COLUMBIA UNIVERSITY
Statistical Inference and Modeling
Quiz 3

Prof. Marco Avella
TA: Ian Kinsella, Shun Xu
March 7, 2019

Name:

UNI:

You have 20 minutes to answer the following 10 questions. Buena suerte!

Question 1

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n \overset{iid}{\sim} N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and consider the estimators $\overline{\mathbf{X}_n} = \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i$ and $\mathbf{S}_n = \frac{1}{n-1}\sum_{i=1}^n (\mathbf{X}_i - \overline{\mathbf{X}_n})(\mathbf{X}_i - \overline{\mathbf{X}_n})^T$. Assume we want to test the hypothesis $H_0 : \boldsymbol{\mu} = \mathbf{0}$ versus $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$. Which of the following statements is NOT correct?

(a) $\overline{\mathbf{X}_n}$ is an unbiased and consistent estimator of $\boldsymbol{\mu}$

(b) Under $H_0 : \boldsymbol{\mu} = \mathbf{0}$ we have that $n\overline{\mathbf{X}_n}^T \mathbf{S}_n^{-1} \overline{\mathbf{X}_n} \xrightarrow[n\to\infty]{\mathcal{D}} \chi^2_d$.

(c) $\mathbf{S}_n$ is a consistent estimator of $\boldsymbol{\Sigma}$

(d) $\hat{\boldsymbol{\mu}}_{ML} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Question 2

Let X and Y be jointly normally distributed as $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

Give the distribution of $\mathbb{E}[Y|X]$

$$\mathbb{E}\left[Y|X\right] = 0 - 1\cdot(2)^{-1}(X-1) = \frac{1}{2}(X-1) \sim N(0, \tfrac{1}{2})$$

Question 3

A Gaussian graphical model was fitted to a data set yielding the following estimates of the covariance and precision matrices respectively

$$\widehat{\mathbf{\Sigma}} = \begin{bmatrix} 3.47 & 1.72 & 2.35 & 1.76 \\ 1.72 & 2.29 & 1.37 & 1.72 \\ 2.35 & 1.37 & 4.36 & 2.29 \\ 1.76 & 1.72 & 2.29 & 5.31 \end{bmatrix} \qquad \widehat{\mathbf{\Theta}} = \begin{bmatrix} 0.59 & -0.31 & -0.22 & 0 \\ -0.31 & 0.78 & 0 & -0.15 \\ -0.22 & 0 & 0.4 & -0.10 \\ 0 & -0.15 & -0.10 & 0.28 \end{bmatrix}$$

Which of the followings statement is NOT correct?

(a) X_1 and X_4 are conditionally independent given (X_2, X_3).

(b) All pairwise correlations between the 4 variables are positive.

(c) The partial correlation of X_2 and X_3 is 0.

(d) X_2 and X_3 are independent.

Question 4

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n \overset{iid}{\sim} N_d(\boldsymbol{\mu}, \mathbf{\Sigma})$, and consider $= \overline{\mathbf{X}_n} = \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i$ and $\mathbf{S}_n = \frac{1}{n-1}\sum_{i=1}^n (\mathbf{X}_i - \overline{\mathbf{X}_n})(\mathbf{X}_i - \overline{\mathbf{X}_n})^T$. Assume we want to test the hypothesis $H_0 : \boldsymbol{\mu} = \mathbf{0}$ versus $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$. Which of the following statements is NOT correct?

(a) $\mathbf{\Sigma}^{-1}\mathbf{S}_n \xrightarrow[n\to\infty]{\mathcal{P}} \mathbf{I}$.

(b) Under H_1 we have that $\overline{\mathbf{X}_n} \sim N_d(\boldsymbol{\mu}, \frac{1}{n}\mathbf{\Sigma})$.

(c) If $\mathbf{\Sigma} = \mathbf{I}$, under $H_0 : \boldsymbol{\mu} = \mathbf{0}$ we have that $n\overline{\mathbf{X}_n}^T\overline{\mathbf{X}_n} \xrightarrow[n\to\infty]{\mathcal{D}} \chi^2_d$.

(d) Under H_1 we have that $\sqrt{n}\overline{\mathbf{X}_n}^T\mathbf{\Sigma}^{-1}\overline{\mathbf{X}_n}^T$ is a pivot.

Question 5

Assume that $Y|X = x$ has a probability density function

$$f(y; x) = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left\{ -\frac{(y - \rho x^2)^2}{2(1-\rho^2)} \right\}.$$

Compute $\mathbb{E}[Y]$ assuming that $X \sim N(1, 2)$.

$\mathbb{E}\{\mathbb{E}[Y|X]\} = \mathbb{E}[\rho X^2] = \rho(2+1^2) = 3\rho$

Question 6

Table 1: Observed counts of a data set with 3 states

	1	2	3	Total
1	362	126	60	548
2	136	89	68	293
3	50	78	124	252

Based Table 1, which of the following statements is most accurate

(a) We cannot fit a first order model using the information in the Table 1.

(b) Table 1 suggests that a zeroth order model is not appropriate for these data.

(c) The likelihood ratio statistic Λ_n for testing a second order model versus a first order model is such that $\Lambda_n \xrightarrow[n\to\infty]{\mathcal{D}} \chi^2_6$

(d) None of the above statements is correct.

Question 7

Based on Table 1, which of the following is NOT correct

(a) $\hat{p}_{11}$ and $\hat{p}_{21}$ are asymptotically independent.

(b) $\hat{p}_{11} = 1 - \hat{p}_{21} - \hat{p}_{31}$.

(c) Using Table 1, the distribution of the MLE of $\mathbb{P}(X_1 = 1|X_0 = 1)$ is approximately $N(\frac{362}{548}, \frac{362\cdot 186}{548^3})$

(d) None of the above statements is correct.

Question 8

Assume first order Markov chain with state space $\{0, 1\}$ with the following transition probability matrix

$$\mathbf{P} = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}$$

Which of the following statements is NOT correct

(a) The stationary distribution is $(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta})$

(b) $\sum_{k=1}^{m} X_{t+k}|X_t = 0 \sim Bin(m, \alpha)$

(c) $\mathbb{P}(X_{n+1} = X_n = \ldots, X_2 = 1, X_1 = 0 | X_0 = 1) = \beta\alpha(1-\beta)^{n-1}$

(d) None of the above statements is correct.

Question 9

The probability of a particular offspring being purely dominant, purely recessive, or hybrid for the trait is given by the table below.

$$\begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0 & 0.5 & 0.5 \end{bmatrix}.$$

Assuming a Markov chain of order one for the state space $\{1, 2, 3\}$, which of the following statements is NOT correct

(a) The stationary distribution is $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$

(b) $\mathbb{P}(X_{t+2} = 3, X_{t+1} = 1, X_t = 1 | X_{t-1} = 2) = 0$

(c) $\mathbb{P}(X_2 = 2, X_3 = 3, X_4 = 1 | X_1 = 1) = 1/\cancel{32}$ $= 0$

(d) Given an initial probability vector $(\frac{1}{2}, 0, \frac{1}{2})$, the probability vector at the next time period is $(1/4, 1/2, 1/4)$

Question 10

Draw a transition graph corresponding to the transition probability matrix of Question 9.