Controlled Discovery and Localization of Astronomical Point Sources via Bayesian Linear Programming (BLiP)

Lucas Janson

Harvard University Department of Statistics

RISE-CHASC Workshop, Aug 3, 2022

Coauthor

Asher Spector (First-year PhD student at Stanford Statistics)

Astronomical point source detection

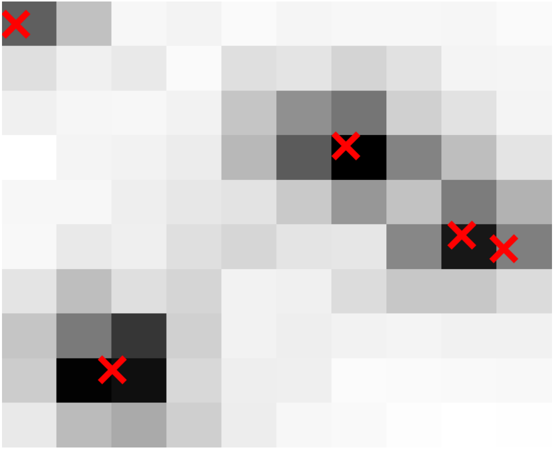


Figure: Cartoon of partial point source data

Informal problem statement

Want method that looks at the data and outputs regions $G_1, \ldots, G_R \subset \mathcal{L}$ so as to:

Informal problem statement

Want method that looks at the data and outputs regions $G_1, \ldots, G_R \subset \mathcal{L}$ so as to:

$$\begin{aligned} \max \quad & \mathbb{E}\left[\text{Power}(G_1, \ldots, G_R)\right] \\ \text{s.t.} \quad & \text{FDR} := \mathbb{E}\left[\frac{\#\{G_r \text{ containing no signal}\}}{\max(1, R)}\right] \leq q, \\ & G_1, \ldots, G_R \subset \mathcal{L} \text{ are disjoint.} \end{aligned}$$

Informal problem statement

Want method that looks at the data and outputs regions $G_1, \dots, G_R \subset \mathcal{L}$ so as to:

$$\begin{aligned} \max \quad & \mathbb{E}\left[\text{Power}(G_1, \dots, G_R)\right] \\ \text{s.t.} \quad & \text{FDR} := \mathbb{E}\left[\frac{\#\{G_r \text{ containing no signal}\}}{\max(1, R)}\right] \leq q, \\ & G_1, \dots, G_R \subset \mathcal{L} \text{ are disjoint.} \end{aligned}$$

What does high Power() look like?

Informal problem statement

Want method that looks at the data and outputs regions $G_1, \ldots, G_R \subset \mathcal{L}$ so as to:

$$\begin{aligned} \max \quad & \mathbb{E}\left[\text{Power}(G_1, \ldots, G_R)\right] \\ \text{s.t.} \quad & \text{FDR} := \mathbb{E}\left[\frac{\#\{G_r \text{ containing no signal}\}}{\max(1, R)}\right] \leq q, \\ & G_1, \ldots, G_R \subset \mathcal{L} \text{ are disjoint.} \end{aligned}$$

What does high Power() look like?

- As many (true) discovered regions G_r as possible

Informal problem statement

Want method that looks at the data and outputs regions $G_1, \ldots, G_R \subset \mathcal{L}$ so as to:

$$
\begin{aligned}
\max \quad & \mathbb{E}\left[\text{Power}(G_1, \ldots, G_R)\right] \\
\text{s.t.} \quad & \text{FDR} := \mathbb{E}\left[\frac{\#\{G_r \text{ containing no signal}\}}{\max(1, R)}\right] \le q, \\
& G_1, \ldots, G_R \subset \mathcal{L} \text{ are disjoint.}
\end{aligned}
$$

What does high Power() look like?

- As many (true) discovered regions G_r as possible
- Discovered regions G_r should be as small as possible

Informal problem statement

Want method that looks at the data and outputs regions $G_1, \ldots, G_R \subset \mathcal{L}$ so as to:

$$\begin{aligned} \max \quad & \mathbb{E}\left[\text{Power}(G_1, \ldots, G_R)\right] \\ \text{s.t.} \quad & \text{FDR} := \mathbb{E}\left[\frac{\#\{G_r \text{ containing no signal}\}}{\max(1, R)}\right] \le q, \\ & G_1, \ldots, G_R \subset \mathcal{L} \text{ are disjoint.} \end{aligned}$$

What does high Power() look like?

- As many (true) discovered regions G_r as possible
- Discovered regions G_r should be as small as possible

Existing work: no formalization of what "power" means, so cannot optimize it

Valuing discovered regions

Define a weighting function $w(G)$ that measures value of discovering a group

Valuing discovered regions

Define a weighting function $w(G)$ that measures value of discovering a group

- Should penalize larger groups

Valuing discovered regions

Define a weighting function $w(G)$ that measures value of discovering a group

- Should penalize larger groups
- A canonical choice is inverse-size weighting: $w(G) = 1/|G|$

Valuing discovered regions

Define a weighting function $w(G)$ that measures value of discovering a group

- Should penalize larger groups
- A canonical choice is inverse-size weighting: $w(G) = 1/|G|$
- If G are circles on a sky survey, $w(G) = 1/\text{radius}(G)$ natural

Valuing discovered regions

Define a weighting function $w(G)$ that measures value of discovering a group

- Should penalize larger groups
- A canonical choice is inverse-size weighting: $w(G) = 1/|G|$
- If G are circles on a sky survey, $w(G) = 1/\text{radius}(G)$ natural
- If want to precisely know the *number* of sources in each G:

Valuing discovered regions

Define a weighting function $w(G)$ that measures value of discovering a group

- Should penalize larger groups
- A canonical choice is inverse-size weighting: $w(G) = 1/|G|$
- If G are circles on a sky survey, $w(G) = 1/\text{radius}(G)$ natural
- If want to precisely know the *number* of sources in each G:
    - Pair each G with a $J \subset \mathbb{N}$ representing possible numbers of sources in G

Valuing discovered regions

Define a weighting function $w(G)$ that measures value of discovering a group

- Should penalize larger groups
- A canonical choice is inverse-size weighting: $w(G) = 1/|G|$
- If G are circles on a sky survey, $w(G) = 1/\text{radius}(G)$ natural
- If want to precisely know the *number* of sources in each G:
  - Pair each G with a $J \subset \mathbb{N}$ representing possible numbers of sources in G
  - Set $w(G, J) = 1/|J|$ (we call this the "separation-based" weight function)

Optimizing resolution-adjusted power

Sum weights of true rejections to get Power():

$$\text{Power}(G_1, \ldots, G_R) = \sum_{r=1}^{R} I_{G_r} w(G_r),$$

where I_G is the indicator that G contains a signal (i.e., is a true discovery)

Optimizing resolution-adjusted power

Sum weights of true rejections to get Power():

$$\text{Power}(G_1, \ldots, G_R) = \sum_{r=1}^{R} I_{G_r} w(G_r),$$

where I_G is the indicator that G contains a signal (i.e., is a true discovery)

Then the power of a Bayesian method that discovers $G_1, \ldots, G_R$ is

$$\mathbb{E}[\text{Power}(G_1, \ldots, G_R) \mid \text{Data}] \;=\; \mathbb{E}\left[\sum_{r=1}^{R} I_{G_r} w(G_r) \,\middle|\, \text{Data}\right] \;=\; \sum_{G \subseteq \mathcal{L}} p_G w(G) x_G,$$

- $x_G \in \{0, 1\}$ is indicator that G is one of the method's discoveries
- $p_G = \mathbb{E}[I_G \mid \text{Data}]$ is *posterior inclusion probability* (PIP)

Posterior optmization

Optimal Bayesian method would solve:

$$\max_{\{x_G\}_{G \subseteq \mathcal{L}}} \text{Power} = \sum_G p_G w(G) x_G$$

$$\text{s.t.} \qquad \text{FDR} := \mathbb{E}\left[\frac{\#\{\text{false discoveries}\}}{\#\{\text{discoveries}\}} \,\middle|\, \text{Data}\right] = \frac{\sum_G (1 - p_G) x_G}{\sum_G x_G} \leq q$$

$$\sum_{G \ni \ell} x_G \leq 1 \quad \forall \ell \qquad \text{(all discoveries are disjoint)}$$

$$x_G \in \{0, 1\} \quad \forall G.$$

Posterior optmization

Optimal Bayesian method would solve:

$$
\begin{aligned}
\max_{\{x_G\}_{G \subseteq \mathcal{L}}} \quad & \text{Power} = \sum_G p_G w(G) x_G \\
\text{s.t.} \quad & \text{FDR} := \mathbb{E}\left[\frac{\#\{\text{false discoveries}\}}{\#\{\text{discoveries}\}} \;\middle|\; \text{Data} \right] = \frac{\sum_G (1 - p_G) x_G}{\sum_G x_G} \leq q \\
& \sum_{G \ni \ell} x_G \leq 1 \quad \forall \ell \qquad \text{(all discoveries are disjoint)} \\
& x_G \in \{0, 1\} \quad \forall G.
\end{aligned}
$$

- Problem is large and non-convex

Posterior optmization

Optimal Bayesian method would solve:

$$\max_{\{x_G\}_{G\subseteq\mathcal{L}}} \text{Power} = \sum_G p_G w(G) x_G$$

$$\text{s.t.} \quad \text{FDR} := \mathbb{E}\left[\frac{\#\{\text{false discoveries}\}}{\#\{\text{discoveries}\}} \,\middle|\, \text{Data}\right] = \frac{\sum_G (1-p_G)x_G}{\sum_G x_G} \leq q$$

$$\sum_{G\ni\ell} x_G \leq 1 \quad \forall \ell \qquad \text{(all discoveries are disjoint)}$$

$$x_G \in \{0,1\} \quad \forall G.$$

- Problem is large and non-convex
- But can be approximated by a linear program (fast!)

Posterior optmization

Optimal Bayesian method would solve:

$$\max_{\{x_G\}_{G\subseteq\mathcal{L}}} \text{Power} = \sum_G p_G w(G) x_G$$

$$\text{s.t.} \qquad \text{FDR} := \mathbb{E}\left[\frac{\#\{\text{false discoveries}\}}{\#\{\text{discoveries}\}} \,\middle|\, \text{Data}\right] = \frac{\sum_G (1-p_G)x_G}{\sum_G x_G} \leq q$$

$$\sum_{G\ni\ell} x_G \leq 1 \quad \forall\ell \qquad \text{(all discoveries are disjoint)}$$

$$x_G \in \{0,1\} \quad \forall G.$$

- Problem is large and non-convex
- But can be approximated by a linear program (fast!)
- Solution provably controls FDR and has computable bound on suboptimality

Posterior optmization

Optimal Bayesian method would solve:

$$
\begin{aligned}
\max_{\{x_G\}_{G\subseteq\mathcal{L}}} \quad & \text{Power} = \sum_G p_G w(G) x_G \\
\text{s.t.} \quad & \text{FDR} := \mathbb{E}\left[\frac{\#\{\text{false discoveries}\}}{\#\{\text{discoveries}\}} \,\middle|\, \text{Data} \right] = \frac{\sum_G (1-p_G) x_G}{\sum_G x_G} \le q \\
& \sum_{G\ni\ell} x_G \le 1 \quad \forall \ell \qquad \text{(all discoveries are disjoint)} \\
& x_G \in \{0,1\} \quad \forall G.
\end{aligned}
$$

- Problem is large and non-convex
- But can be approximated by a linear program (fast!)
- Solution provably controls FDR and has computable bound on suboptimality
- Only search over $G =$ circles (of any radius and center)

Bayesian Linear Programming (BLiP)

Just needs posterior inclusion probabilities p_G as input

- From any Bayesian algorithm for computing/approximating the posterior,
- E.g., MCMC (average over posterior samples whether G contains a signal)
- E.g., variational inference

Bayesian Linear Programming (BLiP)

Just needs posterior inclusion probabilities p_G as input

- From any Bayesian algorithm for computing/approximating the posterior,
- E.g., MCMC (average over posterior samples whether G contains a signal)
- E.g., variational inference

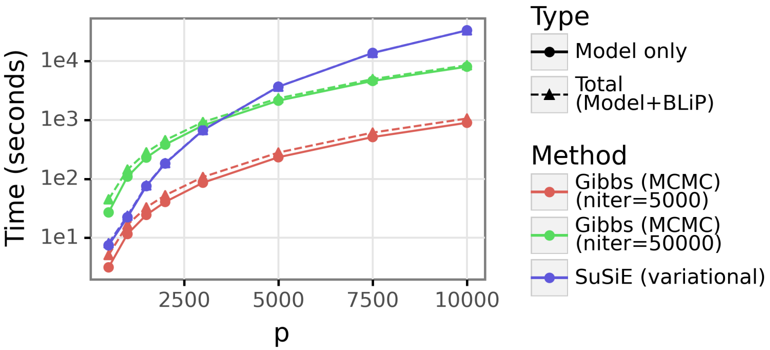


Figure: p denotes dimension of linear model being fit, with $n = p/2$

Point-source detection

100 x 100 pixel sub-image of Messier 2 star cluster from Sloan Digital Sky Survey

- Ground truth available from much more powerful Hubble Space Telescope

Point-source detection

100 x 100 pixel sub-image of Messier 2 star cluster from Sloan Digital Sky Survey

- Ground truth available from much more powerful Hubble Space Telescope
- StarNet (Liu et al., 2021): variational approx.'s MAPs + 0.5-pixel slack

Point-source detection

100 x 100 pixel sub-image of Messier 2 star cluster from Sloan Digital Sky Survey

- Ground truth available from much more powerful Hubble Space Telescope
- StarNet (Liu et al., 2021): variational approx.'s MAPs + 0.5-pixel slack
- *continuous* space of locations $\mathcal{L}$: BLiP takes < 10 min for 15 FDRs

Point-source detection

100 x 100 pixel sub-image of Messier 2 star cluster from Sloan Digital Sky Survey

- Ground truth available from much more powerful Hubble Space Telescope
- StarNet (Liu et al., 2021): variational approx.'s MAPs + 0.5-pixel slack
- *continuous* space of locations $\mathcal{L}$: BLiP takes < 10 min for 15 FDRs

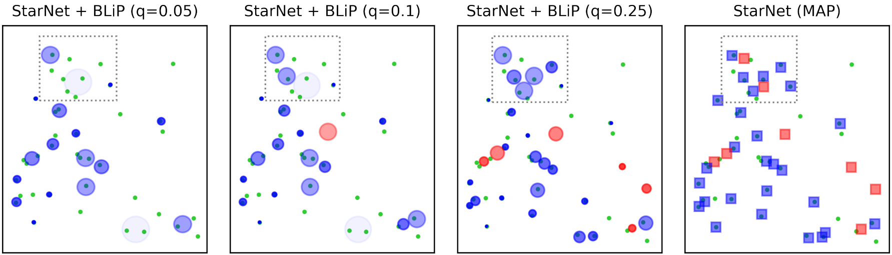
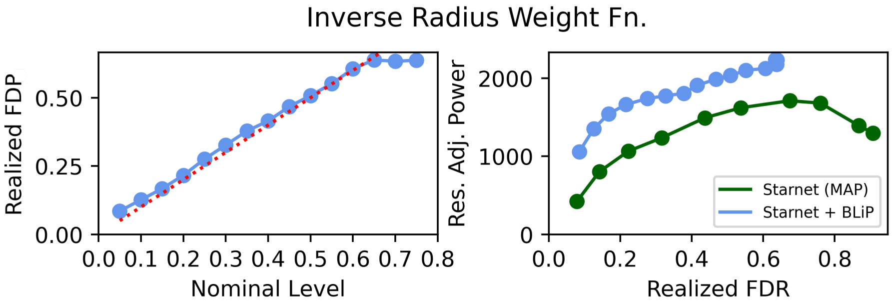
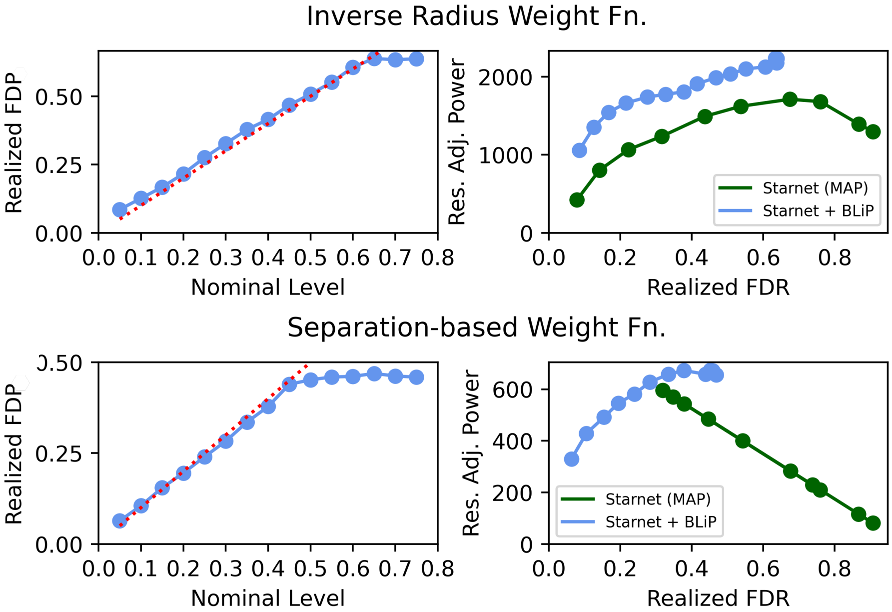


Figure: 20 x 20 pixel sub-image; green dots = ground truth, red regions = false discoveries, blue regions = true discoveries

Point-source detection (contd)



Point-source detection (contd)



Conclusion

BLiP is a powerful, principled, efficient, and flexible method for resolution-adaptive signal discovery

- Flexible objective function

Conclusion

BLiP is a powerful, principled, efficient, and flexible method for resolution-adaptive signal discovery

- Flexible objective function
- Provable error control and verifiable near-optimality

Conclusion

BLiP is a powerful, principled, efficient, and flexible method for resolution-adaptive signal discovery

- Flexible objective function
- Provable error control and verifiable near-optimality
- Substantial power gains in minutes on point-source detection

Conclusion

BLiP is a powerful, principled, efficient, and flexible method for resolution-adaptive signal discovery

- Flexible objective function
- Provable error control and verifiable near-optimality
- Substantial power gains in minutes on point-source detection
- Software packages `pyblip` (Python) and `blipr` (R)

Conclusion

BLiP is a powerful, principled, efficient, and flexible method for resolution-adaptive signal discovery

- Flexible objective function
- Provable error control and verifiable near-optimality
- Substantial power gains in minutes on point-source detection
- Software packages `pyblip` (Python) and `blipr` (R)
- Potential for other signal discovery problems with spatial structure

Conclusion

BLiP is a powerful, principled, efficient, and flexible method for resolution-adaptive signal discovery

- Flexible objective function
- Provable error control and verifiable near-optimality
- Substantial power gains in minutes on point-source detection
- Software packages `pyblip` (Python) and `blipr` (R)
- Potential for other signal discovery problems with spatial structure

paper available at: `https://arxiv.org/abs/2203.17208`
all code posted at: `https://github.com/amspector100`

Conclusion

BLiP is a powerful, principled, efficient, and flexible method for resolution-adaptive signal discovery

- Flexible objective function
- Provable error control and verifiable near-optimality
- Substantial power gains in minutes on point-source detection
- Software packages `pyblip` (Python) and `blipr` (R)
- Potential for other signal discovery problems with spatial structure

paper available at: `https://arxiv.org/abs/2203.17208`
all code posted at: `https://github.com/amspector100`

Thank you!
`http://lucasjanson.fas.harvard.edu`
ljanson@fas.harvard.edu

References

Katsevich, E., Sabatti, C., and Bogomolov, M. (2021). Filtering the rejection set while preserving false discovery rate control. *Journal of the American Statistical Association*, 0(0):1–12.

Lee, Y., Luca, F., Pique-Regi, R., and Wen, X. (2018). Bayesian multi-snp genetic association analysis: Control of fdr and use of summary statistics. *bioRxiv*.

Liu, R., McAuliffe, J. D., and Regier, J. (2021). Variational inference for deblending crowded starfields.

Wang, G., Sarkar, A., Carbonetto, P., and Stephens, M. (2020). A simple new approach to variable selection in regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(5):1273–1300.

Yekutieli, D. (2008). Hierarchical false discovery rate-controlling methodology. *Journal of the American Statistical Association*, 103(481):309–316.