Chapter 1. Fundamentals of Measure and Integration Theory

Note. We assume as background a knowledge of measure theory consistent with that given in Chapter 17 ("General Measure Spaces: Their Properties and Construction") and Chapter 18 ("Integration Over General Measure Space") of Royden and Fitzpatricks *Real Analysis*, 4th edition (Prentice Hall, 2010). So we largely skip the first 3 chapters of Ash and Doleans-Dade's book. However, the results of Section 1.4 are not in the background material, so we now cover this section which will play an important role in measure theory based probability.

Section 1.4. Lebesgue-Stieltjes Measures and Distribution Functions

Note. In this section, we define a measure, the Lebesgue-Stieltjes measure, on the Borel sets $\mathcal{B}(\mathbb{R})$ using a particular type of function, a distribution function, and conversely show that a Lebesgue-Stieltjes measure on $\mathcal{B}(\mathbb{R})$ can be used to define a distribution function. We then do the same for the Borel sets on $\mathbb{R}^n$, $\mathcal{B}(\mathbb{R}^n)$. In Chapter 4 we'll use these ideas to define a probability measure induced by a random variable (see Section 4.6, "Random Variable," and Definition 4.6.1).

Note. Ash and Doleans-Dade refer to the *extended real numbers* as the "two-point compactification" $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. (The set $\overline{\mathbb{R}}$ is compact under the *order topology*; see page 149, "Compactification," of John L. Kelley's *General Topology*, Van Nostrand Company, 1955.) They adopt the following acceptable rules of arithmetic on $\overline{\mathbb{R}}$:

$$a + \infty = \infty + a = \infty \text{ and } a - \infty = -\infty + a = -\infty \text{ for all } \mathbb{R},$$

$$\infty + \infty = \infty, \quad -\infty - \infty = -\infty$$

$$b \cdot \infty = \infty \cdot b = \begin{cases} \infty & \text{if } b \in \overline{\mathbb{R}}, b > 0 \\ -\infty & \text{if } b \in \overline{\mathbb{R}}, b < 0, \end{cases}$$

$$\text{and } \frac{a}{\infty} = \frac{a}{-\infty} = 0 \text{ for all } a \in \mathbb{R}.$$

They leave "$\infty - \infty$" and "∞/∞" undefined. They make the dubious convention that $0 \cdot \infty = \infty \cdot = 0$. They emphasize that $\overline{\mathbb{R}}$ is not a field under these operations (because, for example, neither ∞ nor $-\infty$ have multiplicative nor additive inverses; notice that this makes the convention that $a/\infty = a/(-\infty) = 0$ for $a \in \mathbb{R}$ also dubious). As seen in Royden and Fitzpatrick, we only need to deal with extended real numbers so that we can consider the pointwise limit of a sequence $\{f_n\}$ of real valued functions (in the Monotone Convergence Theorem and the Lebesgue Dominated Convergence Theorem, for example). We have seen that it is unnecessary to claim "$0 \cdot \infty = \infty \cdot 0 = 0$" since our development of the integral ultimately buried the infinities in a limiting process (usually a supremum or infimum).

Definition 1.4.1. A *Lebesgue-Stieltjes measure* on $\mathbb{R}$ is a measure μ on $\mathcal{B}(\mathbb{R})$ such that $\mu(I) < \infty$ for each bounded interval I. A *distribution function* on $\mathbb{R}$ is a map $F : \mathbb{R} \to \mathbb{R}$ that is increasing and right continuous (that is, $\lim_{x \to x_0^+} F(x) = F(x_0)$).

Theorem 1.4.2. Let μ be a Lebesgue-Stieltjes measure on $\mathbb{R}$. Let $F : \mathbb{R} \to \mathbb{R}$ be defined up to an additive constant, by $F(b) - F(a) = \mu(a,b]$. Then F is a distribution function.

Definition. For F a distribution function on $\mathbb{R}$, define the *distribution function extended to* $\overline{\mathbb{R}}$ as $F(\infty) = \lim_{x\to\infty} F(x)$ and $F(-\infty) = \lim_{x\to-\infty} F(x)$. (Since F is, by definition, monotone then $F(\infty)$ and $F(-\infty)$ are defined, though they may be ∞ or $-\infty$ themselves.)

Note. For F a distribution function on $\overline{\mathbb{R}}$, define the set function μ (which we will show is a Lebesgue-Stieltjes measure) on the set of intervals of the form $(a,b]$ where $a,b \in \overline{\mathbb{R}}$ and $a < b$ as $\mu((a,b]) = F(b) - F(a)$. Also define $\mu([-\infty,b]) = \mu([(-\infty,b]) = F(b) - F(-\infty)$ (we include a definition of μ on intervals of the form $[-\infty,b]$ since these are complements of intervals of the form (b,∞) and we are about to consider families of sets closed under unions, intersections, and complements).

Definition. Let X be a set. The collection of subsets $\mathcal{F}_0$ of set X is a *field* (or *algebra*) if it contains X and it is closed under finite unions and complements (and finite intersections, by De Morgan's Laws). The collection of subsets $\mathcal{F}_1$ of set X is a σ*-field* (or σ*-algebra*) if it contains X and it is closed under countable unions and complements (and countable intersections, by De Morgan's Laws). If $\mathcal{F}$ is a collection of subsets of X then the *minimal* σ*-field over* $\mathcal{F}$, denoted $\sigma(\mathcal{F})$, is the intersection of all σ-fields containing $\mathcal{F}$ (this is called the "the smallest σ-algebra that contains $\mathcal{F}$" or the σ-algebra generated by $\mathcal{F}$; see their Proposition 1.13).

Note. We consider the field $\mathcal{F}_0(\overline{\mathbb{R}})$ of finite disjoint unions of right-semiclosed intervals of $\overline{\mathbb{R}}$ (you are asked to show that this is in fact a field in Exercise 1.4.A). If $I_1, I_2, \ldots, I_k$ are disjoint right-semiclosed intervals of $\overline{\mathbb{R}}$, define set function μ on $\mathcal{F}_0(\overline{\mathbb{R}})$ as an extension of μ defined above as $\mu\left(\cup_{j=1}^k I_j\right) = \sum_{j=1}^k \mu(I_j)$. Notice that, by definition, μ is finitely additive on $\mathcal{F}_0(\overline{\mathbb{R}})$. We now show that μ is in fact countably additive on $\mathcal{F}_0(\overline{\mathbb{R}})$ in the following.

Lemma 1.4.3. For μ defined above on field $\mathcal{F}_0(\overline{\mathbb{R}})$, μ is countably additive. That is, for $A_1, A_2, \ldots$ disjoint sets in $\mathcal{F}_0(\overline{\mathbb{R}})$ with $\cup_{n=1}^\infty A_n \in \mathcal{F}_0(\overline{\mathbb{R}})$ we have $\mu\left(\cup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty \mu(A_n)$.

Note. We now show that a distribution function on $\mathbb{R}$ induces a unique Lebesgue-Stieltjes measure.

Theorem 1.4.4. Let F be a distribution function on $\mathbb{R}$, and let $\mu((a,b]) = F(b) - F(a)$ where $a < b$. Then there is a unique extension of μ to a Lebesgue-Stieltjes measure on $\mathcal{B}(\mathbb{R})$.

Note. Next, we consider Lebesgue-Stieltjes measures on $\mathbb{R}^n$. We start by defining sets analogous to the right semi-closed sets in $\mathbb{R}^n$.

Definition 1.4.6. If $a = (a_1, a_2, \ldots, a_n)$ and $b = (b_1, b_2, \ldots, b_n)$ are in $\mathbb{R}^n$, then the *interval* $(a, b]$ is defined as

$$(a, b] = \{x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \mid a_i < x_i \leq b_i \text{ for all } i = 1, 2, \ldots, n\}.$$

The *interval* (a, ∞) is defined as

$$(a, \infty) = \{x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \mid x_i > a_i \text{ for all } i = 1, 2, \ldots, n\}.$$

The *interval* $(-\infty, b]$ is defined as

$$(-\infty, b] = \{x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \mid x_i \leq b_i \text{ for all } i = 1, 2, \ldots, n\}.$$

Note. We can similarly define other intervals in $\mathbb{R}^n$ (such as $[a, b)$, $[a, b]$, (a, b), $[a, \infty)$, $(-\infty, b)$, and $(-\infty, \infty)$). We can also define the intervals in $\overline{\mathbb{R}}$ of $[-\infty, b)$, $[-\infty, b)$, $(a, \infty]$, $[a, \infty]$, and $[-\infty, \infty]$.

Definition. The smallest σ-field containing all intervals $(a, b]$ where $a, b \in \mathbb{R}^n$ is the class of *Borel sets of* $\mathbb{R}^n$, denoted $\mathcal{B}(\mathbb{R}^n)$. The smallest σ-field containing all intervals $(a, b]$ where $a, b \in \overline{\mathbb{R}}^n$ is the class of *Borel sets of* $\overline{\mathbb{R}}^n$, denoted $\mathcal{B}(\overline{\mathbb{R}}^n)$.

Note. A Lebesgue-Stieltjes measure on $\mathbb{R}^n$ is a measure μ on $\mathcal{B}(\mathbb{R}^n)$ such that $\mu(I) < \infty$ for each bounded interval I.

Note. To motivate the next definition, we consider the special case of $n = 3$. We would expect that, for given finite measure μ, we could define

$$F(x_1, x_2, x_3) = \mu(\{\omega = (\omega_1, \omega_2, \omega_3) \in \mathbb{R}^3 \mid \omega_1 \leq x_1, \omega_2 \leq x_2 m \omega_3 \leq x_3\} \qquad (*)$$

and that F would be a distribution function corresponding to μ (so that $\mu((a,b]) = F(b) - F(a)$). We'll see in Theorem 1.4.8 that the computations are more complicated than this, though we use $(*)$ to define F for a given finite measure μ on $\mathbb{R}^3$.

Definition. Define the *difference operator* for $G : \mathbb{R}^n \to \mathbb{R}$ as

$$\Delta_{b_i a_i} G(x_1, x_2, \ldots, x_n) = G(x_1, x_2, \ldots, x_{i-1}, b_i, x_{i+1}, x_{i+2}, \ldots, x_n)$$
$$-G(x_1, x_2, \ldots, x_{i-1}, a_i, x_{i+1}, x_{i+2}, \ldots, x_n).$$

Note. Notice that the difference operator is linear. In particular,

$$\Delta_{b_i a_i}(G \pm F) = \Delta_{b_i a_i} G \pm \Delta_{b_i a_i} F.$$

Lemma 1.4.7. Let $a, b \in \mathbb{R}^3$. If $a \leq b$ (that is, the coordinates of a and b satisfy $a_i \leq b_i$ for $i = 1, 2, 3$), then

(a) $\mu((a,b]) = \mu(\{\omega = (\omega_1, \omega_2, \omega_3) \in \mathbb{R}^3 \mid a_1 < \omega_1 \leq b_1, a_2 < \omega_2 \leq b_2, a_3 < \omega_3 \leq b_3\}) = \Delta_{b_1 a_1} \Delta_{b_2 a_2} \Delta_{b_3 a_3} F(x_1, x_2, x_3)$ where

(b) $\Delta_{b_1 a_1} \Delta_{b_2 a_2} \Delta_{b_3 a_3} F(x_1, x_2, x_3) = F(b_1, b_2, b_3) - F(a_1, b_2, b_3) - F(b_1, a_2, b_3) - F(b_1, b_2, a_3) + F(a_1, a_2, b_3) + F(a_1, b_2, a_3) + F(b_1, a_2, a_3) - F(a_1, a_2, a_3)$.

Note. Of course it is easy to see how Lemma 1.4.7(a) would extend to n dimensions. Notice that in Lemma 1.4.7(b), the terms with -1 as a coefficient are those where F has an odd number of A_i's in it. This pattern will also hold in n dimensions so that we can extend Lemma 1.4.7 to $\mathbb{R}^n$ as follows.

Theorem 1.4.8. Let μ be a finite measure on $\mathcal{B}(\mathbb{R}^n)$ and define

$$F(x) = \mu((-\infty, x]) = \mu(\{\omega = (\omega_1, \omega_2, \ldots, \omega_n) \in \mathbb{R}^n \mid \omega_i \leq x_i, i = 1, 2, \ldots, n\}).$$

If $a \leq b$ (that is, the coordinated of a and b satisfy $a_i \leq b_i$ for $i = 1, 2, \ldots, n$), then

(a) $\mu((a, b]) = \Delta_{b_1 a_1} \Delta_{b_2 a_2} \cdots \Delta_{b_n a_n} F(x_1, x_2, \ldots, x_n)$ where

(b) $\Delta_{b_1 a_1} \Delta_{b_2 a_2} \cdots \Delta_{b_n a_n} F(x_1, x_2, \ldots, x_n) = F_0 - F_1 + F_1 - \cdots + (-1)^n F_n$. Here, F_i is the sum of all $\binom{n}{i}$ terms of the form $F(c_1, c_2, \ldots, c_n)$ with $c_k = 1_k$ for exaclty i integers in $\{1, 2, \ldots, n\}$ and $c_k = b_k$ for the remaining $n - i$ integers.

Revised: 3/23/2019