets is a σ-ring.
Corollary 1.3.1: If A and B are measurable and $A\subset B$, then $\mu(A)\leq\mu(B)$.
Corollary 1.3.2: If A and B are measurable and $\mu(A\cap B)<\infty$, $\mu(A\cup B)=\mu(A)+\mu(B)-\mu(A\cap B)$.
Note 1.4: Unless stated otherwise, all subsets of all sets of measure zero are measurable.

Number of items, length, area, volume, and mass are all examples of measures. Measure is a unifying concept that allows a single mathematical framework to cover many areas. σ-additivity, non-negativity, and its existence for all sets on a σ-ring allows computation of measures for many sets from measures for a small set of basic sets.

Definition 1.5: Measure is unique means that it exists and is uniquely determined on the σ-ring generated (by including countable unions and differences into the collection; and repeating the inclusion recursively) from the initially given measurable sets.
Theorem 1.6: Application of note 1.4 to a given measure gives a unique measure.
Definition/Theorem 1.7: There is a unique measure on $\mathbb{R}$, called *standard measure* (length), that satisfies $\mu(\{x|a<x<b\})=b\text{-}a$ when $b>a$.
Definition 1.8: Let X and Y be sets with measure. Unless stated otherwise, measure on X×Yis such that $\mu(A\times B)=\mu(A)*\mu(B)$ if A and B are measurable, $A\subset X$, and $B\subset Y$.
Note 1.8.1: This definition is a generalization of the formula for area of rectangle.
Theorem 1.9: If measures on X and Y are given, then the measure on $X\times Y$ exists. It is unique if every measurable set on X and Y (X and Y do not have to be measurable) is a union of a countable collection of sets of finite measure.

Example 1.10: Since $\mathbb{R}^n$ can be viewed as $\overbrace{\mathbb{R}\times\mathbb{R}\times\ldots\times\mathbb{R}}^{n}$, its standard measure is defined by $\mu(\{\boldsymbol{x}|a_i<x_i<b_i \text{ for } i=1, 2,...,n\}) = \prod_{i=1}^{n}(b_i-a_i)$ if $b_i>a_i$ for $i=1,2,...,n$.
By theorems **1.7** and **1.9**, the standard measure on $\mathbb{R}^n$ is unique.
Note 1.11: The standard measure, also called Lebesgue measure, on $\mathbb{R}^2$ is area and on $\mathbb{R}^3$ is volume.
Theorem 1.12: $X\subset\mathbb{R}^n$ is measurable (under the standard measure) if and only if for every $\epsilon>0$, there is a closed B and an open A such that $B\subset X\subset A$ and $\mu(A)\text{-}\mu(B)<\epsilon$.
Note 1.13: In practice, almost all subsets of a measurable set are measurable. All open and closed sets on $\mathbb{R}^n$ are measurable. It is impossible to define a non-measurable set on $\mathbb{R}^n$. However, non-

measurable sets on ℝ (and thus $\mathbb{R}^n$) do exist. Unfortunately, making all sets measurable would be contradictory; instead, measure is defined on a σ-ring so that elementary operations on measurable sets always lead only to measurable sets.
Caution: $A \subset B$ and $\mu(A)=\infty$ does not mean that B is measurable.

Integration

Definition 2.1: Let *f* be a function on X that returns nonnegative real numbers or ∞.
Then, integral of *f* over X, $\int_X f\,d\mu \equiv \mu\{(x,y) | x \in X$, f(x) exists, and $0<y<f(x)\}$.
Interpretation 2.1.1: The integral is the measure of the region in X between 0 and f(x) (a subset of the X×ℝ). If μ is standard, and X⊂ℝ, the measure is area, if $X \subset \mathbb{R}^2$, the measure is volume.
Note 2.1.2: Measure is often computed as an integral and vice versa.
Note 2.1.3: In the integral, *f*(*x*) can be written instead of *f*.
Corollary 2.1.4: $\int_A d\mu = \mu(X)$.
Corollary 2.1.5: If $F(A)=\int_A f d\mu$ for all measurable $A \subset X$ and $f(x) \geq 0$, then *F* is a measure.
Definition 2.2: Let *f* be a function on X that returns real numbers, ∞, or -∞. Then,
$\int_X f\,d\mu = \int_X f^+\,d\mu - \int_X f^-\,d\mu$ where $f^+(x)=\sup\{f(x), 0\}$ and $f^-(x)=\sup\{-f(x), 0\}$.
Interpretation 2.2.1: Measure of the region above *f* but below *0* (where $f(x)<0$) is subtracted from measure of the region above *0* but below *f* (where $f(x)>0$).
Theorem 2.2.2: If defined, the integral is unique in value. The integral is defined if and only if $\{x \mid f(x)>y$ and $f(x)\neq 0\}$ is measurable for every real number *y* and not $\int_X f^+ d\mu = \infty = \int_X f^- d\mu$.
(In that case, the difference of the two integrals is ∞-∞, which is undefined).
Definition 2.3: Let *f* be a vector-valued function and *f*(*s*) be its component. Then,
$$\left(\int_X f\,d\mu\right)(s) = \int_X f(s)\,d\mu$$ (Vectors are integrated component-wise.)
Note 2.3.1: The set of all complex numbers (ℂ) is a vector space over ℝ, and every vector space over ℂ is a vector space over ℝ (every component can be viewed as a real and an imaginary component). Thus, integrals are defined for complex-valued functions.
Definition 2.4: (convenient notation) **(a)** $\int_a^b f dm = \int_X f dm$ where $X=\{x | a<x<b\}$ if a≤b; otherwise, $\int_a^b f dm = -\int_b^a f dm$ **(b)** $\int_a^b f(x)\,dx = \int_a^b f dm$ where *m* is the standard measure.
(Pronunciation: "integral from *a* to *b* of ...".)
Definition 2.5: If μ(X) is nonzero and finite, the average value of *f*(*x*) over X is $\int_X f d\mu \,/\, \mu(X)$.
Note 2.5.1: Thus, **2.5** gives another interpretation of the integral.
Definition 2.6: Almost everywhere means everywhere except possibly on a set of measure zero.
Theorem 2.7: (Approximating Measures and Integrals) Let *X* and *A* be measurable, $A \subset X$, *f* be defined almost everywhere in *X*, $\int_X f d\mu$ be defined, and range of *f* be in ℝ. Partition *X* into a countable collection of pairwise-disjoint measurable sets $\{X_i | i \in I\}$. Then,
(a) $\sum_{X_i \subset A} \mu(X_i) \leq \mu(A) \leq \sum_{i \in J} \mu(X_i)$ if $A \subset \bigcup_{i \in J} X_i$ and $J \subset I$
(b) $\sum_{i \in I} \inf_{x \in X_i} f(x)\mu(X_i) \leq \int_X f\,d\mu \leq \sum_{i \in I} \sup_{x \in X_i} f(x)\mu(X_i)$
Note 2.7.1: Make a drawing to easily interpret the theorem. Proof of **(a)** follows from the σ-additivity and non-negativity of measure. **(b)** follows from **(a)** if *f* is nonnegative. The proof is completed by applying **2.2**.
Theorem 2.8: (Fundamental Theorem of Calculus)

(a) $\int_a^b \frac{dF}{dx}dx = F(b) - F(a)$ if the integral and *dF/dx* are defined.

(b) If $F(t) = \int_a^t f(x)\,dx$, then *f(t) = dF/dt* almost everywhere.

Note 2.8.1: **2.8(a)** states that the total change equals to the integral of the rate of change, which equals (by **2.5**) to the average rate of change times time.

Functions *f* are typically integrated by finding *F* such that *F'=f.*

Theorem 2.9: Let *T* be a function on *X* such that $m(A)=\mu(T^{-1}(A))$ for all measurable $A \subset T(X)$. Then, $\int_X f\,d\mu = \int_{T(X)} fdm$.

Note 2.9.1: The proof follows from equivalence of *X* and *T*(*X*) in terms of measurement.

Theorem 2.10: Integral of all *f* is σ-additive (of the region) unless one component is ∞-∞.

Theorem 2.11: Integral is a linear operator on the functions integrated.

Theorem 2.12: (Dominated Convergence Theorem) If, almost everywhere, $\lim_{n\to\infty} f_n(x) = f(x)$ and $\int_X \sup_{n\in\mathbb{N}} \|f_n(x)\|\,d\mu$ is finite, then $\lim_{n\to\infty}\int_X f_n\,d\mu = \int_X f\,d\mu$.

Corollary 2.13: Integral is a "continuous" operator. When the condition in **2.12** is met for the partial sums (f_n), integral of a sum of a countable number of functions is the sum of the integrals of the functions. The finiteness in **2.12** is true if and only if *g*(*x*) exists such that $\int_X g\,d\mu$ is finite and for all *n*, $\|f_n(x)\| \le g(x)$.

Note 2.13.1: **2.10**, **2.11**, and **2.12** are useful for computing integrals. Because of **2.13**, functions can usually be represented by sums (such as a power series) and integrated term by term.

Theorem 2.14: Let *f(x, y)* be a function defined almost everywhere on *X*×*Y*, m_X be a measure on *X*, and m_Y be a measure on *Y*. Assume that $m_{X\times Y}$, the corresponding measure on *X*×*Y*, is unique. Then, if defined, $\int_{X\times Y} fdm_{X\times Y} = \int_Y \int_X fdm_X dm_Y = \int_X \int_Y fdm_Y dm_X$.

Note: $\int_X fdm_x$ is a function of *y* with *f(x) = f(x, y)* at each $y\in Y$.

Note 2.14.1: Integrals over $\mathbb{R}^n$ are typically computed by applying theorem **2.14** *n-1* times (with *Y* = ℝ) and then evaluating the integrals (over ℝ) using the fundamental theorem of calculus.

Theorem 2.15: Let *X* be a union of a countable collection of sets of finite measure. Let distance between $A\subset X$ and $B\subset X$ be $m((A-B)\cup(B-A))$ where *m* is a measure on *X*.

Let *F* be σ-additive, and *F(A)* be finite if *m(A)* is finite.

F is defined to be continuous if as distance→*0*, (change in *F*)→*0*.

(a) *F* is continuous if and only if for some *f*, $F(A)=\int_A fdm$ for all measurable $A\subset X$.

(b) There are unique σ-additive *G* and *H* such that *F=G+H* where *G* is continuous and *H* is singular, that is there is *E* such that *m(E)=0* and $H(A)=H(E\cap A)$ for all measurable $A\subset X$.

Definition 2.16: If *m(X)>0*, *f=dF/dm* means $F(A)=\int_A fdm$ for all measurable $A\subset X$.

dF/dm is called density of *F*, or derivative of *F*.

Theorem 2.17: **(a)** If *f=dF/dm* and *g=dF/dm*, then *f(x)=g(x)* almost everywhere.

(b) If *f=dF/dm* and *f=dG/dm*, then F=G.

Theorem 2.18: Assume that *F* is nonnegative and *dG/dF* and *dF/dm* are defined (by **2.16**).

(a) $\int_A g\frac{dF}{dm}dm = \int_A g\,dF$ **(b)** $\frac{dG}{dF}\frac{dF}{dm} = \frac{dG}{dm}$ almost everywhere.

Note 2.18.1: **2.18(a)** followed by **2.9** is frequently used to compute integrals by simplifying the function being integrated (and then changing to the standard measure).

Definition 2.19: On $\mathbb{R}^n$, $\dfrac{dF}{dm}(\boldsymbol{x}) \equiv \lim_{r\to 0^+} \dfrac{F(O_r(\boldsymbol{x}))}{m(O_r(\boldsymbol{x}))}$ where $O_r(\boldsymbol{x})=\{\boldsymbol{y} \mid \|\boldsymbol{y}-\boldsymbol{x}\|<r\}$ and m is standard.

Theorem 2.20: If $g=dF/dm$ by **2.16**, and $f=dF/dm$ by **2.19**, then $f(x)=g(x)$ almost everywhere.

Interpretation 2.20.1: Thus, **2.16** and **2.19** define the same quantity except that **2.19** is limited to $\mathbb{R}^n$ but is more general and specific on $\mathbb{R}^n$.

Example 2.21: Mass (M) is the integral of mass density ρ (ρ≥0) over volume (V). Thus, mass equals average density times volume. $dM/dV=\rho$. Center of mass is $\int_X \boldsymbol{r}dM / M(X) = \int_{X-E} \boldsymbol{r}dM / M(X) + \int_E \boldsymbol{r}dM / M(X) = \int_{X-E} \boldsymbol{r}\rho dV / \int_X \rho dV + \int_E \boldsymbol{r}dM / M(X)$ where $\boldsymbol{r}$ is location and mass is continuous (satisfied if density is finite) everywhere on X-E.

The fundamentals of integration are now completely presented. However, practice, some techniques, and differentiation tables are required to effectively evaluate integrals.

Measurable Functions are Vectors

Definition 3.1:

1. Inner product of complex-valued functions f and g over X, $\langle f,g\rangle \equiv \int_X fg^c\, d\mu$ where $g^c(x)$ is the complex conjugate of $g(x)$. **Note:** f and g must be defined almost everywhere.
2. f and g are called equivalent ($f \sim g$) if $f(x)=g(x)$ almost everywhere.
3. $\mathscr{L}^2(\mu)$ is the set of all f (or equivalence classes of f) such that $<f, f>$ is finite.

Theorem 3.2: $f\sim g \Leftrightarrow \left(\int_A f d\mu = \int_A f d\mu \text{ for all measurable } A\subset X\right) \Leftrightarrow \int_X |f(x)-g(x)|^2 d\mu = 0$

Note 3.3: Below in this section, by a function, we may mean an equivalence class of functions: We do not distinguish between f and g when $f \sim g$ as integrals do not distinguish between equivalent functions.

Theorem 3.4: $\mathscr{L}^2(\mu)$ is an inner-product vector space (over $\mathbb{C}$).

Note 3.4.1: Because of the theorem, theorems for inner-product vector spaces apply to $\mathscr{L}^2(\mu)$ and greatly simplify analysis of functions. For example, $\mathscr{L}^2(\mu)$ is a complete metric space, with distance $d(f,g)=\|f-g\|=\int_x |f-g|^2 d\mu$ where norm (or magnitude) of f, $\|f\|=\sqrt{\langle f,f\rangle}$.

Definition 3.5: $(\mathscr{L}^2)^n$ is the set of all f (or equivalence classes of f) whose domain is $\mathbb{R}^n$ and range is $\mathbb{C}$ such that $\|f\|$ (under the standard measure) is finite. $\mathscr{L}^2 \stackrel{\text{def}}{=} (\mathscr{L}^2)^1$.

Theorem 3.6: $\{f \mid f(\boldsymbol{x})=e^{i\boldsymbol{N}\cdot\boldsymbol{x}}/(2\pi)^{n/2}$ where $\mathbf{N}$ is an ordered n-tuple of integers$\}$ is an orthonormal basis of $(\mathscr{L}^2)^n$ if we restrict functions to $-\pi \le x_j \le \pi$ for $j=1,2,...,n$.

Corollary 3.6.1: $\{f \mid f(x)=e^{inx}/\sqrt{2\pi} \text{ where } n \text{ is an integer}\}$ is an orthonormal basis of $\mathscr{L}^2$ if we restrict functions to $-\pi \le x \le \pi$.

Corollary 3.6.2: $\{f \mid f(x)=\dfrac{e^{i\boldsymbol{N}\cdot\boldsymbol{x}}}{(2\pi)^{n/2}}$ if $2\pi m_j \le x_j \le 2\pi(m_j+1)$ for $j=1, 2,..., n$ and $f(\boldsymbol{x})$ is zero elsewhere, where $\boldsymbol{N}$ is an ordered n-tuple of integers and all m_j are integers$\}$ is an orthonormal basis of $(\mathscr{L}^2)^n$.

Corollary 3.6.3: $(\mathscr{L}^2)^n$ is equivalent to the infinite-dimensional complex Euclidean space.

Note 3.7: Linear partial differential equations are typically solved by representing the initial conditions (f) in terms of a basis (often the one in theorem 3.6; the basis (usually orthogonal) is chosen to make the equation easy to solve): $f=\sum_{i\in I} c_i f_i$; solving the equation for each vector in the basis: $\boldsymbol{f_i(t)}$; and adding the solutions $f(t)=\sum_{i\in I} c_i f_i(t)$.

More Information on Measure

Example 4.1: Probability has all of the properties of measure. Consider a set of possible events such that at most one event can happen. A measure on such set is the probability that one of the events will happen. The set of all possible events has measure 1. For example, if you toss a fair coin, the set of events is {head, tail}. Both of these events have measure and probability of ½.

Procedure 4.2 Constructing measures: **(a)** Define a nonnegative real or infinite $m(X)$.
(b) Partition X into a countable collection of pair-wise disjoint sets $\{X_i \mid i\in I\}$. Define a nonnegative real or infinite $m(X_i)$ for all $i\in I$ such that $\sum_{i\in I} m(X_i)=m(X)$ and $m(\emptyset)=0$. Repeat step **(b)** (recursively) for each X_i in place of X.
Theorem 4.3: Assume that **4.2** is performed. The measure m is unique on X.
Note 4.4: Procedure 4.2 should be set up to make the measures, except for one point sets, approach zero as **4.2(b)** is repeated many times. Otherwise, some of the sets that we would like to consider measurable will not be measurable.
Theorem 4.5: For every real and monotonically increasing function f, there is a unique measure on $\mathbb{R}$, that satisfies $\mu(\{x|a<x<b\})=f(b^-)-f(a^+)$ if $b>a$.
(Advanced) Theorem 4.6: On a Riemann n-dimensional manifold, there is a unique measure μ such that $\lim_{c\to 0^+}\frac{1}{c^n}\mu(\{\boldsymbol{x}|a_i<x_i<a_i+c \text{ where i=1, 2, ..., n}\})=\sqrt{\text{determinant of the metric}}$.
The two theorems and thus **1.7** and **1.10** can be proved by constructing the measures with **4.2**.

Sometimes, it is essential to define lower dimension measures, such as length of curves.

Definition 4.7: Outer r-dimensional measure ($r\geq 0$) on a metric space (such as $\mathbb{R}^n$),

$$m_r^*(E)\equiv\frac{(\pi/4)^{r/2}}{\int_0^\infty x^{r/2}\mathrm{e}^{-x}dx}*\sup_{d>0}\inf\left\{\sum_{i\in\mathbb{N}}(\operatorname{diam}(A_i))^r \middle| \text{for all } i\in\mathbb{N}, \operatorname{diam}(A_i)<d, \text{and} \bigcup_{i\in\mathbb{N}} A_i\supset E\right\}.$$

Interpretation 4.7.1: Outer measure is computed by covering the set with open balls in the most efficient way and then counting balls with adjustment for the size of each ball. For accuracy, the size of the balls must be small, hence the restriction on the diameter. The multiplier on the left normalizes the measure. Some practice will make the definition intuitive.
Definition 4.8: r-dimensional measure m_r on X (such as $\mathbb{R}^n$) is such that
$m_r(E)=m_r^*(E)$ if for all $A\in X$, $m_r^*(A)=m_r^*(A-E)+m_r^*(A\cap E)$.
Theorem 4.9: $\boldsymbol{m_r}$ **is unique.** All open and closed subsets of $\mathbb{R}^n$ are measurable with m_r. All subsets of sets of m_r measure zero are measurable, but if (on $\mathbb{R}^n$) $s>r>0$, not all sets of m_s measure 0 are measurable with m_r. Every definable subset of $\mathbb{R}^n$ is measurable for every m_r.
Theorem 4.10: On $\mathbb{R}^n$, m_n is the standard measure.
Theorem 4.11: Let T ($\mathbb{R}^m\to\mathbb{R}^n$) be such that $a\|x-y\| \leq \|T(x)-T(y)\| \leq b\|x-y\|$ for all x and y in $\mathbb{R}^m$, and that $T(S)$ is measurable if S is measurable. Assume that $a\geq 0$ and $a>0$ if $r=0$.
Then, for all measurable $S\subset\mathbb{R}^m$, $a^r m_r(S)\leq m_r(T(S))\leq b^r m_r(S)$.
Warning: The author is not certain that **4.11** is a theorem.
Corollary 4.11.1: On $\mathbb{R}^n$, expanding a set in k times increases its m_r in k^r times. A congruence

transformation does not change the measure.

Theorem 4.12: Let T ($\mathbb{R}^m \to \mathbb{R}^n$, m≤n) be a differentiable one-to-one map defined on X⊂$\mathbb{R}^m$. Let T' be the derivative of T. Let A⊂X and let $g(T(x))=f(x)$.
Then, $\int_{T(A)} g\,dm_m = \int_A |det(T')| f\,dm_m$ if the right side is defined.
[see Appendix A for definition of det]
Corollary 4.12.1: If T ($\mathbb{R}^m \to \mathbb{R}^n$) is linear, $m_m(T(S)) = |det\ T|\ m_m(S)$ for all measurable S⊂$\mathbb{R}^m$.
Note 4.12.2: Theorem 4.12 allows easy computation of lengths of curves, areas of surfaces, integrals over curves and surfaces, and more. To do so, an appropriate T is found, and the problem is transferred to $\mathbb{R}^m$ with the standard measure.

Definition 4.13: Dimension of $X \subset \mathbb{R}^n$, dim $X \equiv \inf\{r \mid m_r(X)=0\}$ if for all $s \geq 0$ $m_s(X)$ is defined.
Theorem 4.14: If dimension of X is r, then $m_s(X)=0$ if $r<s$ and $m_s(X)=\infty$ if $s<r$.
Note 4.15: Thus, sets of lower dimension are always smaller than sets of higher dimension. However, it is possible for a set of dimension 0 to have an uncountable number of points in every interval. Moreover, for any $r>0$, dim X = r does not mean $m_r(X)>0$ even if X is closed.
Note 4.16: dim($A \times B$)≥dim(A)+dim(B) if A and B are nonempty. Unfortunately, dim($A \times B$) may be larger than dim(A)+dim(B).
A point has dimension zero. By **4.7** and **4.8**, $m_0(X)$ is the number of points in X.

Although **4.7** and **4.8** may appear strange, $\boldsymbol{m_r}$ has the properties we expect for a measure on r-dimensional sets. Sets with non-integer dimensions (and some other sets) are called fractals. A simple fractal is the Cantor set: $\{x \mid 0 \leq x \leq 1$ and base 3 representation of x is either finite or has no ones$\}$. Fractals are often beautiful and have many applications. For example, a coastline or a surface of a mountain is best represented by a fractal. We can now find dimension and measure of fractals.

Appendix A

Notation and Sets: $a \stackrel{\text{def}}{=} b$ means that a is defined to equal b. A set is any collection of elements; however, all of the elements must be defined before the set can be created. For example, $\{A, B\}$ contains A and B. $s \in S$ means that s is an element (also called member) of set S; s belongs to S, and s is in S are synonymous with $s \in S$. $S=T$ means that for all s, $s \in S \leftrightarrow s \in T$ -- the order of the elements and the number of instances of an element are irrelevant. $\varnothing$ is the empty set--it has no elements. $\{x \mid P(x)\}$ is the set of all x such that $P(x)$ holds. A is a subset of B, $A \subset B$, means that all elements of A are elements of B; $B \supset A \leftrightarrow A \subset B$. $A \subset A$. $\cup$ is called union; $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$. $\cap$ is called intersection; $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$. A set is called countable if it is possible to count its elements so that for each element, the element will be counted at some point. (A finite or empty set is countable in this paper.) A collection of sets is called pairwise-disjoint if for every two sets X and Y in the collection, $X \cap Y$ is empty. For sets, $X \times Y \stackrel{\text{def}}{=} \{(x, y) \mid x \in X \text{ and } y \in Y\}$. $\mathbb{R}$ is the set of all real numbers; $\mathbb{N}$ is the set of all positive integers. $\inf\limits_{P(i)} X_i$ is the largest value that is smaller or equal than all X_i for which $P(i)$ holds. Similarly, $\sup\limits_{P(i)} X_i$ is the smallest value that is larger or equal than all X_i for which $P(i)$ holds. $\sum\limits_{P(i)} f(i)$ is the sum of all $f(i)$ for which $P(i)$ holds. If the set of valid i is infinite, take the limit of the partial sums as the number of elements chosen approach infinity. (For the sum to be defined, the limit may not depend on the way of counting if no i is counted twice.) $\sum\limits_{n=1}^{N} f(n)$ is the sum of all $f(n)$ for which n is an integer from 1 to N. Similarly, $\prod\limits_{n=1}^{N} f(n)$ is the product of all $f(n)$ for which n is an integer from 1 to N. $\bigcup\limits_{P(i)} A_i$ is $\{x \mid x \in A_i$ for some i that satisfies $P(i)\}$.

Diameter of X, diam(X) $\sup\{d(x, y) \mid x \in X$ and $y \in X)$ where d(x, y) is distance between x and y.

Operations on functions: $(f+g)(x)=f(x)+g(x)$, $(cf)(x)=cf(x)$, $(f-g)(x)=f(x)-g(x)$; usually, $(f*g)(x)=f(x)*g(x)$.

Unless defined otherwise, $f(A)=\{y \mid y=f(a) \text{ and } a \in A\}$ when A is a set.

Unless defined otherwise, $(-1)^{1/2}=i$ and $(a+bi)^c=a-bi$ if a and b are real numbers.

Operations on Infinities: Let $r \in \mathbb{R}$; s, t, and u be real numbers, ∞ (infinity), or $-\infty$(minus infinity). Then, $s-t=s+-t$, $s+t=t+s$, $-\infty+-\infty=-\infty$, $-\infty+r=-\infty$, $r+\infty=\infty$, $\infty+\infty=\infty$, but $-\infty+\infty$ is undefined. $-s=-1*s$, $st=ts$, $s(tu)=(st)u$, $\infty*\infty=\infty$; $r>0 \Rightarrow r*\infty=\infty$. $-\infty<r<\infty$. $s/r=(1/r)*s$, $r/\infty=r/(-\infty)=0$; but $s/0$, ∞/∞, $(-\infty)/\infty$, and $\infty/(-\infty)$ are undefined.

Review of vector spaces: A vector space V over a field F (F is usually $\mathbb{R}$ or $\mathbb{C}$) is (or is isomorphic to) a set of functions (vectors) that for every element from a set S return an element of F such that

- If $f \in V$, $g \in V$, and $s \in S$, then $(f+g)(s)=f(s)+g(s)$ and $(f+g) \in V$
- If $f \in V$, $c \in F$, and $s \in S$, then $(cf)(s)=c*f(s)$ and $(cf) \in V$

L is a linear operator means $L(x+cy)=L(x)+cL(y)$ (c is a scalar) when the right side is defined.

If V is $\mathbb{C}$ or $\mathbb{R}$, a vector space over V is called an inner product vector space if an inner product ($\langle f, g\rangle$) is defined such that: magnitude of f, $\|f\|=(\langle f, f\rangle)^{1/2} \geq 0$ and is zero only if $f=\mathbf{0}$; $\langle f, g\rangle \in V$, $\langle f, g\rangle=\langle g, f\rangle^c$, and $\langle f, g\rangle$ is linear in terms of f.

Note: $f=\mathbf{0}$ means that for all $s \in S$, $f(s)=0$. Note: $f/c=(1/c)*f$.

Theorem A1: (a) Always, $|\langle f, g\rangle| \leq \|f\|*\|g\|$, and $|\langle f, g\rangle|=\|f\|*\|g\|$ if and only if $g=cf$ or $f=\mathbf{0}$.

(b) $\langle \boldsymbol{u}, \boldsymbol{v}\rangle=0$ if and only if $\|\boldsymbol{u}+\boldsymbol{v}\|^2=\|\boldsymbol{u}\|^2+\|\boldsymbol{v}\|^2$.

Anything that is equivalent (except for possible additional operations) to a vector space is a vector space. For example, a plane is a vector space since each point can be represented by a function from coordinate number that returns the coordinate value. $\mathbb{C}$ is a vector space over $\mathbb{R}$ since we can associate 1 with (1, 0) and i with (0, 1).

n-dimensional Euclidean space ($\mathbb{R}^n$) is the set of all ordered *n*-tuples ($x_1, x_2, \ldots, x_n$) of real numbers (the numbers are called coordinates); for n=1, it is a line; for n=2, it is a plane; $\mathbb{R}^3$ is a good model for our space. If x and y belong to $\mathbb{R}^n$, then $\langle x, y \rangle \equiv x \cdot y \equiv \sum_{i=1}^{n} x_i y_i$. A complex Euclidean space is the same as a real Euclidean space except that the coordinates are complex numbers and $<x, y> = x^* y^c$. An infinite dimensional Euclidean space is the same as a finite dimensional Euclidean space except that a point $\boldsymbol{r}$ consists of an infinite sequence of numbers such that $\|\boldsymbol{r}\|$ is finite. All Euclidean spaces are inner-product vector spaces.

A set of vectors $\{\boldsymbol{r}_i \mid i \in I\}$ is called orthogonal (mutually perpendicular) if $<\boldsymbol{r}_i, \boldsymbol{r}_j>=0$ when $i \neq j$ and orthonormal if $\|\boldsymbol{r}_i\|=1$ and $<\boldsymbol{r}_i, \boldsymbol{r}_j>=0$ when $i \neq j$. A set of vectors $\{\boldsymbol{r}_i \mid i \in I\}$ is called linearly independent when $\sum_{i \in I} c_i \boldsymbol{r}_i \neq 0$ unless for all i, $c_i=0$. (All orthonormal sets are linearly independent. In all linearly dependent sets, some vector is a linear combination of the other vectors.) The span of $\{\boldsymbol{r}_i \mid i \in I\}$ is $\{r \mid r = \sum_{i \in I} c_i \boldsymbol{r}_i \text{ and each } c_i \in F\}$, that is the set of all linear combination of $\boldsymbol{r}_i$. A set of vectors is called a basis if it is linearly independent and spans the entire vector space.

Theorem A2: If $\{\boldsymbol{r}_i \mid i \in I\}$ is an orthogonal basis of V, then for all $\boldsymbol{r} \in$ V, $\boldsymbol{r} = \sum_{i \in I} \frac{\langle \boldsymbol{r}, \boldsymbol{r}_i \rangle}{\langle \boldsymbol{r}_i, \boldsymbol{r}_i \rangle}$.

Note: The theorem allows computation of representation of functions in any given orthogonal basis. For an orthonormal basis, **A2** simplifies to $\boldsymbol{r} = \sum_{i \in I} \langle \boldsymbol{r}, \boldsymbol{r}_i \rangle$.

Corollary to **A2** and **A1**: $\|\boldsymbol{r}\|^2 = \sum_{i \in I} |\langle \boldsymbol{r}, \boldsymbol{r}_i \rangle|^2$ if $\{\boldsymbol{r}_i \mid i \in I\}$ is an orthonormal basis.

Let T ($\mathbb{R}^m \to \mathbb{R}^n$, $m \leq n$) be linear. There exists a linear S ($\mathbb{R}^m \to \mathbb{R}^m$) such that always, $\|T(y)-T(x)\| = \|S(y)-S(x)\|$. Magnitude of the determinant of T, |det T| = |det S|. Determinant is a continuous real function. If $T(x)(i)=x(i)$ when i is from 1 to m and $T(x)(i)=0$ for all other valid i, then det $T=1$. If S and T are the same except that $S(x)(j)=cT(x)(j)+dT(x)(i)$ (i, j, c and d are fixed), then det $S=c$*det T. **Note:** $(XY)(s)=X(Y(s))$.

Theorem A3: Every linear transformation ($\mathbb{R}^m \to \mathbb{R}^n$, $m \leq n$) has a unique determinant, which is nonzero if and only if the transformation is one-to-one. det (XY) = det X*det Y

If the domain and range of T are vector spaces, derivative of T at point x is a linear transformation D such that $\lim_{\|\boldsymbol{a}\| \to 0} \frac{\|T(\mathbf{x}+\boldsymbol{a}) - T(\mathbf{x}) - D(\boldsymbol{a})\|}{\|\boldsymbol{a}\|} = 0$.

Informally, for a small $\boldsymbol{a}$, $T(\mathbf{x}+\mathbf{a}) \approx T(\mathbf{x}) + D(\mathbf{a})$.

Derivative of T can be viewed as the function T with $T(x) = D$ at x.

Theorem A4: If defined, the derivative is unique.

Appendix B: Extensions of the Integral

Theorem B1.1: If m is σ-additive, is defined on all sets, including X, of a σ-ring on X, and returns real numbers, ∞, or -∞, then there exists $Y \subset X$ such that m is a measure on Y and $-m$ is a measure on X-Y.

Definition B1.2: Assume the terminology of **B1.1**. When it is desirable to do so,

$\int_X f dm \equiv \int_Y f dm - \int_{X-Y} f dm_1$ where $m_1(A)=-m(A)$.

Theorem B1.3: If defined, the integral in **B1.2** has a unique value.

B1.4: (Reason for the definition **B1.2**) Sometimes it is convenient to integrate over σ-additive quantities that can assume negative values, such as electric charge or heat flow.

Theorem/Definition B1.5: **2.3**, **2.9**, **2.10** , **2.11** and **2.16** apply to **B1.2**. **2.18(a)** applies to **B1.2** even without the assumption of non-negativity. **2.13** essentially applies to **B1.2**.

Definition B2.1: When it is desirable to do so,

$\int_a^b f(x)\,dx = \lim_{c \to a^+} \lim_{d \to b^-} \int_c^d f(x)\,dx$ ($a<b$; a can be -∞ and b can be ∞.)

Integral over a set of measure zero is 0, and and integral over a union of a finite number of pair-wise disjoint intervals is equal to the sum of the integrals over the intervals.

Theorem B2.2: **B2.1** is non-contradictory. The integral is the same as the standard integral when the standard integral is defined. Otherwise, the integral is not σ-additive (even when the integral of the union is defined), and the region can be split into two regions such that a component of the integral over the first region is ∞, and the component of the integral over the second region is -∞. Theorems **2.7**, **2.8**, **2.11**, and **2.12** apply to the integral. **2.9** and **2.18** apply if we restrict the range to $\mathbb{R}$.

Explanation B2.3: Integral of the rate of change over time should equal to the total change even if the integral of the magnitude of the rate of change diverges.
(For example, let $f(x)=\sin x/x$ and the interval from 1 to ∞.)

Uses of this Paper

This paper assumes knowledge of limits and differentiation and a solid pre-calculus background.

The purpose of this paper is to give all basic knowledge on the important topic of measure and integration in a single self-contained compact document. As such, this paper can be used to either review or study the topic.

Measure, such as length, is extremely important, so all people should understand it mathematically. The mathematics of measure is very general without being complex. Defining integration as measure (instead of through Riemann sums) is simple and intuitive, but at the same time very powerful and general. Students uncomfortable with generality can always imagine specific examples. Since the generality does not add complexity (the formulas in the paper are simple and all material is covered in only six pages), calculus textbooks should accept the generality and use this paper as a 'backbone'-- expanded with proofs, additional examples, exercises, applications, integration techniques, and additional explanations--to replace their inferior and incomplete coverage.

REFERENCE

Measure Theory by D. H. Fremlin, currently available at http://www.essex.ac.uk/maths/staff/fremlin/mtcont.htm, was used to learn and verify much information about measure and integration.