Homework Assignment 0

Statistical Methods for Analysis with Missing Data, Winter 2019

Instructor: Mauricio Sadinle, Department of Biostatistics, U. of Washington – Seattle

Submit your solutions via Canvas. Feel free to hand-write your solutions and submit a scanned copy. Due by 12:00pm (noon) on Jan 16, 2019.

The grade for this assignment will be either 1 (satisfactory) or 0. This assignment contains four problems, each of which contains subproblems, totaling 17 of them. Your solutions will be graded as 1 if you have at least 10 correct subproblems, including one correct subproblem from each problem.

1. Let the random variables X, Y, Z, W have a joint distribution with density p, such that

$$p(x, y, z, w) = p_X(x)p_{Y|X}(y \mid x)p_{Z|X}(z \mid x)p_{W|X,Z}(w \mid x, z),$$

where p_X, $p_{Y|X}$, $p_{Z|X}$, and $p_{W|X,Z}$ are the densities of the distributions of X, $(Y \mid X)$, $(Z \mid X)$, and $(W \mid X, Z)$, respectively.

(a) Determine which of the following independence statements are true, and provide proofs for the true statements:

i. $Y \perp\!\!\!\perp W$; ii. $Y \perp\!\!\!\perp W \mid X$; iii. $W \perp\!\!\!\perp (X, Z)$;

iv. $Z \perp\!\!\!\perp Y \mid X$; v. $Z \perp\!\!\!\perp (Y, W) \mid X$.

(b) Write down the conditional density of:

i. $X \mid Y, Z, W$; ii. $X \mid Y, Z$; iii. $X \mid Y, W$

(c) Write down the expression for the expected value of XW, $E(XW)$, in terms of the density p.

(d) Say we use the simpler notation

$$p(x, y, z, w) = p(x)p(y \mid x)p(z \mid x)p(w \mid x, z)$$

to denote the p density above. Why is this notation inaccurate?

2. Let $X_1, \ldots, X_K \overset{i.i.d.}{\sim}$ Bernoulli(π).

   (a) Write down the probability mass function of the random vector $X = (X_1, \ldots, X_K)$, denoted $p(x \mid \pi)$, where x is a generic realization of X

   (b) Write down the likelihood function of π. What is the difference between the likelihood function and the probability mass function derived in (a)?

   (c) Derive the maximum likelihood estimator of π, and call it $\hat{\pi}$

   (d) Compute the expected value of $\hat{\pi}$

   (e) Compute the expectation $E_X[\log p(X \mid \pi')]$, $\pi' \neq \pi$

   (f) Find the value that maximizes $E_X[\log p(X \mid \pi')]$ as a function of π'

   (g) Suppose that K is an even number. Let $Y_1 = (X_1, \ldots, X_{K/2})$ and $Y_2 = (X_{K/2+1}, \ldots, X_K)$. Derive the distribution of $W = Y_1 Y_2^T$. T denotes the transpose operator.

   (h) Derive the conditional distribution of X given XX^T

3. Let $X = (X_1, \ldots, X_K) \sim$ MultivariateNormal(μ, Σ). Let A be a $K \times B$ matrix, $B \leq K$, such that its (i,j)th entry $a_{ij} \in \{0, 1\}$, $\sum_{j=1}^{B} a_{ij} \leq 1$ for all $i = 1, \ldots, K$, and $\sum_{i=1}^{K} a_{ij} \geq 1$ for all $j = 1, \ldots, B$. For this problem you can use commonly known facts about the multivariate normal distribution.

   (a) What is the distribution of $Y = XA$? Explain.

   (b) What is the distribution of $(Y_1, Y_2) \mid (Y_3, \ldots, Y_B)$. Explain.

4. Let $X = (X_1, \ldots, X_K) \sim$ Multinomial$\left(N, (\pi_1, \ldots, \pi_K)\right)$.

   (a) Derive the distribution of $(X_1 + X_2, X_3, \ldots, X_K)$.

   (b) What is the distribution of $Y = XA$, where A is a matrix as in problem 3., except that here $\sum_{j=1}^{B} a_{ij} = 1$ for all $i = 1, \ldots, K$. No need to write down the proof.

   (c) Suppose that K is an even number. Let $Y_1 = (X_1, \ldots, X_{K/2})$ and $Y_2 = (X_{K/2+1}, \ldots, X_K)$. Derive the distribution of $X = (Y_1, Y_2)$ given $Y_1 Y_1^T$ and $Y_2 Y_2^T$. T denotes the transpose operator.

Note: 4(c) is complicated in general. Acceptable solutions include (in order of difficulty): i. solve for $N = 1$; ii. simply write generic expression for the conditional distribution for generic N. iii. solve for $K = 4$ and generic N.