Quantitative Information Flow (QIF) **DCC/ICEx/UFMG**
Prof. Mário S. Alvim **2021/1**

SOLUTION OF PROBLEM SET

Discrete Probability
(Based on slide-set)

Necessary reading for this assignment:

- *Slide-set of Lecture 00.B - Discrete Probability*:
  - *An Introduction to Discrete Probability*
  - *Probability Theory*
  - *Bayes' Theorem*
  - *Expected Value and Variance*

Review questions.

1. (Rosen Review Question 7-2)

   (a) What conditions should be met by the probabilities assigned to the outcomes from a finite sample space?

   Instructor's solution: The probability distribution $p : S \to \mathbb{R}$ on a sample space S must satisfy:

   i) $0 \leq p(s) \leq 1$, for all $s \in S$, and
   ii) $\sum_{s \in S} p(s) = 1$.

   (b) What probabilities should be assigned to the outcome of heads and the outcome of tails if heads comes up three times as often as tails?

   Instructor's solution: The sample space is $S = \{H, T\}$, where H represents heads and T represents tails.
   We are told that $p(H) = 3p(T)$. Since it must be the case that $p(H) + p(T) = 1$, we deduce that $3p(T) + p(T) = 1$, and hence $p(T) = 1/4$. From that we derive that $p(H) = 3/4$.

2. (Rosen Review Question 7-3)

   (a) Define the conditional probability of an event E given an event F.

   Instructor's solution: The probability of event E given event F is given by $p(E \mid F) = |E \cap F|/|F|$.

   (b) Suppose E is the event that when a die is rolled it comes up an even number, and F is the event that when a die is rolled it comes up 1, 2, or 3. What is the probability of F given E?

Instructor's solution: The sample space S is the set of all outcomes of the roll of a die, i.e., $S = \{1, 2, 3, 4, 5, 6\}$, and hence $|S| = 6$.
We are interested in $p(F \mid E) = |E \cap F|/|E|$.
The event $E \cap F$ is the event of the die coming up both an even number and a number in $\{1, 2, 3\}$, hence $E \cap F = \{2\}$ and $|E \cap F| = 1$.
The event E is the set $E = \{2, 4, 6\}$, hence $|E| = 3$.
Then we can calculate that $p(F \mid E) = |E \cap F|/|E| = 1/3$.

3. (Rosen Review Question 7-4)

   (a) When are two events E and F independent?

   Instructor's solution: Two events E and F in a sample space S are independent when $p(E) = p(E \mid F)$.
   Equivalently, we can say they are independet when $p(E \cap F) = p(E)p(F)$.

   (b) Suppose E is the event that an even number appears when a fair die is rolled, and F is the event that a 5 or 6 comes up. Are E and F independent?

   Instructor's solution: The sample space S is the set of all outcomes of the roll of a die, i.e., $S = \{1, 2, 3, 4, 5, 6\}$, and hence $|S| = 6$.
   The event E is the set $E = \{2, 4, 6\}$, hence $p(E) = |E|/|S| = 3/6 = 1/2$.
   The event F is the set $F = \{5, 6\}$, hence $p(F) = |F|/|S| = 2/6 = 1/3$.
   The event $E \cap F$ is the set $\{6\}$, hence $p(E \cap F) = |E \cap F|/|S| = 1/6 = 1/6$.
   We can verify that $p(E)p(F) = 1/2 \cdot 1/3 = 1/6$ is equal to $p(E \cap F) = 1/6$, hence E and F are independent events.

4. (Rosen Review Question 7-5)

   (a) What is a random variable?

   Instructor's solution: A random variable is a function $X : S \to \mathbb{R}$ from the sample space S of an experiment to the reals.
   A random variable maps each possible result of an experiment to a real number.

   (b) What are the possible values assigned by the random variable X that assigns to a roll of two dice the larger number that appears on the two dice?

   Instructor's solution: The possible values of X are $1, 2, 3, 4, 5, 6$.

5. (Rosen Review Question 7-6)

   (a) Define the expected value of a random variable X.

   Instructor's solution: The expected value of a random variable X in a sample space S is

   $$E(X) = \sum_{s \in S} p(s)X(s) = \sum_{r \in X(S)} p(X = r) \cdot r.$$

   (b) What is the expected value of the random variable X that assigns to a roll of two dice the larger number that appears on the two dice?

Instructor's solution: The sample space of rolling two dice can be represented by the following table, where the entry (i, j) is the result where the first die (d_1) comes up i and the second die (d_2) comes up j. Each entry (i, j) in the table contains the value of the random variable $X((i, j)) = \max(i, j)$.

d_1/d_2	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	2	3	4	5	6
3	3	3	3	4	5	6
4	4	4	4	4	5	6
5	5	5	5	5	5	6
6	6	6	6	6	6	6

Because each of the 36 outcomes is equally likely, $p((i, j)) = 1/36$ for every pair in the table, and the expectation of X can be calculated as

$$
\begin{aligned}
E(XY) =& \sum_{i=1}^{6} \sum_{j=1}^{6} p((i, j)) \cdot \max(i, j) \\
=& \frac{1}{36}(1 + 3 \cdot 2 + 5 \cdot 3 + 7 \cdot 4 + 9 \cdot 5 + 11 \cdot 6) \\
=& \frac{161}{36} \\
\approx& 4.47
\end{aligned}
$$

6. (Rosen Review Question 7-8)

   (a) What is meant by a "*Bernoulli trial*"?

   Instructor's solution: A Bernoulli trial is an expirement with two possible outcomes (often generally referred to as "success" and "failure").

   (b) What is the probability of k successes in n independent Bernoulli trials?

   Instructor's solution: If p is the probability of success in one Bernoulli trial, the probability of k successes in n independent Bernoulli trials is $C(n, k)p^k(1 - p)^{n-k}$.

   (c) What is the expected value of the number of successes in n independent Bernoulli trials?

   Instructor's solution: The expected value of the number of successes in n independent Bernoulli trials is np, where p is the probability of success in one Bernoulli trial.

7. (Rosen Review Question 7-9)

   (a) What does the linearity of expectations of random variables mean?

   Instructor's solution: The linearity of expecations of random variables means that expecatation is a linear function, in the sense that: (i) if X_i are random variables $(1 \leq i \leq n)$ then $E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$; and (ii) if X is a random variable and a and b are real numbers, then $E(aX + b) = aE(X) + b$.

8. (Rosen Review Question 7-11) State Bayes' theorem and use it to find $p(F \mid E)$ if $p(E \mid F) = 1/3$, $p(E \mid \overline{F}) = 1/4$, and $p(F) = 2/3$, where E and F are events from a sample space S.

Instructor's solution: Bayes' Theorem says that if E and F are events in a sample space S, then

$$p(F \mid E) = \frac{p(F)p(E \mid F)}{p(E)} = \frac{p(F)p(E \mid F)}{p(F)p(E \mid F) + p(\overline{F})p(E \mid \overline{F})}.$$

Noting that $p(\overline{F}) = 1 - p(F) = 1/3$, we can substitute the given values in the above equation to obtain

$$\begin{aligned} p(F \mid E) &= \frac{p(F)p(E \mid F)}{(p(F)p(E \mid F) + p(\overline{F})p(E \mid \overline{F}))} \\ &= \frac{2/3 \cdot 1/3}{2/3 \cdot 1/3 + 1/3 \cdot 1/4} \\ &= \frac{8}{11}. \end{aligned}$$

9. (Rosen Review Question 7-13)

   (a) What is the variance of a random variable?

   Instructor's solution: The variance of a random variable is a measure of how much, in average, its values spread around the expected value. The variance of a random variable X in a sample space S is given by $V(X) = \sum_{s \in S} p(s)[X(s) - E(X)]^2 = E(X^2) - [E(X)]^2$.

Exercises.

10. (Rosen 7.2-11) Suppose that E and F are events such that $p(E) = 0.7$ and $p(F) = 0.5$. Show that $p(E \cup F) \geq 0.7$ and $p(E \cap F) \geq 0.2$.

    Instructor's solution: First, note that since $E \subseteq E \cup F$, we have that $p(E) \leq p(E \cup F)$, and hence $p(E \cup F) \geq 0.7$.

    Second, note that $p(E \cup F) = p(E) + p(F) - p(E \cap F)$, and we can write

$$\begin{aligned} p(E \cap F) =& p(E) + p(F) - p(E \cup F) \\ =& 0.7 + 0.5 - p(E \cup F) \\ =& 1.2 - p(E \cup F) \\ \geq& 1.2 - 1 && \text{(since } p(E \cup F) \leq 1) \\ =& 0.2. \end{aligned}$$

11. (Rosen 7.2-17) If E and F are independent events, prove or disprove that $\overline{E}$ and F are necessarily independent events.

    Instructor's solution: We will prove that if E and F are independent events, $\overline{E}$ and F are necessarily independent events.

    We want to show that $p(\overline{E} \cap F) = p(\overline{E})p(F)$, under the assumption that E and F are independent.

$$\begin{aligned} p(\overline{E} \cap F) =& p(F)p(\overline{E} \mid F) && \text{(by def. of conditional probability)} \\ =& p(F)(1 - p(E \mid F)) && \text{(since } p(E \mid F) + p(\overline{E} \mid F) = 1) \\ =& p(F)(1 - p(E)) && \text{(since } E \text{ and } F \text{ are independent, } p(E \mid F) = p(E)) \\ =& p(F)p(\overline{E}) && (p(\overline{E}) = 1 - p(E)) \end{aligned}$$

12. (Rosen 7.2-25) What is the conditional probability that a randomly generated bit string of length four contains at least two consecutive 0s, given that the first bit is a 1? (Assume the probabilities of a 0 and a 1 are the same.)

    Instructor's solution: The sample space S is the set of all binary strings of size 4, that is, $|S| = 2^4 = 16$.

    Let E be the event that the string starts with a 1, and F be the event that the string has two consecutive zeros. We are interested in calculating $p(F \mid E) = |E \cap F|/|E|$.

    Let E be the event that the string starts with a 1, hence $|E| = 2^3 = 8$ strings (all possibilities for the last 3 bits, since the first one is fixed).

    The event $E \cap F$ is the event that a string starts with a 1 and has two consecutive 0s. It is easy to see that $E \cap F = \{1000, 1001, 1100\}$.

    Hence we can calculate $p(F \mid E) = |E \cap F|/|E| = 3/8$.

13. (Rosen 7.2-27(b)) Let E and F be the events that a family of 4 children has children of both sexes and has at most one boy, respectively. Are E and F independent?

    Instructor's solution: E and F are independent iff $p(E \cap F) = p(E)p(F)$.

    The sample space S is the set of all possible ways a family can have 4 kids according to sex, hence $|S| = 2^4 = 16$.

    Let B represent a boy and G represent a girl.

    Let E be the event that the family has children of both sexes. It means that $E = S - \{BBBB, GGGG\}$, hence $|E| = 16 - 2 = 14$ and $p(E) = 14/16 = 7/8$.

    Let F be the event that the family has at most one boy. It means that $F = \{GGGG, BGGG, GBGG, GGBG, GGGB\}$, hence $|F| = 5$ and $p(F) = 5/16$.

    The event $E \cap F$ is that in which the family has at most one boy, and children of both sexes, hence $E \cap F = \{BGGG, GBGG, GGBG, GGGB\}$, leading to $|E \cap F| = 4$ and $p(E \cap F) = 4/16 = 1/4$

    Now we can verify that $p(E)p(F) = 7/8 \cdot 5/16 = 70/256$, which is different from $p(E \cap F) = 1/4$, and hence E and F are not independent.

14. (Rosen 7.3-9) Suppose that 8% of the patients tested in a clinic are infected with HIV. Furthermore, suppose that when a blood test for HIV is given, 98% of the patients infected with HIV test positive and that 3% of the patients not infected with HIV test positive. What is the probability that

    (a) a patient testing positive for HIV with this test is infected with it?

        Instructor's solution: First let us formally model the problem, and calculate all the values needed to apply Bayes' theorem.
        Let I be the event that a patient is infected, and $\overline{I}$ be the event a patient is not infected.
        The prior probability is that $p(I) = 0.08$, hence $p(\overline{I}) = 0.92$.
        Let P be the event that the HIV test comes up positive, and $\overline{P}$ the event that HIV test comes up negative.
        We know that $p(P \mid I) = 0.98$, and hence $p(\overline{P} \mid I) = 0.02$. Also, we know that $p(P \mid \overline{I}) = 0.03$, and hence $p(\overline{P} \mid \overline{I}) = 0.97$.

Now we calculate the probability of a patient testing positive and negative for HIV.

$$\begin{aligned} p(P) =& p(I)p(P \mid I) + p(\overline{I})p(P \mid \overline{I}) = 0.008 \cdot 0.98 + 0.92 \cdot 0.03 = 0.106 \\ p(\overline{P}) =& p(I)p(\overline{P} \mid I) + p(\overline{I})p(\overline{P} \mid \overline{I}) = 0.008 \cdot 0.02 + 0.92 \cdot 0.97 = 0.894 \end{aligned}$$

Now we are ready to apply Bayes' theorem to calculate the a patient testing positive for HIV when this patient is actually infected with the virus as follows.

$$p(I \mid P) = \frac{p(I)p(P \mid I)}{p(P)} = \frac{0.08 \cdot 0.98}{0.106} \approx 0.740$$

(b) a patient testing positive for HIV with this test is not infected with it?

Instructor's solution: Using the results from the previous item, we can calculate the following.

$$p(\overline{I} \mid P) = \frac{p(\overline{I})p(P \mid \overline{I})}{p(P)} = \frac{0.92 \cdot 0.03}{0.106} \approx 0.260$$

(c) a patient testing negative for HIV with this test is infected with it?

Instructor's solution: Using the results from the previous item, we can calculate the following.

$$p(I \mid \overline{P}) = \frac{p(I)p(\overline{P} \mid I)}{p(\overline{P})} = \frac{0.08 \cdot 0.02}{0.894} \approx 0.002$$

(d) a patient testing negative for HIV with this test is not infected with it?

Instructor's solution: Using the results from the previous item, we can calculate the following.

$$p(\overline{I} \mid \overline{P}) = \frac{p(\overline{I})p(\overline{P} \mid \overline{I})}{p(\overline{P})} = \frac{0.92 \cdot 0.97}{0.894} \approx 0.998$$