Probability and Random Variables

Bahman Moraffah
School of Electrical and Computer Engineering
06/15/2020

Readings:

- Probability, Statistics, and Random Process, B. Moraffah, textbook in statistics, Springer, 2021(Chapter 2)
- Stochastic Processes: Theory for Applications, R. Gallager, Cambridge, 2013(Chapter 2)

Probability and Random Variables: Outline

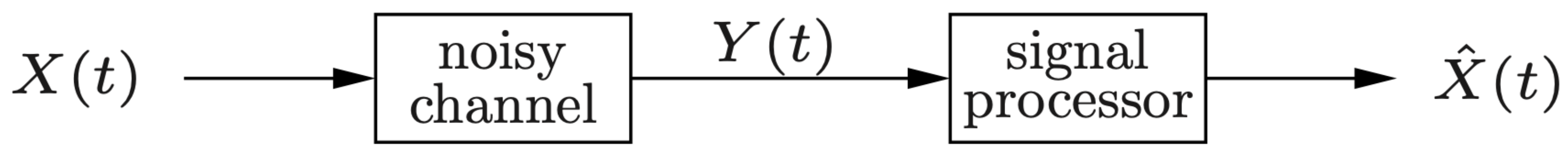
- Probability in real world
- Statistical signal processing
- Probability space
- Axioms of probability theory
- Conditional probability and Bayes rule

Probability in Real World

- Games of chance started in ancient civilizations. Along with a propensity to gamble.
  - people have an intuitive sense of likelihood and average behavior. E.g., Board game since childhood
  - Games of chances are Repeatable!
- Most of life decision involves uncertainly!
  - People learn to associate a sense of likelihood with uncertain possibilities.
  - Ph.D. program! Uncertainty about whether you enjoy it or not!
- Probability is most useful when repeatability under essentially the same conditions occur!
- Important questions involving uncertainly are far harder to make sense of. E.g., what is the probability that google stocks will double in the next 5 years?

Statistical Signal Processing

- In most signal processing, we deal with deterministic (linear) systems, however in this course we study statistical properties of systems.
- Statistical signal processing focuses on extracting useful information from noisy observations.
- A signal processing problem often is



  - X(t) = signal: images, audios, videos, medial, sensor, etc.
  - Noisy channel: is statistical model: wireless, optical, satellite, etc.
  - Noise is a random processes (a sequence of random variables indexed by time).

How do we understand a statistical problem?

Probability Space

- Probability theory is a set of rules which assign probability to random experiments.
  - E.g. Flipping a coin, roll a die, etc.
- A probability space is the triple $(\boldsymbol{\Omega}, \mathcal{F}, \mathbb{P})$ where:
  - Sample Space ($\boldsymbol{\Omega}$): Set of all possible outcomes of a random experiment.
    - Sample space can be discrete or continuous.
  - Set of Events ($\mathcal{F}$): Set of (all?) subsets of sample space.
    - $\mathcal{F}$ must satisfy the σ-field conditions
      - $\mathcal{F}$ is a σ-field if $\boldsymbol{\Omega} \in \mathcal{F}$ and is closed with respect to complement and infinite countable union.
  - Probability Measure ($\mathbb{P}$): Probability measure is a function over the set of events
    - Probability measure must satisfy "Axioms of Probability".

Axioms of Probability

- A probability measure must satisfy:
  1. $\mathbb{P}(\mathbf{A}) \geq \mathbf{0}$, for all events
  2. $\mathbb{P}(\mathbf{\Omega}) = \mathbf{1}$
  3. If $\mathbf{A_1, A_2, \ldots}$ are a sequence of disjoint events -- $\mathbf{A_i} \cap \mathbf{A_j} = \emptyset$ for all i, j- then

$$\mathbb{P}\big(\cup_{\mathbf{i}} \mathbf{A_i}\big) = \sum_{\mathbf{i}} \mathbb{P}(\mathbf{A_i})$$

Remark 1: Probability function is a measure the same sense as mass, area, volume, etc. which satisfy axioms of (1)-(3).

Remark 2: Probability measure is bounded (Axiom 2).

Remark 3: This provides some intuition about probability but not enough to completely specify independence or conditioning.

Discrete Probability Space

- A sample space is discrete if it is countable.
- For a discrete sample, the set of events, $\mathcal{F}$, is typically the power set, $\mathbf{2^{\Omega}}$.
- Example 1 : Tossing a coin: $\mathbf{\Omega = \{H, T\}}$ and

$$\mathcal{F} = \{\emptyset, \{\mathbf{H}\}, \{\mathbf{T}\}, \{\mathbf{\Omega}\}\}$$

- Example 2 : Rolling a die: $\mathbf{\Omega = \{1, 2, \ldots, 6\}}$ and $\mathcal{F} = \mathbf{2^{\Omega}}$

Note: Set of events is not necessarily the entire power set.

For discrete space probability measure can be defined as:

$$\mathbb{P}(\{\omega\}) \geq \mathbf{0} \quad , \forall \omega \in \mathbf{\Omega}$$
$$\sum_{\omega \in \mathbf{\Omega}} \mathbb{P}(\{\omega\}) = \mathbf{1}$$

- Remark: Based on the definition of probability measure for discrete space, for any event A, we have

$$\mathbb{P}(\mathbf{A}) = \sum_{\omega \in \mathbf{A}} \mathbb{P}(\{\omega\})$$

- Example 3: In example 2, we can define $\mathbb{P}(\{\mathbf{i}\}) = \frac{\mathbf{1}}{\mathbf{6}}, \mathbf{i = 1, 2, \ldots, 6}$

Continuous Probability Space

- A sample space is Continuous if it has uncountable number of elements.
- Example 1 : Random Numbers between 0 and 1, i.e. $\mathbf{\Omega = [0, 1)}$
- Example 2: Arrival of a packet, i.e. $\mathbf{\Omega} = \mathbb{R}^+$

For continuous sample space, we cannot take the set of events F to be the power set of Ω. Also, the set of events cannot be an arbitrary collection of subsets of Ω.

For any open interval in R, the σ-field is typically defined as the family of sets obtained by starting from open (closed) intervals and taking countable unions, intersections, and complements. The resulting σ-field is called **Borel** field.

Continuous Probability Space

- Can all the subsets of R = $(-\infty,\infty)$ be written as intersection and union open (closed) subsets?

Continuous Probability Space

- Can all the subsets of R = $(-\infty,\infty)$ be written as intersection and union open (closed) subsets?
  - NO! There are subsets of R that cannot be constructed this way.
- Constructing a σ-field is not an easy task to do and requires the knowledge of measure theory.

- Probability measure is defined to satisfy the axioms of probability. E.g., for continuous sample space: the uniform probability measure on (0,1) is defined by assigning the probability measures to open intervals, meaning for $0 \le a < b \le 1$; $P((a,b)) = b-a$ where $b-a$ is the length of interval.

Useful Probability Laws

- Union Bounds:

$$\mathbb{P}(\cup_{\mathbf{n=1}}^{\mathbf{N}} \mathbf{A_n}) \leq \sum_{\mathbf{i=1}}^{\mathbf{N}} \mathbb{P}(\mathbf{A_n})$$

- Law of Total Probability: Assume $A_1, A_2, \ldots$ to be the events that partition Ω, i.e., disjoint (Ai ∩ Aj = Ø for $\mathbf{i} \neq \mathbf{j}$) and $\cup_{\mathbf{i}} \mathbf{A_i}$= Ω. Then for any event B

$$\mathbb{P}(\mathbf{B}) = \sum_{\mathbf{i}} \mathbb{P}(\mathbf{B} \cap \mathbf{A_i}) = \sum_{\mathbf{i}} \mathbb{P}(\mathbf{B}, \mathbf{A_i})$$

  - Law of total probability is a useful trick to find the probability of sets.

- Probability of Union

$$\mathbb{P}(\mathbf{A} \cup \mathbf{B}) = \mathbb{P}(\mathbf{A}) + \mathbb{P}(\mathbf{B}) - \mathbb{P}(\mathbf{A} \cap \mathbf{B})$$

Conditional Probability

- Suppose B is an event such that P(B)$\neq$0. The Conditional probability of event A given event B is

$$\mathbb{P}(\mathbf{A}|\mathbf{B}) = \frac{\mathbb{P}(\mathbf{A} \cap \mathbf{B})}{\mathbb{P}(\mathbf{B})} = \frac{\mathbb{P}(\mathbf{A}, \mathbf{B})}{\mathbb{P}(\mathbf{B})}$$

  - "Chain Rule": Definition of conditional probability immediately implies that:

$$\mathbb{P}(\mathbf{A}, \mathbf{B}) = \mathbb{P}(\mathbf{A}|\mathbf{B})\mathbb{P}(\mathbf{B}) = \mathbb{P}(\mathbf{B}|\mathbf{A})\mathbb{P}(\mathbf{A})$$

  Conditional probability can be written as:

$$\mathbb{P}(\mathbf{A}|\mathbf{B}) = \frac{\mathbb{P}(\mathbf{A}, \mathbf{B})}{\mathbb{P}(\mathbf{B})} = \frac{\mathbb{P}(\mathbf{B}|\mathbf{A})}{\mathbb{P}(\mathbf{B})}\mathbb{P}(\mathbf{A})$$

  - $\mathbb{P}(\cdot|\mathbf{B})$ is a probability measure, i.e. it satisfies the axioms of probability

Bayes Rule

- Let $\mathbf{A_1}, \mathbf{A_2}, \ldots$ be nonzero probability events that partition Ω, and let B be a nonzero probability event, we know for all i,

$$\mathbb{P}(\mathbf{A_i}|\mathbf{B}) = \frac{\mathbb{P}(\mathbf{B}|\mathbf{A_i})}{\mathbb{P}(\mathbf{B})}\mathbb{P}(\mathbf{A_i})$$

Law of Total probability implies

$$\mathbb{P}(\mathbf{B}) = \sum_{\mathbf{i}} \mathbb{P}(\mathbf{A_i}, \mathbf{B}) = \sum_{\mathbf{i}} \mathbb{P}(\mathbf{B}|\mathbf{A_i})\mathbb{P}(\mathbf{A_i})$$

Substituting we obtain:

$$\mathbb{P}(\mathbf{A_i}|\mathbf{B}) = \frac{\mathbb{P}(\mathbf{B}|\mathbf{A_i})}{\sum_{\mathbf{i}} \mathbb{P}(\mathbf{B}|\mathbf{A_i})\mathbb{P}(\mathbf{A_i})}\mathbb{P}(\mathbf{A_i})$$

- This formula is known as "Bayes Rule".

Bayes Rule Example

- Dynamic State Space Model
  - Assume the unknown states $\mathbf{X_k}$, and observations $\mathbf{Z_k}$, at time k
  - In practice, we model this problem as

$$\mathbf{X_k} = \mathbf{f}(\mathbf{X_{k-1}}) + \mathbf{U_k}$$

$$\mathbf{Z_k} = \mathbf{h}(\mathbf{X_k}) + \mathbf{W_k}$$

Where. $\mathbf{U_k}, \mathbf{W_k}$ are Gaussian error.

GOAL: We need to obtain $\mathbb{P}(\mathbf{X_k}|\mathbf{Z_k})$ using transition equation (predict) and measurement equation (update)

Using Bayes Rule:

$$\mathbb{P}(\mathbf{X_k}|\mathbf{Z_k}) = \frac{\mathbb{P}(\mathbf{Z_k}|\mathbf{X_k})\mathbb{P}(\mathbf{X_k})}{\mathbb{P}(\mathbf{Z_k})}$$

- There is an efficient way to recursively implement this. This is known as "Kalman Filter".