Optimal Inference After Model Selection (Fithian, Sun & Taylor 2017).

target : test | the test is performed. ⟹ broader than model selection.

$P_{M,H_0}(\text{reject } H_0 \mid (M, H_0) \text{ selected}) \overset{\text{if}}{=\!=} P_{M,H_0}(\text{reject } H_0)$

selective inference

achieved by data splitting.

Q: use of M here for underlying probability measure?

selective inference (data carving). : $\underset{\text{selection}}{\mathcal{F}_0 \subseteq \mathcal{F}(1_A(Y))} \subseteq \underset{\text{inference}}{\mathcal{F}(Y)}$. ~~information.~~

Condition on more information = finer A (e.g. LASSO)

⇓

longer interval, smaller power.

data splitting : $\mathcal{F}_0 \subseteq \underset{\text{selection}}{\mathcal{F}(Y_1)} \subseteq \underset{\text{inference}}{\mathcal{F}(Y_1, Y_2)}$.

In reality, often use finer $S(Y)$ rather than $1_A(Y)$. [for computational benefit].

⟹ leftover information (used for inference) $-E[\nabla^2 l(\theta; Y|S) \mid S]$ total information of θ in Y, excluding information of θ in S(Y)

$= \nabla^2 l(\theta; Y) - \nabla^2 l(\theta; S)$.

Justification for controlling selective type I error rate.:

① larger power than data splitting (especially when selection event has reasonable probability)
(meaning it has some information of θ)

② better control of error rate than FWER.

③ the assumption of true FEM. is of no problem.

since it's not really asking for a parametric form of distribution.

e.g. for linear regression, the model M has inference ~~tag~~ target $\eta_M^T Y$, the only thing needed is still just P_Y.

(note: the robustness considered in this paper refers to F is wrong, rather than F ∉ M).

Methodology & results.

$\boxed{\text{Exponential Family}}$ $Y \sim f_\theta(y) = \exp\{\theta' T(y) - \psi(\theta)\} f_0(y)$

$\Rightarrow Y \mid Y \in A \sim \exp\{\theta' T(y) - \psi_A(\theta)\} f_0(y) 1_A(y)$ same parameter space & sufficient stat

consider the framework of $\theta \to \theta$, ζ nuisance parameters.

$$Y \sim f_{\theta,\zeta}(y) = \exp\{\theta' T(y) + \zeta' U(y) - \psi(\theta, \zeta)\} f_0(y)$$

* proposed method for arbitrary selection

otherwise, rejection sampling.

one always needs this step as the existence of multiple parameters in selected model.

Consider the law $L(T \mid U, A) \Rightarrow$ still exponential family. $U = P_\eta^T Y$.

Gaussian, stepwise AIC .... the conditional law has closed-form. (truncated normal).

Main result: under the assumed model, tests based on data splitting is inadmissible.
(in the sense that. "better" power can be obtained from a selective test).

other questions left: ① connection / explaination to simultaneous coverage.
② intuition of randomized selection? how is it dropping infinity length property?