ECE541: Stochastic Signals and Systems	Fall 2019
Lecture 1 - September 5, 2019	
Prof. Salim El Rouayheb	*Scribe: Peiwen Tian, Lu Liu, Ghadir Ayache*

Chapter 1: Introduction to Probability Theory

1 Axioms of Probability

Definition 1 (Probability Space). *The* Probability space *is defined by a triplet* $(\Omega, \mathcal{F}, \mathcal{P})$*,*

where:

Ω *is the Sample space*

$\mathcal{F}$ *is the set of Events*

$\mathcal{P}$ *is the Probability function*

Definition 2 (Sample space). *The* Sample space*,* Ω*, is the set of all possible outcomes of a random experiment.*

Example 1. *When we toss a coin, all the possible outcomes are Heads or Tails. Therefore, the sample space of a coin tossing is* $\Omega = \{Head,\ Tail\}$*.*

Example 2. *When we toss a die, one of the 6 faces is going to come up. Therefore, the sample space of a die tossing is* $\Omega = \{1, 2, 3, 4, 5, 6\}$*.*

Example 3. *Suppose that we want to measure the temperature a Thursday afternoon in September. Then* $\Omega = \mathbb{R}$*.*

Definition 3 (Event). *An event* E *is a subset of the sample space, i.e.,* $E \subseteq \Omega$*.*

Example 4. ***Coin Tossing****: The Event of getting a Head is* $E = \{H\}$

Example 5. ***Die Tossing****: The Event of getting a "multiple of 3" is* $E = \{3, 6\}$

Example 6. ***Temperature measurement****: The Event of getting a temperature between* 70^oF *and* 90^oF *is* $E = [70, 90]$

Example 7. *If we toss a fair coin twice, then the sample space is* $\Omega = \{HH,\ HT,\ TH,\ TT\}$*. Consider the event* A *"at least one Head occurs"; then, the event is* $A = \{HH,\ HT,\ TH\}$*.*

Let B *be the event of tossing the coin repeatedly until a Head occurs. Then,* $B = \{H,\ TH,\ TTH, \dots\}$*. Let* C *be the event of tossing the coin an even number of times until a Head occurs. Then,* $C = \{TH,\ TTTH, \dots\}$*.*

Definition 4 ($\mathcal{F}$)**.** *$\mathcal{F}$ is the set of all events.*

Typically when Ω is countable, $\mathcal{F}$ is the set of all subsets of Ω, i.e., the power set of Ω denoted by 2^{Ω}. But this is not the case for Ω uncountable, where $\mathcal{F}$ is going to be too large and there will often be sets to which it will be impossible to assign a unique measure like in $\Omega = \mathbb{R}$. (Check Definition 5. and Remark 2.)

Remark 1. *$\mathcal{F}$ must be a $\sigma - algebra$ such that*

1. *$\Omega \in \mathcal{F}$,*
2. *If $A \in \mathcal{F}$, then its complement set $A^C \in \mathcal{F}$,*
3. *if $A_i \in \mathcal{F}\, for\, all\, i = 1, 2\ldots, then\, \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$*

Corollary. *From De Morgan's Law and the previous property we get that if $A_i \in \mathcal{F}\, for\, all\, i = 1, 2\ldots, then\, \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$*

Example 8. ***Temperature measurement*** *How to prove that $(a,\, b)\, is\, an\, Event\, when\, [a,\, b] \in \mathcal{F}$?*

If $E = [a,\, b] \in \mathcal{F},\, then\, a\, \&\, b \in \mathcal{F}. Then\, [a, b] \bigcap \{\bar{a}\} \bigcap \left\{\bar{b}\right\} \in \mathcal{F}\, which\, is\, (a,\, b)$.

Definition 5 (Borel Set)**.** *Borel sets are the sets that can be constructed from open or closed sets by repeatedly taking countable unions and intersections. Formally, Borel algebra is the smallest $\sigma - algebra$ that makes all open sets measurable.*

Remark 2. *Not all subsets of Ω are events. You can define sets that have no probability. In such a case, we have to use the smallest $\sigma - algebra$ called Borel algebra that contains all closed intervals*

For this class, any subset of Ω is an event.

Definition 6 (Axioms of probability)**.** *A probability measure P on Ω is a function*

$$\begin{aligned} P :& \mathcal{F} \to [0,\ 1], \\ & E \to P(E), \end{aligned}$$

such that it satisfies the following properties:

(1) $P(\emptyset) = 0$.

(2) $P(\Omega) = 1$.

(3) If $A_1,\ A_2,\ A_3 \ldots$ are disjoint subsets of Ω,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_i P(A_i).$$

Lemma 1. *Let A and B be two subsets of Ω. We define $\bar{A}$ to be the complement of A in Ω, we have:*

(a) $P(\bar{A}) = 1 - P(A)$.

(b) If $A \subseteq B$, then $P(A) \leq P(B)$.

(c) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof. For part(a),

$$\begin{aligned} &P(A \cup \bar{A}) = P(\Omega) = 1 \text{ and } A,\, \bar{A} \text{ are disjoint} \\ &\Rightarrow P(A) + P(\bar{A}) = 1 \\ &\Rightarrow P(\bar{A}) = 1 - p(A) \end{aligned}$$

For part(b),

$$\begin{aligned} B &= A \cup (B \setminus A) \\ \Rightarrow P(B) &= P(A) + P(B \setminus A) \\ &\geq P(A) \end{aligned}$$

For part(c),

$$P(A \cup B) = P(A) + P(B \setminus A) = P(A) + P(B \setminus A \cap B).$$

Now,

$$(A \cap B) \subseteq B \Rightarrow P(B \setminus A \cap B) = P(B) - P(A \cap B).$$

□

Lemma 2 (Union bound)**.** *Let A and B be two subsets of Ω, then*

$$P(A \cup B) \leq P(A) + P(B).$$

In general,

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i).$$

Example 9 (Tossing a Die (a))**.** *A_1: The result number is a multiple of 2. A_2: The result number is a multiple of 3.*

$A_1 = \{2, 4, 6\}$, $P(A_1) = \frac{1}{2}$. $A_2 = \{3, 6\}$, $P(A_2) = \frac{1}{3}$.

$$P(A_1 \cup A_2) \leq P(A_1) + P(A_2) = \frac{1}{2} + \frac{1}{3} = \frac{5}{6}$$

In fact, $A_1 \cup A_2 = \{2, 3, 4, 6\}$ and $P(A_1 \cup A_2) = \frac{2}{3}$.

Example 10 (Tossing a Die (b))**.** *A_1: The result number is greater than or equal to 3. A_2: The result number is prime.*

$A_1 = \{3, 4, 5, 6\}$, $P(A_1) = \frac{2}{3}$. $A_2 = \{2, 3, 5\}$, $P(A_2) = \frac{1}{2}$.

$$P(A_1 \cup A_2) \leq P(A_1) + P(A_2) = \frac{2}{3} + \frac{1}{2} = \frac{7}{6} > 1$$

In fact, $A_1 \cup A_2 = \{2, 3, 4, 5, 6\}$ and $P(A_1 \cup A_2) = \frac{5}{6}$.

1.1 Conditional Probability

Example 11. *Consider the experiment of tossing two fair dice. Let A be the event that their total sum is greater than 6.*

(a) Find $P(A)$. The set of all events Ω is given by the following set:

$$\Omega = \left\{ \begin{array}{cccc} (1,1), & (1,2), & \dots & (1,6), \\ (2,1), & (2,2), & \dots & (2,6), \\ \vdots & \vdots & \ddots & \vdots \\ (6,1), & (6,2), & \dots & (6,6). \end{array} \right\}.$$

Now, we need to find $P(A)$. All the possible outcomes of A (total exceeds 6) are:

$$\begin{aligned} A = \{&(1,6) \\ &(2,5),\ (2,6) \\ &(3,4),\ (3,5),\ (3,6) \\ &(4,3),\ (4,4),\ (4,5),\ (4,6) \\ &(5,2),\ (5,3),\ (5,4),\ (5,5),\ (5,6) \\ &(6,1),\ (6,2),\ (6,3),\ (6,4),\ (6,5),\ (6,6)\}. \end{aligned}$$

Therefore

$$P(A) = \sum_{e \in A} P(e) \overset{fair\,dice}{=} \frac{21}{36}$$

(b) Let B the event that the first dice is 3. Find $P(B)$.

All the possible outcomes of event B are:

$$B = \{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\}. \tag{1}$$

Then all the possible outcomes of event A given B are the events in equation (1) satisfying A (total exceeds 6), hence

$$(A \cap B) = \{(3,4),(3,5),(3,6)\}.$$

(c) What is the probability of "Total exceeds 6 given that the first dice is 3"

$$P(A|B) = \frac{3}{6},$$

we can find that by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Definition 7 (Conditional probability)**.** *We define the conditional probability of an event A given that event B happened (with $P(B) > 0$) by:*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Definition 8 (Independent events)**.** *Two events A and B are independent iff*

$$P(A \cap B) = P(A)P(B).$$

In general,

$$\begin{aligned} P(A \cap B) &= P(A)P(B|A) \quad (2) \\ &= P(B)P(A|B). \quad (3) \end{aligned}$$

We can also say that the events A and B are independent iff

$$\begin{aligned} P(A|B) &= P(A), \quad (P(B) \neq 0) \\ P(B|A) &= P(B), \quad (P(A) \neq 0). \end{aligned}$$

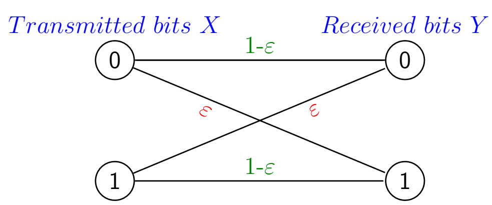


Figure 1: Binary Symmetric Channel (BSC) with probability of error $P_e = \varepsilon$.

Example 12 (Binary symmetric channel)**.**
In the BSC of Fig. 8 the bits are flipped with probability ε (ε is called crossover probability), we can write

$$\begin{aligned} \varepsilon &= P(Y = 0|X = 1) \\ &= P(Y = 1|X = 0). \end{aligned}$$

Suppose the bits '0' and '1' are equal likely to be sent, i.e.,

$$P(X = 0) = P(X = 1) = 0.5,$$

Q. *Find the probability of sending a '0' and receiving a '0'.*

Ans.

$$\begin{aligned} P(X = 0, Y = 0) &= P(X = 0)P(Y = 0|X = 0) \\ &= 0.5(1 - \varepsilon). \end{aligned}$$

Example 13 (Random Graphs)**.** *Consider the graph $\mathcal{G} = (V, E)$ over 4 vertices, given in Figure 1, where $V = \{1, 2, 3, 4\}$ is the vertex set and $E = \{\{1, 2\}, \{2, 3\}, \{3, 4\}\}$ is the edge set.*

A random graph $\mathcal{G}$ defined over the vertex set V is a graph where an edge between any two vertices exists with a probability p. If we take a graph on n vertices and the edge exists between 2 vertices

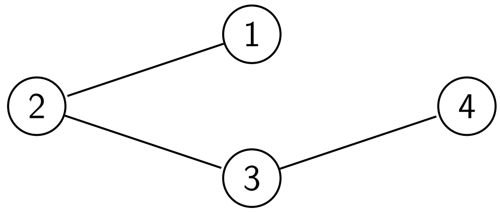
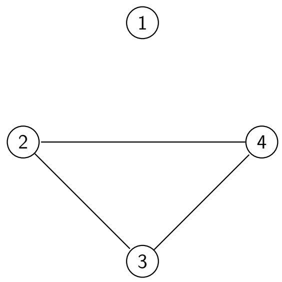
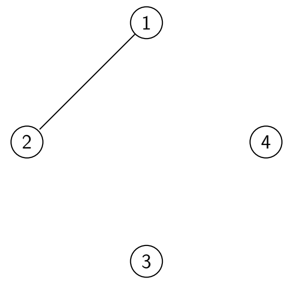


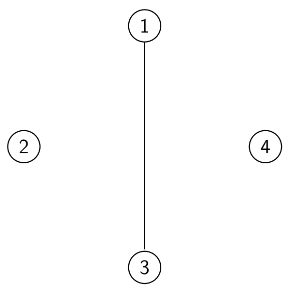
Figure 2: Graph connection.



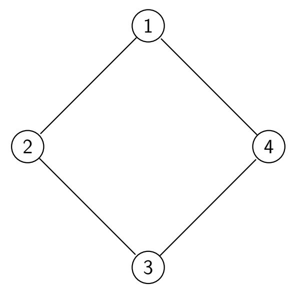
(a) vertices 2,3,4 are connected to each other



(b) vertices 1,2 are connected to each other



(c) vertices 1,3 are connected to each other



(d) vertices 1,2,3,4 connected to each other

Figure 3: Graph connection for 4 vertices

with probability=p=0.5. Then the number of subsets of V *of size* $2 = \dfrac{n(n-1)}{2} = \binom{n}{2}$. *The number of subsets of* V *of size* $k = \binom{n}{k}$.

If we have 4 vertices in a graph. What is the probability that vertex 1 is connected to k other nodes?

Let N *be the neighbors of vertex 1,* $N = \phi$ *in fig(a),* $N = \{1,2\}$ *in fig(b),* $N = \{1,3\}$ *in fig(c),* $N = \{2,3,4\}$ *in fig(d).*
Then we define the event A_N *is that the vertex 1 is connected to the vertices in* N

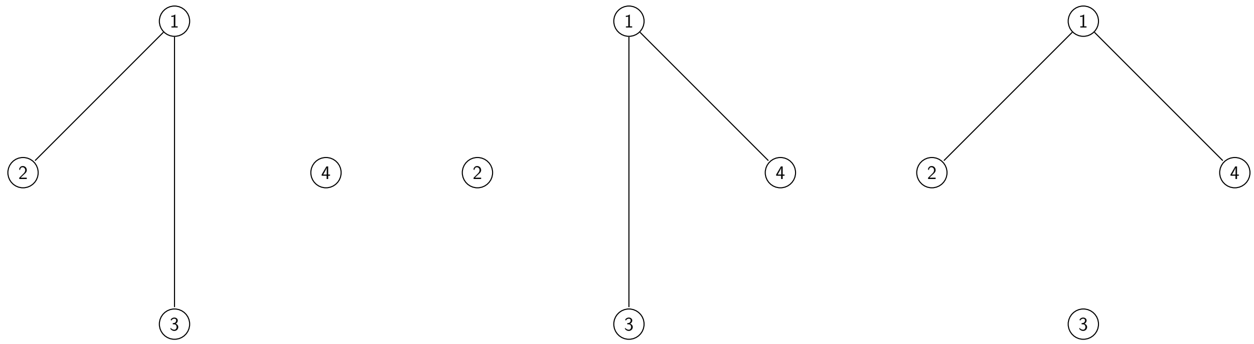
We say vertex 1 is connected to k other vertices, if k =2, all the possible graph are as (Figure 3).

Define event A vertex 1 is connected to 2 other vertices,therefore:

$$A = A_{\{2,3\}} \cup A_{\{3,4\}} \cup A_{\{2,4\}}.$$

The probability of this event A is

$$P(A) = P(A_{\{2,3\}}) + P(A_{\{3,4\}}) + P(A_{\{2,4\}}).$$



(a): Node 1 connected to 2 and 3 (b): Node 1 connected to 3 and 4. (c): Node 1 connected to 2 and 4.

Figure 4: vertex 1 is connected to two vertices

The probability of vertex 1 is connected to vertex 2 and 3 is

$$P(A_{\{2,3\}}) = (\frac{1}{2})^3 = p^2(1-p) = P(A_{\{3,4\}}) = P(A_{\{2,4\}}),$$

therefore,

$$P(A) \quad = \quad 3p^2(1-p).$$

In general, the probability vertex 1 *is connected to* k *specific vertices is*

$$P(A_N) \quad = \quad p^k(1-p)^{n-1-k}.$$

The probability vertex 1 *is connected to* k *other vertices is*

$$\begin{aligned} P(A) &= \Sigma P(A_N), \\ &= \binom{n-1}{k} p^k(1-p)^{n-1-k}. \end{aligned}$$

1.2 Total Law of Probability

Theorem 1. *Let* $A_1, A_2, \ldots, A_n$ *be* n *mutually disjoint events such that*

$$\Omega \quad = \quad \bigcup_{i=1}^{n} A_i \ (P(A_i) \neq 0), \tag{4}$$

then for any event $B \subseteq \Omega$ *we have*

$$P(B) \quad = \quad P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + ... + P(A_n)P(B|A_n).$$

Proof. For n=2

$$B \quad = \quad (B \cap A_1) \cup (B \cap A_2), \tag{5}$$

$$P(B) \quad = \quad P(B \cap A_1) + P(B \cap A_2), \tag{6}$$

$$= \quad P(A_1)P(B|A_1) + P(A_2)P(B|A_2). \tag{7}$$

□

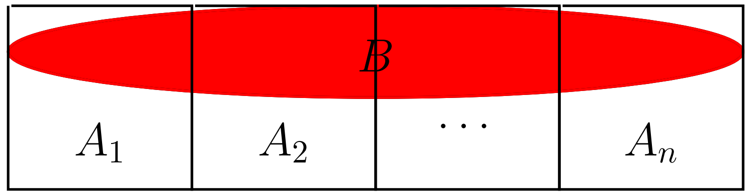


Figure 5: Total law of probability.

Example 14. *(BSC) Consider a BSC in Fig. 6 with crossover probability $\varepsilon = 0.1$. The probability of sending '0' is 0.4 and the probability of sending '1' is 0.6.*

Q. *Find the probability of receiving a '0'.*

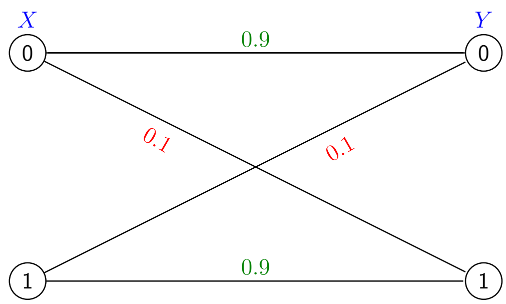


Figure 6: Binary Symmetric Channel with probability of error $P_e = 0.1$.

Ans. *The probability of sending '1' is $P(X = 1) = 0.6$, and the probability of sending '0' is $P(X = 0) = 0.4$. Then if we want to know the probability of receiving '0', we can use the total law of probability to calculate $P(Y = 0)$,*

$$\begin{aligned} P(Y=0) &= P(X=0)P(Y=0|X=0) + P(X=1)P(Y=0|X=1), \\ &= (0.4) \times (0.9) + (0.6) \times (0.1) = 0.42. \end{aligned}$$

1.3 Birthday paradox

Question: What is the probability that at least 2 students in class have the same birthday.
E:at least 2 students have the same birthday.
Number of days per year is n, number of students in class is m.
$\bar{E}$: each student has distinct birthday.
Answer:

$$P(\bar{E}) = 1 \times (1 - \frac{1}{n}) \times (1 - \frac{2}{n}) \times \cdots \times (1 - \frac{m-1}{n}).$$

We know that

$$1 - \frac{k}{n} \approx e^{-\frac{k}{n}}, \quad k \ll n.$$

Then,

$$
\begin{aligned}
P(\bar{E}) &= e^{-\frac{1}{n}} \times e^{-\frac{2}{n}} \times \cdots \times e^{-\frac{m-1}{n}}, \\
&= exp(-\frac{1}{n}(1+2+\cdots+m-1)), \\
&= e^{-\frac{m(m-1)}{2n}}, \\
&\approx e^{-\frac{m^2}{2n}}.
\end{aligned}
$$

Now we have student $m = 50$, and number of birthdays $n = 365$.

$$
\begin{aligned}
P(E) &\approx 1 - e^{-\frac{50^2}{2 \times 365}}, \\
&\approx 96.7\%.
\end{aligned}
$$

Question: How big the class should be if the probability of 2 students have same birthday is larger than 50%?
Answer:

$$
P(E) = \frac{1}{2}.
$$

Then

$$
1 - e^{-\frac{m^2}{2n}} = \frac{1}{2},
$$

so

$$
\frac{m^2}{2n} = \ln 2,
$$

$$
\begin{aligned}
m &= \sqrt{2 \ln 2} \times \sqrt{n}, \\
&\approx 23.
\end{aligned}
$$

So we need approximately 23 students in same class to make the probability that at least 2 students have the same birthday is larger than $\frac{1}{2}$.

Theorem 2 (Baye's Theorem)**.**

$$
P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}. \tag{8}
$$

Example 15 (BSC)**.** *In this case we have* $P(X = 0) = P(X = 1) = \frac{1}{2}$ *(0s and 1s are equal likely transmitted)*
Suppose we observe $Y = 1$*. What value of* X *should we decode?*

$$
P(X = 1|Y = 1) = \frac{P(X = 1, Y = 1)}{P(Y = 1)}.
$$

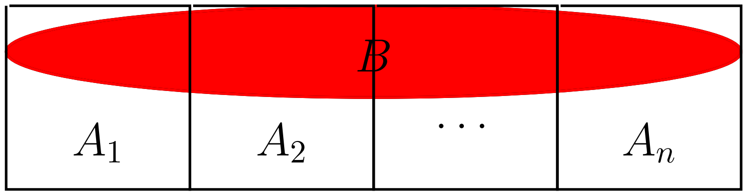


Figure 7: Baye's theorem.

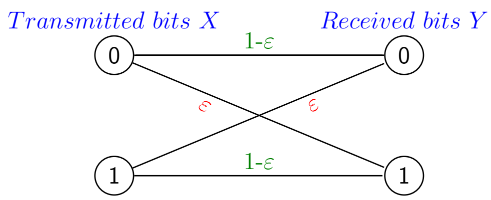


Figure 8: Binary Symmetric Channel (BSC) with probability of error $P_e = \epsilon$.

According to the Baye's theorem

$$
\begin{aligned}
P(X=1|Y=1) &= \frac{P(X=1)P(Y=1|X=1)}{P(X=0)P(Y=1|X=0)+P(X=1)P(Y=1|X=1)}, \\
&= \frac{0.5(1-\varepsilon)}{0.5\varepsilon+0.5(1-\varepsilon)}, \\
&= 1-\varepsilon.
\end{aligned}
$$