CS206 **Review Sheet 1** **September 8, 2018**

1. **Set Theory Review** We recall some concepts (and notations) from elementary set theory that will be used throughout the course.

   (a) $x \in A$ means "x is an element of the set A"; $x \notin A$ means x is not an element of A.

   (b) relations between sets: The notation

      i. $A \subseteq B$ means that if $x \in A$ then also $x \in B$ (we say "A is a subset of B").

      ii. $A \supseteq B$ means $B \subseteq A$ ("A is a superset of B").

      iii. $A \subset B$ means $x \in A \Rightarrow x \in B$ and $\exists y \in B : y \notin A$ (proper subset).

      iv. $A \supset B$ means $B \subset A$ (proper superset)..

      v. $A = B$ means $A \subseteq B$ and $B \subseteq A$.

   (c) operations on sets: (From now on we assume A and B are both subsets of a given set S (the "universe"); i.e., $A, B \subseteq S$)

      i. $A^c \equiv \{x \in S : x \notin A\}$ (complement). [$\equiv$ means "equal, by definition"].

      ii. $\phi = S^c$ (the empty set).

      iii. $A \cap B = \{x \in S : x \in A \text{ and } x \in B\}$ (intersection).

      iv. $A \cup B = \{x \in S : x \in A \text{ or } x \in B \text{ (or both)}\}$ (union).

      v. $A \setminus B = \{x \in S : x \in A \text{ and } x \notin B\}$, ($= A \cap B^c$, the complement of B in A).

   (d) set identities

      i. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

      ii. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
      (these are the distributive laws for union and intersection)

      iii. $(A \cap B)^c = (A^c) \cup (B^c)$

      iv. $(A \cup B)^c = (A^c) \cap (B^c)$ [prove i.-iv. using 1.(b) v.]

      v. $(\bigcup_{i=1}^n A_i)^c = \bigcap_{i=1}^n (A_i^c)$ and $(\bigcap_{i=1}^n A_i)^c = \bigcup_{i=1}^n (A_i^c)$
      ((iii) and (iv) are known as de Morgan's laws. Try to prove (v) using induction on the deMorgan laws).

2. **Probability Theory - basic ingredients**

   (a) Random Experiment $\mathcal{E}$: an *idealized* or *conceptual* experiment that could be repeated infinitely often, always under identical conditions.

   (b) Sample Space S: the set of possible outcomes (elementary events) of a random experiment.
   (many examples of experiments and their sample spaces were given in class)

   (c) An event A is a subset of the sample space S of an experiment $\mathcal{E}$.

   If $\mathcal{E}$ is performed and the outcome $x \in S$ is observed we say "x occurs". In addition if $x \in A$ we say "A occurs" and if $x \notin A$ we say "A does not occur".

   Probability theory describes certain sets using the "language of events". Some of this terminology is shown in the following table.

Notation	Name in Set Theory (or equivalent expression)	Name in Probability
A^c	complement	not A
$A = S$	universe	certain event
$A = \phi$	empty	impossible event
$A \cap B$	intersection	both
$\bigcap_{i=1}^n A_i$	intersection	all
$(A^c) \cap (B^c)$	$(A \cup B)^c$	neither
$\bigcap_{i=1}^n (A_i^c)$	$(\bigcup_{i=1}^n A_i)^c$	none
$A \cup B$	union $(A^c \cap B^c)^c$	at least 1 (not neither)
$\bigcup_{i=1}^n A_i$	union $(\bigcap_{i=1}^n A_i^c)^c$	at least 1 (not none)
$(A^c) \cup (B^c)$	$(A \cap B)^c$	not both
$\bigcup_{i=1}^n (A_i^c)$	$(\bigcap_{i=1}^n A_i)^c$	not all
$A \cap B = \phi$	disjoint	mutually exclusive
$A \subseteq B$	inclusion	$A \Rightarrow B$

As an exercise draw a Venn diagram of a sample space S with three events A, B, C and locate those outcomes where (i) exactly one event occurs; (ii) exactly two occur; (iii) all occur; (iv) only A occurs; (v) only B and C occur.

(d) Probability measure P: a real-valued, non-negative function on events in S which satisfies the following two axioms

i. $P(S) = 1$

ii. $P(A \cup B) = P(A) + P(B)$ whenever $A \cap B = \phi$ (additivity)

The pair (S, P) is called the probability space of $\mathcal{E}$.

(e) Facts about P

i. P does not decrease: $P(A) \leq P(B)$ whenever $A \subseteq B$ (implies that $P(A) \leq 1$, all A; for a proof, take $B = S$ and use Axiom i.).

ii. $P(\bigcup_{i=1}^n A_i) = P(A_1) + \cdots + P(A_n)$ if the A_i are mutually exclusive; i.e., $A_i \cap A_j = \phi$ if $i \neq j$ (we say P is finitely additive). The proof is by induction, using Axiom ii (from (d), above).

iii. $P(A^c) = 1 - P(A)$.

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

(f) A third axiom about P. We also assume that P is countably additive; i.e., that

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

when the A_i are mutually exclusive.

i. (Fact v. about P:) To assign a non-negative function to the subsets of sample space S so that the three axioms hold, it suffices to assign a value $P(w) \geq 0$ to *each outcome* $w \in S$ in such a way that

$$\sum_{w \in S} P(w) = 1.$$

In this way

$$\boxed{P(A)} = P(\bigcup_{w \in A} w) = \boxed{\sum_{w \in A} P(w)}$$

(by Axiom iii, if A is infinite, or by finite additivity otherwise), and the that Axioms i and ii hold is now easily checked.

3. **Conditional Probability** An experiment $\mathcal{E}$ with probability space (S, P) is performed. The outcome was seen by an observer who gives you the HINT: event "A has occurred". You want to revise the probability measure P on S to account for this new information. A natural way is given by the following definition.

   If $P(A) > 0$, P_A, the conditional probability measure given A, is defined by

$$P_A(w) = \begin{cases} 0 & \text{if } w \notin A \\ P(w)/P(A) & \text{if } w \in A \end{cases}$$

   Some consequences of this definition are

   (a) P_A is a probability measure [i.e., it satisfies Axioms 1-3] and points in A have the same relative probability with P_A as they did with P.

   (b) $P_A(B) = P(A \cap B)/P(A)$.
   This equation is known as *the conditional probability formula.*

   (c) Let $H_1, \ldots, H_n$ be events that partition S; i.e., they are mutually exclusive ($H_i \cap H_j = \phi$ when $i \neq j$) and exhaustive (i.e., $S = \cup_{i=1}^n H_i$). We call them *hypotheses.* Then

$$P(A) = \sum_{i=1}^{n} P(H_i) P_{H_i}(A).$$

   (d) (Bayes Rule) Again $H_1, \ldots, H_n$ are hypotheses that partition the sample space S. Then for each i,

$$P_A(H_i) = \frac{P_{H_i}(A) P(H_i)}{\sum_{j=1}^{n} P(H_j) P_{H_j}(A)}.$$

   Think of the equation as telling you how to revise the original assessment of the probability of a hypothisis (H_i) after you have some evidence from the experiment, namely that the event A has occurred.

4. **Independence:** Events A and B (both with positive probability) are said to be *independent* iff

$$P_A(B) = P(B).$$

   This means that the probability of B given the information that A has occurred (the left-hand side) is the original probability of B, so A gives no new information about B's probability. Using the conditional probability formula (3b, above) for the left-hand side we see that $P(B) =$

$P(A \cap B)/P(A)$, and multiplying both sides of this equation by $P(A)$ we get the *product law* for independent events:

$$P(A \cap B) = P(A)P(B). \tag{1}$$

This is sometimes taken as the *definition* of independence. If you divide this equation by $P(B)$ you see that if A and B are independent then also $P_B(A) = P(A)$, which means that in addition, B gives no information about A. Independence is a very important aspect of Probability Theory.

When there are more than two events of interest the situation gets (much) more complicated.

(a) A family $A_1, \ldots, A_n$ of n events is pairwise independent if each pair is. Thus for each $i \neq j$ $P_{A_i}(A_j) = P(A_j)$, or equivalently, $P(A_i \cap A_j) = P(A_i)P(A_j)$.

Take $S = \{1, 2, 3, 4\}$ as the sample space and use equally likely probability. Let $A_1 = \{1, 2\}$, $A_2 = \{1, 3\}$, and $A_3 = \{1, 4\}$. This family is pairwise independent. But $P_{A_1 \cap A_2}(A_3) = 1 \neq P(A_3)$. Thus, although no single event gives information about the probability of any other, two, together, do give information. It is clear then, that we need a stronger notion of independence to distinguish possible relations between events when we are considering more than two of them.

(b) Take an integer $k \in \{2, \ldots, n\}$. A family $A_1, \ldots, A_n$ of n events is k-wise independent if every group of k of them satisfies the product law analogous to (1). Thus for any choice of pointers $i_1, \ldots, i_k$, $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, we require that

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}). \tag{2}$$

(c) The family $A_1, \ldots, A_n$ is mutually independent if it is k-wise independent for every k, $2 \leq k \leq n$.

(d) Here are some interesting examples that illustrate:

- The example in (a) is pairwise independent (k-wise with $k = 2$) but not 3-wise.
- Let $S = \{1, \ldots, 8\}$ under equally likely probability and $A_1 = \{1, 2, 3, 5\}$, $A_2 = \{1, 2, 4, 6\}$, $A_3 = \{1, 3, 4, 7\}$, and $A_4 = \{2, 3, 4, 8\}$. This family is k-wise independent for $k < 4$ but not mutually independent.
- Using the same probability space as above, let $A_1 = \{1, 2, 3, 4\}$, $A_2 = \{1, 2, 5, 6\}$, and $A_3 = \{1, 3, 7, 8\}$. These are 3-wise independent but not pairwise independent.