BIOSTAT 802 Homework #1

Due on January 29, 2018 in class

Problem 1: (Item Response Model) Suppose n students each take an examination with k items. Each item is scored as correct or incorrect, and Y_{ij} equals 1 if the i-th subject correctly answers the j-th item and 0 otherwise. Assume $Y_{ij} \overset{ind}{\sim} \text{Bernoulli}(\pi_{ij})$ with

$$\text{logit}(\pi_{ij}) = \text{logit}(\pi_{ij}|\theta_i, \alpha_j, \beta_j) = \alpha_j \theta_i - \beta_j$$

where the two regression coefficients α_j, β_j may be interpreted as follows. Parameter α_j measures the discriminatory power (to discriminate between strong and weak students) of the j-th item, and parameter β_j represents the difficulty of the j-th item. The quantity θ_i represents the unobserved "ability" of the i-th student, which may be assumed to follow a certain distribution under density $\phi(\theta)$. Thus, when $\alpha_j > 0$, this model implies that stronger students are more likely to correctly answer the j-th item. The objective is to obtain the MLE of α_j and β_j. Solve the following problems:

(a) Set up the probability space.

(b) Set up the action space.

(c) Give the decision function based on the method of MLE.

(d) Give the risk function under the squared error loss function.

Problem 2 Consider two independent normal populations $X \sim N(\mu_1, 1)$ and $Y \sim N(\mu_2, 1)$, with $\mu_1 \in (-\infty, \infty)$ and $\mu_2 \in (-\infty, \infty)$. The objective is to establish a decision rule for a comparison between the two means. Suppose the action space $A = \{a_0, a_1\}$ where a_0 corresponds to the case for $\mu_1 \le \mu_2$ and otherwise, a_1. The loss function is given as follows:

$$\begin{aligned} L((\mu_1, \mu_2), a_0) &= \begin{cases} 0, & \text{if } \mu_1 \le \mu_2 \\ \mu_1 - \mu_2, & \text{if } \mu_1 > \mu_2; \end{cases} \\ L((\mu_1, \mu_2), a_1) &= \begin{cases} \mu_2 - \mu_1, & \text{if } \mu_1 \le \mu_2 \\ 0, & \text{if } \mu_1 > \mu_2; \end{cases} \end{aligned}$$

Now given random samples (*iid*) $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ from the above normal populations, respectively, define a decision rule of the following form:

$$\delta_c(X_1, \ldots, X_m; Y_1, \ldots, Y_n) = \begin{cases} a_0, & \text{if } \bar{X} \le \bar{Y} + c, \\ a_1, & \text{otherwise}, \end{cases}$$

where $c \in (-\infty, \infty)$ is a constant, and $\bar{X}$ and $\bar{Y}$ are the respective sample means.

(a) Let $\theta = \mu_1 - \mu_2$. Write the expression of the loss function $L(\theta, \delta_c(x_1, \ldots, x_m; y_1, \ldots, y_n))$.

(b) Derive the risk function of the decision rule δ_c under the above loss function in part (a).

(c) Consider a class of decision rules given by $D = \{\delta_c(x_1, \ldots, x_m; y_1, \ldots, y_n), c \in (-\infty, \infty)\}$. Show that in this class D every decision rule is inadmissible.

Problem 3 Consider a probability space $(\mathcal{X}, \mathcal{B}_x, \mathcal{P} = \{P_\theta, \theta \in \Theta\})$. Show the following statements.

(a) If $\delta_0(x)$ is the unique minimax decision rule with respect to the loss function $L(\theta, a)$, then δ_0 is admissible.

(b) If $\delta_0(x)$ is an admissible decision rule and has a constant risk function on Θ, then δ_0 is the minimax decision rule.