homework1(concept)

Cuiqing Li(cli92@jhu.edu)

September 2016

1 Questions for programming part

My answer:

(1) X2's confidence interval is wider than the X1's confidence interval. In my opinion, if we check the matrix Σ, you will find the second element in the diagonal is 3, and the first element of the diagonal of the matrix is 1. It means the sampled data set's second variable is more likely to have larger variance then the first variable(feature). Larger variable indicates wider confidence interval, that's why X2's confidence interval is wider than X1's confidence interval in this case.

(2)Based on the coefficients of the linear regression, we can find that the constant's coefficient is nearly 1.5. Also, based on the coefficients of the features, we also can know that depress2' level will be more likely to be affected by depress1 and job_seek feature than any other feature in the list, since the depress1 and job_seek's coefficients has larger absolute value than other features' coefficients. Also, we can know that the depress2 is not very relevant to age feature, since the age feature has very small coefficient(nearly 0) compared to other features' coefficients.

2 Analytic Part

(1) Example1: event A: the professor is sick today. event B: the class is cancelled today. We can know that event A causes event B. I think the event that the professor is ill today can be possible to cause the event that the class is cancelled at today. Example2: event A: Google invests more money into technology innovation. event B: Google gets much more profits from technology innovation. In this example, we can say that since Google invests more money into the technology innovation, so Google gets much more profits from the technology innovation. On the other hand, we also can say that since Google get more profits from tech innovation, so Google decides to invest more money into the tech innovation. That's why I think A and B are common causes

(2)As for the version 1,I don't think I need to give the drug. Since the patients who take drugs' total recovery rate is 78%, and the patients who didn't take drugs(take placebo) is 83%, we can know that patients who didn't take drugs are more likely to recover than people having taken drugs.Also, we couldn't see obvious increase of the recovery rate if that patient have taken drugs no matter it is whether that person is in the group of men or women. That's why I think we shouldn't give drugs in terms of version 1. Also, as for the version2, I still couldn't give the drugs since I couldn't see obvious increase of the recovery rate if that person haveing taken drugs no matter whether that person is in the group of patients who have side effects and the group of people who don't have side effects. Also, the total recovery rate of the people took placebo is more likely to recover than people who took drugs. That's why I shouldn't give drugs in terms of version 2.

(3)As for Simpson's paradox, it means that one conclusion based on several variables can be possibly reversed if we regroup those data. Specifically, a data set has group A and B, and we get positive conclusion. However, the data set also has group C , D and E, if we regroup them based on C,D and E, it is possible to get reversed sign(negative conclusion). That's why Simpson's paradox is a paradox. I think the reason is that we didn't take other contributing factors into consideration, if so, it will be likely to result in Simpson's paradox. Or, we didn't change a perspective to think about those questions and neglect the implicit factors or causes. If so, the paradox is also likely to happen.

(4)

$$P(\lambda) = nlog(\lambda) - \lambda(x_1 + x_2... + x_n)$$

If we differentiate the function above with respect to λ, then we will get the maximum likelihood estimate $\lambda = \hat{\lambda}$, which is $\hat{\lambda} = n/(x_1 + x_2... + x_n)$

Then, we need to compute $E[X]$. Based on $E[X] = \int f[X]XdX$,then we can use calculus knowledge to get

$$E[X] = 1/\lambda$$

(using integration by part, we can get this answer)

(5)show that conditional independence satisfies the semi-graphoid axioms Given two conditional independence envent A and B, symmetry:Based on definition, we know that $P(A, B|C) = P(A|C)p(B|C) = p(B|C)p(A|C) = P(B, A|C)$, thus, the symmetry property holds

chain rule: Firstly, assume that $(A \perp B \cup D)|C$,then certainly we have known that D is irrelevant to A,so it is necessary for us to know D if we need to know the relation between A and B, so $(A \perp B \cup D)|C) \rightarrow (A \perp D)|C$, also we have $(A \perp B \cup D)|C \rightarrow (A \perp B)|C \cup D$. Then, let assume that $(A \perp B)|C$ and $(A \perp B)|C \cup D$,then we can know that knowing information D is irrelevant to predict A or B, that's why we have $(A \perp B)|C$ and $(A \perp B)|C \cup D \rightarrow (A \perp B \cup D)|C$

3 extra credit

question1: Proof: Firstly,we know that $p(C, A, Y)$is a distribution, then we know that $q_a(Y, C) = p(Y|A = a, C)p(C) = p(Y, C|A = a) = P(Y, C|A = a)p(A =$

$a)/P(A=a) = p(Y, A=a, C)/p(A=a)$,thus we get

$$q_a(Y, C) = p(Y, A=a, C)/p(A=a)$$

.Since $p(Y, A=a, C)$ is distribution we have known, and $p(A=a)$ is a constant,so $q_a(Y, C)$ is also a distribution.

question2: first of all, for any c in C,we have $q_{a_1}(Y|C=c) < q_{a_2}(Y|C=c)$,thus we have $q_{a_1}(Y|C=c)p(C=c) < q_{a_2}(Y|C=c)P(C=c) \to q_{a_1}(Y, C=c) < q_{a_2}(Y, C=c)$. Thus $\sum_C q_{a_1}(Y, C) < \sum_C q_{a_2}(Y, C)$, so we have $q_{a_1}(Y) < q_{a_2}(Y)$