The material in Sections 1-6 below serves two purposes: first, to introduce some notation and terminology, with which one should be familiar, and second, to provide a compendium of mostly elementary but commonly used facts. This material is also covered in Chapter 3 of [W]. A short summary would be that a "nice" function of a set of random variables is a random variable, and the same is true when one considers various kinds of limits of a sequence of random variables.

1 Borel σ-fields

- Recall that the σ-field of **Borel sets** in $(0, 1]$ (also known as the Borel σ-field) is the σ-field generated by the collection of all intervals of the form $(a, b]$, where $0 \leq a \leq b \leq 1$. We will use $\mathcal{B}$ to denote it.
- A Borel σ-field can also be defined when dealing with the entire real line, $\Re$. (It is denoted again by $\mathcal{B}$, by a slight abuse of notation.) This is the σ-field generated by the collection of intervals of the form $(a, b]$, but without any restrictions on a, b.
- Sometimes, instead of dealing with the set $\Re$ of real numbers, we may want to deal with the set $\overline{\Re} = [-\infty, \infty]$ of **extended real numbers.** The Borel σ-field on $\overline{\Re}$ can be defined as the smallest σ-field that contains all Borel subsets of $\Re$, as well as the sets $\{-\infty\}$ and $\{\infty\}$.
- The σ-field of Borel subsets of $\Re^k$, denoted by $\mathcal{B}^k$ can be defined in several ways, which turn out to be equivalent:
  - It is the smallest σ-field that contains all sets of the form $\{(x_1, \ldots, x_k) \mid x_1 \leq c_1, \ldots, x_k \leq c_k\}$, where $c_1, \ldots, c_k$ can be any real numbers.
  - It is the smallest σ-field that contains all sets of the form $A_1 \times A_2 \times \cdots \times A_k$, where each A_i is a Borel subset of $\Re$.
- The second definition above is the one given in p. 15 of [GS]. But the first definition is sometimes easier to use.
- The notation $\mathcal{B}^k$ is misleading. The set $\mathcal{B}^k$ is **not** the Cartesian product of k copies of $\mathcal{B}$.
- Let λ_1 be the Lebesgue measure on $(\Re, \mathcal{B})$. We then define the Lebesgue measure λ on $(\Re^k, \mathcal{B}^k)$ as follows. If $A_i \in \mathcal{B}$ for each i, then we let $\lambda(A_1 \times \cdots \times A_k) = \lambda_1(A_1) \cdots \lambda_1(A_k)$. (This corresponds to the elementary formula for the volume of a "rectangle"). We then use the extension theorem to obtain the measure $\lambda(A)$ for any $A \in \mathcal{B}$.

2 Measurable functions

- Let $(\Omega, \mathcal{F})$ be a sample space together with a σ-field of subsets of Ω. Such a pair is called a **measurable space**, and the elements of $\mathcal{F}$ are **$\mathcal{F}$-measurable sets**. When $\mathcal{F}$ is clear from the context, we just call them "measurable sets." If we also have a probability measure on $(\Omega, \mathcal{F})$, then we have a **probability space** $(\Omega, \mathcal{F}, \mathbf{P})$.
- **Definition of a random variable.** Let X be a function from Ω to $\Re$, the set of real numbers. We say that X is a **random variable**, or that X is **$\mathcal{F}$-measurable**, if the set $\{\omega \mid X(\omega) \leq c\}$ is $\mathcal{F}$-measurable for every $c \in \Re$.
- We are sometimes interested in functions from Ω that can also take infinite values. In such cases, we repeat the above definition but with $(\Re, \mathcal{B})$ replaced by $(\overline{\Re}, \mathcal{B})$, and we say that X is an **extended-valued** random variable.

- Because the intervals of the form $(-\infty, c]$ "generate" the Borel σ-field in $\Re$, it can be shown that if X is a random variable, then for any Borel-measurable set B, the set $X^{-1}(B) = \{\omega \mid X(\omega) \in B\}$ is $\mathcal{F}$-measurable set. Because of this, the probability $\mathbf{P}(X \in B) = \mathbf{P}(\{\omega \mid X(\omega) \in B\})$ is well-defined, and is actually a probability measure on $(\Re, \mathcal{B})$.

- **A more elegant definition of a random variable.** Because of the previous remark, we might as well define a random variable as a function from Ω to $\Re$ such that $X^{-1}(B) \in \mathcal{F}$ for every $B \in \mathcal{B}$. Still, whenever we need to check this condition, we only need to check it for sets B of the special form $B = (-\infty, c]$.

- **The distribution function of a random variable X.** The function $F : \Re \to [0, 1]$ defined by $F(c) = \mathbf{P}(X \leq c)$ is called the **distribution function** of X, or sometimes the **cumulative distribution function of X (CDF)**.

- For every Borel subset B of the real line, we define $\mathbf{P}_X(B) = \mathbf{P}(X \in B)$. [Note that $\mathbf{P}$ is a measure on $(\Omega, \mathcal{F})$, whereas $\mathbf{P}_X$ is a measure on $(\Re, \mathcal{B})$.] This probability measure is called the **probability law** of X, or sometimes the **distribution** of X.

3 Facts about measurability

- If A is a subset of Ω, we define its **indicator function** $I_A : \Omega \to \Re$ by $I_A(\omega) = 1$ if $\omega \in A$, and $I_A(\omega) = 0$ if $\omega \notin A$.

- If $A_1, \ldots, A_n$ are $\mathcal{F}$-measurable sets, and $\lambda_1, \ldots, \lambda_n$ are real numbers, the function $X = \sum_{i=1}^{n} \lambda_i I_{A_i}$, or in more detail,

$$X(\omega) = \sum_{i=1}^{n} \lambda_i I_{A_i}(\omega), \qquad \forall \omega \in \Omega,$$

  is $\mathcal{F}$-measurable, i.e., a random variable. Such a random variable is called a **simple** random variable.

- Let each f_n be a function from Ω into $\Re$. Consider a new function $f = \inf_n f_n$ defined by $f(\omega) = \inf_n f_n(\omega)$ for every $\omega \in \Omega$. The functions $\sup_n f_n$, $\liminf_n f_n$, and $\limsup_n f_n$ are defined similarly. Note that even if the f_n are everywhere finite, the above defined functions may turn out to be extended-valued.

- **Measurability of limits:** If each f_n is $\mathcal{F}$-measurable, then $\inf_n f_n$, $\sup_n f_n$, $\liminf_n f_n$, and $\limsup_n f_n$ are $\mathcal{F}$-measurable (extended-valued) functions. Furthermore, if $\lim_{n\to\infty}(f_n(\omega))$ exists for all ω, then the limit function f defined by $f(\omega) = \lim_{n\to\infty}(f_n(\omega))$ is also measurable. This limit function is called the **pointwise limit** of the sequence of functions f_n, because we consider separately each different "point" ω and take the limit as $n \to \infty$ to obtain $f(\omega)$. For example, suppose that $\omega \in [0, 1]$ and that $f_n(\omega) = \omega^n$. Then, the pointwise limit f exists, with $f(1) = 1$ and $f(\omega) = 0$ for $\omega \in [0, 1)$.

- According to the previous item, a pointwise limit of simple functions is measurable. It turns out that the converse is also true. Namely, every measurable function is the pointwise limit of simple functions. And every nonnegative measurable function is the pointwise limit of a nondecreasing sequence of simple functions (that is, $f_n(\omega) \leq f_{n+1}(\omega)$ for every n and ω). These facts will be revisited when we study expected values of random variables.

- A function $g : \Re \to \Re$ is called measurable if $g^{-1}(B) \in \mathcal{B}$ for every $B \in \mathcal{B}$. (The proper term would be "Borel-measurable," but the Borel σ-field is the default choice unless the contrary is stated).

- **Composition of measurable functions.** If $f : \Omega \to \Re$ is $\mathcal{F}$-measurable, and $g : \Re \to \Re$ is measurable, then the composition $g \circ f$ of these two functions, defined by $(g \circ f)(\omega) = g(f(\omega))$, is also $\mathcal{F}$-measurable.

4 Measurability and continuity

- **Continuity implies measurability.** If f is a continuous function from $\Re$ to $\Re$, then f is measurable.
- Remark: We can define continuity of f by requiring that whenever $x_n \to x$, we must also have $f(x_n) \to f(x)$. It turns out that this is equivalent to the following more abstract condition: for every open subset U of $\Re$, the set $f^{-1}(U)$ is also open. Since every open subset of $\Re$ is Borel (Homework 1), this discussion outlines a proof of the preceding claim, that continuity implies measurability.
- Since continuity implies measurability, and since the composition of measurable functions is measurable, we obtain the following. If X is a random variable and f is a continuous function from $\Re$ into itself, then $f(X)$ is also a random variable.

5 Multiple random variables

- If X and Y are random variables, then so are $X+Y$, XY, $\max\{X,Y\}$.
- If $Y(\omega) \neq 0$ for every ω, then X/Y is also a random variable.
- Generalizing the first fact of this section, the following is true. If $X_1, \ldots, X_n$ are random variables and f is a continuous function from $\Re^n$ to $\Re$, then $f(X_1, \ldots, X_n)$ is a random variable.

6 A more economical development

Most of the above discussion and facts can be expressed in a more economical manner, as follows. The statements below are crisper, but more abstract, and somewhat harder to parse.

Suppose that we have three measurable spaces $(\Omega_1, \mathcal{F}_1)$, $(\Omega_2, \mathcal{F}_2)$, $(\Omega_3, \mathcal{F}_3)$.

- A function $f : \Omega_1 \to \Omega_2$ is called $\mathcal{F}_1/\mathcal{F}_2$-measurable if $f^{-1}(B) \in \mathcal{F}_1$ for every $B \in \mathcal{F}_2$.
- With this definition, if $\Omega_2 = \Re$, then a random variable is just a $\mathcal{F}_1/\mathcal{B}$-measurable function.
- The composition of a $\mathcal{F}_1/\mathcal{F}_2$-measurable function and a $\mathcal{F}_2/\mathcal{F}_3$-measurable function is $\mathcal{F}_1/\mathcal{F}_3$-measurable.
- If $(\Omega_2, \mathcal{F}_2) = (\Re^k, \mathcal{B}^k)$, $(\Omega_3, \mathcal{F}_3) = (\Re^m, \mathcal{B}^m)$, and $g : \Omega_2 \to \Omega_3$ is continuous, then g is $\mathcal{F}_2/\mathcal{F}_3$-measurable.
- If $X_1, \ldots, X_k$ are random variables, then the mapping $f : \Omega_1 \to \Re^k$ defined by $f(\omega) = (X_1(\omega), \ldots, X_k(\omega))$ is $\mathcal{F}_1/\mathcal{B}^k$-measurable.
- By combining the last two facts, we recover the last fact of the previous section.

7 The Cantor set and a singular measure

- Every number $x \in [0,1]$ has a ternary expansion of the form

$$x = \sum_{i=1}^{\infty} \frac{x_i}{3^i}, \qquad \text{with } x_i \in \{0,1,2\}. \tag{1}$$

  This expansion is not unique. For example, $1/3$ admits two expansions, namely $.10000\cdots$ and $.022222\cdots$. Nonuniqueness occurs only for those x that admit an expansion that ends with an infinite sequence of 2's. The set of such unusual x is countable, and therefore has Lebesgue measure zero.

- The **Cantor set** C is the set of all x that have a ternary expansion that uses only 0's and 2's (no 1's allowed). The set C is constructed as follows. Start with the interval $[0,1]$ and remove the "middle third" $(1/3, 2/3)$. Then, from each of the remaining closed intervals $[0,1/3]$ and $[2/3,1]$ remove their middle thirds, $(1/9, 2/9)$ and $(7/9, 8/9)$, resulting in four closed intervals, and continue similarly. Note that C is measurable, since it is constructed by removing a countable sequence of intervals. Also, the length (Lebesgue measure) of C is 0, since at each stage its length is mutliplied by a factor of $2/3$.

- Consider an infinite sequence of rolls of a 3-sided die, whose faces are labeled 0, 1, and 2. Assume that rolls are independent, and that at each roll each of the three possible results has probability $1/3$. If we use the sequence of these rolls to form a number x, then the probability law of the resulting random variable is the Lebesgue measure (i.e., picking a ternary expansion "at random" leads to a uniform random variable).

- The Cantor set can be identified with those roll sequences in which a "1" never occurs. (These roll sequences have zero probability which is consistent with the fact that C has zero Lebesgue measure.)

- The set C has the same cardinality as the set $\{0,2\}^\infty$, which is uncountable.

- Consider an infinite sequence of coin tosses. If the ith flip results in tails, record $x_i = 0$; if it results in heads, record $x_i = 2$. Use the x_i's to form a number x, using Eq. (1). This defines a random variable X whose range is the set C. The probability law of this random variable is therefore concentrated on the "zero-length" set C. At the same time $\mathbf{P}(X = x) = 0$ for every x, because any particular sequence of heads and tails has zero probability.

8 Decomposition of measures on $(\Re, \mathcal{B})$

- A probability measure $\mathbf{P}$ on $(\Re, \mathcal{B})$ is **discrete** if there is a countable set $A \subset \Re$ such that $\mathbf{P}(A) = 1$.

- A probability measure $\mathbf{P}$ has **no point mass** if $\mathbf{P}(\{x\}) = 0$ for every x.

- A probability measure $\mathbf{P}$ on $(\Re, \mathcal{B})$ is **absolutely continuous** if $\mathbf{P}(A) = 0$ for every Borel set A that has Lebesgue measure zero.

- It turns out that absolutely continuous probability laws admit a density f, i.e., there exists a measurable function f so that $\mathbf{P}(X \le c) = \int_{-\infty}^{c} f(x)\,dx$, for every c. This is a deep and difficult result. Furthermore, it requires a proper definition of an integral, when the integrand is measurable but discontinuous.

- A probability measure $\mathbf{P}$ on $(\Re, \mathcal{B})$ is **singular** if there is a set A that has zero Lebesgue measure and such that $\mathbf{P}(A) = 1$. (A discrete distribution is singular. The probability law in the last item of the preceding section, which is concentrated on the Cantor set, is also singular.)

- If $\mathbf{P}_1, \mathbf{P}_2$ are two probability measures on $(\Re, \mathcal{B})$, and λ_1, λ_2 are nonnegative numbers that sum to 1, we define $\mathbf{P}(A) = \lambda_1 \mathbf{P}_1(A) + \lambda \mathbf{P}_2(A)$, for every Borel set A. It is easily checked that $\mathbf{P}$ is also a probability measure. It is referred to as a **mixture** of $\mathbf{P}_1$ and $\mathbf{P}_2$.

- **Lebesgue decomposition theorem.** Every probability measure $\mathbf{P}$ on $(\Re, \mathcal{B})$ is of the form $\mathbf{P} = \lambda_1 \mathbf{P}_1 + \lambda_2 \mathbf{P}_2 + \lambda_3 \mathbf{P}_3$, where $\mathbf{P}_1$ is discrete, $\mathbf{P}_2$ is absolutely continuous, and $\mathbf{P}_3$ is singular without any point mass, and the λ_i are nonnegative numbers that sum to 1. Furthermore, this decomposition is unique.

- The decomposition theorem is also true for measures on $(\Re^k, \mathcal{B}^k)$.