# 1 Exercises

**Exercise 1.1** *Sufficient Statistics. Suppose that you are playing with a 4-sided die and you suspect that it is biased. In the past 60 times that it was rolled, 20 times it came up 1, 15 times it came up 2, 15 times 3, and 10 times 4. Let* $\theta_1$ *be the true probability of the die landing on* 1*, and similarly for* $\theta_2$*,* $\theta_3$*, and* $\theta_4$*. To estimate these parameters from these past 60 rolls using a simple multinomial model, which is a sufficient statistic for this data?*

a) *The sum of all die rolls.*

b) *The total number of times that the dice was rolled.*

c) *A vector with with four components, where each component is the number of times a specific side was observed.*

d) *None of these are sufficient statistics.*

**Exercise 1.2** *Maximum likelihood estimation. In the previous question, which is the unique Maximum Likelihood Estimate (MLE) of the parameter* $\theta_1$*?*

**Exercise 1.3** *Likelihood Functions. For a Naive Bayes model with a parent variable,* $C$*, and 3 descriptive variables,* $\{X_1, X_2, Y_3\}$ *, which is a correct expression for the likelihood, decomposed in terms of the local likelihood functions?*

a) $L(\theta : D) = \prod_{m=1}^{M} P(c^{(m)}, x_1^{(m)}, x_2^{(m)}, x_3^{(m)} : \theta)$

b) $L(\theta : D) = \prod_{m=1}^{M} P(c^{(m)}|x_1^{(m)}, x_2^{(m)}, x_3^{(m)} : \theta)P(x_1^{(m)}, x_2^{(m)}, x_3^{(m)} : \theta)$

c) $L(\theta : D) = \prod_{m=1}^{M} P(c^{(m)}|x_1^{(m)}, x_2^{(m)}, x_3^{(m)} : \theta_C)P(x_1^{(m)}|c^{(m)} : \theta_{X_1|C})P(x_2^{(m)}|c^{(m)} : \theta_{X_2|C})P(x_3^{(m)}|c^{(m)} : \theta_{X_3|C})$

d) $L(\theta : D) = \prod_{m=1}^{M} P(c^{(m)} : \theta_C)P(x_1^{(m)}|c^{(m)} : \theta_{X_1|C})P(x_2^{(m)}|c^{(m)} : \theta_{X_2|C})P(x_3^{(m)}|c^{(m)} : \theta_{X_3|C})$

**Exercise 1.4** *MLE for Naive Bayes. Using a Naive Bayes model for spam classification with the vocabulary* $V$ *={"Secret", "Offer", "Low", "Price", "Valued", "Customer", "Today", "Dollar", "Million", "Sports", "Is", "For", "Play", "Healthy", "Pizza"}.*

*Given a set of spam messages and another set of normal messages:*
*Spam = {"Million Dollar Offer", "Secret Offer Today", "Secret Is Secret"}*
*Regular = {"Low Price For Valued Customer", "Play Secret Sports Today", "Sports Is Healthy", "Low Price Pizza"}*

*Create a multinomial naive Bayes model for the data given above. This can be modeled as a parent node taking values "Spam" and "Regular" and a child node for each word in the vocabulary. The parameters are estimated based on the number of times that a word appears in the vocabulary. Give the MLE for* $\theta_{Spam}$*:*

**Exercise 1.5** *MLE for Naive Bayes. Given the model from the previos exercise, give the MLE for* $\theta_{Secret|Spam}$*:*

**Exercise 1.6** *MLE for Naive Bayes. Given the model from the previos exercise, give the MLE for* $\theta_{Secret|Regular}$*:*

**Exercise 1.7** *Parameter Learning in MNs or BNs. Compared to learning parameters in Bayesian networks, learning in Markov networks is generally...*

a) *more difficult because we cannot push in sums to decouple the likelihood function, allowing independent parallel optimizations, as we can in Bayesian networks with CPDs.*

b) *equally difficult, as both require an inference step at each iteration.*

c) *equally difficult, though MN inference will be better by a constant factor difference in the computation time as we do not need to worry about directionality.*

d) *less difficult because we must separately optimize decoupled portions of the likelihood function in a Bayesian network, while we can optimize portions together in a Markov network.*

## Answers

***Ex. 1.1***: c

***Ex. 1.2***: 1/3

***Ex. 1.3***: d

***Ex. 1.4***: 0.4286

***Ex. 1.5***: 1/3

***Ex. 1.6***: 1/15

***Ex. ??***: a