BST 675 — Fall 2010 — Dr. Charnigo

Solutions to Written Assignment 1

1. Let $x \in A \cap (B \cup C)$. Then $x \in A$ and $x \in B \cup C$. Since $x \in B \cup C$, either $x \in B$ or $x \in C$. In the first case, $x \in A \cap B$. In the second case, $x \in A \cap C$. Hence $x \in (A \cap B) \cup (A \cap C)$. This shows that $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$.

Let $x \in (A \cap B) \cup (A \cap C)$. Then either $x \in A \cap B$ or $x \in A \cap C$. In the first case, $x \in A$ and $x \in B$. In the second case, $x \in A$ and $x \in C$. Hence $x \in A$ and $x \in B \cup C$. So $x \in A \cap (B \cup C)$. This shows that $(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$.

2. Put $A_1 := A$, $A_2 := A^c$, and $A_j := \emptyset$ for $j \in \{3, 4, \ldots\}$. Since $A \cup A^c = S$ and $A \cap A^c = \emptyset$, we have that $\cup_{j=1}^{\infty} A_j = S$ and $A_i \cap A_j = \emptyset$ for $i \neq j$. As such, axiom 3 yields

$$P(S) = P(\cup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} P(A_j).$$

By axiom 2 $P(S) = 1$ and by result 4 $P(A_j) = P(\emptyset) = 0$ for $j \in \{3, 4, \ldots\}$. We are left with

$$1 = P(A_1) + P(A_2) = P(A) + P(A^c).$$

Since $P(A^c) \geq 0$ by axiom 1, we conclude that $P(A) \leq 1$.

Remark: Note that this argument proves results 5 and 6 simultaneously.

3. We have

$$P(A \cup B \cup C) = P(A \cup D) = P(A) + P(D) - P(A \cap D)$$

from result 8 for calculating probabilities. Exercise 1 showed that

$$A \cap D = A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$

so that invoking result 8 once again yields

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

A final invocation of result 8 yields

$$P(D) = P(B \cup C) = P(B) + P(C) - P(B \cap C).$$

Putting all of these pieces together, we find that

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

4. A good Venn diagram would depict $A \cup B \cup C$ as the union of seven mutually exclusive sets ($A \cap B \cap C$, $A \cap B \cap C^c$, $A \cap B^c \cap C$, $A \cap B^c \cap C^c$, $A^c \cap B \cap C$, $A^c \cap B \cap C^c$, $A^c \cap B^c \cap C$) and indicate that the formula from exercise 3 provides a net count of one for each of these sets. For example, the formula from exercise 3 counts $A \cap B \cap C$ in each of its seven terms, but four of the terms have positive signs and three have negative signs, while $A \cap B \cap C^c$ is counted in the first, second, and fourth terms, of which two have positive signs and one has a negative sign.

5. Answers will vary. One possibility replaces infant mortality, smoking, and low birthweight with lung cancer, drinking, and smoking respectively. Lung cancer has a strong positive association with smoking, smoking has a strong positive association with drinking, and (hence) lung cancer has a strong positive association with drinking. Yet, when we look at smokers only, there is little if any association between lung cancer and drinking.

6. With a cohort study you could estimate $P(A|B)$ as the fraction of smokers who developed lung cancer. Such an estimate would not make sense for a case-control study, in which the fraction of subjects with lung cancer would be fixed a priori. However, you could estimate $P(B|A)$ as the fraction of cases who smoked and $P(B|A^c)$ as the fraction of controls who smoked. With an estimate of $P(A)$ available, you could then apply Bayes' Theorem to estimate $P(A|B)$ via the formula

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)[1 - P(A)]}.$$

7. Let A denote the event that a person has HIV and B the event that a person tests positive for HIV. Applying Bayes' Theorem, we have

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + [1 - P(B^c|A^c)][1 - P(A)]}.$$

Substituting 0.007 for $P(A)$, 0.90 for $P(B|A)$, and $P/100$ for $P(B^c|A^c)$, we obtain

$$P(A|B) = \frac{0.0063}{0.0063 + [1 - P/100]0.993}.$$

If $P(A|B)$ is set to 0.5, then we can solve for P. We have $0.0063 + [1 - P/100]0.993 = 0.0126$, then $[1 - P/100].993 = 0.0063$, then $[1 - P/100] = 0.0063/0.993$, then $[P/100] = 1 - 0.0063/0.993$, and finally $P = 100 - 0.63/0.993 = 99.4$. The test must have almost perfect specificity for a positive predictive value as modest as 0.5!