Revisiting Subset Selection

David J. Eckman
Matthew Plumlee
Barry L. Nelson

Northwestern University, IEMS

Winter Simulation Conference
December 17, 2020

Supported by NSF grants DMS-1854562 and DMS-1953111.

Introduction

Finite number of simulated **systems** under consideration.

Each system has an associated scalar **performance** (response).

- **Examples:** *expected* cost, *expected* completion time.
- Estimated by running replications of the simulation.

Ranking and Selection

Select the system(s) with the best performance.

Subset Selection

Motivation

Many systems may be clearly inferior. We would like to *cheaply* remove them, retaining a **subset** of near-optimal systems.

Approach:

1. Take replications at *all* systems.
2. Infer the *relative* performance of each system.
3. If inferior, remove from consideration (i.e., *screen out*).

Returned **subset** can be used for

- post hoc analysis, or
- input to a second-stage algorithm.

Subset Selection

A workhorse screening method with software implementations.

Development of subset-selection procedures has been ad hoc.

Popular **probability of correct selection (PCS)** guarantee.

- Subset contains the best system with *high probability*.

Modern procedures:

- Handle unknown/unequal variances and unequal sample sizes.
- Construct a subset based on **pairwise comparisons**.

This Talk

Introduce a broader subset-selection framework for delivering PCS.

Setup

There are k systems under consideration.

- Obtain n_i replications at each System $i = 1, 2, \ldots, k$.
- Outputs from System i are i.i.d. $\mathcal{N}(\mu_i, \sigma_i^2)$.
- Outputs from different systems are independent.

We assume the variances are **known**.

- Methods can be modified to handle unknown variances.

Vector $\mu = (\mu_1, \mu_2, \ldots, \mu_k)$ denotes the **problem instance**.

- Estimated by vector of sample means, $\widehat{\mu} = (\widehat{\mu}_1, \widehat{\mu}_2, \ldots, \widehat{\mu}_k)$.

Ordering systems by performance: $\mu_{[1]} \leq \mu_{[2]} \leq \cdots \leq \mu_{[k]}$.

- Large performance better $\Longrightarrow$ System $[k]$ is (one of) the best.

PCS Guarantee

Procedure returns a subset of systems, S.

- S is *random*, depending on $\widehat{\mu}$.

PCS Guarantee

For specified $1 - \alpha$,

$$\text{PCS} \equiv \mathbb{P}([k] \in S) \geq 1 - \alpha \text{ for all } \mu \in \mathbb{R}^k.$$

If ties for the best, a tied system is arbitrarily tagged as $[k]$.

- PCS guarantee: each tied system is in S with high probability.

Discrepancy-Based Framework

Define $d(m, \widehat{\mu})$, the **standardized discrepancy** between $\widehat{\mu}$ and a vector of performances $m = (m_1, \ldots, m_k)$. *(Examples soon.)*

Approach: Minimize $d(m, \widehat{\mu})$ subject to the constraint that a given system is the *best according to m*.

Define the **minimum standardized discrepancy** of System i as

$$D_i(\widehat{\mu}) \equiv \min_{m \in \mathrm{M}_i} d(m, \widehat{\mu}),$$

where $\mathrm{M}_i \equiv \{m : m_i \geq m_j \text{ for all } j \neq i\}$.

- Measures how well the data agree with the hypothesis that *System i is one of the best*.
- Smaller values indicate closer agreement.

Discrepancy-Based Framework

As a function of $\widehat{\mu}$, $D_i(\widehat{\mu})$ is a *random variable*.

- Distribution depends on the unknown problem instance μ.

Consider a deterministic **cutoff** D_i satisfying

$$\mathbb{P}(D_i(\widehat{\mu}) \leq \mathsf{D}_i) \geq 1 - \alpha \text{ for all } \mu \in \mathsf{M}_i.$$

Interpretation

For any problem instance in which System i is one of the best, its index will be less than its cutoff with high probability.

For such cutoffs, the PCS guarantee is delivered by the subset

$$S = \{i : D_i(\widehat{\mu}) \leq \mathsf{D}_i\}.$$

Standardized Discrepancies: Examples

Weighted variations of ℓ_1, ℓ_2 and ℓ_∞ distance functions:

$$d^1(m, \widehat{\mu}) \equiv \sum_{j=1}^{k} \frac{\sqrt{n_j}}{\sigma_j} |\widehat{\mu}_j - m_j|,$$

$$d^2(m, \widehat{\mu}) \equiv \sum_{j=1}^{k} \frac{n_j}{\sigma_j^2} (\widehat{\mu}_j - m_j)^2, \text{ and}$$

$$d^\infty(m, \widehat{\mu}) \equiv \max_{j=1,\ldots,k} \frac{\sqrt{n_j}}{\sigma_j} |\widehat{\mu}_j - m_j|.$$

Pairwise differences:

$$d^{\mathsf{P}}(m, \widehat{\mu}) \equiv \max_{j,\ell=1,\ldots,k} \frac{|\widehat{\mu}_j - \widehat{\mu}_\ell - m_j + m_\ell|}{\sqrt{\sigma_j^2/n_j + \sigma_\ell^2/n_\ell}}.$$

Computation

For $i = 1, 2, \ldots, k$, computing $D_i(\widehat{\mu})$ entails

- optimizing a univariate convex function or
- enumerating a discrete set of cardinality k.

$$D_i^1(\widehat{\mu}) \equiv \min_{m \in \mathrm{M}_i} d^1(m, \widehat{\mu}) = \min_{\overline{m} \in \mathbb{R}} \sum_{j \neq i} \frac{\sqrt{n_j}}{\sigma_j} [\widehat{\mu}_j - \overline{m}]^+ + \frac{\sqrt{n_i}}{\sigma_i} |\widehat{\mu}_i - \overline{m}|,$$

$$D_i^2(\widehat{\mu}) \equiv \min_{m \in \mathrm{M}_i} d^2(m, \widehat{\mu}) = \min_{\overline{m} \in \mathbb{R}} \sum_{j \neq i} \frac{n_j}{\sigma_j^2} \left([\widehat{\mu}_j - \overline{m}]^+\right)^2 + \frac{n_i}{\sigma_i^2} (\widehat{\mu}_i - \overline{m})^2,$$

$$D_i^\infty(\widehat{\mu}) \equiv \min_{m \in \mathrm{M}_i} d^\infty(m, \widehat{\mu}) = \max_{j=1,\ldots,k} \frac{\widehat{\mu}_j - \widehat{\mu}_i}{\sigma_j/\sqrt{n_j} + \sigma_i/\sqrt{n_i}}, \quad \text{and}$$

$$D_i^\mathrm{P}(\widehat{\mu}) \equiv \min_{m \in \mathrm{M}_i} d^\mathrm{P}(m, \widehat{\mu}) = \max_{j=1,\ldots,k} \frac{\widehat{\mu}_j - \widehat{\mu}_i}{\sqrt{\sigma_j^2/n_j + \sigma_i^2/n_i}}.$$

Connections: Frequentist Subset Selection

Gupta's Procedure for common, known variance σ^2:

$$\begin{aligned} S^{\text{Gupta}} &\equiv \{i : \widehat{\mu}_i \geq \widehat{\mu}_j - W_{ij} \text{ for all } j \neq i\} \\ &= \left\{ i : \max_{j=1,\dots,k} \frac{\widehat{\mu}_j - \widehat{\mu}_i}{\sqrt{\sigma^2/n + \sigma^2/n}} \leq h \right\} \\ &= \{i : D_i^{\text{P}}(\widehat{\mu}) \leq h\} \end{aligned}$$

Extended Screen-to-the-Best (ESTTB) *(known variances case)*:

$$\begin{aligned} S^{\text{ESTTB}} &\equiv \{i : \widehat{\mu}_i \geq \widehat{\mu}_j - W_{ij} \text{ for all } j \neq i\} \\ &= \left\{ i : \max_{j=1,\dots,k} \frac{\widehat{\mu}_j - \widehat{\mu}_i}{\sqrt{\sigma_j^2/n_j + \sigma_i^2/n_i}} \leq z_\beta \right\} \\ &= \{i : D_i^{\text{P}}(\widehat{\mu}) \leq z_\beta\}. \end{aligned}$$

Connections: Bayesian Subset Selection

A subset S^{Bayes} for which the **posterior probability** that S^{Bayes} includes *at least one of the optimal systems* exceeds $1 - \alpha$.

1. For Systems $i = 1, \ldots, k$, calculate pPCS_i.
2. Sort systems in descending order by pPCS_i.
3. Add systems to S^{Bayes} until the sum of the pPCS_i terms for $i \in S^{\text{Bayes}}$ first exceeds $1 - \alpha$.

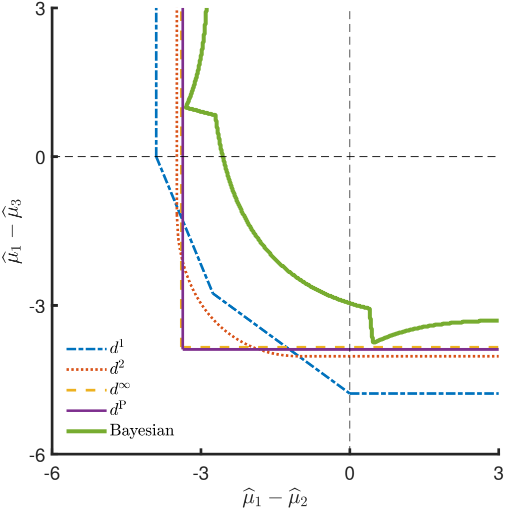
Connections to

$$f(m; \widehat{\mu}) = \prod_{j=1}^{k} \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n_j}{\sigma_j^2}} \exp\left(-\frac{(m_j - \widehat{\mu}_j)^2}{2\sigma_j^2/n_j}\right),$$

when regarded as a function of m.

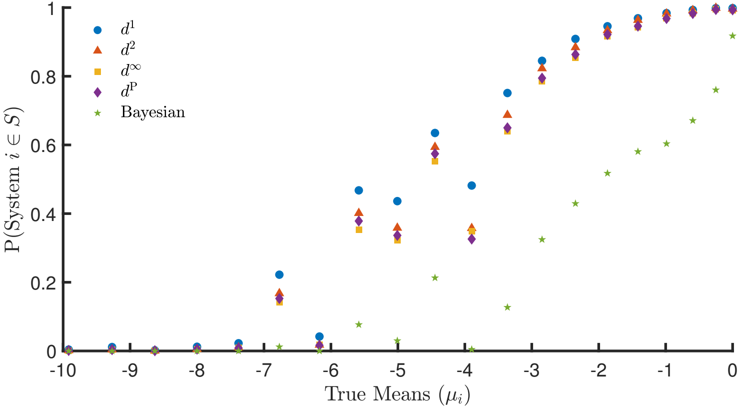
- pPCS_i: *integrate* $f(m, \widehat{\mu})$ over M_i.
- $D_i^2(\widehat{\mu})$: *maximize* $f(m, \widehat{\mu})$ over M_i.

Acceptance Regions



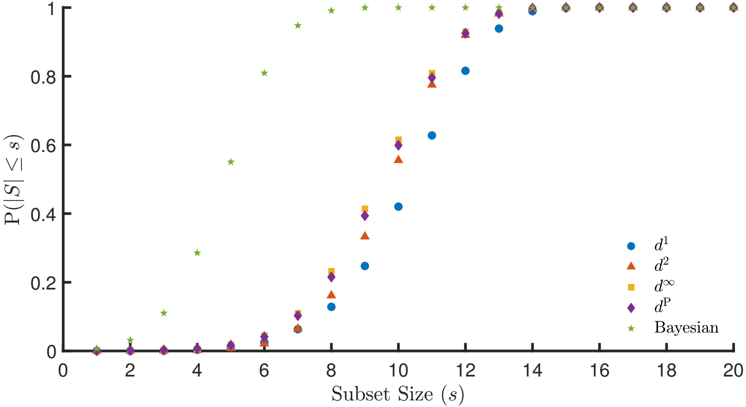
$k = 3$ systems. Regions for including System 1 in S.

Inclusion Probabilities



$k = 20$ systems and $\mu_i = -(1/4)(i-1)^{5/4}$ for $i = 1, 2, \ldots, 20$.

Empirical CDF of Subset Size



$k = 20$ systems and $\mu_i = -(1/4)(i-1)^{5/4}$ for $i = 1, 2, \ldots, 20$.

Conclusion

Takeaway

A generalized framework for subset selection based on *minimizing a standardized discrepancy* for each solution.

- Insightful connections to existing subset-selection methods.

Future Work

- Handling common random numbers and unknown variances.
- Finding tighter cutoffs.