Graphical Models: Worksheet 0 MT 2023

This sheet is designed to revise some bits from previous courses that will be particularly useful for Graphical Models.

1. **Sufficient Statistics.**

   Let $X_i \sim \text{Binom}(n_i, \theta)$, $i = 1, \ldots, k$ be independent binomial random variables with known sizes n_i, and unknown $\theta \in [0, 1]$.

   (a) Write down the log-likelihood for θ, and find a sufficient statistic.

   (b) Find the MLE for θ and its asymptotic distribution.

   (c) What would constitute a conjugate prior for θ?

   (d) Suppose you have $n_1 = n_2 = 100$, and data $X_1 = 48$, $X_2 = 52$. Build a confidence interval for θ using your answer to (b).

   (e) Now suppose $X_1 = 10$ and $X_2 = 90$. How does your answer differ?

2. **Conditional Distributions.**

   Suppose that X, W are independent Exponential(λ) random variables. Define $Y = X + W$. Find the joint density of X and Y. Are X and Y independent?

   Find the conditional density of X given Y.

3. **Conditional Events**

   Let X, Y, and Z be discrete random variables taking values in the sets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$.

   (a) Write down and briefly justify the law of total probability for discrete random variables X and Y.

   (b) Prove Bayes' Formula:

   $$P(Y = y \mid X = x) = \frac{P(X = x \mid Y = y) \cdot P(Y = y)}{\sum_{y' \in \mathcal{Y}} P(X = x \mid Y = y') \cdot P(Y = y')}.$$

   (c) Express $P(Z = z)$ in terms of probabilities of the form $P(X = x), P(Y = y \mid X = x), P(Z = z \mid X = x, Y = y)$. In terms of the sizes of the sets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$, how many calculations (additions, subtractions, multiplications, divisions) are required to evaluate it for all $z \in \mathcal{Z}$?

   (d) What difference does it make if $P(Z = z \mid X = x, Y = y) = P(Z = z \mid Y = y)$?

4. **Contingency Tables.**

   Let (X_i, Y_i, Z_i), $i = 1, \ldots, n$ be i.i.d. vectors of categorical variables such that $P(X = x, Y = y, Z = z) = \pi_{xyz}$. Define

   $$n_{xyz} = \sum_{i=1}^{n} \mathbb{1}\{X_i = x, Y_i = y, Z_i = z\}.$$

   The array $(n_{xyz})_{x,y,z}$ is called a *contingency table* (see Part A Stats).

   (a) Write down the likelihood for $\boldsymbol{\pi} = (\pi_{xyz})_{x,y,z}$.

(b) We say X is *conditionally independent* of Y given Z if we can write

$$P(X = x, Y = y, Z = z) = P(Z = z) \cdot P(X = x \mid Z = z) \cdot P(Y = y \mid Z = z)$$

for all x, y, z. Show that, the MLE of $\boldsymbol{\pi}$ under this restriction is

$$\hat{\pi}_{xyz} = \frac{n_{x+z} \cdot n_{+yz}}{n_{++z} \cdot n},$$

where, for example, $n_{x+z} = \sum_y n_{xyz}$. *[Hint: this is similar to the two-dimensional independence case from Part A stats.]*

5. **Multivariate Normal Distributions.**

[This is harder, but do-able.]

Let $M = (m_{ij})$ be a $p \times p$-matrix and $C \subseteq \{1, \dots, p\}$; let $D = \{1, \dots, p\} \setminus C$. We say that

$$M_{DD \cdot C} \equiv M_{DD} - M_{DC}(M_{CC})^{-1} M_{CD}$$

is the *Schur complement* of M with respect to C, and its entries are

$$m_{ij \cdot C} \equiv m_{ij} - M_{iC}(M_{CC})^{-1} M_{Cj} \qquad \text{for } i, j \in D.$$

Now let $X_V \sim N_p(\mu, \Sigma)$ have a multivariate normal distribution, meaning that it has Lebesgue density

$$f(x_V; \mu, \Sigma) = \frac{1}{(2\pi)^{p/2} (\det \Sigma)^{1/2}} \exp\left\{ -\frac{1}{2}(x_V - \mu)^T \Sigma^{-1} (x_V - \mu) \right\}, \quad x_V \in \mathbb{R}^p,$$

for $\mu \in \mathbb{R}^p$ and a symmetric positive definite matrix Σ.

(a) Let Σ be partitioned as

$$\Sigma = \begin{pmatrix} \Sigma_{CC} & \Sigma_{CD} \\ \Sigma_{DC} & \Sigma_{DD} \end{pmatrix}$$

with $|C| = p_1$, $|D| = p_2$. Show that

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_{CC \cdot D}^{-1} & -\Sigma_{CC \cdot D}^{-1} \Sigma_{CD} \Sigma_{DD}^{-1} \\ -\Sigma_{DD}^{-1} \Sigma_{DC} \Sigma_{CC \cdot D}^{-1} & \Sigma_{DD}^{-1} + \Sigma_{DD}^{-1} \Sigma_{DC} \Sigma_{CC \cdot D}^{-1} \Sigma_{CD} \Sigma_{DD}^{-1} \end{pmatrix}$$

[Note that Σ_{DD}^{-1} means $(\Sigma_{DD})^{-1}$.]

(b) By considering the terms in the density which depend upon x_C, show that

$$X_C \mid X_D = x_D \;\sim N_{p_1}\big(\mu_C + \Sigma_{CD} \Sigma_{DD}^{-1} (x_D - \mu_D),\, \Sigma_{CC \cdot D}\big).$$

where $\Sigma_{CC \cdot D} = \Sigma_{CC} - \Sigma_{CD} \Sigma_{DD}^{-1} \Sigma_{DC}$.

(c) Hence show that the marginal distribution $X_D \sim N_{p_2}(\mu_D, \Sigma_{DD})$.