Inference on Selected Subgroups in Clinical Trials (JASA, 2020)

Xinzhou Guo and Xuming He

Chao Cheng

School of Statistics and Management

Shanghai University of Finance and Economics, China

February 27, 2022

Outline

Outline

Introduction

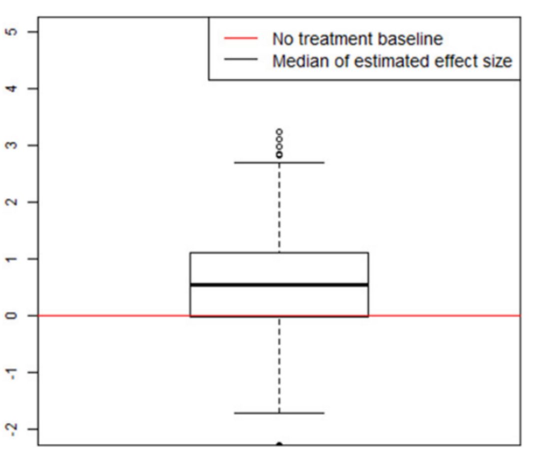
Clinical Trials

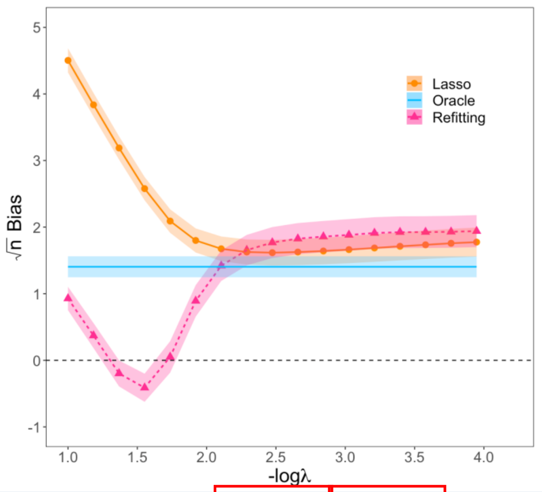
- Idealy: effective for the overall population
- Often reality: effective only for some subgroups, which are even identified ad-hoc.
  - subgroup identification
  - subgroup comfirmation: statistical inference and further clinical trial

Subgroup comfirmation fails frequently, e.g MONET1

- The identification of the subgroup leads to an overly optimistic evaluation, hence the selection bias.

Plug-in method is over-optimism





Introduction

Contribution

- Statistical inference on the best selected subgroup
- An approximately de-biased estimate of the subgroup treatment effect as well as a confidence bound
- Model-free, easy to compute, asymptotically sharp

Outline

Problem Setting

Assume we have n subjects and k possibly overlapped subgroups.

- k is fixed.
- β_i and $\hat{\beta}_i$ are the true and observed/estimated effect size of the ith subgroup, respectively.
- There are n_i subjects in the ith subgroup hence $\sum_{i=1}^{k} n_i \geq n$. Also n_i/n is bounded away from 0 and 1 as $n \to \infty$.
- Without loss of generality, we assume a larger value of β_i means a better treatment effect.

Quantities of Interest

Let $[k] = \{1, \cdots, k\}$ be the index set.

Quantities of interest

- The effect of the best selected subgroup:

$$\beta_{\hat{s}} \text{ , where } \hat{s} = \mathrm{argmax}_{i \in [k]} \hat{\beta}_i;$$

- The best subgroup effect:

$$\beta_{max} = \max_{i \in [k]} \beta_i.$$

The proposed method works for both of these quantities.

Bootstrap method

- The data consists of independent observations $\{D_j, Z_j\}$ from $j = 1, \cdots, n$ subjects where D_j represents the treatment and response measures and $Z_j \subset [k]$ indicates the subgroup relationship.
- $\left\{D_j^\star, Z_j^\star\right\}, j = 1, \cdots, n$ for the bootstrap sample.
- $\hat{\beta}_i^\star, i = 1, \cdots, k$ for the estimated treatment effect based on the bootstrap sample.

Lower confidence limit for β_{max}

Algorithm 1 Lower confidence limit for β_{max}

Require: Choose $r \in (0, 0.5)$.

1: For $i = 1, \cdots, k$ set $d_i = (1 - n^{r-0.5})\left(\hat{\beta}_{max} - \hat{\beta}_i\right)$.

2: **for** $b = 1, \cdots, B$ **do**

3: For bootstrap sample b; calculate the subgroup effect size $\hat{\beta}^{\star}_{i,b}$ and the modified bootstrap estimate of β_{max} as

$$\beta^{\star}_{max,mod,b} = \max_{i \in [k]}\left(\hat{\beta}^{\star}_{i,b} + d_i\right).$$

Then

$$T^{\star}_b = \sqrt{n}\left(\beta^{\star}_{max,mod,b} - \hat{\beta}_{max}\right).$$

4: **end for**

5: Let $c_\alpha = \text{quantile}\left(T^{\star}_b, 1 - \alpha\right)$. Then level $1 - \alpha$ lower confidence limit for β_{max} is

$$\hat{\beta}_{max} - c_\alpha / \sqrt{n}.$$

Assumptions

1.1 (Asymptotic normality)

$$\sqrt{n}\left(\hat{\beta}_1 - \beta_1, \cdots, \hat{\beta}_k - \beta_k\right)$$

is asymptotically normal.

1.2 (Bootstrap consistency) The asymptotic distribution of

$$\sqrt{n}\left(\hat{\beta}_1^\star - \hat{\beta}_1, \cdots, \hat{\beta}_k^\star - \hat{\beta}_k\right)$$

is the same as the limiting distribution in Assumption 1.1 in probability, which means the bootstrap consistency.

Theoretical Results

Theorem

Under Assumption 1.1 and 1.2, and for any $r \in (0, 0.5)$, *we have*

$$\sup_{x \in \mathcal{R}} \left| P^\star\left(\sqrt{n}\left(\beta^\star_{max,mod} - \hat{\beta}_{max}\right) \leq x\right) - P\left(\sqrt{n}\left(\hat{\beta}_{max} - \beta_{max}\right) \leq x\right) \right| \to 0$$

as $n \to \infty$, *in probability w.r.t.* P.

Corollary

Under Assumption 1.1 and 1.2, and for any $r \in (0, 0.5)$, *we have*

$$\sup_{x \in \mathcal{R}} \left| P^\star\left(\sqrt{n}\left(\beta^\star_{max,mod} - \hat{\beta}_{max}\right) \leq x\right) - P\left(\sqrt{n}\left(\hat{\beta}_{max} - \beta_{\hat{s}}\right) \leq x\right) \right| \to 0$$

as $n \to \infty$, *in probability w.r.t.* P.

Bias-Reduced Estimator

$\mathrm{E}\left(\sqrt{n}\left(\hat{\beta}_{max} - \beta_{max}\right)\right)$ is asymptotically equivalent to $\mathrm{E}\left(\max_{i \in H} \sqrt{n}\left(\hat{\beta}_i - \beta_i\right)\right)$ where $H = \{i : \beta_i = \beta_{max}\}$. This $O\left(1/\sqrt{n}\right)$ bias is nonnegligible for inference if $|H| > 1$.

Bias-reduced estimator

$$\hat{\beta}_{max,reduced} = \hat{\beta}_{max} - \mathrm{E}^\star\left(\beta^\star_{max,mod} - \hat{\beta}_{max}\right),$$

where $\mathrm{E}^\star$ denotes the expectaion under the bootstrap distribution.

Assumptions

2.1 (2nd moment bound)

$$\limsup_{n\to\infty} \mathrm{E}\left[\sqrt{n}\left(\hat{\beta}_i - \beta_i\right)\right]^2 < \infty,$$

for $i = 1, \cdots, k$.

2.2 (2nd bootstrap moment)

$$\limsup_{n\to\infty} \mathrm{E}^\star\left[\sqrt{n}\left(\hat{\beta}_i^\star - \hat{\beta}_i\right)\right]^2 < \infty,$$

in probability, for $i = 1, \cdots, k$.

2.3 (2nd bootstrap moment on population)

$$\limsup_{n\to\infty} \mathrm{E}\left\{\mathrm{E}^\star\left[\sqrt{n}\left(\hat{\beta}_i^\star - \hat{\beta}_i\right)\right]^2\right\} < \infty,$$

for $i = 1, \cdots, k$.

Theoretical Results

Theorem

Under Assumption 1.1, 1.2, 2.1 and 2.2, and for any $r \in (0, 0.5)$, we have

$$\left| \mathrm{E}^\star \left[\sqrt{n} \left(\beta^\star_{max,mod} - \hat{\beta}_{max} \right) \right] - \mathrm{E} \left[\sqrt{n} \left(\hat{\beta}_{max} - \beta_{max} \right) \right] \right| \to 0$$

as $n \to \infty$, in probability w.r.t P.

Corollary

Under Assumption 1.1, 1.2, 2.1 and 2.3, and for any $r \in (0, 0.5)$, we have

$$\left| \mathrm{E} \left\{ \mathrm{E}^\star \left[\sqrt{n} \left(\beta^\star_{max,mod} - \hat{\beta}_{max} \right) \right] \right\} - \mathrm{E} \left[\sqrt{n} \left(\hat{\beta}_{max} - \beta_{max} \right) \right] \right| \to 0$$

as $n \to \infty$.

Choice of the Tuning Parameter

- Let $A = \{r_1, \cdots, r_m\}$ denote a set of possible tuning parameters in the range of $(0, 0.5)$ with $r_1 < \cdots < r_m$.
- Choose r to minimize the mean square error between $\hat{\beta}_{max,reduced}(r)$ and β_{max}.
- An approximation of this mean square error is necessary since β_{max} is unknown.

Theorem

Under Assumption 1.1, 1.2, 2.1 and 2.3, and given the set A, there exists an integer N_A such that for any $n > N_A$ and $r \in A$, we have

$$\mathrm{E}\left[\hat{\beta}_{max,reduced,1}(r) - \beta_{max}\right]^2 = \min_{i \in [k]} \mathrm{E}\left[\hat{\beta}_{max,reduced,1}(r) - \beta_i\right]^2.$$

Choice of the Tuning Parameter

1. Partition the data into v (approximately) equalsized subsamples.
2. For $j = 1, \cdots, v$, use the jth subsample as the reference data and the rest as the training data.
    1. Use the training data to obtain the bias-reduced estimator $\hat{\beta}_{max,reduced,j}(r_l)$, with r_l as the tuning parameter, $l = 1, \cdots, m$.
    2. Use the reference data to estimate the effect $\hat{\beta}_{i,j}$ and its standard error $\hat{\sigma}_{i,j}$.
    3. Evaluate the accuracy

$$h_{i,j}(r_l) = \left(\hat{\beta}_{max,reduced,j}(r_l) - \hat{\beta}_{i,j}\right)^2 - \hat{\sigma}_{i,j}^2.$$

3. The tuning parameter is chosen to be

$$\underset{r_l}{\operatorname{argmin}} \left\{ \min_{i \in [k]} \left[\sum_{j=1}^{v} h_{i,j}(r_l) / v \right] \right\}.$$

Outline

Asymptotically Sharp Inference

When the best subgroup is post-hoc identified by searching over many (possibly infinitely many) subgroups.

- Let $\{S(c) : c \in \mathcal{D}\}$ denote the family of subgroups:
  - $S(c)$ is a subgroup indexed by c.
  - $\mathcal{D}$ is a compact set in a Euclidean space.
- $\beta(c)$ and $\hat{\beta}(c)$ represents the true and estimated effect size of $S(c)$, respectively.
- To distinguish from the previous section:
  - $\gamma_{max} = \sup_{c \in \mathcal{D}} \beta(c)$ and $\hat{\gamma}_{max} = \sup_{c \in \mathcal{D}} \hat{\beta}(c)$.
  - $\gamma_{\hat{s}}$ where $\hat{s} = \operatorname{argmax}_{c \in \mathcal{D}} \hat{\beta}(c)$.
- $\max_{c \in \mathcal{D}} \hat{\beta}(c)$ exists almost surely.

Lower Confidence Limit for γ_{max}

Algorithm 2 Lower confidence limit for γ_{max}

1: For $c \in \mathcal{D}$, let $d(c) = (1 - n^{r-0.5})\left(\hat{\gamma}_{max} - \hat{\beta}(c)\right)$

2: **for** $b = 1, \cdots, B$ **do**

3: For bootstrap sample b, calculate effect sizes $\hat{\beta}^{\star}_{b}(c)$ for $c \in \mathcal{D}$ and

$$\gamma^{\star}_{max,mod,b} = \sup_{c \in \mathcal{D}} \left(\hat{\beta}^{\star}_{b}(c) + d(c)\right).$$

Then

$$T^{\star}_{b} = \sqrt{n}\left(\gamma^{\star}_{max,mod} - \hat{\gamma}_{max}\right).$$

4: **end for**

5: Let $c_{\alpha} = \mathrm{quantile}\left(T^{\star}_{b}, 1 - \alpha\right)$, then the level α lower confidence limit is

$$\hat{\gamma}_{max} - c_{\alpha}/\sqrt{n}.$$

Assumptions

3.1 (Asymptotically Gaussian process)

$$\sqrt{n}\left(\hat{\beta}(\cdot) - \beta(\cdot)\right) \xrightarrow{\mathrm{d}} G(\cdot)$$

in $l_\infty(\mathcal{D})$, where $G(\cdot)$is a Gaussian process with continuous sample path in probability.

3.2 (Bootstrap consistency)

$$\sqrt{n}\left(\hat{\beta}^\star(\cdot) - \hat{\beta}(\cdot)\right) \xrightarrow{\mathrm{d}} G(\cdot)$$

in $l_\infty(\mathcal{D})$ in probability.

3.3 (Continuous mapping) $c \to \beta(c)$ is a continuous mapping in $\mathcal{D}$.

Theoretical Results

Theorem

Under Assumption 3.1–3.3 and for any $r \in (0, 0.5)$, we have as $n \to \infty$

$$\sup_{x \in \mathcal{R}} \left| P^{\star}\left(\sqrt{n}\left(\gamma^{\star}_{max,mod} - \hat{\gamma}_{max}\right) \leq x\right) - P\left(\sqrt{n}\left(\hat{\gamma}_{max} - \gamma_{max}\right) \leq x\right) \right| \to 0$$

and

$$\sup_{x \in \mathcal{R}} \left| P^{\star}\left(\sqrt{n}\left(\gamma^{\star}_{max,mod} - \hat{\gamma}_{max}\right) \leq x\right) - P\left(\sqrt{n}\left(\hat{\gamma}_{max} - \gamma_{\hat{s}}\right) \leq x\right) \right| \to 0$$

in probability w.r.t P.

A Bias-reduced Estimator of γ_{max}

Similarly, we have a bias-reduced estimator of γ_{max} as

$$\hat{\gamma}_{max,reduced} = \hat{\gamma}_{max} - \mathrm{E}^{\star}\left(\hat{\gamma}^{\star}_{max,mod} - \hat{\gamma}_{max}\right).$$

Outline

Date Generation

- Focus on the censored outcomes where the treatment effect is measured by the log hazard ratio from the proportional hazard model.
- D denote the treatment indicator. k denote the number of subgroups. Sample size $n = 200k$.
- Hazard function $\lambda(t) = \lambda_0(t) \exp(\beta_i D)$ for $i = 1, \cdots, k$. $\lambda_0(t)$ is Weibull$(1, 1)$.
- Subject has equal probability falling to each subgroup. The treatment assignment is also random with equal probability.
- The response is randomly censored from the right by C where $\log C \sim \mathrm{unif}(-1.25, 1.0)$.

$k = 2$, β_1 fixed at 0.

Empirical converage rate of 95% lower confidence bound of $\beta_{\hat{S}}$

	$r = 1/3$	$1/12$	$1/21$	$1/30$	Naive	Adaptive
$\beta_2 = 0$	0.933	0.950	0.952	0.952	0.896	0.943
1/10	0.926	0.945	0.947	0.947	0.912	0.936
2/10	0.928	0.949	0.951	0.951	0.910	0.939
3/10	0.941	0.957	0.959	0.959	0.919	0.947
4/10	0.939	0.955	0.956	0.957	0.927	0.945
5/10	0.952	0.965	0.965	0.966	0.934	0.953

$k = 2$, β_1 fixed at 0.

Average distance between 95% lower confidence bound and $\beta_{\hat{s}}$

	$r = 1/3$	1/12	1/21	1/30	Naive	Adaptive
$\beta_2 = 0$	0.248	0.265	0.266	0.266	0.213	0.258
1/10	0.252	0.269	0.270	0.270	0.218	0.262
2/10	0.267	0.285	0.288	0.287	0.233	0.277
3/10	0.290	0.311	0.313	0.314	0.258	0.302
4/10	0.301	0.326	0.328	0.329	0.273	0.313
5/10	0.310	0.339	0.342	0.343	0.286	0.323

$k = 2$, β_1 fixed at 0.

Empirical bias: $\hat{\beta}_{\hat{s}} - \beta_{\hat{s}}$

	$r = 1/3$	1/12	1/21	1/30	Naive	Adaptive
$\beta_2 = 0$	0.028	0.008	0.007	0.006	0.107	0.018
1/10	0.024	0.002	0.000	−0.001	0.100	0.012
2/10	0.005	−0.021	−0.022	−0.023	0.077	−0.008
3/10	−0.003	−0.045	−0.036	−0.037	0.061	−0.018
4/10	−0.018	−0.063	−0.065	−0.066	0.029	−0.042
5/10	−0.027	−0.067	−0.070	−0.071	0.022	−0.040

Various ks, β_is are all fixed at 0

The coverage rate of 95% l.c.b and empirical bias

		$r = 1/3$	1/12	1/21	1/30	Naive	Adaptive
$k = 2$	Cover	0.929	0.952	0.953	0.953	0.900	0.939
	Bias	0.029	0.006	0.004	0.004	0.105	0.014
6	Cover	0.891	0.941	0.943	0.945	0.739	0.930
	Bias	0.060	0.011	0.009	0.008	0.240	0.029
10	Cover	0.866	0.944	0.949	0.950	0.594	0.927
	Bias	0.066	0.009	0.006	0.005	0.290	0.031
12	Cover	0.860	0.946	0.950	0.950	0.543	0.925
	Bias	0.062	0.003	0.001	0.001	0.302	0.026

Data Generation

- D denote the treatment indicator and W denote a continuous variable used to define the subgroups. Sample size $n = 400$.
- Proportional hazard model $\lambda_0(t)\exp(b(W)D)$, where $\lambda_0(t)$ is Weibull$(1,1)$, $D \sim \mathrm{Bernoulli}(1, 0.5)$ and $W \sim \mathrm{Unif}(0, 80)$.
- $b(w) = \begin{cases} \beta_1 & w > 30 \\ \beta_2 & w \leq 30 \end{cases}$.
- Post-hoc identified subgroups: $S(c) = \{W \leq c\}$. Let $\beta(c)$ denote the subgroup effect of $S(c)$ for $c \in [30, 60]$.
- $\beta(c)$ is usually not equal to $b(c)$.

Post-Hoc Identified Subgroups, β_1 is fixed at 0.

Empirical coverage rate of 95% lower bound of $\gamma_{\hat{s}}$

	$r = 1/3$	1/12	1/21	1/30	Naive
$\beta_2 = 0$	0.947	0.961	0.962	0.962	0.872
1/10	0.960	0.972	0.972	0.972	0.879
2/10	0.958	0.966	0.967	0.967	0.890
3/10	0.959	0.969	0.970	0.970	0.895
4/10	0.962	0.968	0.968	0.968	0.906
5/10	0.964	0.972	0.973	0.973	0.901

MONET1

- Original trial data showed that for the East Asian subgroup the treatment has the hazard ratio of $HR = 0.669$ and $p_v = 0.0223$.
  - Data were proprietary.
  - Predefined subgroups were used in the identification.
  - No additional information on how many and which subgroups are considered.
- Synthesize the data that mimic MONET1 pattern.
  - 8 categorical variables are considered: East Asian patient, Stage IIIB, male, age over 65, etc.
  - There are 16 overlapping subgroups.
  - Subgroups are homogeneous and no treatment effect in any one.

Synthetic Data

The synthetic data is close to MONET1

	Hazard ratio	*p*-value
Synthetic data	0.663	0.019
MONET1	0.669	0.022

The bias-reduced estimator and 95% upper confidence bound

No. of subgroups	2	4	8	10	16	Naive
Upper bound	0.894	0.947	1.012	1.013	1.024	0.883
Hazard ratio	0.711	0.747	0.781	0.790	0.818	0.663

Synthetic Data

Empirical converage rate of the 95% upper bound

r	1/3	1/9	1/30	Naive	Adaptive
Coverage	0.917	0.946	0.950	0.805	0.935

- Naive method is over-optimism.

Outline

Acknowledgement

Thank you all for your attention!