Stat 512
Emerson, Fall 15
Homework #1
September 30, 2015

Written solutions to the homework problems are due on Wednesday October 7, 2015 at the beginning of class.

The homework problems are divided into "regular" and "more involved" problems. In order to facilitate multiple graders, you should hand in these categories of problems separately. That is, hand in one paper that contains only the "regular" problems, and another paper that contains only the "more involved" problems.
As noted on the syllabus and discussed during the first class, copying of homework solutions is not allowed and, when detected, will be investigated as an infraction of the academy integrity policy of the University of Washington. While it is permissible to discuss problems with other students, TAs, or the instructor in order to learn how to solve a problem, your written solutions must be prepared without directly referencing any notes or solutions derived from other students or sources found on the internet.

REGULAR PROBLEMS

1. Let A, B, and C be measurable events for some probability measure P on sample space Ω. Further suppose $P(A) > 0$, $P(B) > 0$, and $P(C) > 0$, with none of the pairs of events mutually exclusive.
   (a) Show $P(AB|C) = P(A|BC) \cdot P(B|C)$.
   (b) Show $P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$.
   (c) Show $P(ABC) = P(A|BC) \cdot P(B|C) \cdot P(C)$.
   (d) What general condition is sufficient for $P(ABC) = P(A|BC) \cdot P(B|AC) \cdot P(C|AB)$ to be true. Is that condition necessary?

2. Let A and B be independent measurable events for some probability measure P on sample space Ω.
   (a) Show A and B^c are independent measurable events.
   (b) Show A^c and B are independent measurable events.
   (c) Show A^c and B^c are independent measurable events.

3. Let $A_1, A_2, \ldots, A_n$ be measurable events for some probability measure P on sample space Ω.
   (a) Show $P\left(\cap_{i=1}^n A_i^c\right) = P\left(\left[\cup_{i=1}^n A_i\right]^c\right)$.
   (b) Show $P\left(\cup_{i=1}^n A_i^c\right) = P\left(\left[\cap_{i=1}^n A_i\right]^c\right)$.

4. Let A and B be measurable events for some probability measure P on sample space Ω.
   (a) Assume $P(A) > 0$ and $P(B) > 0$. Show that if A and B are mutually exclusive events, they are not independent.
   (b) Assume $P(A) > 0$ and $P(B) > 0$. Show that if A and B are independent events, they are not mutually exclusive.
   (c) Find necessary and sufficient conditions on A and B that they would be both mutually exclusive and independent.

5. Let A and B be measurable events for some probability measure P on sample space Ω. Further suppose $P(A) = 0.3$, $P(A \cup B) = 0.8$.
   (a) Find $P(B)$ if A and B are mutually exclusive events.
   (b) Find $P(B)$ if A and B are independent events.
   (c) Find $P(B)$ if $P(A|B) = 0.3$.
   (d) Can $P(A|B) = 0.5$? If so, find $P(B)$. If not, explain why not.

MORE INVOLVED PROBLEMS

6. Consider an experiment consisting of n independent rolls of a fair die. Let X_k be the random variable measuring the number showing on top of the die on the kth roll. Hence, the sample space for X_k is $\Omega_X = \{1, 2, 3, 4, 5, 6\}$, with $P_X(X_k = \omega) = \frac{1}{6}$ for all $\omega \in \Omega_X$ and for all $k = 1, \ldots, n$.

   Define $Y_k = \min(X_1, \ldots, X_k)$ and $W_k = \max(X_1, \ldots, X_k)$ for $k = 1, \ldots, n$.
   (a) Provide formulas for the cumulative distribution functions $F_{Xk}(x) = Pr(X_k \leq x)$, $F_{Yk}(y) = Pr(Y_k \leq y)$, and $F_{Wk}(w) = Pr(W_k \leq W)$.
   (b) Provide plots of F_{Xk}, F_{Yk}, and F_{Wk} for $k = 1$, 2, and 5.
   (c) Comment on how this problem might be indicative of the "multiple comparison problem" in which the smallest of multiple p values is used to define an association.

7. Let A, B, and C be measurable events for some probability measure P on sample space Ω. Further suppose $P(A) > 0$, $P(B) > 0$, and $P(C) > 0$, with none of the pairs of events mutually exclusive.
   Simpson's paradox states that it is possible to have

$$\begin{aligned} P(A \,|\, BC) &> P(A \,|\, B^c C) \\ P(A \,|\, BC^c) &> P(A \,|\, B^c C^c) \end{aligned}$$

   but

$$P(A|B) < P(A|B^c).$$

(a) Illustrate how Simpson's paradox might lead to erroneous conclusions when investigating smoking effects on lung cancer deaths across two different countries. Consider analyses that merely compare smokers to nonsmokers and finds that smoking is protective against lung cancer death versus an analysis that compares smokers to nonsmokers within each country. Explicitly provide probabilities for each event that would lead to such a setting. (It is sufficient to consider dichotomized smoking behavior.)

(b) Prove the following (including showing that the conditions are not necessary). For events A, B, and C, with

$$\begin{aligned} P(A \mid BC) &> P(A \mid B^cC) \\ P(A \mid BC^c) &> P(A \mid B^cC^c) \end{aligned}$$

then either of the following conditions

- B and C are independent, OR
- A and C are independent when conditioned on B (so $P(AC \mid B) = P(A \mid B)P(C \mid B)$)

are sufficient (but not necessary) to guarantee

$$P(A \mid B) > P(A \mid B^c)$$

thus avoiding Simpson's Paradox.