

# Statistical Decision Framework

akim1

February 17, 2015

As described in lecture, there are five components to the statistical decision theoretical problem. One example of such is described below for the problem of clustering nodes in a graph.

**Sample space** The sample space describes all possible arrangements of edges, nodes, and labels.

$$\mathcal{G} = (V, E, Y)$$

**Model** The model is given by the two stochastic blocks and can be described as follows:

$$P = SBM_n^2(\rho, \beta) \text{ where } \rho \in \Delta_2, \beta \in (0, 1)^{2 \times 2}$$

**Action space** The action space is the assignment given by the clustering algorithm.

$$A = \{y \in \{0, 1\}^n\}$$

**Decision rule class** The decision rule class is given by k-means square clustering.

**Loss function** The loss function is given by the following using the adjusted rand index:

$$l = \sum_{i=1}^n \Theta(\hat{y}_i = y_i)$$

**Risk function** The risk function is simply the expected value of the loss function:

$$R = E\{l\}$$