Elements of Information Theory Update: Friday 7^{th} February, 2020

Notes 1: Preliminaries on Random Variables and Convexity

By: Behrooz Razeghi

In this note we will review some of the main definitions and properties of probability and convex analysis, and discuss Gaussian random variables and vectors.

1 Notation

A quick summary of the notation

1. **Random Variable (object):** capital letter, e.g. X, Y
2. **Deterministic Value (object):** small letter, e.g. x, y
3. **Random Vector (object):** capital bold letter, e.g. $\mathbf{X}$, $\mathbf{Y}$
4. **Deterministic Vectors (objects):** small bold letters, e.g. $\mathbf{x}$, $\mathbf{y}$
5. **Alphabets (Sets):** Calligraphic font, e.g. $\mathcal{B}, \mathcal{X}, \mathcal{Y}, \mathcal{E}, \mathcal{F}$, or $\mathscr{E}, \mathscr{B}, \mathscr{F}, \mathscr{I}$
6. **Specific quantity/value:** x, y, C, D, P, Ω

2 Measure Spaces

The material of this section is mostly inspired by the nice text [1]. Before we discuss abstract measure measure spaces, let us to define the class of *elementary sets* and corresponding *elementary measure*, to get some intuition for more abstract measure spaces.

Definition 1. *(Intervals, boxes and elementary sets) An interval is a subset of $\mathbb{R}$ of the form $[a,b] := \{x \in \mathbb{R} : a \le x \le b\}$, $[a,b) := \{x \in \mathbb{R} : a \le x < b\}$, $(a,b] := \{x \in \mathbb{R} : a < x \le b\}$, or $(a,b) := \{x \in \mathbb{R} : a < x < b\}$, where $a \le b$ are real numbers. The length $\mathscr{I}$ of an interval $\mathscr{I} = [a,b], [a,b), (a,b], (a,b)$ to be $|\,\mathscr{I}\,| := b - a$. A box in $\mathbb{R}^d$ is a Cartesian product $\mathscr{B} := \mathscr{I}_1 \times \mathscr{I}_2 \times ... \times \mathscr{I}_d$ of d intervals $\mathscr{I}_1, ..., \mathscr{I}_d$. The volume $|\mathscr{B}| := |\mathscr{I}_1| \times ... \times |\mathscr{I}_d|$. An elementary set is any subset of $\mathbb{R}^d$ which is the union of a finite number of boxes.*

Definition 2. *(Boolean closure for elementary sets) If $\mathscr{E}_1$ and $\mathscr{E}_2 \subset \mathbb{R}^d$ are elementary sets, then the union $\mathscr{E}_1 \cup \mathscr{E}_2$, the intersection $\mathscr{E}_1 \cap \mathscr{E}_2$, and the set theoretic difference $\mathscr{E}_1 \setminus \mathscr{E}_2 := \{x \in \mathscr{E}_1 : x \notin \mathscr{E}_2\}$, and symmetric difference $\mathscr{E}_1 \Delta \mathscr{E}_2 := (\mathscr{E}_1 \setminus \mathscr{E}_2) \cup (\mathscr{E}_2 \setminus \mathscr{E}_1)$ are also elementary.*

Definition 3. *(Boolean algebras) Let Ω be a set. A Boolean algebra on Ω is a collection $\mathcal{B}$ of Ω which obeys the following properties:*

(i) ***(Empty Set)*** $\emptyset \in \mathcal{B}$

(ii) ***(Complement)*** *If $\mathscr{E} \in \mathcal{B}$, then the complement $\mathscr{E}^c := \Omega \setminus \mathscr{E}$ also lies in $\mathcal{B}$.*

(iii) ***(Finite Unions)*** *$\mathscr{E}_1, \mathscr{E}_2 \in \mathcal{B}$, then $\mathscr{E}_1 \cup \mathscr{E}_2 \in \mathcal{B}$.*

Note that the above definition is only assumed closure under *complement* and *finite union* Boolean operations. However, by using the de Morgan's laws of elementary set theory, one can easily verify other closure properties such as intersection $\mathscr{E}_1 \cap \mathscr{E}_2$, set difference $\mathscr{E}_1 \setminus \mathscr{E}_2$, and symmetric difference $E_1 \triangle \mathscr{E}_2$.

Definition 4. *(σ-algebras) Let Ω be a set. A σ-algebra on Ω is a collection $\mathcal{B}$ of Ω which obeys the following properties:*

(i) ***(Empty Set)*** $\emptyset \in \mathcal{B}$

(ii) ***(Complement)*** *If $\mathcal{E} \in \mathcal{B}$, then the complement $\mathcal{E}^c := \Omega \setminus \mathcal{E}$ also lies in $\mathcal{B}$.*

(iii) ***(Countable Unions)*** *$\mathcal{E}_1, \mathcal{E}_2, ... \in \mathcal{B}$, then $\cup_{n=1}^{\infty} \mathcal{E}_n \in \mathcal{B}$.*

Terminology: The pair $(\Omega, \mathcal{B})$ of a set Ω together with a σ-algebra on that set, is refer as a *measurable space*. The subsets belonging to $\mathcal{B}$ are said to be $\mathcal{B}$-measurable.
Note: Using de Morgan's law, we see that σ-algebras are closed under countable intersections as well as countable unions.
Note: By padding a finite union into a countable union by using the empty set, we see thata every σ-algebra is automatically a Boolean algebra.
Example: For a generic set Ω, there are two extremes:

- $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

  This is the smallest possible σ-field of Ω (also known as trivial σ-field, or trivial space).

- $\mathcal{F} = \{\text{set of all subsets of } \Omega\}$.

  This is the largest possible σ-field of Ω (also known as complete σ-field, or power set).

Definition 5. *(Finitely additive measure) Let $\mathcal{B}$ be a Boolean algebra on a space Ω. An (unsigned) finitely additive measure μ on $\mathcal{B}$ is a 'map' $\mu : \mathcal{B} \to [0, +\infty]$ that obeys the following axioms:*

(i) ***(Empty Set)*** $\mu(\emptyset) = 0$.

(ii) ***(Finite Additivity)*** *Whenever $\mathcal{E}_1, \mathcal{E}_2 \in \mathcal{B}$ are disjoint, then $\mu(\mathcal{E}_1 \cup \mathcal{E}_2) = \mu(\mathcal{E}_1) + \mu(\mathcal{E}_2)$.*

Definition 6. *(Countably additive measure) Let $(\Omega, \mathcal{B})$ be a measurable space. An (unsigned) countably additive measure μ on $\mathcal{B}$, or measure for short, is a 'map' $\mu : \mathcal{F} \to [0, +\infty]$ that obeys the following axioms:*

(i) ***(Empty Set)*** $\mu(\emptyset) = 0$.

(ii) ***(Countable Additivity)*** *Whenever $\mathcal{E}_1, \mathcal{E}_2, ... \in \mathcal{B}$ are countable sequence of disjoint measurable sets, then $\mu(\cup_{n=1}^{\infty} \mathcal{E}_n) = \sum_{n=1}^{\infty} \mu(\mathcal{E}_n)$.*

A triplet $(\Omega, \mathcal{B}, \mu)$, where $(\Omega, \mathcal{B})$ is a measurable space and $\mu : \mathcal{B} \to [0, +\infty]$ is a countably additive measure, is known as a *measure space*.
Note: A measure space and a measurable space are distinct. A measurable space has the *capability* to be equipped with a measure, but a measure space is actually *equipped* with a measure.

3 Probability Spaces

Probability spaces is an important special type of measure spaces.

Definition 7. *(Probability Space) A probability space is a measurable space $(\Omega, \mathcal{B}, \mathsf{P})$ of total measure 1, i.e., $\mathsf{P}(\Omega) = 1$. The measure P is known as a probability measure.*

Let us to interpret the components $\Omega, \mathcal{B}$ and P in probability theory:

- The space Ω is known as *sample space*, and is interpreted as the set of all possible states (outcomes) $\omega \in \Omega$ that a random system (nature) could be in.
- The *event space* $\mathcal{B}$ is a σ-algebra, and is interpreted as the set of all possible events $\mathcal{E} \in \mathcal{B}$ that one can measure.
- The *probability measure* $\mathsf{P}(\cdot)$ of an event is a mapping that assigns to each event a real number that follows the foundational axioms of measure and probability theory.

4 Random Variables (Objects)

Given a measurable space $(\Omega, \mathcal{B})$, let $(\mathcal{X}, \mathcal{F})$ denote another measurable space. The first space can be thought of as *original space* and the second as *target space.* A *random variable* (or measurable space) is a mapping (or function) $X : \Omega \to \mathcal{X}$, i.e., a measurable function defined on $(\Omega, \mathcal{B})$ and taking values in $(\mathcal{X}, \mathcal{F})$, with the property that:

$$\text{if } \mathcal{E} \in \mathcal{F}, \text{ then } X^{-1}(\mathcal{E}) = \{\omega : X(\omega) \in \mathcal{E}\} \in \mathcal{B}. \tag{1}$$

Terminology: The random variable is sometimes called $\mathcal{F}$-measurable. The name random object includes random variables, random vectors and random processes. However, usually, the term random variable is used as random object.

Let us to define the induced probability measure P_X as:

$$\mathsf{P}_X(\mathcal{E}) = \mathsf{P}\left(X^{-1}(\mathcal{E})\right) = \mathsf{P}\left(\{\omega : X(\omega) \in \mathcal{E})\}\right), \forall \mathcal{E} \in \mathcal{X}. \tag{2}$$

The induced probability measure P_X is called distribution of the random variable X. We shall sue the common notation P_X for induced probability measure P_X. Now, let us to discuss this concept in terms of discrete and continuous random variables.

4.1 Discrete Random Variables

In discrete probability theory, the sample space is finite ($\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$) or at most countable ($n = \infty$). The *atomic events* are the events that contain only a single sample point, e.g. $\{\omega_i\}$. In other words, the smallest elements contained in a σ-field are called the *atoms* of the σ-field. Note that a σ-field with n atoms has 2^n elements. Let denote by p_i the numbers (probabilities) that the probability measure $\mathsf{P}(\cdot)$ assigns to the atomic event $\{\omega_i\}$, i.e. $p_i = \mathsf{P}(\{\omega_i\})$, for $i = 1, 2, ..., n$. It turns out that p_i can completely determine the probabilities of all events.

A discrete random variable is a mapping from the sample space Ω into the real numbers. For example, let $\Omega = \{\omega_1, \omega_2, ..., \omega_6\}$, we might define $X(\omega_i) = i$, $i = 1, ..., 6$. So, the set of possible values of the random variables is $\mathcal{X} := X(\Omega) = \{1, 2, ..., 6\}$. The set $\mathcal{X}$ is called alphabet of X.

Example: Let $\mathcal{X} = \{1, 2, ..., 6\}$. The following are σ-algebra on $\mathcal{X}$:

- $\mathcal{F}_1 = \{\emptyset, \{1\}, \{2, ..., 6\}, \mathcal{X}\}$.
- $\mathcal{F}_2 = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \mathcal{X}\}$.
- $\mathcal{F}_3 = \{\emptyset, \{1, 2, 3\}, \{4, 5, 6\}, \mathcal{X}\}$.

Example: Let $\mathcal{X} = \{1, 2, ..., 6\}$ and:

- Let $\mathcal{G}_1 = \{\{1\}\}$, which is subset of $\mathcal{X}$. The σ-filed generated by $\mathcal{G}_1$, denoted by $\sigma(\mathcal{G}_1)$, is $\mathcal{F}_1$.
- Let $\mathcal{G}_2 = \{\{1, 3, 5\}\}$. Then, $\sigma(\mathcal{G}_2) = \mathcal{F}_2$.
- Let $\mathcal{G}_3 = \{\{1, 2, 3\}\}$. Then, $\sigma(\mathcal{G}_3) = \mathcal{F}_3$.
- Let $\mathcal{G} = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$. Then, $\sigma(\mathcal{G}) = \{\text{set of all subsets of } \mathcal{X}\}$, i.e., the power set of $\mathcal{X}$.[1]

The probability distribution of a random variable X, denoted by $P_X(\cdot)$, is the mapping from $X(\Omega)$ into the interval $[0, 1]$ such that

$$P_X(x) \triangleq \mathsf{P}\left(\{\omega : X(\omega) = x\}\right), \quad \forall x \in X(\Omega). \tag{3}$$

Usually, P_X is called the probability mass function (p.m.f) of X. For ease of notation, p.m.f is simply expressed as:

$$P_X(x) = \mathsf{P}\left[X = x\right]. \tag{4}$$

[1]This is a particular example to show that a σ-field is always generated by the collection of its atoms.

Let us $X_1, X_2, ..., X_n$ are random variables with alphabets $\mathcal{X}_1, \mathcal{X}_2 ..., \mathcal{X}_n$, respectively. Their joint probability distribution (or joint p.m.f) is defined as the mapping from $\mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_n$ into interval $[0,1]$ such that

$$P_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) = \mathsf{P}[X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n]. \tag{5}$$

From the definition, $P_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) \geq 0$ and $\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} P_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) = 1$. If $X_1, X_2, ..., X_n$ are statistically independent, then $P_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) = P_{X_1}(x_1) \cdot P_{X_2}(x_2) \cdots P_{X_n}(x_n)$, $\forall x_i \in \mathcal{X}_i$, $i = 1, 2, ..., n$.

4.2 Continuous Random Variables

Example: Let $\mathcal{X} = [0,1]$ and $\mathcal{E}_1, ..., \mathcal{E}_n$ be a family of disjoint intervals in Ω such that $\mathcal{E}_1 \cup \mathcal{E}_2 \cup ... \cup \mathcal{E}_n = \mathcal{X}$ [2]. The following is a σ-algebra on $\mathcal{X}$:

$$\mathcal{F}_4 = \{\emptyset, \mathcal{E}_1, ..., \mathcal{E}_n, \mathcal{E}_1 \cup \mathcal{E}_2, ..., \mathcal{E}_1 \cup E_2 \cup \mathcal{E}_3, ..., \Omega\}.$$

Note that there are 2^n events in total in $\mathcal{F}_4$. Also, note that the σ-field generated by a collection of events, is much larger than the collection of events itself.

5 Gaussian Random Variables

6 Gaussian Vectors

7 Convexity

References

[1] Terence Tao. *An introduction to measure theory*, volume 126.

[2] The set $\{\mathcal{E}_1, \mathcal{E}_2, ..., \mathcal{E}_n\}$ is also called a partition of $\mathcal{X}$.