Assignment 1 - CSC/DSC 262/462 - FALL 2017

All questions are worth equal marks. Note that graduate students only are required to submit question **Q5**.

Q1: The Inclusion-Exclusion Principle. If events $A_1, \ldots, A_n$ are mutually exclusive, the probability of the union is

$$P(\cup_{i=1}^{n} A_i) = P(A_1) + \ldots + P(A_n).$$

If they are not mutually exclusive, then calculation of the probability of their union can become quite complex. For two events, we have

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 A_2).$$

This can be extended to n events using the *inclusion-exclusion identity*

$$\begin{aligned} P(\cup_{i=1}^{n} A_i) \quad = \quad & \sum_i P(A_i) \\ & - \sum_{i<j} P(A_i A_j) \\ & + \sum_{i<j<k} P(A_i A_j A_k) \\ & \vdots \\ & -1^{n+1} P(A_1 A_2 \ldots A_n). \end{aligned} \tag{1}$$

(a) Write explicitly the *inclusion-exclusion identity* for $n = 3$.

(b) A dice game has the following rules. Three dice are tossed, and the three values from 1 to 6 are noted. Points are awarded for the following combinations:

(i) The sum of the values is at least 16.
(ii) The three values are the same.
(iii) The three values are each even numbers.

A dice toss may match more than one of these combinations, and would receive additional points accordingly. If the dice toss does not match any of these combinations, no points are awarded. Use the inclusion-exclusion principle to determine the probability that no points are awarded for a given toss.

HINT: A number of examples of this type of problem, with solutions, can be found in Practice Problem Set 1.

Q2: A lunch counter has 16 stools in a single row. Currently, 5 stools are occupied by customers. You notice that no two customers sit next to each other. Assume that customers select stools at random (that is, all selections of 5 stools are equally likely to be occupied).

(a) What is the probability that no two adjacent stools are occupied?

(b) What is the probability that all 5 customers sit in adjacent stools (that is, there are no empty stools between any two customers)?

HINT: A selection of 5 from 16 stools can be represented as an arrangement of 5 `O`'s and 11 `E`'s, such as `EOEEOEEOEEOEOEEE` or `EEEOOEOEEOEEOEEE`. In the second example, but not the first, there are two customers sitting next to each other. Then, an arrangement can be constructed in the following way. Make 'slots' on each side of the 11 `E`'s:

_E_E_E_E_E_E_E_E_E_E_E_

There are 12 slots (they should also be placed at the ends of the arrangement). Then, the `O`'s can be placed in the slots according to some relevant rule.

Q3: A standard 52 card playing deck assigns a unique combination of 13 ranks (2,3,4,5,6,7,8,9,10,J,Q,K,A) and 4 suits (Clubs, Diamonds, Hearts, Spades) to each card ($13 \times 4 = 52$). Suppose 10 cards are selected at random. Derive the probability that each of the following events occurs.

(a) Only 1 suit is represented.

(b) Exactly 2 suits are represented in equal number.

(c) No rank is represented more than once.

Carefully list the *tasks* used in the application of the *rule of product*.

HINT: A number of examples of this type of problem, with solutions, can be found in Practice Problem Set 1.

Q4: In genetics, a genotype consists of two genes, each of which is one of (possibly) several types of alleles. Suppose we consider only two alleles, `r` and `R`. Furthermore, suppose the allele `r` exists in a population with frequency q. Under Hardy-Weinberg equilibrium, genotypes are essentially random samples of 2 alleles, one sampled from each parent. Since the genes are usually not ordered (because we don't know which is maternal and which is paternal), the probability of each possible genotype is

$$\begin{aligned} P(\mathtt{rr}) &= q^2, \\ P(\mathtt{rR}) &= 2(1-q)q, \\ P(\mathtt{RR}) &= (1-q)^2. \end{aligned}$$

We then note that genotypes of unrelated individuals are independent, but genotypes of related individuals are not independent. For example, when two organisms mate, each passes one allele, selected at random, to the offspring, forming that offspring's genotype (this predicts Mendels' Law of Inheritance). If G_o is an offspring genotype then

$$\begin{aligned} P(G_o = \mathtt{rr}) &= q^2, \text{ but} \\ P(G_o = \mathtt{rr} \mid \text{ both parents have genotype } \mathtt{rR}) &= 1/4. \end{aligned}$$

Now, suppose `r` is a *recessive* allele, meaning that it only determines a trait when the genotype is `rr`. Such a trait is a *recessive trait*. Typically, a genetic disease is a recessive trait, and `r` is a *rare allele*, meaning q is very small.

Next, suppose we may determine without error whether or not an individual possesses a recessive trait (and therefore has genotype `rr`). Let G_m, G_f, G_1, G_2 be the genotypes of a mother, father and two offspring (ie. siblings). Define events

$$\begin{aligned} R_m &= \{ \text{ Mother has recessive trait } \}, \\ R_f &= \{ \text{ Father has recessive trait } \}, \\ R_1 &= \{ \text{ Offspring 1 has recessive trait } \}, \\ R_2 &= \{ \text{ Offspring 2 has recessive trait } \}. \end{aligned}$$

(a) Determine the conditional probability $P(R_1 \mid R_f^c \cap R_m^c)$, the probability that an offspring possesses the recessive trait given that neither parent does.

(b) Determine the conditional probability $P(R_1 \mid R_2 \cap R_f^c \cap R_m^c)$, the probability that an offspring possesses the recessive trait given that a sibling does, and that neither parent does.

(c) Does the probability defined in Part (b) depend on q? Give a brief explanation for this.

HINT: A number of examples of this type of problem, with solutions, can be found in Practice Problem Set 1.

Q5: [For Graduate Students] The formal axioms of probability can be stated as follows. Let the set S be the sample space. Let $\mathcal{F}$ be some collection of subsets of S. Not all subsets of S need be in $\mathcal{F}$, but S and $\emptyset$ must be. We assign a probability $P(E)$ to each $E \in \mathcal{F}$. The three axioms are as follows:

Axiom 1. For any $E \in \mathcal{F}$, $P(E) \geq 0$.

Axiom 2. $P(S) = 1$.

Axiom 3. Suppose $E_1, E_2, \ldots, E_i, \ldots$ is a countable collection of mutually exclusive sets from $\mathcal{F}$. Suppose that the union $\cup_{i=1}^{\infty} E_i$ is also in $\mathcal{F}$. Then

$$P\left\{\cup_{i=1}^{\infty} E_i\right\} = \sum_{i=1}^{\infty} P(E_i).$$

Axiom 3 is referred to as *countable additivity.* For most probability models, we assume that if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$ as well. Furthermore if the collection of sets E_i are in $\mathcal{F}$, then so is any union or intersection of any combination of these sets. We will make these assumptions below.

(a) Suppose in the following statements A,B are in $\mathcal{F}$. Prove that Axioms 1-3 imply each of the following statements.

   (i) $P(\emptyset) = 0$.
   (ii) $A \cap B = \emptyset$ implies $P(A \cup B) = P(A) + P(B)$.
   (iii) $P(A^c) = 1 - P(A)$.
   (iv) $A \subset B$ implies $P(A) \leq P(B)$.

   HINT: The set $\emptyset$ is disjoint to all other sets, including $\emptyset$ itself. This means $S = S \cup \{\cup_{i=1}^{\infty} \emptyset\}$. Also note that countable additivity and finite additivity (*ie* statement (ii) above) are distinct statements.

(b) Let $E_1, E_2, \ldots$ be a countable collection of sets in $\mathcal{F}$. It is sometimes useful to construct an associated collection of sets, denoted

$$\begin{aligned} \bar{E}_1 &= E_1, \\ \bar{E}_i &= E_i \cap E_{i-1}^c \cap \ldots \cap E_1^c, \;\; i \geq 2. \end{aligned}$$

   Verify the following properties of $\bar{E}_i$:

   (i) $\bar{E}_i \subset E_i$ for all $i \geq 1$.
   (ii) The sets $\bar{E}_1, \bar{E}_2, \ldots$ are mutually exclusive.
   (iii) $\cup_{i=1}^{\infty} E_i = \cup_{i=1}^{\infty} \bar{E}_i$.

(c) Prove **Boole's Inequality**: Let $E_1, E_2, \ldots$ be a countable collection of sets. Then $P\left\{\cup_{i=1}^{\infty} E_i\right\} \leq \sum_{i=1}^{\infty} P(E_i)$.

(d) Suppose world records for a given sport are compiled annually. For convenience, label the years $i = 1, 2, \ldots$ with $i = 1$ being the first year records are kept. Define the events

$$E_i = \{ \text{ World record broken in year } i \ \}, \;\; i \geq 1.$$

   Then, let Q_i be the probability that from year i onwards, the current world record is never broken. Prove that if $\sum_{i=1}^{\infty} P(E_i) < \infty$, then $\lim_{i \to \infty} Q_i = 1$. Verify that, in particular, that if $P(E_i) \leq c/i^k$ for some finite constants $c > 0$ and $k > 1$ then $\lim_{i \to \infty} Q_i = 1$.