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.

> **Solution:** If $\mathbf{1}\{\text{at least one false discovery}\} = 0$, then no false discovery has been made, in which case the FDP is clearly 0. If $\mathbf{1}\{\text{at least one false discovery}\} = 1$, then there is at least one false discovery, so $\text{FDP} = \frac{\#\text{ false disc.}}{\#\text{discoveries}}$, but since the number of discoveries is at least as big as the number of false discoveries, $\text{FDP} \leq 1$.

(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):

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

> **Solution:** Due to monotonicity of expectations, we take expectations on both sides of the inequality from part (a) to get:
>
> $$\mathbb{E}[\mathbf{1}\{\text{at least one false discovery}\}] \geq \mathbb{E}[\text{FDP}] \implies \text{FWER} \geq \text{FDR}.$$

(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.

> **Solution:** We use the usual union-bound argument:
>
> $$\text{FWER} \leq \sum_{t \in \text{nulls}} \mathbb{P}(P_t \leq \alpha_t) = \sum_t \left(\frac{1}{2}\right)^t \alpha = \alpha.$$
>
> Therefore, the rule does indeed control the FWER.

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

Solution: From part (b), we know that FDR $\leq$ FWER, so if the FWER is under α, then so is the FDR.

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)

Solution: Following the pointwise minimization strategy, for any particular value of $X = x$ we find the value $a^* = \delta(x)$ that solves

$$a^* = \min_{a \in \mathbb{R}} \mathbb{E}_\theta[(1/2)(\theta - a)^2 \mid X = x].$$

To do this, we take the derivative with respect to a and set it to zero, since $f(a) = \mathbb{E}_\theta[(1/2)(\theta - a)^2 \mid X = x]$ is a convex function in a (try to prove this as a quick exercise; see below for solution). Swapping the differentiation and expectation operators and applying the chain rule gives

$$f'(a) = \mathbb{E}_\theta[a - \theta \mid X = x]$$

and setting the derivative to zero gives

$$f'(a) = 0 \implies a^* = \mathbb{E}[\theta | X = x].$$

That is, for any particular value of $X = x$, we should take $\delta^*(x) = \mathbb{E}[\theta \mid X = x]$. That means that the decision rule that minimizes the Bayes risk for the squared error loss is $\delta^*(X) = \mathbb{E}[\theta \mid X]$, the posterior expectation (the expectation of the posterior distribution)!

To show that $f(a)$ is a convex function, for any $a_1, a_2 \in \mathbb{R}$ and $t \in [0, 1]$, we have that

$$\begin{aligned}
f(ta_1 + (1-t)a_2) &= \mathbb{E}_\theta[(1/2)(\theta - [ta_1 + (1-t)a_2])^2 \mid X = x] \\
&\leq \mathbb{E}_\theta[(1/2)t(\theta - a_1)^2 + (1/2)(1-t)(\theta - a_2)^2 \mid X = x] \\
&= t\mathbb{E}_\theta[(1/2)(\theta - a_1)^2 \mid X = x] + (1-t)\mathbb{E}_\theta[(1/2)(\theta - a_2)^2 \mid X = x] \\
&= tf(a_1) + (1-t)f(a_2)
\end{aligned}$$

where the second line due to the convexity of the function $g(a) = (\theta - a)^2$ and monotonicity of expectations, and the third line is due to linearity of expectations.

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

Solution: We use the same strategy as Part (a). For a given value $X = x$, we assign $\delta^*(x)$ to be the value

$$\begin{aligned}
\underset{a}{\operatorname{argmin}}\, \mathbb{E}_{\theta \sim \mathbb{P}(\theta|X=x)}[\mathbf{1}[\theta \neq a]] &= \underset{a}{\operatorname{argmin}}\, \mathbb{P}(\theta \neq a|X = x) \\
&= \underset{a}{\operatorname{argmin}}(1 - \mathbb{P}(\theta = a|X = x)) \\
&= \underset{a}{\operatorname{argmax}}\, \mathbb{P}(\theta = a|X = x) \\
&= \underset{\theta}{\operatorname{argmax}}\, \mathbb{P}(\theta|X = x).
\end{aligned}$$

That is, the decision rule that minimizes the Bayes risk for the zero-one loss is $\delta^*(X) = \operatorname{argmax}_\theta \mathbb{P}(\theta \mid X)$ the *posterior mode* (the mode of the posterior distribution).