Elements of Information Theory Update: Wednesday 5th February, 2020

Note 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. **X**, **Y**
4. **Deterministic Vectors (objects):** small bold letters, e.g. **x**, **y**
5. **Alphabets (Sets):** Calligraphic font, e.g. $\mathcal{X}, \mathcal{Y}$
6. **Specific quantities/value:** x, y, C, D, P

2 Measure Spaces

Definition 1. *(Boolean algebras) Let Ω be a set. A Boolean algebra on Ω is a collection $\mathcal{F}$ of Ω which obeys the following properties:*

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

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

(iii) ***(Finite Unions)*** *$E_1, E_2 \in \mathcal{F}$, then $E_1 \cup E_2 \in \mathcal{F}$.*

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 $E_1 \cap E_2$, set difference $E_1\ E_2$, and symmetric difference $E_1 \triangle E_2$.

Definition 2. *(σ-algebras) Let Ω be a set. A σ-algebra on Ω is a collection $\mathcal{F}$ of Ω which obeys the following properties:*

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

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

(iii) ***(Countable Unions)*** *$E_1, E_2, ... \in \mathcal{F}$, then $\cup_{n=1}^{\infty} E_n \in \mathcal{F}$.*

The pair $(\Omega, \mathcal{F})$ of a set Ω together with a σ-algebra on that set, is refer as a *measurable space.*
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.

Definition 3. *(Finitely additive measure) Let $\mathcal{F}$ be a Boolean algebra on a space Ω. An (unsigned) finitely additive measure μ on $\mathcal{F}$ is a 'map' $\mu : \mathcal{F} \to [0, +\infty]$ that obeys the following axioms:*

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

(ii) ***(Finite Additivity)*** *Whenever $E_1, E_2 \in \mathcal{F}$ are disjoint, then $\mu(E_1 \cup E_2) = \mu(E_1) + \mu(E_2)$.*

Definition 4. *(Countably additive measure) Let $(\Omega, \mathcal{F})$ be a measurable space. An (unsigned) countably additive measure μ on $\mathcal{F}$, 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 $E_1, E_2, ... \in \mathcal{F}$ are countable sequence of disjoint measurable sets, then $\mu(\cup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \mu(E_n)$.*

A triplet $(\Omega, \mathcal{F}, \mu)$, where $(\Omega, \mathcal{F})$ is a measurable space and $\mu : \mathcal{F} \to [0, +\infty]$ 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 5. *(Probability Space) A probability space is a measurable space $(\Omega, \mathcal{F}, \mathsf{P})$ 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 $\Omega, \mathcal{F}$ 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* $\mathcal{F}$ is a σ-algebra, and is interpreted as the set of all possible events $E \in \mathcal{F}$ 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.

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

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

Example: Let $\Omega = [0, 1]$ and $E_1, ..., E_n$ be a family of disjoint intervals in Ω such that $E_1 \cup E_2 \cup ... \cup E_n = \Omega$ [1]. The following is a σ-algebra on Ω:

$$\mathcal{F}_4 = \{\emptyset, E_1, ..., E_n, E_1 \cup E_2, ..., E_1 \cup E_2 \cup E_3, ..., \Omega\}.$$

Note that there are 2^n events in total in $\mathcal{F}_4$.

[1] The set $\{E_1, E_2, ..., E_n\}$ is also called a partition of Ω.

4 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. $\{\omega_i\}$. 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 $X(\Omega) = \{1, 2, ..., 6\}$. The set $X(\Omega)$ is called alphabet of X.

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

$$P_X(x) \triangleq \Pr(\{\omega : X(\omega) = x\}), \quad \forall x \in X(\Omega). \tag{1}$$

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:

$$P_X(x) = \Pr[X = x]. \tag{2}$$

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

$$P_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) = \Pr[X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n]. \tag{3}$$

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 $P_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) = P_{X_1}(x_1) \cdot P_{X_2}(x_2) \cdots P_{X_n}(x_n)$, $\forall x_i \in \mathcal{X}_i$, $i = 1, 2, ..., n$.

5 Continuous Random Variables

6 Gaussian Random Variables

7 Gaussian Vectors

8 Convexity