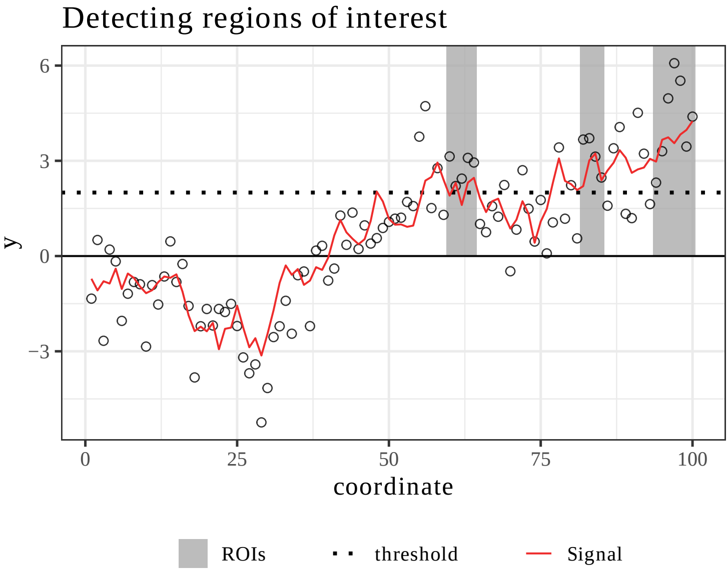
Bayes-optimal post-selection inference in spatial modeling

Spencer Woody* James Scott

Department of Statistics and Data Sciences
University of Texas at Austin

28 June 2018


Detecting regions of interest
6
3
0
−3
y
0
25
50
75
100
coordinate
ROIs
threshold
Signal

Example: Differentially methylated regions

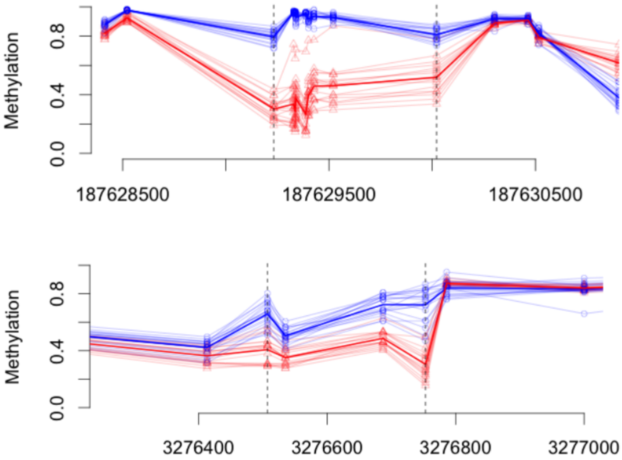


Figure: From Benjamini et al. (2016)

Other examples

- “Bump hunting” in high-energy physics problems to find energy regions of high event activity
- Detecting regions of neural activity in fMRI scans
- Finding environmental contamination areas

Our method

Spatial selection-adjusted FAB intervals

- Correctly adjusts for selection
- Retains nominal coverage across the parameter space
- Incorporates hierarchical modeling for “information borrowing”
- Bayes-optimal w.r.t. expected length of intervals

Set up

Observe a vector y associated with a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with a latent spatial signal θ,

$$(y_v | \theta_v) \sim \mathcal{N}(\theta_v, \sigma^2),\ v \in \mathcal{V}$$

Detecting regions of interest (ROIs)

Denote an ROI as R, found following a three-step process:

(i) *Smooth* the noisy observations (optional), e.g. with a linear smoother,

$$\tilde{y} := Hy.$$

(ii) *Threshold* the smoothed observations at some value t.

(iii) *Merge* together contiguous regions where smoothed observations fall above the threshold. With a chain graph,

$$R = (a, a+1, \ldots, b-1, b) \text{ s.t. } \tilde{y}_i > t \; \forall \; i \in R$$

Key fact: Restrict inference to R *conditioned* on

$$\tilde{y}_R > t \Leftrightarrow H_R y > t$$

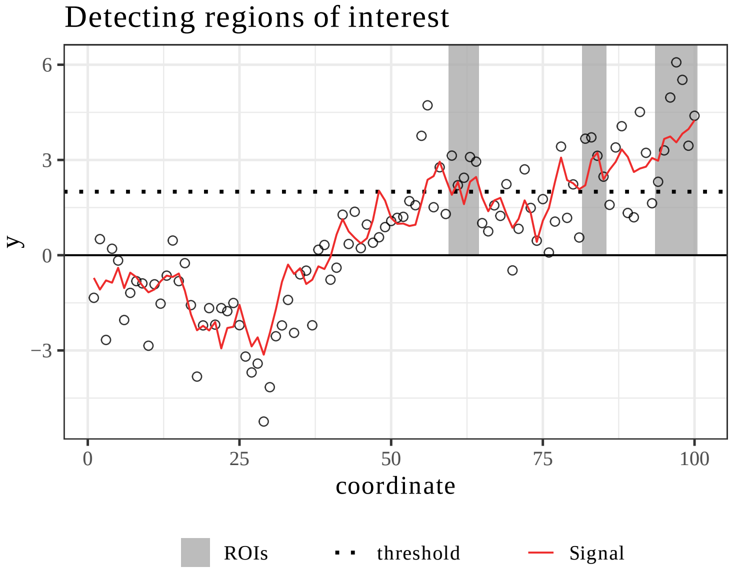


Figure: Threshold & merge

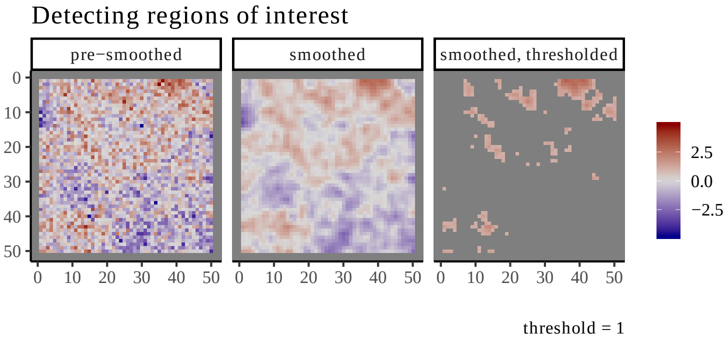


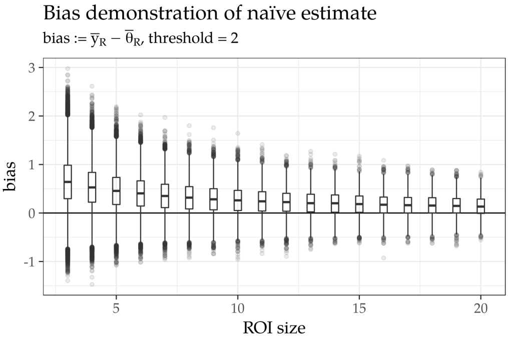
Figure: Smooth, threshold & merge

Target of inference

After detecting a region R, the goal is to provide inference for

$$\bar{\theta}_R := \frac{1}{|R|} \sum_{i \in R} \theta_i,$$

i.e. the *mean signal for the ROI*. The naïve estimate $\bar{y}_R$ will be *biased upwards*.



Selection-adjusted inference

Appropriate inference must condition on the selection event. The selection-adjusted likelihood is

$$f_S(y \mid \theta) = \frac{f(y \mid \theta) \cdot \mathbf{1}(y \in S)}{\int_{y \in S} f(y \mid \theta) dy},$$

or equivalently, the likelihood truncated to the selection event S.

See, e.g.,

- Yekutieli (2012), selection-adjusted Bayesian inference
- Fithian, Sun & Taylor (2014), selective frequentist performance

In our case,

$$f_S(y \mid \theta) = \frac{\mathcal{N}(y \mid \theta, \sigma^2 \mathcal{I}) \cdot \mathbf{1}(H_R y > t)}{\int_{H_R y > t} \mathcal{N}(y \mid \theta, \sigma^2 \mathcal{I}) dy}.$$

Bayesian inference

We use the centered ICAR prior for θ,

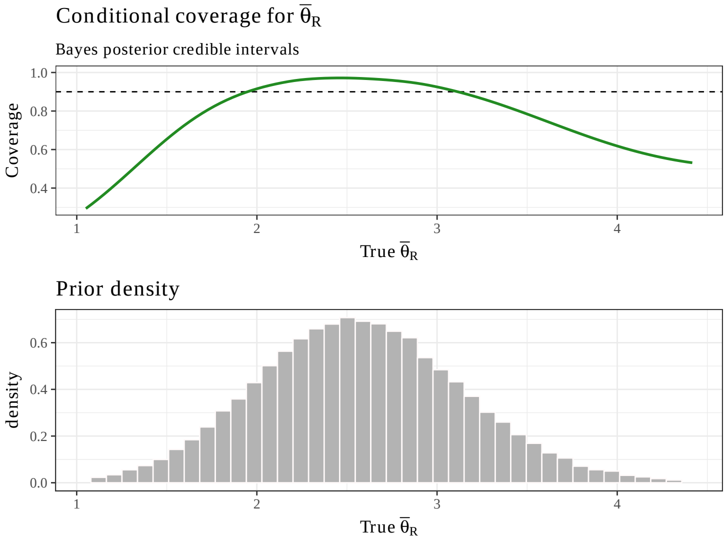
$$\pi(\theta) \propto \exp\left[-\frac{1}{2\tau^2}\sum_{(v,w)\in\mathcal{E}}(\theta_v - \theta_w)^2\right]\cdot\exp\left[-\frac{1}{2\lambda^2}\bar{\theta}^2\right],$$

where $\bar{\theta}$ is the mean of the components of θ.

The sampling model is

$$(y_v \mid \theta_v) \sim \mathcal{N}(\theta_v, \sigma^2),\ v \in \mathcal{V}$$

for the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.


Conditional coverage for $\overline{\theta}_R$
Bayes posterior credible intervals
Coverage
1.0
0.8
0.6
0.4
1
2
3
4
True $\overline{\theta}_R$
Prior density
density
0.6
0.4
0.2
0.0
1
2
3
4
True $\overline{\theta}_R$

Selection-adjusted confidence interval

Construct hypothesis tests around the the sampling distribution for the statistic $f_S(\bar{y}_R|\bar{\theta}_R)$. See Benjamini et al. (2016).

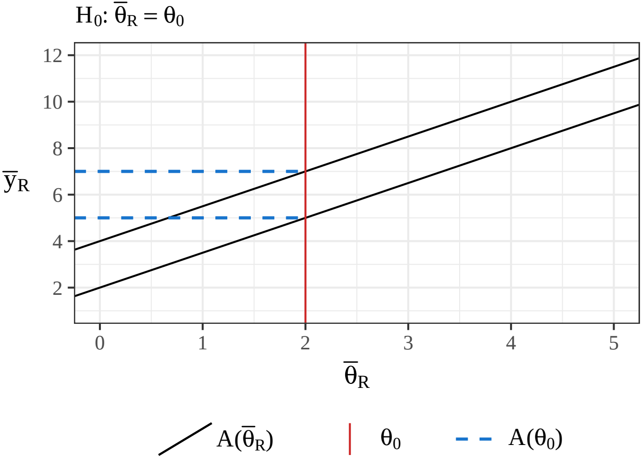
For $F_S(\bar{y}_R;\ \bar{\theta}_R)$ the CDF for $\bar{y}_R \sim f_S(\bar{y}_R;\ \bar{\theta}_R)$, the acceptance region for the α-level uniformly most powerful (UMP) test of $H_0 : \bar{\theta}_R = \theta_0$ is

$$\begin{aligned} A(\theta_0) &= \{\bar{y}_R : F_S^{-1}(\alpha/2;\ \theta_0) \le \bar{y}_R \le F_S^{-1}(1-\alpha/2;\ \theta_0)\} \\ &= \{\bar{y}_R : L(\theta_0) \le \bar{y}_R \le U(\theta_0)\}. \end{aligned}$$

Inversion yields the $1-\alpha$-level **universally most accurate unbiased (UMAU)** confidence interval for $\bar{\theta}_R$,

$$C(\bar{y}_R) = \{\bar{\theta}_R : \bar{y}_R \in A(\bar{\theta}_R)\}.$$

Inverting a family of tests

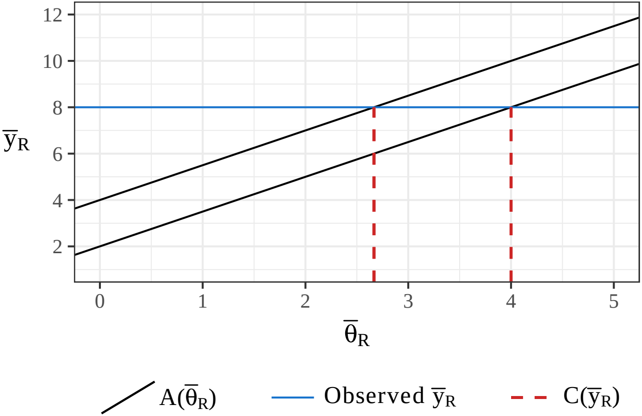


$$A(\theta_0) = (L(\theta_0), U(\theta_0))$$

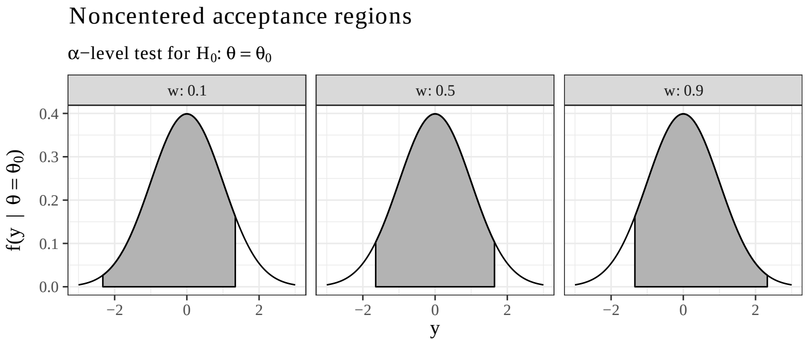
$$L(\theta_0) = F_S^{-1}(\alpha/2;\ \theta_0), \quad U(\theta_0) = F_S^{-1}(1-\alpha/2;\ \theta_0)$$

Inverting a family of tests

H_0: $\bar{\theta}_R = \theta_0$



$$A(\theta_0) = (L(\theta_0), U(\theta_0))$$
$$C(\bar{y}_R) = \{\bar{\theta}_R : \bar{y}_R \in A(\bar{\theta}_R)\}$$


Noncentered acceptance regions
α−level test for H0: θ = θ0
w: 0.1
w: 0.5
w: 0.9
f(y | θ = θ0)
0.4
0.3
0.2
0.1
0.0
−2
0
2
y

Background: FAB procedure

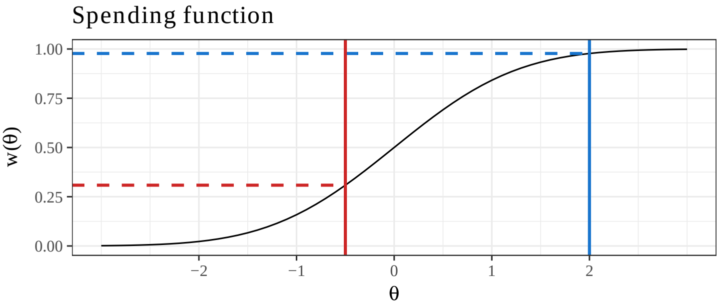
In general, inversion of

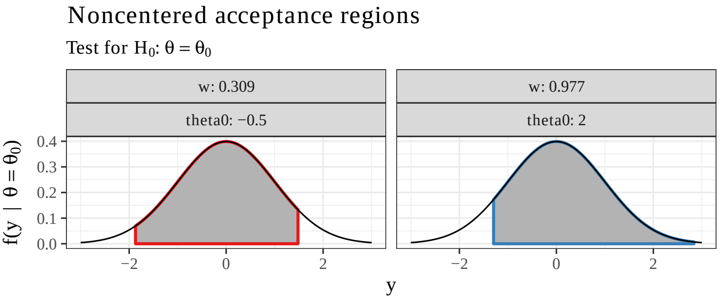
$$\begin{aligned} A_w(\theta_0) &= \{y : F^{-1}(\alpha w;\ \theta_0) \leq \bar{y} \leq F^{-1}(\alpha w + 1 - \alpha;\ \theta_0)\} \\ &= \{y : L_w(\theta_0) \leq y \leq U_w(\theta_0)\} \end{aligned}$$

for any $0 \leq w \leq 1$ will yield a confidence interval procedure

$$C_w(y) = \{\theta : y \in A_w(\theta)\}$$

which retains nominal coverage for θ.


Spending function
w(θ)
1.00
0.75
0.50
0.25
0.00
−2
−1
0
1
2
θ


Noncentered acceptance regions
Test for H0: θ = θ0
w: 0.309
theta0: −0.5
w: 0.977
theta0: 2
f(y | θ = θ0)
0.4
0.3
0.2
0.1
0.0
−2
0
2
y

FAB procedure

Frequentist assisted by Bayes (FAB) procedure

- Key idea from Pratt (1963)
- Extended by Yu and Hoff (2018) for confidence intervals for group-level means

Goal: Find $w(\theta)$ which minimizes the expected size of the confidence set under a prior $\pi(\theta)$.

Define the **risk** of a confidence interval procedure to be its expected Lebesgue measure,

$$L(\theta, w) = \int \int \mathbf{1}(y \in A_w(\tilde{\theta})) f(y|\theta) d\tilde{\theta} dy.$$

FAB procedure

Introduce a prior $\theta \sim \pi(\theta)$. Then the **Bayes risk** for the confidence interval procedure is

$$\begin{aligned}
L(\pi, w(\theta)) &= \int L(\theta, w(\theta))\pi(\theta)d\theta \\
&= \int \left[\int \int \mathbf{1}(y \in A_w(\tilde{\theta}))f(y|\theta)d\tilde{\theta}dy\right] \pi(\theta)d\theta \\
&= \int \left[\int \int \mathbf{1}(y \in A_w(\tilde{\theta}))f(y|\theta)\pi(\theta)dyd\theta\right] d\tilde{\theta} \\
&= \int \Pr(Y \in A_w(\tilde{\theta}))d\tilde{\theta}.
\end{aligned}$$

FAB procedure

Let $M(y)$ be the CDF for the marginal distribution $m(y) = \int f(y|\theta)\pi(\theta)$.

The Bayes-optimal interval is found by choosing $w(\theta)$ to minimize the objective function

$$\begin{aligned} \Pr(Y \in A(\theta)) &= M(U_w(\theta)) - M(L_w(\theta)) \\ &= M\left[F^{-1}(\alpha w + 1 - \alpha;\ \theta)\right] - M\left[F^{-1}(\alpha w;\ \theta)\right]. \end{aligned}$$

Spatial selection-adjusted FAB procedure

(i) Specify the truncated likelihood $f_S(\bar{y}_R;\ \bar{\theta}_R)$ and spatial prior $\pi(\theta)$

(ii) Construct the spending function by solving

$$w(\bar{\theta}_R) = \arg\min_w M_S\left[F_S^{-1}(\alpha w + 1 - \alpha;\ \bar{\theta}_R)\right] - M_S\left[F_S^{-1}(\alpha w;\ \bar{\theta}_R)\right]$$

(iii) Invert the family of tests specifed by $w(\bar{\theta}_R)$ and $f_S(\bar{y}_R;\ \bar{\theta}_R)$,

$$A_w(\bar{\theta}_R) = \{y : F_S^{-1}(\alpha w(\bar{\theta}_R);\ \bar{\theta}_R) \le y \le F_S^{-1}(\alpha w(\bar{\theta}_R) + 1 - \alpha;\ \bar{\theta}_R)\}.$$

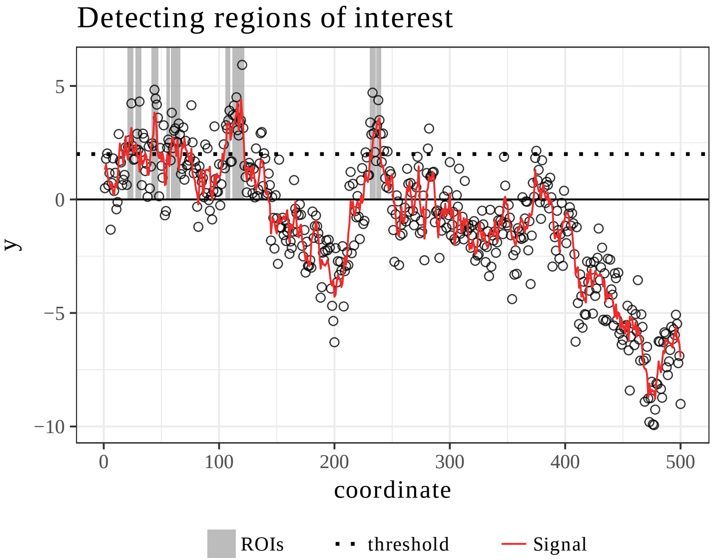
Use this to give Bayes-optimal selection-adjusted confidence regions for $\bar{\theta}_R$ which retain coverage for entire parameter space.

Simulation study

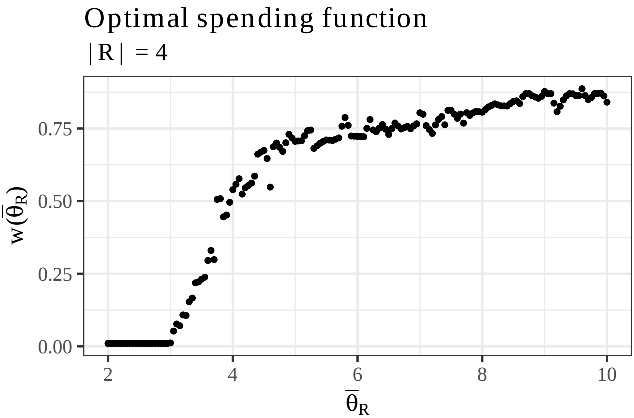
50,000 simulations performed as follows:

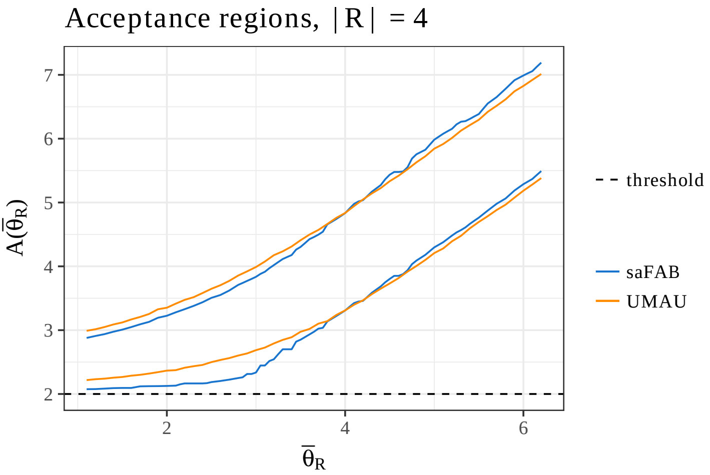
- Chain graph of length 500
- θ generated from ICAR prior with $\tau^2 = 0.25$ and $\lambda^2 = 1$
- $(y|\theta) \sim \mathcal{N}(\theta, \mathcal{I})$
- Threshold for detecting ROIs set to $t = 2$
- No smoothing step involved ($H = \mathcal{I}$)

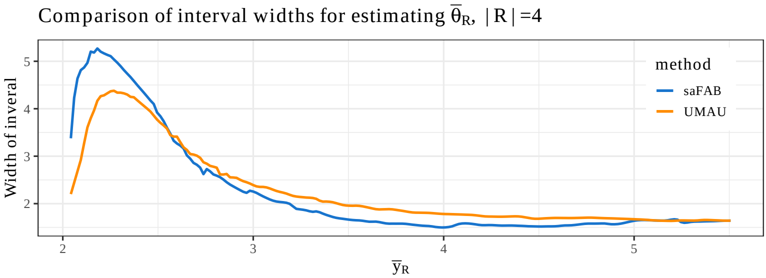
$F_S(\bar{y}_R \mid \bar{\theta}_R)$ and $M_S(\bar{y}_R)$ are approximated via Monte Carlo.

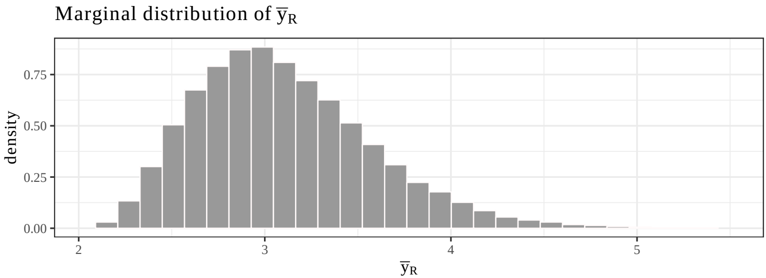

Detecting regions of interest
5
0
−5
−10
y
0
100
200
300
400
500
coordinate
ROIs
threshold
Signal

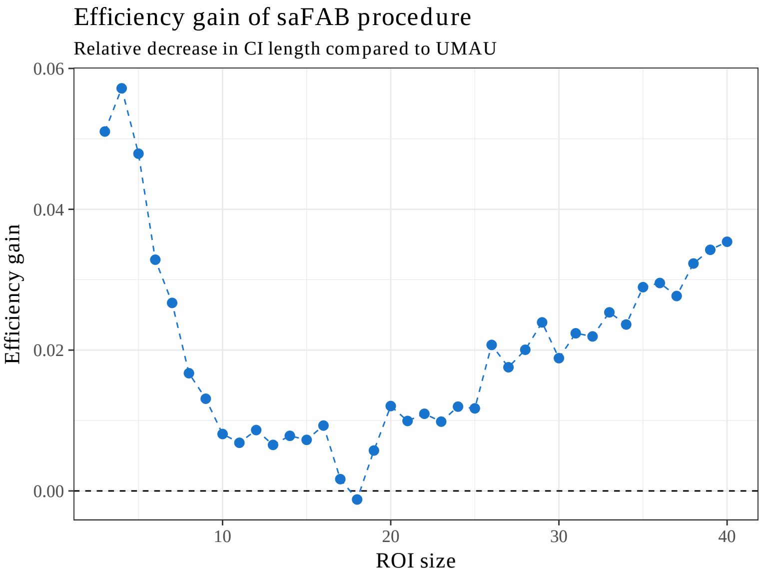
The spending function

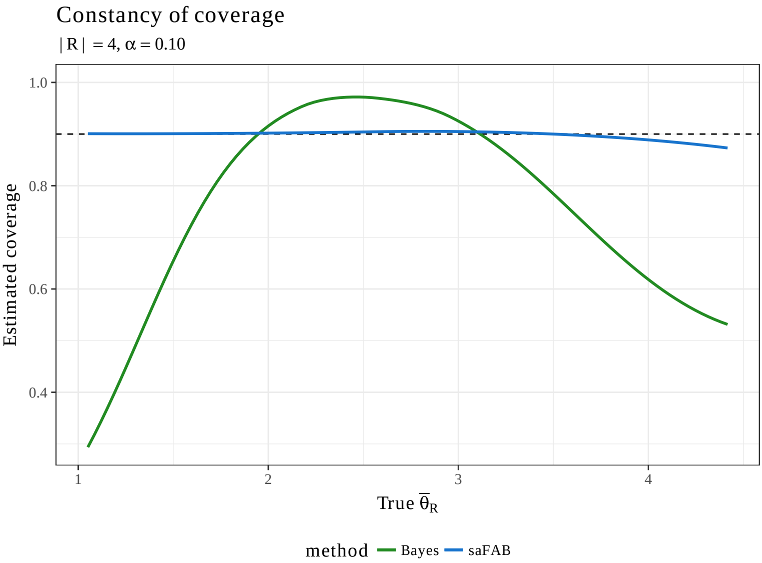



Acceptance regions, |R| = 4
$A(\overline{\theta}_R)$
7
6
5
4
3
2
2
4
6
$\overline{\theta}_R$
threshold
saFAB
UMAU


Comparison of interval widths for estimating $\bar{\theta}_R$, $|R|=4$
Width of inveral
method
saFAB
UMAU
$\bar{y}_R$
2
3
4
5


Marginal distribution of $\bar{y}_R$
density
0.00
0.25
0.50
0.75
$\bar{y}_R$
2
3
4
5


Efficiency gain of saFAB procedure
Relative decrease in CI length compared to UMAU
0.06
0.04
0.02
0.00
Efficiency gain
10
20
30
40
ROI size


Constancy of coverage
|R| = 4, α = 0.10
1.0
0.8
0.6
0.4
Estimated coverage
1
2
3
4
True $\bar{\theta}_R$
method
Bayes
saFAB

Conclusion

Slides:

spencerwoody.github.io/talks

Email:

spencer.woody@utexas.edu

References I

Yuval Benjamini, Jonathan Taylor, and Rafael A Irizarry. Selection corrected statistical inference for region detection with high-throughput assays. *bioRxiv*, 2016. doi: 10.1101/082321. URL https://www.biorxiv.org/content/early/2016/10/23/082321.

W. Fithian, D. Sun, and J. Taylor. Optimal Inference After Model Selection. *ArXiv e-prints*, October 2014.

John W. Pratt. Shorter confidence intervals for the mean of a normal distribution with known variance. *Ann. Math. Statist.*, 34(2):574–586, 06 1963. doi: 10.1214/aoms/1177704170. URL https://doi.org/10.1214/aoms/1177704170.

Daniel Yekutieli. Adjusted Bayesian inference for selected parameters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3): 515–541, 2012. doi: 10.1111/j.1467-9868.2011.01016.x. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2011.01016.x.

C Yu and P D Hoff. Adaptive multigroup confidence intervals with constant coverage. *Biometrika*, 105(2):319–335, 2018. doi: 10.1093/biomet/asy009. URL http://dx.doi.org/10.1093/biomet/asy009.