STAT 520: Introduction to Statistical Inference

Homework 2. Based on Chapter 3 of Trosset's textbook. Due Thursday, September 9th. For questions that require `R` code, you must turn in your `R` code on Canvas. Your should comment your code using # to at least denote the problem if not the reason you are doing what you are doing.

Handwritten questions:

Question 1: Define the sample space S by $S = \{A, B, C, D\}$ and the field (algebra) $\mathcal{C}$ of subsets of S by $\mathcal{C} = \mathcal{P}(S)$ where $\mathcal{P}(S)$ is the powerset of S (the set of all subsets of S). Define the probability function P by $P(\{A\}) = P(\{B\}) = 0.3$ and $P(\{C\}) = P(\{D\}) = 0.2$ and extending to all of $\mathcal{C}$ using the sum and complement rules.

a. What are $P(\{A, B\})$ and $P(\{B, C\})$? Justify your answers.

b. Is the event $\{A, B\}$ independent of the event $\{B, C\}$? Justify your answer.

c. Is the event $\{A, C\}$ independent of the event $\{C, D\}$? Justify your answer.

d. Is the event $\{A, B, C\}$ independent of the event $\{B, C, D\}$? Justify your answer.

Question 2: Define the sample space S by $S = \{A, B, C\} \times \{1, 2, 3\}$, which is a collection of pairs of the form (Letter, Number) for Letter $\in \{A, B, C\}$ and Number $\in \{1, 2, 3\}$. Define the field (algebra) $\mathcal{C}$ of subsets of S by $\mathcal{C} = \mathcal{P}(S)$ where $\mathcal{P}(S)$ is the powerset of S (the set of all subsets of S). Define the probability function P by

$$\begin{aligned}
P(\{(A,1)\}) &= 0.20 \\
P(\{(A,2)\}) &= 0.09 \\
P(\{(A,3)\}) &= 0.10 \\
P(\{(B,1)\}) &= 0.20 \\
P(\{(B,2)\}) &= 0.09 \\
P(\{(B,3)\}) &= 0.05 \\
P(\{(C,1)\}) &= 0.10 \\
P(\{(C,2)\}) &= 0.12 \\
P(\{(C,3)\}) &= 0.05
\end{aligned}$$

and extending to all of $\mathcal{C}$ using the sum and complement rules.

a. Let D be the event $D = \{(A,1), (A,2), (A,3), (B,1), (B,2), (B,3)\}$. Compute $P(D)$ and $P(\{s\}|D)$ for all $s \in S$. Justify your answers.

b. Is the event $\{(A,1), (A,2), (A,3)\}$ independent of the event $\{(A,3), (B,3)\}$ when conditioning on D? Justify your answer.

c. Let E be the event $E = \{(A,1), (A,2), (B,1), (B,2)\}$. Compute $P(E)$ and $P(\{s\}|E)$ for all $s \in S$. Justify your answers.

d. Is the event $\{(A,1), (A,2)\}$ independent of the event $\{(A,1), (B,1)\}$ when conditioning on E? Justify your answer.

e. Let F be the event $F = \{(A,1), (A,2), (B,1), (B,2), (C,1), (C,2)\}$. Compute $P(F)$ and $P(\{s\}|F)$ for all $s \in S$. Justify your answers.

f. Is the event $\{(A,1),(A,2)\}$ independent of the event $\{(A,1),(B,1),(C,1)\}$ when conditioning on F? Justify your answer.

Question 3: Define the sample space S by $S=(0,1)$, which is the open the interval from 0 to 1. Define the $\sigma-$field ($\sigma-$algebra) $\mathcal{C}$ as being the smallest $\sigma-$field containing all sets of the form (a,b) for $0<a<b<1$. In the following, the notation $[(a,b)]$ means an interval that is open or closed on either side (formally, it represents the indifference of the definition of the probability to the sets (a,b), $[a,b)$, $(a,b]$, and $[a,b]$ - the probability is defined as being the same for any of these sets). The collection of sets $\mathcal{C}$ contains all such sets.

a. Does defining the function P by $P([(a,b)])=b-a$ for $0<a<b<1$ (and extending to all of $\mathcal{C}$ using the sum and complement rules) define a probability measure on $(S,\mathcal{C})$? Explain your reasoning.

b. Does defining the function P by $P([(a,b)])=\frac{1}{b-a}$ for $0<a<b<1$ (and extending to all of $\mathcal{C}$ using the sum and complement rules) define a probability measure on $(S,\mathcal{C})$? Explain your reasoning.

c. Does defining the function P by $P([(a,b)])=\frac{2(b-a)}{a+b}$ for $0<a<b<1$ (and extending to all of $\mathcal{C}$ using the sum and complement rules) define a probability measure on $(S,\mathcal{C})$? Explain your reasoning.

d. Does defining the function P by $P([(a,b)])=0.5(b-a)(b+a)$ for $0<a<b<1$ (and extending to all of $\mathcal{C}$ using the sum and complement rules) define a probability measure on $(S,\mathcal{C})$? Explain your reasoning.

e. Does defining the function P by $P([(a,b)])=\frac{2b}{b+1}-\frac{2a}{a+1}$ for $0<a<b<1$ (and extending to all of $\mathcal{C}$ using the sum and complement rules) define a probability measure on $(S,\mathcal{C})$? Explain your reasoning.

f. Does defining the function P by $P([(a,b)])=\frac{1-a}{a+1}-\frac{1-b}{b+1}$ for $0<a<b<1$ (and extending to all of $\mathcal{C}$ using the sum and complement rules) define a probability measure on $(S,\mathcal{C})$? Explain your reasoning.

Hint: To determine whether P is a probability function check non-negativity, bounded above by 1, and that P behaves correctly for intersections, unions, and complements of intervals of the form $[(a,b)]$ and $[(c,d)]$.

Computational questions:

Question 4: Define the sample space S in `R` as the vector `S=c(1:12)` and define the probability vector p on the singletons $\{1\},\ldots,\{12\}$ by the vector `p=c(1/2^c(1:11),1/2^11)`. Use this `p` to extend to a probability measure on $\mathcal{C}=\mathcal{P}(S)$ where $\mathcal{P}(S)$ is the powerset of S (the set of all subsets of S). Use these `R` objects to answer the following questions.

a. Is the event $\{3,6,9\}$ independent of the event $\{3,6,10,11,12\}$?

b. Is the event $\{3,6,9\}$ independent of the event $\{1,6,9,11,12\}$?

c. Is the event $\{1,6,9\}$ independent of the event $\{1,6,9,11,12\}$?

d. What is the conditional probability $P(\{s \in S : s < 9\}|\{s \in S : s > 5\})$?

e. What is the conditional probability $P(\{s \in S : 3 < s < 9\}|\{s \in S : s < 6\})$?

f. What is the conditional probability $P(\{s \in S : 3 < s < 9\}|\{s \in S : s \geq 2\})$?

Hints: We can subset S by using lines like `S[S>3]` (which pulls out the values of S that are larger than 3). Multiple conditions can be combined using `&` (for logical `and`) and `|` (for logical `or`). For instance, if I wanted to subset S to its values between larger than 3 and less than 7, then I could use the line `S[S>3 & S<7]`. If I wanted to subest S by its values less than or equal to 3 or greater than or equal to 7, then I could use the line `S[S<=3 | S>=7]`. I could similarly use these logicals to pull out the corresponding values in p. For example, `p[S<=3 | S>=7]`. If I wanted the sum of these probabilities, then I could use the `sum` function. For example, `sum(p[S<=3 | S>=7])`.