Homework 1

S520, Spring 2016

Due at the beginning of class, Thursday 21st January

All students should answer all questions. Trosset question numbers refer to the hardcover textbook. Show all working and include any graphs you are asked to draw.

1. *In class, we discussed the randomized controlled experiment that showed the Salk vaccine was successful at preventing polio. The same year, there was another large-scale test of the Salk vaccine, carried out by the National Foundation for Infantile Paralysis (NFIP). In that study, there was no randomization—all second-graders at participating schools were offered the vaccine. However, many second-graders did not receive the vaccine because their parents objected or for other reasons.*

   The results of the randomized controlled experiment were:

   - *Of 201,000 students who received the vaccine, 59 developed polio.*
   - *Of 201,000 students who received the placebo, 142 developed polio.*

   The results of the NFIP study were:

   - *Of 222,000 vaccinated second-graders, 56 developed polio.*
   - *Of 124,000 unvaccinated second-graders, 54 developed polio.*

   (a) *For the randomized controlled experiment, find the percentage that developed polio out of the students who got the vaccine, and the percentage that developed polio out of the students who got the placebo.*

   Vaccine: 0.029%; placebo: 0.071%.

   (b) *For the NFIP study, find the percentage of the vaccinated second-graders that developed polio and the percentage of the unvaccinated second-graders that developed polio.*

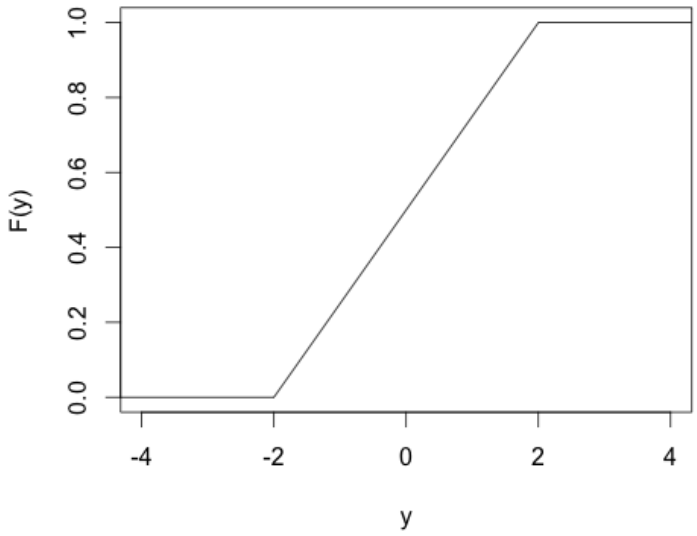
   Vaccine: 0.025%; placebo: 0.044%.

   (c) *Why is the polio percentage in the placebo group so different from the polio percentage in the NFIP unvaccinated group? Which of the two estimates is strongly biased, and why?*

   One study has treatments assigned by randomization and the other has treatments assigned nonrandomly, so we should not expect the two studies to agree. Since the NFIP study is the one that is nonrandom, we would expect to give a biased estimate of the effect of the vaccine. (What in fact happened is that children from lower socio-economic groups were less likely to get consent to take the vaccine, while children from lower socio-economic groups were less likely to develop polio. So the NFIP study underestimated the effect of the vaccine.)

2. (a) *Draw a graph of the following piecewise function $F(y)$:*

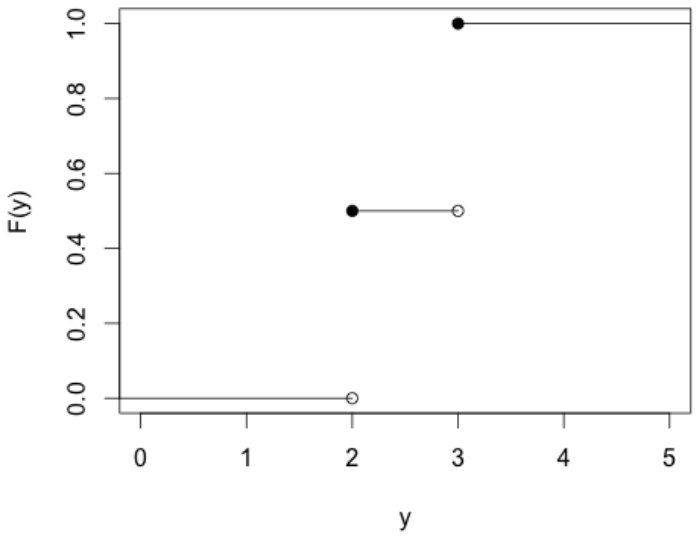
$$F(y) = \begin{cases} 0 & y < -2 \\ \frac{y+2}{4} & -2 \leq y < 2 \\ 1 & y \geq 2 \end{cases}$$



You may draw the graph by hand, and drawing this plot in R is beyond the scope of this course, but here's some R code anyway:

```
x = 0
y = 0
plot(x, y, xlab="y", ylab="F(y)", xlim=c(-4,4), ylim=c(0,1), type="n")
### type="n" creates a blank plot with the specified axes
### Add line segment from (-5, 0) to (-2, 0)
segments(-5, 0, -2, 0)
### Add line segment from (-2, 0) to (2, 1)
segments(-2, 0, 2, 1)
### Add line segment from (2, 1) to (5, 1)
segments(2, 1, 5, 1)
```

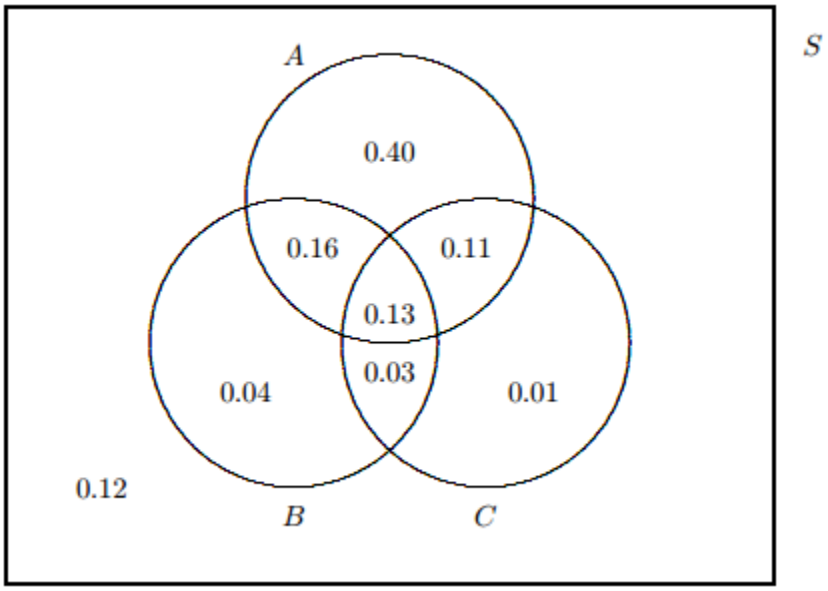
(b) *Write down a formal mathematical expression for the piecewise function $F(y)$ pictured in the graph below:*



$$F(y) = \begin{cases} 0 & y < 2 \\ 0.5 & 2 \leq y < 3 \\ 1 & y \geq 3 \end{cases}$$

3. *Trosset exercise 3.7.1 (Venn diagram and probabilities.)*

(a)



(b) $P(A \cap B \cap C^c) = 0.16$

(c) $P(A \cap B^c \cap C^c) = 0.40$

(d) $P(A^c \cap B^c \cap C^c) = 0.12$

(e) $P(A^c \cap [B \cup C]) = 0.04 + 0.03 + 0.01 = 0.08$

4. *Trosset exercise 3.7.5 (four fair dice.)*

(a) $6 \times 6 \times 6 \times 6 = 1296$

(b)
$$\frac{6 \cdot 5 \cdot 4 \cdot 3}{6^4} = \frac{5}{18}$$

(c) There are four possible ways for this to happen: $2111, 1211, 1121, 1112$. Hence the probability is
$$\frac{4}{6^4} = \frac{1}{324}$$

(d) Let A denote the event that at least one odd number appears. Then
$$P(A) = 1 - P(A^c) = 1 - \frac{\#(A^c)}{\#(S)}.$$
If A^c occurs, then an even number appears on each die. There are 3 even numbers available for each die, so $\#(A^c) = 3^4$, and therefore
$$P(A) = 1 - \frac{3^4}{6^4} = \frac{15}{16}.$$

(e) There are 3 ways to choose the one odd number, $C(4,3)$ ways to choose the three dice on which the odd number appears, and 3 ways to choose the even number on the remaining die. Hence the answer is $3 \cdot C(4,3) \cdot 3/6^4 = 1/36$.

5. *Trosset exercise 3.7.6 (dreidl.) Note: Parts of this question are extremely hard, and you are advised to go to the TA's or the lecturer's office hours for help.*

(a) There are 4^{10} equally likely ordered sequences of ten dreidl spins. Of these, $C(10,5) = 252$ such sequences have five gimmels and five hehs. The probability is thus $C(10,5)/4^{10} \approx 0.00024$.

(b) There are 2^{10} ordered sequences with no nuns or shins. The probability is thus $2^{10}/4^{10} = 1/2^{10} \approx 0.00098$.

(c) There are $2^{10}-2$ ordered sequences with just gimmels and hehs (we subtract two because we don't count all gimmels or all hehs). But there are the same number with just gimmel/shin, with just gimmel/nun, just heh/shin, just heh/nun, and just shin/nun. (In other words, there are $C(4,2) = 6$ ways of choosing two and only two of the four possible dreidl results.) So the number of ways of having two letters present and two absent is $6 \cdot (2^{10} - 2) = 6132$. The probability is $6132/4^{10} = 0.005848$.

(d) "At least two letters absent" means either two letters present or one letter present. There are 6132 sequences with two letters and 4 sequences with one letter. The probability is $(6132 + 4)/4^{10} = 0.5852$.