Notation: In the following, $P(A)$ denotes the probability of the event or proposition $A$. A random variable $X$ is said to be discrete if its range is finite, or infinite but countable. In this assignment you may assume that all random variables are discrete. The probability mass function (pmf) of a discrete random variable $X$ is the function $p(x) = P(X = x)$. Later in the course we will use this notation quite liberally, and assume that which random variable $p$ refers to is inferred from its argument. For example, we might talk about $p(x)$ and $p(y)$, and understand that while these look like the same function, they are really two different pmf’s referring to two different random variables, $X$ and $Y$. However, in this first assignment, except for question 10, we are going to be pedantically explicit, and always write out probabilities as $P(X = x)$, etc..

1. Let $A$ and $B$ be two propositions, and let $A$ and $B$ be the corresponding events. Recall that the event corresponding to $A \land B$ (read “$A$ and $B$”) is $A \setminus B$, the event corresponding to $A \lor B$ (read “$A$ or $B$”) is $A \cup B$, and the event corresponding to $\neg A$ (read “not $A$”) is $\overline{A}$, where $\overline{A}$ is the complement of $A$. Use De Morgan’s laws to show that
   
   \begin{align*}
   (a) \quad P(A \cap B) &= 1 - P(\overline{A} \cup \overline{B}) \\
   (b) \quad P(A \cup B) &= 1 - P(\overline{A} \cap \overline{B}).
   \end{align*}

2. Approximately one third of twins are identical twins and the remaining two thirds are fraternal. Identical twins are necessarily of the same sex, whereas fraternal twins might or might not be of the same sex (you can assume that in the case of fraternal twins each child is independently equally likely to be a boy or a girl). Your friend finds out that she is pregnant and the ultrasound tech tells her that he can see two boys on the ultrasound. What is the probability that they are identical twins?

3. The sensitivity of a medical test $T$ for some condition $C$ is the probability that given that the patient has condition $C$, the test $T$ will detect it. The specificity of $T$ is the probability that given that the patient does not have $C$, the test will not falsely indicate that he/she does.

Mammography is a widely used x-ray test for screening women for breast cancer. The sensitivity of screening mammography is about 90%, and its specificity is also about 90%. Some organizations recommend that women have their first mammogram at age 40. At that age, the probability of an average risk, previously unscreened, woman to have breast cancer is estimated to be about 1%.

Assume that Mrs. $P$ is a 40 year old woman who has just had her first mammogram, and the result came back positive. What is the probability that she has cancer?

Note: The sensitivity and specificity figures in this question are in the right ballpark, but only pertain to screening, not the follow-up tests that are normally used to establish a diagnosis. Moreover, in this context there is some controversy as to what constitutes “cancer”. This exercise is intended as an illustration, and does not do justice to the complexity of the statistical and medical issues related to screening.

4. The law of total probability states that if $\{A_1, A_2, \ldots, A_n\}$ are mutually disjoint events that together cover the sample space, then

$$P(B) = \sum_{i=1}^{n} P(A_i, B).$$

Show that in this case (assuming that $P(B) \neq 0$ and $P(A_i) \neq 0$ for any $i$), Bayes’ rule can be written in the form

$$P(A_i | B) = \frac{P(B | A_i) P(A_i)}{\sum_{j=1}^{n} P(B | A_j) P(A_j)}$$

for any $i \in \{1, 2, \ldots, n\}$.  

1
5. Recall that two discrete random variables $X_1$ and $X_2$ are said to be independent (denoted $X_1 \perp X_2$) if and only if
\[ P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1) \cdot P(X_2 = x_2) \]
for any $x_1$ in the domain of $X_1$ and any $x_2$ in the domain of $X_2$. Similarly, a sequence of $n$ random variables $X_1, X_2, \ldots, X_n$ are independent if and only if
\[ P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \cdot \ldots \cdot P(X_n = x_n) \]
for any $x_1, \ldots, x_n$.

(a) Show that if $X_1 \perp X_2$, then for any $x_2$, $P(X_2 = x_2|X_1 = x_1)$ is constant as a function of $x_1$, and is equal to the marginal probability $P(X_2 = x_2)$. You may assume that $P(X_1 = x_1) \neq 0$ for any $x_1$.

(b) Recall that any two random variables $X_1$ and $X_2$ together form a new random variable $(X_1, X_2)$, called the joint. For clarity, let $Z = (X_1, X_2)$. Show that if $X_1 \perp X_2$ and $X_3 \perp Z$, then \{X_1, X_2, X_3\} are independent.

(c) Given an example of three random variables $X_1, X_2, X_3$ that are pairwise independent (i.e., $X_1 \perp X_2$, $X_2 \perp X_3$, and $X_1 \perp X_3$), but not jointly independent.

6. Recall that given a (real valued) random variable $X$, the expected value of $X$ is
\[ \mathbb{E}(X) := \sum_x x P(X = x). \]
Show the linearity of expectation, i.e., that

(a) For any real valued random variable $X$ and any scalar $\alpha \in \mathbb{R}$, $\mathbb{E}(\alpha X) = \alpha \mathbb{E}(X)$.

(b) For any two real valued random variables $X$ and $Y$, $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.

7. Show that if the two real valued random variables $X$ and $Y$ are independent, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

8. Recall that given a (real valued) random variable $X$, the variance of $X$ is
\[ \text{Var}(X) := \mathbb{E}((X - \mathbb{E}(X))^2). \]
Show that

(a) For any scalar $\alpha \in \mathbb{R}$, $\text{Var}(\alpha X) = \alpha^2 \text{Var}(X)$.

(b) $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

9. Recall that given two real valued random variables $X$ and $Y$, their covariance is
\[ \text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))). \]

(a) Show that for any two scalars $\alpha, \beta \in \mathbb{R}$, $\text{Cov}(\alpha X, \beta Y) = \alpha \beta \text{Cov}(X, Y)$.

(b) Show that $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.

(c) Show that if $X$ and $Y$ are independent, then $\text{Cov}(X, Y) = 0$.

10. We now revert to the sloppy “p”-notation, otherwise this question just gets too messy. Thus, $p(x, y)$ will denote $P(X = x, Y = y)$; $p(x|y)$ will denote $P(X = x|Y = y)$, and so on. In this notation, by the definition of conditional probability, $p(x, y) = p(x|y) \cdot p(y)$, assuming that $p(y) \neq 0$.

(a) Show that for any three discrete random variables $X_1, X_2$ and $X_3$,
\[ p(x_1, x_2, x_3) = p(x_1|x_2, x_3) \cdot p(x_2|x_3) \cdot p(x_3). \]

You may assume that $p(x_1, x_2, x_3) \neq 0$ for any $(x_1, x_2, x_3)$.

(b) Show that for any $n$ discrete random variables $X_1, X_2, \ldots, X_n$,
\[ p(x_1, x_2, \ldots, x_n) = p(x_1|x_2, \ldots, x_n) \cdot p(x_2|x_3, \ldots, x_n) \cdot \ldots \cdot p(x_{n-1}|x_n) \cdot p(x_n). \]

You may assume that $p(x_1, x_2, \ldots, x_n) \neq 0$ for any $(x_1, x_2, \ldots, x_n)$. 