DS 102 Discussion 3

Wednesday, September 16, 2020

1. **False Discovery Rate vs. Family-Wise Error Rate.**

   Suppose that we are testing some number of hypotheses. We are making decisions according to some unknown decision rule, where a discovery is indicated by a decision of 1 and no discovery is indicated by a decision of 0.

   (a) Prove that $\mathbf{1}\{\text{at least one false discovery}\} \geq \text{FDP}$, where FDP denotes the false discovery proportion.

   (b) Prove that the family-wise error rate (FWER), *i.e.*, the probability of making at least one false discovery, is at least as big as the false discovery rate (FDR):

   $$\text{FWER} \geq \text{FDR}.$$

   (c) Suppose we want to test possibly infinitely many hypotheses in an online fashion. At time $t = 1, 2, \ldots$, a p-value P_t arrives, and we proclaim a discovery if $P_t \leq \alpha_t$, where $\alpha_t = \left(\frac{1}{2}\right)^t \alpha$. Does this rule control the FWER under α? Give a proof or counterexample.

   (d) Does the rule from part (c) control the FDR under α?

2. **Decision Theory: Computing and Minimizing the Bayes Risk**

For the following two parts, derive the decision procedure δ^* that minimizes the Bayes risk, for the given loss function. That is, provide an expression for

$$\delta^* = \operatorname*{argmin}_{\delta} R(\delta)$$

where the Bayes risk $R(\delta)$ can be written out as

$$R(\delta) = \mathbb{E}_{\theta,X}[\ell(\theta, \delta(X))] = \mathbb{E}_X[\mathbb{E}_\theta[\ell(\theta, \delta(X)) \mid X]].$$

Hint. One strategy to find the decision rule that minimizes the Bayes risk is based on the following rationale. For any given value of the data, $X = x$, the quantity $\delta(x)$ is simply a scalar value. Suppose, for any given value of $X = x$, we can find the scalar value $\delta^*(x) = a^* \in \mathbb{R}$ such that

$$a^* = \operatorname*{argmin}_{a \in \mathbb{R}} \mathbb{E}_\theta[\ell(\theta, a) \mid X = x]$$

(that is, a^* is the scalar value that minimizes the Bayes posterior risk for this particular value of $X = x$). Then, the rule given by this computation of $\delta^*(x)$ (for each value of $X = x$) must also be the one that minimizes the Bayes risk, which just takes an expectation over all possible values of X. This is sometimes referred as a *pointwise minimization* strategy.

(a) $\ell(\theta, \delta(X)) = (1/2)(\theta - \delta(X))^2$ (squared-error loss)

(b) $\ell(\theta, \delta(X)) = \mathbf{1}[\theta \neq \delta(X)]$ (zero-one loss)