CSCE-658 Randomized Algorithms

Lecture #3, January 28, 2016

Lecturer: Professor Jianer Chen

3 Basic probability theory

We give brief introduction to Probability Theory.

Definition 3.1 A *probability space* is a triple $(\Omega, \mathcal{F}, \Pr)$, where
1. Ω is a *sample set*, whose elements can be called *outcomes*;
2. $\mathcal{F}$ is a set of *events*, where each event is a subset of Ω; and
3. Pr is a function from $\mathcal{F}$ to real numbers. Pr is called a *probability measure*.

The event set $\mathcal{F}$ must satisfy the following conditions:
- 2.a the sample space Ω is an event;
- 2.b for an event E, the complement $E^C = \Omega \setminus E$ of E is also an event; and
- 2.c for finite or countable many events E_1, E_2, ... in $\mathcal{F}$, the union $\bigcup_{i\geq 1} E_i$ is also an event.

The probability measure Pr must satisfy the following conditions:
- 3.a for all events E in $\mathcal{F}$, $0 \leq \Pr[E] \leq 1$;
- 3.b $\Pr[\Omega] = 1$; and
- 3.c for finite or countable mutually disjoint events E_1, E_2, ..., $\Pr[\bigcup_{i\geq 1} E_i] = \sum_{ii\geq 1} \Pr[E_i]$.

Remark 1. From 2.a, we know that $\mathcal{F}$ is not empty. From 2.b and 2.c, we derive that $\mathcal{F}$ is a *σ-algebra*, i.e., it is closed under complement, countable union, and countable intersection.

Remark 2. If Ω is finite or countable, we can simply let $\mathcal{F}$ be the power set 2^Ω of Ω, i.e, $\mathcal{F}$ consists of all subsets of Ω. In this case, the probability $\Pr[E]$ of an event E can be defined via the probabilities of the elements included in E, i.e., $\Pr[E] = \sum_{a\in E} \Pr[a]$ (note here we have used Rule 3.c). Such a probability space is called a *discrete probability space*, for which many results become more intuitive with easier proofs. Most of our studies are based on discrete probability space.

Remark 3. The sample set Ω can be uncountable. A typical example is that Ω is the set of all points in the unit circle in the plan. In this case, not all subsets of ω can be events of $\mathcal{F}$. For example, research in Set Theory has shown that there are point sets in the unit circle in the plan that are *unmeasurable*. Moreover, it may become impossible to compute the probability $\Pr[E]$ of an event E based on the probabilities of the elements in E: if E is uncountable, how do we add uncountable many real numbers?

The following theorem gives a few easily verified facts about probability measures.

Theorem 3.1 *The following are true:*
(a) *For any event E, $\Pr[E^C] = 1 - \Pr[E]$;*
(b) *For two events E_1 and E_2 such that $E_1 \subseteq E_2$, $\Pr[E_1] \leq \Pr[E_2]$;*
(c) *For any two events E_1 and E_2, $\Pr[E_1 \cup E_2] + \Pr[E_1 \cap E_2] = \Pr[E_1] + \Pr[E_2]$;*
(d) *For any two events E_1 and E_2, $\Pr[E_1 \cup E_2] \leq \Pr[E_1] + \Pr[E_2]$.*

PROOF. The proofs for (a) and (b) are trivial. We now give a proof for (c). For the events E_1 and E_2, we can decompose $E_1 \cup E_2$ into a union of disjoint subsets:

$$E_1 \cup E_2 = (E_1 \cap E_2) \cup (E_1 \setminus E_2) \cup (E_2 \setminus E_1).$$

Thus (by Rule 3.c),

$$\Pr[E_1 \cup E_2] = \Pr[E_1 \cap E_2] + \Pr[E_1 \setminus E_2] + \Pr[E_2 \setminus E_1]. \tag{5}$$

Now since E_1 can be decomposed into a union of disjoint subsets as $E_1 = (E_1 \cap E_2) \cup (E_1 \setminus E_2)$, we have $\Pr[E_1 \setminus E_2] = \Pr[E_1] - \Pr[E_1 \cap E_2]$. Similarly, $\Pr[E_2 \setminus E_1] = \Pr[E_2] - \Pr[E_1 \cap E_2]$. Bringing these two equalities in (5) gives (c). Finally, (d) follows directly from (c). □

Note that by a simple induction based on (d), we can easily derive that $\Pr[\bigcup_{i=1}^{n} E_i] \leq \sum_{i=1}^{n} \Pr[E_i]$ holds true for any finite number of events $E_1, \ldots, E_n$.

The following equality will be very useful in our probability analysis, which is called the *Principle of Inclusion-Exclusion*.

Theorem 3.2 (Principle of Inclusion-Exclusion) *For any number n of events $E_1, \ldots, E_n$, we have*

$$\begin{aligned}
\Pr\left[\bigcup_{i=1}^{n} E_i\right] &= \sum_{i=1}^{n} \Pr[E_i] - \sum_{i<j} \Pr[E_i \cap E_j] + \sum_{i<j<k} \Pr[E_i \cap E_j \cap E_k] \\
&\quad + \cdots + (-1)^{t+1} \sum_{1 \leq k_1 < \cdots < k_t \leq n} \Pr\left[\bigcap_{h=1}^{t} E_{k_h}\right] + \cdots + (-1)^{n+1} \Pr[E_1 \cap \cdots \cap E_n] \\
&= \sum_{t=1}^{n} (-1)^{t+1} \sum_{1 \leq k_1 < \cdots < k_t \leq n} \Pr\left[\bigcap_{h=1}^{t} E_{k_h}\right] \qquad (6)
\end{aligned}$$

PROOF. Theorem 3.1(c) proves Theorem 3.2 for the case $n = 2$:

$$\Pr[E_1 \cup E_2] = (\Pr[E_1] + \Pr[E_2]) - \Pr[E_1 \cap E_2].$$

We use induction to prove the theorem for general n. Let $n \geq 3$. By the theorem for case $n = 2$, we get

$$\begin{aligned}
\Pr\left[\bigcup_{i=1}^{n} E_i\right] &= \Pr\left[\left(\bigcup_{i=1}^{n-1} E_i\right) \cup E_n\right] = \Pr\left[\left(\bigcup_{i=1}^{n-1} E_i\right)\right] + \Pr[E_n] - \Pr\left[\left(\bigcup_{i=1}^{n-1} E_i\right) \cap E_n\right] \\
&= \Pr\left[\bigcup_{i=1}^{n-1} E_i\right] + \Pr[E_n] - \Pr\left[\bigcup_{i=1}^{n-1} (E_i \cap E_n)\right] \qquad (7)
\end{aligned}$$

By applying induction on $n - 1$, we get

$$\Pr\left[\bigcup_{i=1}^{n-1} E_i\right] = \sum_{t=1}^{n-1} (-1)^{t+1} \sum_{1 \leq k_1 < \cdots < k_t \leq n-1} \Pr\left[\bigcap_{h=1}^{t} E_{k_h}\right] \qquad (8)$$

and

$$\begin{aligned}
\Pr\left[\bigcup_{i=1}^{n-1} (E_i \cap E_n)\right] &= \sum_{t=1}^{n-1} (-1)^{t+1} \sum_{1 \leq k_1 < \cdots < k_t \leq n-1} \Pr\left[\bigcap_{h=1}^{t} (E_{k_h} \cap E_n)\right] \\
&= \sum_{t=1}^{n-1} (-1)^{t+1} \sum_{1 \leq k_1 < \cdots < k_t \leq n-1} \Pr\left[E_n \cap \bigcap_{h=1}^{t} E_{k_h}\right]
\end{aligned}$$

Note that the first term (with $t = 1$) $\sum_{i=1}^{n-1} \Pr[E_i]$ in (8) plus the term $\Pr[E_n]$ in (7) gives exactly $\sum_{i=1}^{n} \Pr[E_i]$, which is the first term (with $t = 1$) in (6). Therefore, in order to prove the theorem, we only need to prove

$$\begin{aligned}
&\sum_{t=2}^{n} (-1)^{t+1} \sum_{1 \leq k_1 < \cdots < k_t \leq n} \Pr\left[\bigcap_{h=1}^{t} E_{k_h}\right] \qquad (9) \\
= &\sum_{t=2}^{n-1} (-1)^{t+1} \sum_{1 \leq k_1 < \cdots < k_t \leq n-1} \Pr\left[\bigcap_{h=1}^{t} E_{k_h}\right] - \sum_{t=1}^{n-1} (-1)^{t+1} \sum_{1 \leq k_1 < \cdots < k_t \leq n-1} \Pr\left[E_n \cap \bigcap_{h=1}^{t} E_{k_h}\right]
\end{aligned}$$

In fact, we have

$$\begin{aligned}
&\sum_{t=2}^{n-1}(-1)^{t+1}\sum_{1\le k_1<\cdots<k_t\le n-1}\Pr\left[\bigcap_{h=1}^{t}E_{k_h}\right] \;-\; \sum_{t=1}^{n-1}(-1)^{t+1}\sum_{1\le k_1<\cdots<k_t\le n-1}\Pr\left[E_n\cap\bigcap_{h=1}^{t}E_{k_h}\right]\\
=\;&\sum_{t=2}^{n-1}(-1)^{t+1}\sum_{1\le k_1<\cdots<k_t\le n-1}\Pr\left[\bigcap_{h=1}^{t}E_{k_h}\right] \;-\; \sum_{t=2}^{n}(-1)^{t}\sum_{1\le k_1<\cdots<k_{t-1}\le n-1}\Pr\left[E_n\cap\bigcap_{h=1}^{t-1}E_{k_h}\right]\\
=\;&\sum_{t=2}^{n-1}(-1)^{t+1}\sum_{1\le k_1<\cdots<k_t\le n-1}\Pr\left[\bigcap_{h=1}^{t}E_{k_h}\right] \;+\; \sum_{t=2}^{n}(-1)^{t+1}\sum_{1\le k_1<\cdots<k_{t-1}\le n-1}\Pr\left[E_n\cap\bigcap_{h=1}^{t-1}E_{k_h}\right]\\
=\;&\sum_{t=2}^{n-1}(-1)^{t+1}\left(\sum_{1\le k_1<\cdots<k_t\le n-1}\Pr\left[\bigcap_{h=1}^{t}E_{k_h}\right] \;+\; \sum_{1\le k_1<\cdots<k_{t-1}\le n-1}\Pr\left[E_n\cap\bigcap_{h=1}^{t-1}E_{k_h}\right]\right)\\
&+\;(-1)^{n+1}\sum_{1\le k_1<\cdots<k_{n-1}\le n-1}\Pr\left[E_n\cap\bigcap_{h=1}^{n-1}E_{k_h}\right]
\end{aligned}$$

It is easy to verify that

$$\sum_{1\le k_1<\cdots<k_t\le n-1}\Pr\left[\bigcap_{h=1}^{t}E_{k_h}\right] + \sum_{1\le k_1<\cdots<k_{t-1}\le n-1}\Pr\left[E_n\cap\bigcap_{h=1}^{t}E_{k_h}\right] = \sum_{1\le k_1<\cdots<k_t\le n}\Pr\left[\bigcap_{h=1}^{t}E_{k_h}\right]$$

and

$$\sum_{1\le k_1<\cdots<k_{n-1}\le n-1}\Pr\left[E_n\cap\bigcap_{h=1}^{n-1}E_{k_h}\right] = \sum_{1\le k_1<\cdots<k_n\le n}\Pr\left[\bigcap_{h=1}^{n}E_{k_h}\right]$$

This proves (9), thus proves the theorem. □

The Principle of Inclusion-Exclusion looks quite tedious, but is easy to remember: the probability of a union of a set of events is equal to the sum of the probabilities of the single events, minus the sum of the probabilities of all possible intersections of two events, plus the sum of the probabilities of all possible intersections of three events, and so on.

Now we introduce conditional probability.

Definition 3.2 Let E and F be two events, where $\Pr[F]\neq 0$. The *conditional probability of E given F* is defined as

$$\Pr[E|F] = \Pr[E\cap F]/\Pr[F].$$

There are also some easy verified facts for conditional probability.

Lemma 3.3 *Let E_1, ..., E_n be a partition of the sample space Ω, i.e, E_1, ..., E_n are pairwise disjoint and $E_1\cup\cdots\cup E_n=\Omega$. Then for any event F,*

$$\Pr[F] = \sum_{i=1}^{n}\Pr[F|E_i]\cdot\Pr[E_i].$$

PROOF. Since $\Pr[F|E_i]\cdot\Pr[E_i] = (\Pr[F\cap E_i]/\Pr[E_i])\Pr[E_i] = \Pr[F\cap E_i]$, the right side of the above equation becomes

$$\sum_{i=1}^{n}\Pr[F|E_i]\cdot\Pr[E_i] = \sum_{i=1}^{n}\Pr[F\cap E_i] = \Pr\left[\bigcup_{i=1}^{n}(F\cap E_i)\right] = \Pr\left[F\cap\bigcup_{i=1}^{n}E_i\right] = \Pr[F\cap\Omega] = \Pr[F],$$

where we have used the disjointness of the events $F\cap E_1$, ..., $F\cap E_n$ (because of the disjointness of the events E_1, ..., E_n), and the fact that the union of E_1, ..., E_n is Ω. □

Finally, we give the definition of a very important concept.

Definition 3.3 Two events E and F are *independent* if $\Pr[E \cap F] = \Pr[E] \cdot \Pr[F]$, or equivalently, if $\Pr[E \mid F] = \Pr[E]$, which is also equivalent to $\Pr[F \mid E] = \Pr[F]$.

By reading the well known textbooks in probability theory, some of them are listed in the reference [2, 5, 6, 9, 10], students can gain more information about the basic concepts and facts of probability theory introduced in this section.