

Maximizing Quantitative Traits in the Mating Design Problem via Simulation-Based Pareto Estimation

Susan R. Hunter

School of Industrial Engineering, Purdue University, West Lafayette, IN 47907, USA, susanhunter@purdue.edu

Benjamin McClosky

Nature Source Genetics, 33 Thornwood Drive, Ithaca, NY 14850, USA bmclosky@naturesourcegenetics.com

Commercial plant breeders improve economically important traits by selectively mating individuals from a given breeding population. Potential pairings are evaluated before the growing season using Monte Carlo simulation, and a mating design is created to allocate a fixed breeding budget across the parent pairs to achieve desired population outcomes. We introduce a novel objective function for this mating design problem that accurately models the goals of a certain class of breeding experiments. The resulting mating design problem is a computationally burdensome simulation optimization problem on a combinatorially large set of feasible points. We propose a two-step solution to this problem: (i) simulate to estimate the performance of each parent pair, and (ii) solve an estimated version of the mating design problem, which is an integer program, using the simulation output. To reduce the computational burden when implementing steps (i) and (ii), we analytically identify a Pareto set of parent pairs that will receive the entire breeding budget at optimality. Since we wish to estimate the Pareto set in step (i) as input to step (ii), we derive an asymptotically optimal simulation budget allocation to estimate the Pareto set that, in our numerical experiments, out-performs Multi-objective Optimal Computing Budget Allocation (MOCBA) in reducing misclassifications. Given the estimated Pareto set, we provide a branch and bound algorithm to solve the estimated mating design problem. Our approach dramatically reduces the computational effort required to solve the mating design problem when compared to naive methods.

Key words: multi-objective simulation optimization, ranking and selection, Pareto frontier, maximize the expected maximum, max k -armed bandit problem

Revised: August 14, 2015.

1. Introduction

Plant breeding is concerned with the evolution of plant populations under selection pressure imposed by humans (Allard, 1999). This process has proceeded for thousands of years as breeders purged undesirable plants and preferentially propagated individuals with desirable characteristics (Allard, 1999; Simmonds, 1979). In a given breeding population, the space of potential outcomes (i.e., progeny) is often too large to produce and evaluate empirically (Simmonds, 1979). Thus historically, breeders have optimized traits over a small subset of feasible progeny. Modern computing allows breeders to explore the space of candidate matings using Monte Carlo simulation, the results of which can accelerate the process of improving economically important traits during natural breeding. Thus the question of how to conduct matings to achieve “maximal” improvement in the

progeny is a question of decision-making under uncertainty. Unlike traditional operations research application areas such as manufacturing, finance, healthcare, reliability, and transportation systems, plant breeding and agriculture have received less attention from the operations research community. Traditionally the domain of statisticians, agriculture is rife with decisions that must be made under uncertainty. Thus agriculture and plant breeding offer impactful and interesting opportunities for the operations research community and, more broadly, the industrial engineering community — one of which we explore in this paper, called the *mating design problem*.

The mating design problem arises when commercial breeders selectively mate individuals to improve quantitative traits, such as height or yield. The mating design problem can be informally described as follows: given a finite parent population and limited resources for breeding — that is, a set number of plants that can be planted, called the breeding budget — find an assignment of the breeding budget to parent pairs that maximizes the expected value of some statistic of the trait observed in the progeny. A solution to the mating design problem, called a mating design, allocates the finite breeding budget to specific mate pairs. Given a predetermined population of parent pairs, our primary objectives in this work are (i) to identify optimal mating designs while (ii) avoiding excessive cloud computing costs due to large-scale scientific simulations, and (iii) to produce accurate predictions of progeny traits expected from the selected mating designs.

Researchers have used operations research techniques such as linear programming (Jansen and Wilton, 1985), integer programming (McConnel and Galligan, 2004), and genetic algorithms (Kinghorn, 2011) to solve variants of the mating design problem. This line of work measures the quality of a mating design with respect to the average trait in the progeny population. However, some argue that mating designs should be evaluated with respect to individuals in the tail of the progeny distribution (Chahota et al., 2007; Snape, 1982). As a result, researchers began using quantiles of the progeny distribution to measure mating design quality (Zhong and Jannink, 2007; Smith and Hammond, 1987; Schnell, 1983). We offer a fundamental improvement to the mating design problem by introducing a novel objective function that directly maximizes the expected maximum progeny trait, rather than the mean or a quantile.

To reduce the computational burden associated with optimizing the new objective function, we analytically characterize and discard the subset of parent pairs guaranteed to receive zero allocation at optimality in the mating design problem, regardless of the breeding budget. The

remaining parent pairs are characterized as the solution to a bi-objective simulation optimization (SO) problem, called the “Pareto set.” Our simulation experiments indicate that this Pareto set is typically small enough to allow efficient construction of the optimal mating design via branch and bound. Therefore, our solution approach effectively reduces the mating design problem to a problem solved in two steps. First, we estimate the Pareto set of parent pairs using Monte Carlo simulation. Then, we solve an estimated version of the mating design problem using a branch-and-bound algorithm on the estimated Pareto set. This procedure dramatically reduces the search space of potential progeny and prevents the need to re-simulate if the breeding budget changes. We present simulation evidence to suggest that our approach does, in fact, identify near-optimal solutions with high predictive accuracy. We view predictive accuracy as a vital performance requirement as predictions of trait selection gains are often used to justify the resource expenditure incurred by proceeding with empirical experiments.

Plant breeding is an application of great consequence, since “... the planet’s population is predicted to reach 9 billion people by 2050. To feed all these people, we must grow plants that deliver higher yield, are more nutritious, use water and nutrients more efficiently, and tolerate more environmental variations ...” (Purdue University, 2013). This work illustrates a successful theoretical advancement and implementation in plant breeding, an important but less visible domain for the operations research and industrial engineering communities.

1.1 Problem Statement and Overview of Approach

We now consider the mating design problem and our solution approach in more detail. For each candidate parent pair, suppose that the trait expression in a child is one realization from the progeny trait distribution corresponding to that parent pair. For example, the height of a child plant is one realization from the distribution of the height of child plants produced by its parent pair. We assume the distributional family for the trait is known, but the parameters of the distribution are unknown. We now define the new *max breeding problem*.

Max breeding problem: Consider a finite number $r > 0$ of breeding parent pairs producing a fixed, finite sample of $b > 0$ children, and let the vector of breeding allocations to the parent pairs be $\mathbf{b} = (b_1, b_2, \dots, b_r)$. For random observations of the children’s traits Y_{ij} , $j = 1, \dots, b_i$ from

parent pair i , we wish to solve:

$$\begin{aligned} \text{Find: } & \arg \max_{b_1, \dots, b_r} \mathbb{E} \left[\max_{i=1, \dots, r} \max_{j=1, \dots, b_i} Y_{ij} \right] \\ \text{s.t. } & \sum_{i=1}^r b_i = b, \quad b_i \in \mathbb{Z}_+ \quad \forall i = 1, \dots, r, \end{aligned} \tag{1}$$

where \mathbb{Z}_+ is the set of nonnegative integers. That is, we wish to find the allocation of the breeding budget b that maximizes the expected maximum of the trait observed in the progeny. The solution to this problem, $\mathbf{b}^* = (b_1^*, b_2^*, \dots, b_r^*)$, is a vector containing the number of child plants to breed from each of the parent pairs in the population.

Since breeders must construct the entire mating design prior to the growing season, at which time the children’s traits are realized simultaneously, we cannot use “real-life” sequential methods to solve this problem. Indeed, the problem in (1) resembles the max k -armed bandit problem described by Cicirello and Smith (2005) and Streeter and Smith (2006a,b), but simultaneous observation of the children’s traits, or “rewards,” in the growing season precludes the application of such sequential methods. Instead, we use Monte Carlo simulation to assess the trait distributions of the parent pairs (see Li et al. (2012) for a discussion of simulation in breeding). Thus the problem in (1) can be considered a ranking and selection (R&S) problem over the space of all possible breeding budget allocations (see, e.g., Kim and Nelson (2006) for an overview of R&S methods). However, the set of feasible points is combinatorially large, rendering R&S methods impractical.

Thus we propose a different, two-step approach. In the first step, called the *simulation step*, we simulate progeny from each parent pair to estimate the trait distribution parameters. In the second step, called the *IP step*, we solve an *estimated* version of the integer program in (1), using estimates of the trait distributions for each parent pair obtained as output from the simulation step. For clarity, we emphasize here that the computational budget in the simulation step, n , is separate and distinct from the physical breeding budget b in (1). The breeding budget b is part of the problem, while the simulation budget n is part of our solution. The final output of the IP step, the estimated optimal breeding allocation, is a function of the simulation budget n . Under mild assumptions, as the simulation budget n tends to infinity, the estimated solution to the mating design problem that results from the IP step converges to the true solution to (1).

Two potential issues remain when solving the max breeding problem with this two-step ap-

proach. First, if the number of parent pairs r is large, both the simulation step and the IP step may still require significant computing effort. Second, the solution to the max breeding problem sometimes allocates all children to a single parent pair, which is unlikely to be implemented in practice. Thus we wish to return to the user a reduced *set* of parent pairs likely to produce extrema, in addition to the allocation resulting from the IP step. The breeder then acts as a human-in-the-loop during the “optimization” process, accounting for factors external to the genetic model.

We address these two issues by analytically identifying and eliminating from consideration parent pairs that will not receive any of the breeding budget at optimality in the mating design problem (1). In particular, when the trait distributions are normal with finite mean μ_i and finite variance σ_i^2 for all $i = 1, \dots, r$, the only parent pairs that receive non-zero breeding budget at optimality in (1) are members of a Pareto set that is characterized by parent pairs with non-dominated means and variances, $\mathcal{P} = \{\text{parent pairs } i : (\mu_i > \mu_k) \cup (\sigma_i^2 > \sigma_k^2) \text{ for all } k = 1, \dots, r; k \neq i\}$. (We further discuss this set in §2). Then instead of solving (1) for r parent pairs, it is equivalent to solve (1) for only parent pairs in \mathcal{P} , and allocate none of the breeding budget for parent pairs outside the Pareto set \mathcal{P} . We then focus our simulation effort on identifying the Pareto set in the simulation step since only parent pairs that are estimated as Pareto will be included in the IP step. Further, we can return the estimated Pareto set to the breeder as part of the mating design process. We note here that the assumption of normal trait distributions is typical for quantitative traits such as yield (Falconer and Mackay, 1996; Lynch and Walsh, 1998), and is based on the infinitesimal model of genetic architecture. While the infinitesimal model is not taken as an exact description of biological reality, countless experiments have led the scientific community to accept the model as a useful and practical approximation of quantitative trait expression. Unless otherwise stated, henceforth we consider normal trait distributions.

1.1.1 Search space reduction in the IP step

How much easier is the max breeding problem in (1) on the reduced search space of parent pairs in \mathcal{P} ? For randomly generated populations of parent pairs using the standard maize genetic map (McMullen et al. (2009); see Appendix A of the online supplement for simulation details), we calculated sample quantiles of the cardinality of the estimated Pareto set across different parent pair population sizes. Given a randomly generated set of parent pairs, 500 children were simulated from each

parent pair to estimate the mean and variance parameters. The cardinality of the estimated Pareto set ($|\hat{\mathcal{P}}|$) was calculated for this parent pair population, using the estimated means and variances. This process was repeated 3,000 times for each parent pair population size. Sample quantiles of $|\hat{\mathcal{P}}|$ are presented in Figure 1.

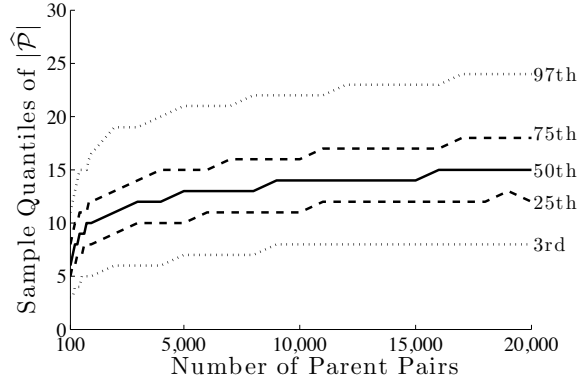


Figure 1: The cardinality of the estimated Pareto set is robustly small relative to the total number of parent pairs.

The data in Figure 1 appear to follow a logarithmic law, and the cardinality of the estimated Pareto set rarely exceeds 25 parent pairs — even when the original population is 20,000 parent pairs. Thus we have drastically reduced the search space of the max breeding problem. For example, assuming $b = 50$ children, $r = 100$ potential parent pairs, and $|\hat{\mathcal{P}}| = 15$, the number of feasible solutions is reduced from approximately 1.34×10^{40} to 4.78×10^{13} — a far smaller number, but still too large to solve (1) with available R&S methods. Further, since the estimated Pareto set is “small,” we can present the entire set to the breeder.

1.1.2 Simulation effort reduction in the simulation step

In addition to reducing the computational burden of the IP step, a focus on identifying the Pareto set in the simulation step can save significant simulation effort. In the case of normal trait distributions, where the Pareto set is characterized by parent pairs with non-dominated means and variances, estimating the Pareto set is a bi-objective SO problem on an unordered finite set.

The body of work on SO in the presence of multiple performance measures is relatively new, with most work posing additional performance measures as constraints (see, e.g., Szechtman and Yücesan, 2008; Andradóttir and Kim, 2010; Batur and Kim, 2010; Lee et al., 2012; Hunter and

Pasupathy, 2013; Pasupathy et al., 2014; Healey et al., 2013; Healey and Andradóttir, 2014; Park and Kim, 2011; Luo and Lim, 2013; Nagaraj and Pasupathy, 2013). Very little work has been done in the area of simultaneous stochastic objectives, apart from that of Ryu et al. (2009); Kim and Ryu (2011b,a) in continuous spaces, and Butler et al. (2001); Lee et al. (2010) on finite spaces. Since Butler et al. (2001) use utility functions to turn the multi-objective problem into a single-objective problem, it is not an appropriate method for our context. Thus the only appropriate method is Multi-objective Optimal Computing Budget Allocation (MOCBA) by Lee et al. (2010), a heuristic sampling framework developed along the lines of the popular Optimal Computing Budget Allocation (OCBA) framework (Chen et al., 2000) for multi-objective SO on finite sets. MOCBA requires the assumption that each objective is estimated by a sample mean constructed from independent and identically distributed (iid) normal random variables.

We implement and evaluate MOCBA’s performance in estimating the Pareto set in the context of our plant breeding application. We also derive, implement, and evaluate a new sequential sampling scheme that is based on the asymptotically optimal sampling allocation. To derive the sampling allocation in a bi-objective context, we use a large deviations (LD) framework. Such a framework has previously been employed in the unconstrained, single-objective context (Glynn and Juneja, 2004), and in the constrained, single-objective context (Hunter and Pasupathy, 2013; Pasupathy et al., 2014). The asymptotically optimal sampling allocation is characterized as the solution to a concave maximization problem, which we progressively estimate through a sequential sampling framework. Sampling terminates when the simulation budget has been expended or some amount of wall-clock time has passed. This paper is the first to derive an asymptotically optimal sampling allocation in a bi-objective context and to show successful implementation of an LD-based allocation when objective estimates are non-normal. Specifically, our second objective is a sample variance, which can be written as a sum of iid squared normal random variables, having a chi-square distribution.

Our implementation reveals that our sequential sampling framework out-performs MOCBA on a variety of performance measures, in part because of the way our sampling framework controls for misclassification errors. Two types of error can occur in estimating a Pareto set: misclassification by exclusion (MCE), where a truly Pareto parent pair is falsely estimated as non-Pareto, and misclassification by inclusion (MCI), where a non-Pareto parent pair is falsely estimated as Pareto.

MOCBA controls MCE and MCI errors by first creating a bound on each type of error, and then allocating to control the type of error whose estimated bound is the largest in each step of the sequential algorithm. As the estimated bounds converge to their true values, the allocation will be based on minimizing the larger bound. In contrast, the sampling framework we present allocates to simultaneously control both types of error, MCE and MCI, at *each* step in the sequential implementation. Further, as the sampling budget tends to infinity, our algorithm will continue to allocate in a way that controls both types of error simultaneously.

The estimated Pareto set resulting from the simulation step is used in the IP step to estimate the optimal breeding budget allocation. Since the estimated Pareto set is constant as a function of the breeding budget, we also save simulation resources in the event that the breeding budget changes. For normal trait distributions, we use a branch and bound method for solving the resulting nonlinear integer program.

1.2 Notation and Assumptions

We assume that an appropriate genetic simulation model, the desired parent pairs, and the trait distributional family are provided as input to the methods we develop. Unless otherwise stated, we consider traits for which the trait distribution is assumed to be normal. Formally, we make the following assumptions; for brevity, we write $i \leq r$ as shorthand for $i = 1, 2, \dots, r$.

Assumption 1. (i) For each parent pair $i \leq r$, the observations of the children’s traits are iid copies from a normal distribution with unknown finite mean μ_i and unknown finite variance $\sigma_i^2 > 0$; (ii) observations of the children’s traits are mutually independent across parent pairs; (iii) $(\mu_i, \sigma_i) \neq (\mu_k, \sigma_k) \forall i, k \leq r$ where $i \neq k$; and (iv) $\mu_i \neq \mu_k$ and $\sigma_i^2 \neq \sigma_k^2 \forall i \in \mathcal{P}, k \leq r, i \neq k$.

Assumption 1(iv) is standard in optimal allocation literature, since it ensures all parent pairs are distinguishable on each objective with a finite sample size.

The probability space we consider is conditional on the fixed and unknown mean and variance of the trait distribution for each parent pair. It may be possible to develop a higher-level Bayesian model that accounts for some belief about the population structure due to pedigree. There are instances in which population structure is minimally-informative — it may be the case that *all* parent pairs share a single common parent, or that each parent pair is distinct, as may result

from each parent breeding with itself. However population structure may be highly informative, as when half the population shares one parent, while the other half does not. The methods we present are general in that they can be used for any population structure, do not require a detailed population model, and provide significant computational gains over previously-used naive methods. Exploiting information about population structure may result in methods that further reduce the computational effort of the simulation step. We consider such methods as future research.

2. The Max Breeding Problem

In this section, we first provide intuition on the nature of optimal breeding budgets. We then formulate the objective function for the max breeding problem under the normality assumption. Throughout this section, we assume that the parameters of the trait distributions are known.

First, consider the case in which the trait is either present or not present in a child. For example, a child may either be resistant or susceptible to a particular disease, and suppose that the resistant state is desirable. Thus the trait distributions under consideration are Bernoulli with parameter θ_i for each parent pair $i \leq r$, where θ_i represents the probability that the child of parent pair i is resistant to the disease. Then the max breeding problem is equivalent to the following problem that minimizes the probability of observing no children with the desirable trait:

$$\text{Find: } \arg \min_{b_1, \dots, b_r} \prod_{i=1}^r (1 - \theta_i)^{b_i} \quad \text{s.t.} \quad \sum_{i=1}^r b_i = b, \quad b_i \geq 0, \quad b_i \text{ integer } \forall i \leq r.$$

The solution to this problem gives all the breeding budget to the parent pair with the largest probability of producing a child with the desirable trait, which we assume is unique. Thus letting $k = \arg \max_{i \leq r} \theta_i$, the solution is $b_k^* = b$ and $b_i^* = 0$ for all $i \neq k$. The Pareto set of parent pairs is $\mathcal{P} = \{\text{parent pairs } i : i = \arg \max_{i \leq r} \theta_i\}$, and identifying the Pareto set is an R&S problem.

Unlike with Bernoulli trait distributions, when the trait distribution is normal, one will not necessarily allocate all of the breeding budget to one parent pair in the max breeding problem. In Example 1 we show that allocating all of the breeding budget to one parent pair may be suboptimal.

Example 1. Let two parent pairs have normal trait distributions with means and standard deviations $\mu_1 = 0, \sigma_1 = 50$, and $\mu_2 = 50, \sigma_2 = 0.01$. Let $\mathbf{b} = (b_1, b_2)$ be the breeding allocation. Table 1

shows allocations and the estimated expected maximum trait value for breeding budget $b = 3$.

Table 1: The expected value of the maximum is larger when the parent pairs share the breeding budget.

b_1	b_2	Estimated expected value of maximum ^a
3	0	42.32
0	3	50.08
2	1	57.98

^a 10,000,000 children simulated (standard errors < 0.02).

When one child is observed ($b = 1$), intuitively, the child should be bred from the parent pair with the largest mean. When the breeding budget is infinite, the children should be bred from the parent pair with the largest variance — as the breeding budget grows, the largest-variance parent pair will overcome the largest-mean parent pair in producing extrema. For finite budgets larger than one, there is a balance in allocating to high-mean and high-variance parent pairs. In this example, allocating one sample to the higher-mean, lower-variance parent pair places a high probability on observing at least one child with a large trait value. The remainder of the samples are then allocated to capitalize on the high-variance parent pair’s ability to produce extrema.

In Proposition 1, we specify the max breeding problem as an integer program, assuming the means and variances are known. Let $M(\mathbf{b})$ be a random variable representing the maximum trait observed among the progeny for some breeding budget \mathbf{b} , and let $E[M(\mathbf{b})]$ be its expected value. That is, $E[M(\mathbf{b})]$ is the objective function of the max breeding problem in (1), where we explicitly note the dependence of the maximum on the breeding budget \mathbf{b} .

Proposition 1. *Let Φ be the standard normal cumulative distribution function (cdf). The expected value of the maximum child characteristic is*

$$E[M(\mathbf{b})] = \int_0^\infty 1 - \prod_{i=1}^r \Phi\left(\frac{y - \mu_i}{\sigma_i}\right)^{b_i} dy - \int_{-\infty}^0 \prod_{i=1}^r \Phi\left(\frac{y - \mu_i}{\sigma_i}\right)^{b_i} dy.$$

Proof. Let the cdf of $M(\mathbf{b})$ be $F_M(y)$. Since all children represent iid draws, $F_M(y) = P\{M(\mathbf{b}) \leq y\} = P\{\max_{i,j} Y_{ij} \leq y\} = \prod_{i=1}^r \Phi\left(\frac{y - \mu_i}{\sigma_i}\right)^{b_i}$, where $\Phi(\cdot)$ is the standard normal cdf. Define random variables $M(\mathbf{b})^+ = \max\{M(\mathbf{b}), 0\}$ and $M(\mathbf{b})^- = -\min\{M(\mathbf{b}), 0\}$. Then $E[M(\mathbf{b})] = E[M(\mathbf{b})^+] - E[M(\mathbf{b})^-] = \int_0^\infty 1 - F_M(y) dy - \int_{-\infty}^0 F_M(y) dy$, and the result follows. \square

To optimally allocate samples to maximize the expected maximum child characteristic observed, we wish to solve the max breeding problem

$$\begin{aligned} \text{Problem } MB : \quad & \text{maximize } \int_0^\infty 1 - \prod_{i=1}^r \Phi\left(\frac{y-\mu_i}{\sigma_i}\right)^{b_i} dy - \int_{-\infty}^0 \prod_{i=1}^r \Phi\left(\frac{y-\mu_i}{\sigma_i}\right)^{b_i} dy \\ & \text{s.t. } \sum_{i=1}^r b_i = b, \ b_i \geq 0 \text{ and } b_i \text{ integer for all } i \leq r. \end{aligned}$$

Solving Problem MB may be unwieldy for large b and r . Thus we identify and discard parent pairs that receive a sample size of zero at optimality.

Loosely speaking, to generate extrema from two normal distributions, where one distribution has larger mean *and* variance than another distribution, all random variables should be generated from the distribution with the larger mean and variance. Therefore intuitively, all parent pairs that are dominated in both mean and variance by some other parent pair should not receive any of the final breeding budget. This intuition is correct: for any breeding budget $b \geq 1$, the only parent pairs that receive nonzero breeding budget in the max breeding problem are a subset of a Pareto set of increasing means and decreasing variances. Theorem 1 states this result.

Theorem 1. *Let the Pareto set \mathcal{P} be the set of all non-dominated parent pairs, $\mathcal{P} = \{\text{parent pairs } i : (\mu_i > \mu_k) \cup (\sigma_i^2 > \sigma_k^2) \text{ for all } k \leq r; k \neq i\}$. Then for all $b \geq 1$, the set of parent pairs receiving nonzero sample at optimality in the max breeding problem, Problem MB , is a subset of \mathcal{P} .*

Proof. See Appendix B of the online supplement. □

While in Example 1 two parent pairs share the allocation, it is not intuitively obvious from Theorem 1 that three or more parent pairs may share the allocation at optimality in Problem MB . Thus there will not necessarily be a “best mean” parent pair and a “best variance” parent pair that will together receive all of the breeding budget, as shown in Example 2.

Example 2. Let three parent pairs have normal trait distributions with means and standard deviations $\mu_1 = 0.80, \sigma_1 = 0.58; \mu_2 = 1.0, \sigma_2 = 0.4025; \mu_3 = 1.15, \sigma_3 = 0.20$. Let $\mathbf{b} = (b_1, b_2, b_3)$ be the breeding allocation. For budget $b = 3$, at optimality in Problem MB , each of the three parent pairs receives an allocation of one child, as shown in Table 2.

Table 2: For $b = 3$, the expected maximum is largest when all three parent pairs share the budget.

b_1	b_2	b_3	$E[M(\mathbf{b})]^a$	b_1	b_2	b_3	$E[M(\mathbf{b})]^a$
3	0	0	1.2908	1	0	2	1.3395
2	1	0	1.3176	0	3	0	1.3406
0	0	3	1.3193	0	2	1	1.3457
1	2	0	1.3325	2	0	1	1.3457
0	1	2	1.3356	1	1	1	1.3466

^a Calculated by numerical integration in MATLAB.

3. Allocating the Simulation Budget in the Simulation Step

We now consider the problem of efficiently allocating a simulation budget to estimate the Pareto set. For normal trait distributions, identifying the Pareto set \mathcal{P} through Monte Carlo simulation is a bi-objective SO problem. The asymptotically optimal sampling allocation that maximizes the rate of decay of the probability of a misclassification is derived in Appendix C of the online supplement, which we outline in §3.1 before presenting a sequential implementation in §3.2. For brevity, the results in §3.1 are presented with an emphasis on knowledge required for implementation, with full details and all proofs in the online supplement.

3.1 Asymptotically Optimal Allocation Strategy

Consider a procedure to estimate the Pareto set that consists of expending a simulation budget n to estimate the means and variances of the trait distributions for each parent pair and returning to the user the estimated set of Pareto parent pairs. Let $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_r)$ be the proportional allocation of the simulation budget n to the parent pairs. For now, we ignore that $n\alpha_i$ is not necessarily an integer. For each parent pair producing $n\alpha_i$ children with characteristics Y_{ij} observed as iid copies from a normal distribution with mean μ_i and variance σ_i^2 for all $j = 1, \dots, n\alpha_i$, let $\hat{Y}_i = \frac{1}{n\alpha_i} \sum_{\ell=1}^{n\alpha_i} Y_{i\ell}$ estimate μ_i and $\hat{\sigma}_i^2 = \frac{1}{n\alpha_i} \sum_{\ell=1}^{n\alpha_i} (Y_{i\ell} - \frac{1}{n\alpha_i} \sum_{\ell=1}^{n\alpha_i} Y_{i\ell})^2$ estimate σ_i^2 . Then the estimated Pareto set is $\hat{\mathcal{P}} = \{\text{parent pairs } i : (\hat{Y}_i > \hat{Y}_k) \cup (\hat{\sigma}_i^2 > \hat{\sigma}_k^2) \forall k \leq r; k \neq i\}$.

Under this sampling paradigm, as the simulation budget tends to infinity, the probability of a misclassification (MC) event, that is, misclassifying a parent pair on its status as Pareto or non-Pareto, decays to zero. In this section, we present the asymptotically optimal sampling allocation that maximizes the rate of decay of the probability of misclassification ($P\{\text{MC}\}$) as a function of

the proportional sampling allocation, α .

In the context of our estimation procedure, an MC event can occur in one of two ways, which we will call misclassification by exclusion (MCE) and misclassification by inclusion (MCI). MCE is the event in which a truly Pareto parent pair is falsely excluded from the estimated Pareto set by being estimated as dominated by another parent pair, be it Pareto or non-Pareto. MCI is the event in which a truly non-Pareto parent pair is falsely included in the estimated Pareto set by being estimated as non-dominated. That is,

$$\text{MCE} := \underbrace{\bigcup_{i \in \mathcal{P}} \bigcup_{k \leq r, k \neq i} (\hat{Y}_i \leq \hat{Y}_k) \cap (\hat{\sigma}_i^2 \leq \hat{\sigma}_k^2)}_{\text{some } i \in \mathcal{P} \text{ dominated by some } k} \quad \text{and} \quad \text{MCI} := \underbrace{\bigcup_{j \notin \mathcal{P}} \bigcap_{k \leq r, k \neq j} (\hat{Y}_j \geq \hat{Y}_k) \cup (\hat{\sigma}_j^2 \geq \hat{\sigma}_k^2)}_{\text{some } j \notin \mathcal{P} \text{ not dominated by any } k}.$$

A straightforward way of writing the MC event is $\text{MC} := \text{MCE} \cup \text{MCI}$, such that the probability of an MC event is $P\{\text{MC}\} = P\{\text{MCE} \cup \text{MCI}\}$, but this probabilistic statement is difficult to analyze due to dependence in the MCI term. Thus we reformulate the MC event for easier analysis.

First, let us label the true Pareto parent pairs by their means, so that $\mu_0 := -\infty < \mu_1 < \dots < \mu_{p-1} < \mu_p$ and $\sigma_1^2 > \sigma_2^2 > \dots > \sigma_p^2 > \sigma_{p+1}^2 := 0$, where p is the cardinality of \mathcal{P} . Then the true Pareto parent pairs are $(\mu_\ell, \sigma_\ell^2)$ for $\ell = 1, \dots, p$, where ℓ uniquely indexes the Pareto parent pairs. We also define the true *phantom* parent pairs, which are the coordinates $(\mu_\ell, \sigma_{\ell+1}^2)$ for $\ell = 0, 1, \dots, p$, where we place phantom parent pairs at $(-\infty, \sigma_1^2)$ and $(\mu_p, 0)$. There are $p + 1$ phantom parent pairs. Assuming the true Pareto set is known, Figure 2 displays the Pareto and phantom Pareto parent pairs.

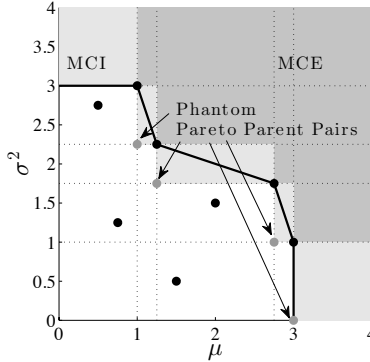


Figure 2: If the Pareto set were known, then to be falsely estimated as Pareto, the non-Pareto parent pairs must be falsely estimated as being in the MCI or MCE “regions,” thereby dominating a phantom parent pair.

Since the true Pareto set is actually unknown, the phantom Pareto parent pairs must be estimated. Define $\hat{Y}_{[\ell]}$ and $\hat{\sigma}_{[\ell]}^2$ as corresponding to the order statistics of the estimated mean and variance values of the true Pareto set, so $\hat{Y}_{[0]} := -\infty < \hat{Y}_{[1]} < \dots < \hat{Y}_{[p-1]} < \hat{Y}_{[p]}$ and $\hat{\sigma}_{[1]}^2 > \hat{\sigma}_{[2]}^2 > \dots > \hat{\sigma}_{[p]}^2 > \hat{\sigma}_{[p+1]}^2 := 0$, and the *estimated phantom parent pairs for the true Pareto set* are $(\hat{Y}_{[\ell]}, \hat{\sigma}_{[\ell+1]}^2)$ for $\ell = 0, 1, \dots, p$. Define misclassification by dominating a phantom parent pair as

$$\text{MCI}_{ph} := \underbrace{\bigcup_{j \notin \mathcal{P}} \bigcup_{\ell=0}^p (\hat{Y}_{[\ell]} \leq \hat{Y}_j) \cap (\hat{\sigma}_{[\ell+1]}^2 \leq \hat{\sigma}_j^2)}_{\text{some } j \notin \mathcal{P} \text{ dominates some phantom system}}.$$

Theorem 2. Let $MC_{ph} := MCE \cup \text{MCI}_{ph}$. Then $P\{MC\} = P\{MC_{ph}\}$.

Since Theorem 2 holds, henceforth we use the notation $P\{MC\}$ to refer to the probability of a misclassification event.

Recall that to efficiently identify the Pareto set, we wish to identify the sample allocation vector that maximizes the rate of decay of the $P\{MC\}$. Before presenting the rate of decay of $P\{MC\}$ in Theorem 3, we first require additional notation and results. By Glynn and Juneja (2004), under Assumption 1, for all $i \in \mathcal{P}, k \leq r$ such that $\mu_k < \mu_i$,

$$G(\alpha_i, \alpha_k) := -\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\hat{Y}_i \leq \hat{Y}_k\} = \frac{(\mu_i - \mu_k)^2}{2(\sigma_i^2/\alpha_i + \sigma_k^2/\alpha_k)}.$$

We present an expression for the large deviations rate function corresponding to the event $\hat{\sigma}_i^2 \leq \hat{\sigma}_k^2$ in Proposition 2.

Proposition 2. Under Assumption 1, for all $i \in \mathcal{P}, k \leq r$ such that $\sigma_k^2 < \sigma_i^2$,

$$H(\alpha_i, \alpha_k) := -\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\hat{\sigma}_i^2 \leq \hat{\sigma}_k^2\} = \frac{\alpha_i}{2} \log \left(\frac{\sigma_i^2/\alpha_i + \sigma_k^2/\alpha_k}{\sigma_k^2(1/\alpha_i + 1/\alpha_k)} \right) + \frac{\alpha_k}{2} \log \left(\frac{\sigma_i^2/\alpha_i + \sigma_k^2/\alpha_k}{\sigma_i^2(1/\alpha_i + 1/\alpha_k)} \right).$$

Since the estimated mean and estimated variance for a normal distribution are independent, the objectives \hat{Y}_i and $\hat{\sigma}_i^2$ are independent for all $i \leq r$. This fact, along with the additional material provided in the online supplement, yields the following Theorem 3.

Theorem 3. Let $\alpha_0 := 1$ and $\alpha_{p+1} := 1$. Under Assumption 1, the rate function of $P\{MC\}$ for

the mating design problem is

$$-\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{MC\} = \min \left(\min_{i \in \mathcal{P}} \min_{k \leq r, k \neq i} \left(G(\alpha_i, \alpha_k) \mathbb{I}_{\{\mu_k < \mu_i\}} + H(\alpha_i, \alpha_k) \mathbb{I}_{\{\sigma_k^2 < \sigma_i^2\}} \right), \right. \\ \left. \min_{j \notin \mathcal{P}} \min_{\ell=0, \dots, p} \left(G(\alpha_\ell, \alpha_j) \mathbb{I}_{\{\mu_j < \mu_\ell\}} + H(\alpha_{\ell+1}, \alpha_j) \mathbb{I}_{\{\sigma_j^2 < \sigma_{\ell+1}^2\}} \right) \right).$$

Thus the overall rate of decay of the probability of a misclassification is the slowest among the rates of decay of many pairwise misclassification events.

To maximize the rate of decay of $P\{MC\}$, we allocate sample by solving

Problem Q : maximize z

$$\begin{aligned} \text{s.t. } & G(\alpha_i, \alpha_k) \mathbb{I}_{\{\mu_k < \mu_i\}} + H(\alpha_i, \alpha_k) \mathbb{I}_{\{\sigma_k^2 < \sigma_i^2\}} \geq z \quad \forall i, k \in \mathcal{P}, k \neq i, \\ & G(\alpha_\ell, \alpha_j) \mathbb{I}_{\{\mu_j < \mu_\ell\}} + H(\alpha_{\ell+1}, \alpha_j) \mathbb{I}_{\{\sigma_j^2 < \sigma_{\ell+1}^2\}} \geq z \quad \forall j \notin \mathcal{P}, \ell = 0, \dots, p, \\ & \sum_{i=1}^r \alpha_i = 1, \quad \alpha_i \geq 0 \quad \forall i \leq r. \end{aligned}$$

Note that the first set of constraints in Problem Q hold for all $i, k \in \mathcal{P}, k \neq i$ instead of for all $i \in \mathcal{P}, k \leq r, k \neq i$. This simplification occurs because anytime some $j \notin \mathcal{P}$ falsely excludes $i \in \mathcal{P}$, it must also dominate a phantom parent pair, an event whose rate function already appears in the second set of constraints in Problem Q . Thus Problem Q has $p(p-1)$ constraints corresponding to controlling the rate of decay of $P\{MCE\}$, and $(r-p)(p+1)$ constraints corresponding to controlling the rate of decay of $P\{MCI_{ph}\}$, where each constraint has at least one nonzero term in the sum. Problem Q is a concave maximization problem in α , which follows from a proof similar to that in Glynn and Juneja (2006).

Since Slater's condition holds for the concave maximization Problem Q , the Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient for global optimality (see, e.g., Boyd and Vandenberghe, 2004). The KKT conditions for Problem Q imply that $\alpha_i > 0$ for all $i \leq r$. While in the case of Glynn and Juneja (2004) and Hunter and Pasupathy (2013), the KKT conditions required equating the rates of the suboptimal systems, the KKT conditions here are somewhat less insightful: the optimal allocation in Problem Q may be determined by various combinations of the pairwise "contests" between the systems. Thus we solve Problem Q with numerical optimization.

Example 3. One hundred parent pairs were randomly generated using the standard maize genetic map. Each parent pair received a sample of 500 children to estimate the mean and variance, which we assume are the true mean and variance in this example. Figure 3 shows the locations of the means and standard deviations for the one hundred parent pairs. The size of the circle is proportional to the amount of sample received when the allocation is optimal under equal allocation and the proposed allocation scheme, respectively. That is, Figure 3(b) provides a pictorial representation of the solution to Problem Q .

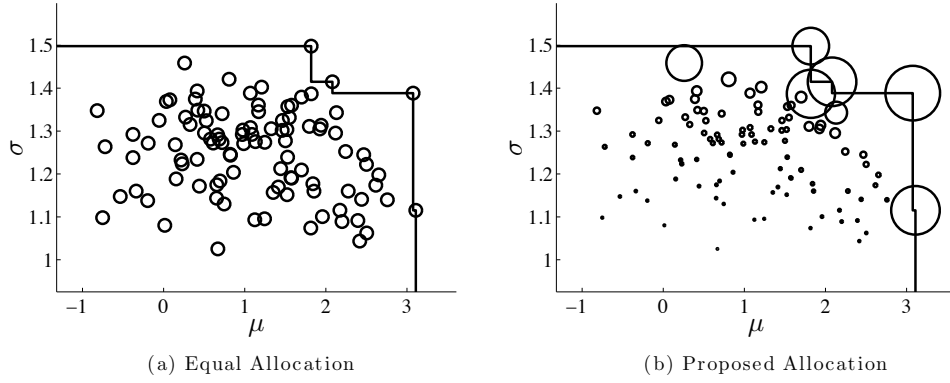


Figure 3: The size of the circle is proportional to the amount of sample received by each of 100 parent pairs.

3.2 Sequential algorithm for estimating the Pareto set

Since μ_i and σ_i^2 must be estimated for all $i \leq r$, Problem Q cannot be implemented as written. Define the following estimated rate functions, in which we replace all unknown quantities with their corresponding estimators:

$$\hat{G}(\alpha_i, \alpha_k) := \frac{(\hat{Y}_i - \hat{Y}_k)^2}{2(\hat{\sigma}_i^2/\alpha_i + \hat{\sigma}_k^2/\alpha_k)}, \quad \text{and} \quad \hat{H}(\alpha_i, \alpha_k) := \frac{\alpha_i}{2} \log \left(\frac{\hat{\sigma}_i^2/\alpha_i + \hat{\sigma}_k^2/\alpha_k}{\hat{\sigma}_k^2(1/\alpha_i + 1/\alpha_k)} \right) + \frac{\alpha_k}{2} \log \left(\frac{\hat{\sigma}_i^2/\alpha_i + \hat{\sigma}_k^2/\alpha_k}{\hat{\sigma}_i^2(1/\alpha_i + 1/\alpha_k)} \right).$$

We now define the estimated Problem Q , called Problem \hat{Q} :

Problem \hat{Q} : maximize z

$$\begin{aligned} \text{s.t. } & \hat{G}(\alpha_i, \alpha_k) \mathbb{I}_{\{\hat{Y}_k < \hat{Y}_i\}} + \hat{H}(\alpha_i, \alpha_k) \mathbb{I}_{\{\hat{\sigma}_k^2 < \hat{\sigma}_i^2\}} \geq z \quad \forall i, k \in \hat{\mathcal{P}}, k \neq i, \\ & \hat{G}(\alpha_{[\ell]}, \alpha_j) \mathbb{I}_{\{\hat{Y}_j < \hat{Y}_{[\ell]}\}} + \hat{H}(\alpha_{[\ell+1]}, \alpha_j) \mathbb{I}_{\{\hat{\sigma}_j^2 < \hat{\sigma}_{[\ell+1]}^2\}} \geq z \quad \forall j \notin \hat{\mathcal{P}}, \ell = 0, \dots, |\hat{\mathcal{P}}|, \\ & \sum_{i=1}^r \alpha_i = 1, \alpha_i \geq 0 \quad \forall i \leq r, \end{aligned}$$

where $[\ell]$ denotes the ℓ th order statistic, and $\alpha_{[\ell]}$ is the allocation corresponding to the ℓ th ordered estimated Pareto parent pair. We propose Algorithm 1 to iteratively estimate the Pareto set and update the optimal allocation scheme by solving Problem \hat{Q} .

Algorithm 1 Sequential algorithm for implementation of the optimal allocation scheme in the mating design problem with normal trait distributions.

Require: Number of pilot samples $\delta_0 > 0$; number of samples between allocation vector updates $\delta > 0$; and a minimum-sample vector $\epsilon = \{\epsilon_1, \dots, \epsilon_r\} > 0$.

- 1: Initialize: collect δ_0 samples from each parent pair $i \leq r$.
 - 2: Set $n = r\delta_0, n_i = \delta_0$. {Initialize total simulation effort and effort for each parent pair.}
 - 3: Update the sample means \hat{Y}_i and sample variances $\hat{\sigma}_i^2$ for all parent pairs $i \leq r$. Update the estimated Pareto set $\hat{\mathcal{P}}$.
 - 4: Solve Problem \hat{Q} to obtain the estimated optimal allocation $\hat{\alpha}^*(n) = (\hat{\alpha}_1(n), \dots, \hat{\alpha}_r(n))$.
 - 5: Collect one sample at each of the parent pairs $X_i, i = 1, 2, \dots, \delta$, where the X_i 's are iid random variates having probability mass function $\hat{\alpha}^*(n)$ on support $\{1, 2, \dots, r\}$; update $n_{X_i} = n_{X_i} + 1$.
 - 6: Set $n = n + \delta$ and update $\bar{\alpha}_n = \{\bar{\alpha}_{1,n}, \dots, \bar{\alpha}_{r,n}\} = \{n_1/n, n_2/n, \dots, n_r/n\}$.
 - 7: **if** $\bar{\alpha}_n > \epsilon$ **then**
 - 8: Set $\delta^+ = 0$.
 - 9: **else**
 - 10: Collect one sample from each parent pair in the set of parent pairs receiving insufficient sample $\mathcal{I}_n = \{i : \bar{\alpha}_{i,n} < \epsilon_i\}$.
 - 11: Update $n_i = n_i + 1$ for all $i \in \mathcal{I}_n$. Set $\delta^+ = |\mathcal{I}_n|$.
 - 12: **end if**
 - 13: Set $n = n + \delta^+$ and go to Step 3.
-

The sequential allocation scheme in Algorithm 1 first takes an initial amount of sample δ_0 from each parent pair and initializes all of the estimators. Then the estimated optimal allocation problem, Problem \hat{Q} , is solved. The resulting estimated optimal allocation is used as a sampling distribution from which to choose the index of the next parent pair to sample, denoted by the random variable X_i in the algorithm. The vector ϵ ensures that each parent pair is sampled infinitely often, and

should be small relative to $1/r$. Once the sample δ has been expended, all estimators are updated, and the process repeated. Algorithm 1 is nonterminating, however the user may wish to terminate the algorithm after some total budget has been expended or some overall amount of computing time has passed. We present Example 4 to demonstrate the sequential algorithm.

Example 4. Let the true means and standard deviations for 100 parent pairs be as specified in Example 3, with the optimal allocation shown in Figure 3(b). We implement Algorithm 1 by assuming the (μ_i, σ_i) pairs in Figure 3(a) are true and sequentially generating children from the appropriate normal distributions. The parameters for Algorithm 1 are $\delta_0 = 100$, $\delta = 2500$, and $\epsilon_i = 10^{-10}$ for all $i \leq r$. A description of the solver used to find the estimated optimal allocation in Step 4 of Algorithm 1 is given in Appendix D.1 of the online supplement. Sample quantiles of the optimality gap of the rate of decay of $P\{\text{MC}\}$ for Algorithm 1, $z(\alpha^*) - z(\bar{\alpha}_n)$, calculated across 100 sample paths, are shown in Figure 4. The optimality gap for equal allocation is also shown.

We also implement MOCBA on this problem. To use MOCBA, we approximate the distribution of the sample variance as $\hat{\sigma}_i^2 \approx N\left(\frac{(n_i-1)}{n_i}\sigma_i^2, \frac{2(n_i-1)}{n_i^2}\sigma_i^4\right)$, where n_i is the sample received by parent pair i and $N(\cdot, \cdot)$ indicates a normal distribution with mean and variance parameters. Therefore if we were to observe iid copies of a random variable “ σ_i^2 ” from which we could construct a sample mean, we would assume $\sigma_i^2 \approx N\left(\frac{(n_i-1)}{n_i}\sigma_i^2, \frac{2(n_i-1)}{n_i^2}\sigma_i^4\right)$, where we scale the distribution of $\hat{\sigma}_i^2$ by n_i . We set the MOCBA parameters as $N_0 = 100$ (initial sample size), $\Delta = 2500$ (increased sample per iteration), and $\delta = 1250$ (the maximum number of samples a parent pair can receive in a given iteration; we set $\delta = \Delta/2$ as in the numerical section of Lee et al. (2010)). Lee et al. (2010) also suggest that the allocation calculations be simplified by using the α_i from the previous iteration. Since it is possible that $\alpha_i = 0$ for some $i \leq r$ in the previous iteration, resulting in operations that divide by zero, we use the fraction of sample allocated to system i so far, $\bar{\alpha}_{i,n} = n_i/n$. Sample quantiles of the optimality gap of the rate of decay of $P\{\text{MC}\}$ for MOCBA, $z(\alpha^*) - z(\bar{\alpha}_{\text{MOCBA},n})$, calculated across 100 sample paths, are also shown in Figure 4.

As expected, in Figure 4, the optimality gap for the implementation of Algorithm 1 reduces as the sample size increases. The optimality gap for MOCBA appears to have a larger initial variance than that of the proposed procedure. For larger sample sizes, MOCBA appears to be a competitive alternative to equal allocation.

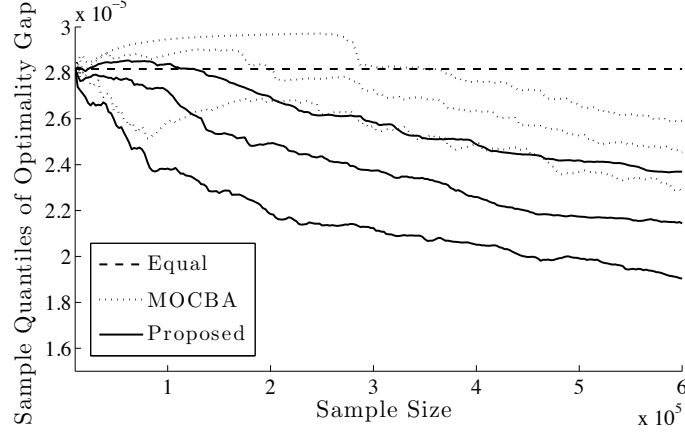


Figure 4: The 25th, 50th, and 75th sample quantiles of the distributions of the optimality gap for the problem in Example 3 using equal allocation, MOCBA, and the proposed Algorithm 1

To better understand the meaning of the optimality gap in Figure 4, recall that the probability of misclassification is $P\{\text{MC}(\alpha)\} \approx e^{-nz(\alpha)}$, where $z(\alpha)$ is the large deviations rate of decay of the probability of misclassification, as a function of the allocation to the parent pairs, α . Thus the optimality gap $z(\alpha^*) - z(\alpha)$, can be interpreted as $z(\alpha^*) - z(\alpha) \approx -\frac{1}{n} \log P\{\text{MC}(\alpha^*)\} + \frac{1}{n} \log P\{\text{MC}(\alpha)\}$, which implies $P\{\text{MC}(\alpha^*)\} \approx e^{-n(z(\alpha^*) - z(\alpha))} P\{\text{MC}(\alpha)\}$. For example, if the optimality gap $z(\alpha^*) - z(\alpha) = 2.8 \times 10^{-5}$ at sample size $n = 1 \times 10^5$, then $P\{\text{MC}(\alpha^*)\} \approx (0.061)P\{\text{MC}(\alpha)\}$. Therefore what appears to be a “small” optimality gap can still lead to significant computational savings over competing methods.

Note that in Algorithm 1, the choice of δ_0 and δ will affect how quickly the optimality gap converges to zero. Methods for choosing δ_0 and δ are future research; we suspect that the rates of convergence of the rate functions $\hat{G}(\cdot)$ and $\hat{H}(\cdot)$ to their true values $G(\cdot)$ and $H(\cdot)$ may provide guidance in this respect. We also speculate that, due to autocorrelation in $\bar{\alpha}_n$, it may be better to err on the side of “larger” δ_0 so that initial poorer estimates of the optimal allocation do not result in a sample path that initially performs worse than equal allocation.

4. Solving the Estimated Max Breeding Problem in the IP Step

Having estimated the Pareto set, the breeder may wish to create the mating design by solving the max breeding problem on the estimated Pareto set. In this section, we develop a branch and bound algorithm for constructing the optimal mating design, given the estimated Pareto set. Thus we

consider

$$\begin{aligned} \text{Problem } \widehat{MB}: \quad & \text{maximize } \int_0^\infty 1 - \prod_{i=1}^{|\widehat{\mathcal{P}}|} \Phi\left(\frac{y - \widehat{Y}_i}{\widehat{\sigma}_i}\right)^{b_i} dy - \int_{-\infty}^0 \prod_{i=1}^{|\widehat{\mathcal{P}}|} \Phi\left(\frac{y - \widehat{Y}_i}{\widehat{\sigma}_i}\right)^{b_i} dy \\ & \text{s.t. } \sum_{i=1}^{|\widehat{\mathcal{P}}|} b_i = b, \quad b_i \geq 0 \text{ and } b_i \text{ integer for all } i \leq |\widehat{\mathcal{P}}|. \end{aligned}$$

We wish to find the optimal mating design $\widehat{\mathbf{b}}^*(n)$ that solves this modified max breeding problem, which is a nonlinear integer program. As the simulation budget n tends to infinity, assuming that we sample from each parent pair infinitely often (as in Algorithm 1), the solution to Problem \widehat{MB} will tend to the solution of Problem MB , \mathbf{b}^* . Henceforth, we suppress the explicit notation that the solution to Problem \widehat{MB} depends on the simulation budget n .

4.1 Lower and upper bounds

To prune the search space in the branch and bound algorithm, we need lower and upper bounds on the performance of designs encountered during enumeration. The max breeding problem is a maximization problem, so any feasible solution produces a valid lower bound on the optimal objective value. We construct an initial feasible solution by finding the mate pair producing the largest objective value when receiving the entire breeding budget b . Thus we compute

$$LB_i := \int_0^\infty 1 - \Phi\left(\frac{y - \widehat{Y}_i}{\widehat{\sigma}_i}\right)^b dy - \int_{-\infty}^0 \Phi\left(\frac{y - \widehat{Y}_i}{\widehat{\sigma}_i}\right)^b dy$$

for each $i \in \widehat{\mathcal{P}}$. We then define $LB := \max_{i \in \widehat{\mathcal{P}}} LB_i$.

At a given node in the search tree, we test whether further branching can produce a solution that exceeds the best solution so far. We propose an upper bound for the performance of designs encountered through further branching in the search space. Let \mathbb{Z}_+ denote the set of non-negative integers, and let $\mathcal{B} := \{\mathbf{b} \in \mathbb{Z}_+^{|\widehat{\mathcal{P}}|} : \sum_{i=1}^{|\widehat{\mathcal{P}}|} b_i \leq b\}$ be the set of feasible mating designs that breed from the parent pairs in the estimated Pareto set. Given a set of ordered pairs $\mathcal{F} = \{(i_1, k_1), (i_2, k_2), \dots\}$ where i_1, i_2, \dots are parent pairs in $\widehat{\mathcal{P}}$ and k_1, k_2, \dots are the fixed breeding budget variables at the current node of the search tree, define $\mathcal{B}(\mathcal{F}) := \{\mathbf{b} \in \mathcal{B} : b_i = k \text{ for all } (i, k) \in \mathcal{F}\}$ as the set containing all mating designs that can be encountered through further branching. To prune the search space, we need an upper bound on the performance of designs in $\mathcal{B}(\mathcal{F})$. Let $\mathcal{L}(\mathcal{F}) = \{j \in$

$\widehat{\mathcal{P}} : j \neq i \text{ for all } (i, k) \in \mathcal{F}$ be the set of indices of free variables, that is, variables not fixed by \mathcal{F} , and let d be the breeding budget remaining after fixing the elements of \mathcal{F} . Suppose $\mathcal{L}(\mathcal{F}) \neq \emptyset$. Consider the piecewise function $f : \mathbb{R} \mapsto \mathbb{R}$, $f(y) := \min_{j \in \mathcal{L}(\mathcal{F})} \Phi\left(\frac{y - \widehat{Y}_j}{\widehat{\sigma}_j}\right)$, and let

$$UB := \int_0^\infty 1 - f(y)^d \prod_{(i,k) \in \mathcal{F}} \Phi\left(\frac{y - \widehat{Y}_i}{\widehat{\sigma}_i}\right)^k dy - \int_{-\infty}^0 f(y)^d \prod_{(i,k) \in \mathcal{F}} \Phi\left(\frac{y - \widehat{Y}_i}{\widehat{\sigma}_i}\right)^k dy.$$

The following proposition states that the expected maximum achieved through further branching is less than or equal to the upper bound.

Proposition 3. *Let $\widehat{E}[M(\mathbf{b})]$ denote the objective value of Problem \widehat{MB} as a function of \mathbf{b} . Then $\widehat{E}[M(\mathbf{b})] \leq UB$ for all $\mathbf{b} \in \mathcal{B}(\mathcal{F})$.*

Proof. Let $\mathbf{b} \in \mathcal{B}(\mathcal{F})$. By analysis similar to that in equation (B.2) (see Appendix B of the online supplement),

$$UB - \widehat{E}[M(\mathbf{b})] = \int_{-\infty}^\infty \prod_{(i,k) \in \mathcal{F}} \Phi\left(\frac{y - \widehat{Y}_i}{\widehat{\sigma}_i}\right)^k \left(\prod_{i \in \mathcal{L}(\mathcal{F})} \Phi\left(\frac{y - \widehat{Y}_i}{\widehat{\sigma}_i}\right)^{b_i} - f(y)^d \right) dy. \quad (2)$$

By construction, $\Phi\left(\frac{y - \widehat{Y}_j}{\widehat{\sigma}_j}\right) \geq f(y) \forall y, j \in \mathcal{L}(\mathcal{F})$. Since $\mathbf{b} \in \mathcal{B}(\mathcal{F})$, then $\sum_{j \in \mathcal{L}(\mathcal{F})} b_j \leq d$, and $\prod_{i \in \mathcal{L}(\mathcal{F})} \Phi\left(\frac{y - \widehat{Y}_i}{\widehat{\sigma}_i}\right)^{b_i} \geq f(y)^d \forall y$. Thus the quantity in line (2) is positive. \square

4.2 Branch and bound algorithm

We present a branch and bound algorithm, Algorithm 2, that uses the bounds derived in §4.1. To initialize the algorithm, we first evaluate LB_i for all $i \in \widehat{\mathcal{P}}$, sort the parent pairs by descending order of LB_i , and store the list as \mathcal{L} . The set \mathcal{L} is an ordered set of mate pairs whose allocation is not fixed. Then an initial feasible solution and lower bound is found by taking $LB = \max_{i \in \widehat{\mathcal{P}}} LB_i$, as discussed in §4.1. The variable \mathbf{b}_{LB} stores the breeding budget allocation that produces the bound LB , that is, the best feasible breeding budget found so far. Then the implicit enumeration function is called.

Each state of the implicit enumeration scheme is characterized by \mathcal{L} ; d , the remaining breeding budget; and \mathbf{b} , a vector containing the values of b_i for all mate pairs whose allocation is fixed. The ENUM function begins by choosing the first free parent pair i in the ordered list \mathcal{L} . If $\mathcal{L} \setminus \{i\} = \emptyset$,

then i receives the remaining breeding budget d , and the incumbent solution is updated if necessary. If $\mathcal{L} \setminus \{i\} \neq \emptyset$, the algorithm cycles through all feasible values for b_i , as determined by the remaining breeding budget d , beginning with zero. For each feasible value of b_i , the algorithm computes UB as described in §4.1, and the recursive call is only made when $UB > LB$. After processing mate pair i , the value for b_i is reset to 0, and the function returns. The algorithm never makes a recursive call when $\mathcal{L} = \emptyset$.

Algorithm 2 Branch and bound algorithm to solve the estimated max breeding problem.

```

1: Initialize: Evaluate  $LB_i$  for all  $i \in \widehat{\mathcal{P}}$ . Sort parent pairs by descending  $LB_i$  and store list as  $\mathcal{L}$ ,
   where ties are broken arbitrarily.
2: Initialize global variables: lower bound  $LB$ , the current best solution  $\mathbf{b}_{LB}$  as the breeding budget
   that produces  $LB$ .
3: Call  $\text{ENUM}(\mathcal{L}, b, \mathbf{0})$ .
4: function  $\text{ENUM}(\mathcal{L}, d, \mathbf{b})$ 
5:   choose the first free parent pair  $i$  in the list  $\mathcal{L}$  and set  $\mathcal{L} \leftarrow \mathcal{L} \setminus \{i\}$ 
6:   if  $\mathcal{L} = \emptyset$  then
7:      $b_i \leftarrow d$ 
8:     if  $E[M(\mathbf{b})] > LB$  then
9:        $LB \leftarrow E[M(\mathbf{b})]$  and  $\mathbf{b}_{LB} \leftarrow \mathbf{b}$ 
10:    end if
11:  else
12:    for  $k = 0$  to  $d$  do
13:       $b_i \leftarrow k$ 
14:      calculate  $UB$ 
15:      if  $UB > LB$  then
16:         $\text{ENUM}(\mathcal{L}, d - k, \mathbf{b})$ 
17:      end if
18:    end for
19:  end if
20:   $b_i \leftarrow 0$ 
21: return
```

Example 5. Let $b = 3$, $|\widehat{\mathcal{P}}| = 3$, $\mathcal{L} = \{1, 2, 3\}$, where $LB_1 = \max_{i \in \{1, 2, 3\}} LB_i$. If none of the search space were pruned, Algorithm 2 would evaluate feasible points in the order

$$\{(0, 0, 3), (0, 1, 2), (0, 2, 1), (0, 3, 0), (1, 0, 2), (1, 1, 1), (1, 2, 0), (2, 0, 1), (2, 1, 0), (3, 0, 0)\}.$$

5. Implementation

To assess the optimality gap of our methods, we require a set of test problems to which we know the solutions — both to the true max breeding problem in (1), as well the true optimal allocation problem, Problem Q . We use simulation to obtain realistic populations for which we know the mean and variance parameters for all of the parent pairs, and that span a wide range of quantitative trait models. Specifically, we first randomly place loci throughout the genome and sample from an empirically-derived distribution of allelic effects (see Buckler et al., 2009). Then, using a standard model of meiosis, we simulate progeny to estimate the population mean and variance, which we assume to be the fixed “true” population values. (See Appendix A of the online supplement for additional simulation details.) Then, we conduct our experiments by generating iid normal random variates from the “true” trait distributions of the parent pairs to simulate the progeny.

In what follows, we fix 98 randomly generated populations of $r = 100$ parent pairs each. The population size $r = 100$ was chosen as the largest number of parent pairs for which we could reliably run numerical experiments. Some plant breeding applications will require techniques for larger populations, however this size is large enough to be useful (Simmonds, 1979). Appendix D.2 of the online supplement shows the 98 populations and their proposed allocations.

Ideally, we would always numerically solve the allocation Problem Q to optimality. However some of the randomly generated populations resulted in an ill-conditioned Problem Q , likely resulting from parent pairs on or near the Pareto front lying too close to each other (such that, for example, the rate functions become shallow and the solver has difficulty estimating a valid Hessian). Rather than eliminate such populations for being ill-conditioned, which a practitioner would not know in advance, we implemented a series of heuristics to find better solutions to these difficult problems. All numerical experiments presented in this section use the solver outlined in Appendix D.2; thus in Step 4 of the sequential Algorithm 1, the estimated optimal allocation is the best solution returned by the solver using the techniques in Appendix D.2.

5.1 Performance of the Sequential Algorithm to Estimate the Pareto Set

To find the estimated Pareto set using the sequential Algorithm 1 for each of the 98 parent pair populations, we fixed the algorithm parameters $\delta_0 = 100$, $\delta = 2500$, and $\epsilon_i = 10^{-10}$ for all $i \leq r$, and

ran one sample path for each population. Every ten thousand samples, the percent of parent pairs misclassified, percent of Pareto parent pairs falsely excluded, and percent of non-Pareto parent pairs falsely included were recorded. Figure 5 shows the average of each of these quantities, taken across the 98 populations, as a function of the simulation budget n . Note that reported average percentages are correlated across the simulation budget values. For comparison, we ran MOCBA and equal allocation in a similar fashion. Parameters for MOCBA are $N_0 = 100$, $\Delta = 2500$, and $\delta = 1250$ (as in Example 4).

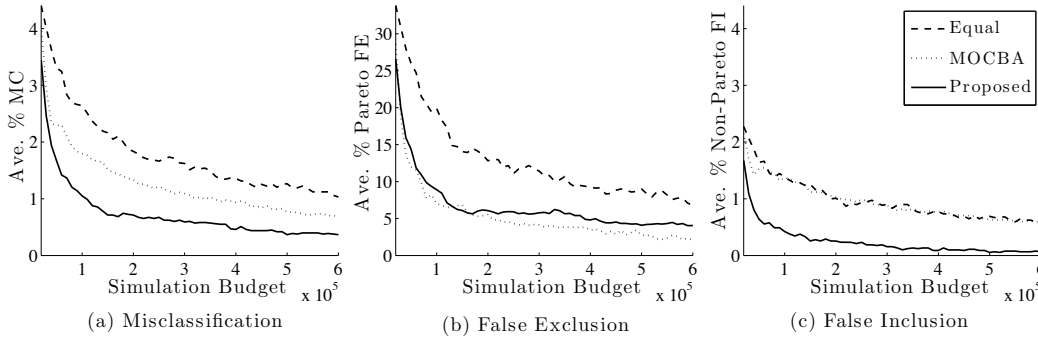


Figure 5: Average percent of parent pairs misclassified (MC), Pareto parent pairs falsely excluded (FE), and non-Pareto parent pairs falsely included (FI) for equal allocation, MOCBA, and the proposed Algorithm 1, calculated across one sample path each for 98 populations

From Figure 5(a), when considering the metric of the overall percent of parent pairs misclassified, the proposed algorithm performs better than MOCBA, which in turn performs better than equal allocation. However interestingly, in Figures 5(b) and 5(c), while the proposed allocation performs better than equal allocation, MOCBA appears to perform similar to the proposed allocation in 5(b) corresponding to false exclusions (FE) and similar to equal allocation in 5(c) corresponding to false inclusions (FI). This behavior occurs because, unlike the proposed algorithm, MOCBA does not allocate in a way that simultaneously controls the probability of MCE and MCI events. Instead, MOCBA initially alternates between allocating to minimize bounds on the probability of MCE and allocating to minimize bounds on the probability of MCI, depending on which estimated bound is larger. At any one iteration, MOCBA allocates to minimize the largest estimated bound. Since the estimated values of the bounds will converge to their true values, in the limit MOCBA allocates to minimize one bound. For the parent pair populations we consider, the estimated bound for the probability of MCE is routinely larger than the estimated bound for the probability of MCI, which

causes the MOCBA algorithm to allocate by minimizing the bound on the probability of MCE.

We note that MOCBA is a “closed-form” allocation scheme, requiring very little time to implement, while the proposed allocation scheme requires solving a convex optimization problem at each iteration in the sequential algorithm. Thus the tradeoff of computation time required to solve for the estimated optimal allocation versus computation time required to simulate a child may affect a practitioner’s choice between the proposed allocation and MOCBA. The benefit of using the proposed allocation scheme may increase as the computation time required to produce a simulated child grows. Further research is required to derive a large-population approximation to the proposed allocation scheme like the one for constrained SO in Pasupathy et al. (2014).

5.2 Performance of the Branch and Bound Algorithm

Table 3 reports how quickly the proposed branch and bound algorithm finds the optimal solution to Problem *MB*, for the 98 populations described in Appendix D.2.

Table 3: Statistics on the number of branch and bound nodes (BBN) visited by Algorithm 2

	Breeding Budget b					
	10	20	50	100	500 ^a	1000
Ave. # BBN	2.0	3.1	4.8	16.5	1,126.6	23.8
% of Poplns. with # BBN > 1	5.1	5.1	3.1	4.1	2.0	3.1
Ave. # BBN, given # BBN > 1	20.0	41.2	123.7	381.0	55,156.0	745.3

^a The table contains statistics from 588 branch and bound problems. Of these, only one (0.2 percent) required extensive computational time: Population 21, $b = 500$ visited 109,787 BBN.

From Table 3, the average number of branch and bound nodes (BBN) visited by Algorithm 2 generally increases as b increases. However Table 3 also shows that, as b increases, the percent of populations with BBN larger than one seems to decrease, while the average number of BBN visited *given* that BBN is greater than one increases. This result is expected: as b tends to infinity, giving all the breeding budget to the parent pair with the largest variance becomes optimal. However for populations where the “all-to-one” strategy is not immediately declared optimal, large b increases the number of feasible solutions, making the optimal solution harder to find.

5.3 Performance of Two-Step Procedure

Throughout the paper we have proposed a two-step procedure to solve the mating design problem, consisting of a simulation step and an IP step. In Figure 6, we evaluate this procedure under different simulation step policies by presenting the optimality gap of the solution to Problem \widehat{MB} as a function of sample size for various breeding budgets. Figure 6 required solving 7,056 branch and bound problems (in addition to the 588 solved in Table 3), of which eight (0.1 percent) required significant computational time. Seven are still running at the time of writing. These problems correspond to runs for populations 11, 21, 25, 49, and 73, which are excluded from the figures in this section. Future research may result in good heuristics for these difficult-to-solve problems. The simulation sample paths for the IP step in Figure 6 are the same as in Figure 5.

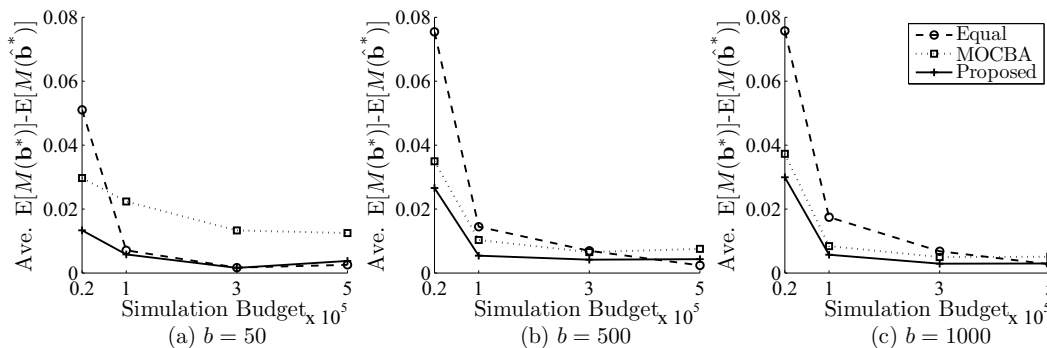


Figure 6: Average optimality gap of the estimated breeding budget as a function of sample size

While equal allocation initially performs worst at simulation budget $n = 20,000$, perhaps surprisingly, equal allocation eventually becomes a competitive algorithm in Figure 6. For smaller b , equal allocation becomes competitive by simulation budget $n = 100,000$. This result emphasizes the fact that our objective in proposing a simulation step that maximizes the rate of decay of the $P\{\text{MC}\}$ is to classify the Pareto parent pairs correctly, which happens at an exponential rate, while the IP step requires correct estimation of the means and variances of parent pairs *in* the Pareto set, which happens at rate $O(1/\sqrt{n_i})$ for parent pair i receiving n_i samples (see also Glynn and Juneja (2004)). Easily-classified parent pairs may receive insufficient simulation budget to meet the precision requirements of the IP step. While the proposed algorithm consistently performs well, one may imagine further splitting the simulation step into two: a first step for Pareto set classification, and a second step to ensure the means and variances of the Pareto set are estimated with high

precision. This procedure will also save computational resources since the estimation step can be conducted with naive allocations on only the estimated Pareto set.

In practice, the estimate of the expected maximum that breeding decision-makers will have is the objective value of Problem \widehat{MB} , $\hat{E}[M(\hat{\mathbf{b}}^*)]$. Thus we require an estimate of the predictive value of $\hat{E}[M(\hat{\mathbf{b}}^*)]$ for $E[M(\mathbf{b}^*)]$, the expected maximum trait observed in the progeny when implementing the estimated optimal breeding budget $\hat{\mathbf{b}}^*$. Figure 7 shows that the objective value of the estimated max breeding problem tends to overestimate the true expected maximum. This result might be expected since we conjecture that the first inequality of $E[\hat{E}[M(\hat{\mathbf{b}}^*)]] \geq E[M(\mathbf{b}^*)] \geq E[M(\hat{\mathbf{b}}^*)]$ holds (see Mak et al. (1999)). The proposed allocation often yields better predictions than MOCBA and equal allocation, particularly for smaller sample sizes. As in Figure 6, the predictive value will be better for more accurate mean and variance estimates.

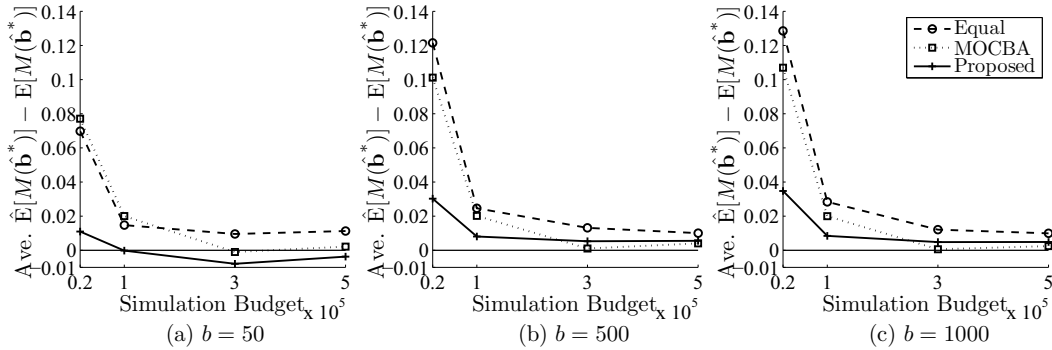


Figure 7: Average prediction gap of the objective of Problem \widehat{MB} as a function of sample size

The results in this section suggest that the use of the proposed algorithm can provide substantial gains in the performance of the max breeding problem, especially for smaller sample sizes.

6. Concluding Remarks

This paper provides an example of the collaboration opportunities available between the plant breeding and industrial engineering communities. Fruitful collaborations provide an avenue for advances and insights in both domains. Further, our paper contributes to the literature as follows.

- (i) We introduce the breeding objective of maximizing the expected trait value of the single best individual in the progeny population, a beneficial objective for plant breeders, and we provide implementable methods to design a breeding allocation based on this objective.
- (ii) We show

that when the trait distributions are normal, only parent pairs with non-dominated means and variances will receive nonzero sample at optimality in the max breeding problem. Thus we make the max breeding problem tractable by exploiting its structure to reduce the search space. (iii) We provide an implementable sequential sampling framework based on the asymptotically optimal sample allocation that maximizes the rate of decay of the probability of misclassification. Through numerical experiments, we show that this framework out-performs the current standard allocation scheme, MOCBA, in our application. (iv) We provide the first known application to successfully implement an LD-based sequential sampling scheme with non-normal rate functions.

We also re-emphasize the following practical insight from this work. A broad class of popular efficient sampling methods, including OCBA-like methods (Chen et al., 2000; Lee et al., 2012, 2010) and LD-based methods (Glynn and Juneja, 2004; Szechtman and Yücesan, 2008; Hunter and Pasupathy, 2013; Pasupathy et al., 2014), capitalize on the fact that the probability of misclassification decays to zero at an exponential rate as sample size increases. However, practitioners may also wish to accurately estimate the performance measure of interest for the best systems, which happens at a slower rate. Thus one might initially simulate using an OCBA-like or LD-based method to quickly classify systems, but then employ naive sampling allocations on the estimated-best systems to ensure an accurate solution to a second optimization step.

Finally, we observe that the branch and bound algorithm converges extremely fast in practice and allows the user to assert optimality of the resulting mating design, subject to estimation errors. Extensions of this work in plant breeding include developing analogous methods that (i) exploit a population structure based on pedigree, and (ii) consider multiple traits, for example, maximizing expected yield subject to fruit weight remaining in a desired range.

Acknowledgments. The authors thank Shane Henderson and Raghu Pasupathy for helpful discussions and suggestions, and for commenting on a draft version of this paper. In particular, Raghu Pasupathy suggested Lemma B.1 and Helmert’s transformation. The first author was supported in part by National Science Foundation Grant CMMI-1200315.

References

Allard, R.W. 1999. *Principles of Plant Breeding*. 2nd ed. J. Wiley & Sons.

- Andradóttir, S., S.-H. Kim. 2010. Fully sequential procedures for comparing constrained systems via simulation. *Naval Research Logistics* **57** 403–421. doi:10.1002/nav.20408.
- Batur, D., S.-H. Kim. 2010. Finding feasible systems in the presence of constraints on multiple performance measures. *ACM Transactions on Modeling and Computer Simulation* **20** 13:1–26.
- Boyd, S., L. Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press, New York.
- Buckler, E. S., et al. 2009. The genetic architecture of maize flowering time. *Science* **325** 714–718.
- Butler, J. C., D. J. Morrice, P. Mullarkey. 2001. A multiple attribute utility theory approach to ranking and selection. *Management Science* **47** 800–816.
- Chahota, R. K., et al. 2007. Predicting transgressive segregants in early generation using single seed descent method-derived *micro-macrosperma* genepool of lentil (*Lens culinaris Medikus*). *Euphytica* **156** 305–310.
- Chen, C.-H., J. Lin, E. Yücesan, S. E. Chick. 2000. Simulation budget allocation for further enhancing the efficiency of ordinal optimization. *Discrete Event Dynamic Systems* **10** 251–270. doi:10.1023/A:1008349927281.
- Cicirello, V. A., S. F. Smith. 2005. The max k -armed bandit: A new model of exploration applied to search heuristic selection. *Proc. of the Twentieth National Conference on Artificial Intelligence*. American Assoc. for Artificial Intelligence, 1355–1361.
- Falconer, D. S., T. F. Mackay. 1996. *Introduction to Quantitative Genetics*. 4th ed. Longman, Harlow, Essex, United Kingdom.
- Glynn, P. W., S. Juneja. 2004. A large deviations perspective on ordinal optimization. R. G. Ingalls, M. D. Rossetti, J. S. Smith, B. A. Peters, eds., *Proc. of the 2004 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers, Inc., Piscataway, NJ, 577–585. doi:10.1109/WSC.2004.1371364.
- Glynn, P. W., S. Juneja. 2006. Ordinal optimization: A large deviations perspective. Tech. rep., Indian School of Business. Working Paper Series.
- Healey, C. M., S. Andradóttir. 2014. Selection procedures for simulations with multiple constraints under independent and correlated sampling. *ACM Transactions on Modeling and Computer Simulation* **24** 25. doi:10.1145/2567921.
- Healey, C. M., S. Andradóttir, S.-H. Kim. 2013. Efficient comparison of constrained systems using dormancy. *European Journal of Operational Research* **224** 340–352. doi:10.1016/j.ejor.2012.08.012.
- Hunter, S. R., R. Pasupathy. 2013. Optimal sampling laws for stochastically constrained simulation optimization on finite sets. *INFORMS Journal on Computing* **25** 527–542. doi:10.1287/ijoc.1120.0519.
- Jansen, G. B., J. W. Wilton. 1985. Selecting mating pairs with linear programming techniques. *J Dairy Sci* **68** 1302–1305.

- Kim, S., J. Ryu. 2011a. The sample average approximation method for multi-objective stochastic optimization. S. Jain, R. R. Creasey, J. Himmelspace, K. P. White, M. Fu, eds., *Proc. of the 2011 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers, Inc., Piscataway, NJ, 4026–4037. doi:10.1109/WSC.2011.6148092.
- Kim, S., J. Ryu. 2011b. A trust-region algorithm for bi-objective stochastic optimization. *Procedia Computer Science* **4** 1422–1430. doi:10.1016/j.procs.2011.04.153.
- Kim, S.-H., B. L. Nelson. 2006. Selecting the best system. S. G. Henderson, B. L. Nelson, eds., *Simulation*. Handbooks in Operations Research and Management Science, Volume 13, Elsevier, 501–534.
- Kinghorn, B. P. 2011. An algorithm for efficient constrained mate selection. *Genetics Selection Evolution* **43**.
- Lee, L. H., E. P. Chew, S. Teng, D. Goldsman. 2010. Finding the non-dominated pareto set for multi-objective simulation models. *IIE Transactions* **42** 656–674. doi:10.1080/07408171003705367.
- Lee, L. H., N. A. Pujowidianto, L.-W. Li, C.-H. Chen, C. M. Yap. 2012. Approximate simulation budget allocation for selecting the best design in the presence of stochastic constraints. *IEEE Transactions on Automatic Control* **57** 2940–2945. doi:10.1109/TAC.2012.2195931.
- Li, X., C. Zhu, J. Wang, J. Yu. 2012. Computer simulation in plant breeding. *Advances in Agronomy* **116** 219–264.
- Luo, Y., E. Lim. 2013. Simulation-based optimization over discrete sets with noisy constraints. *IIE Transactions* **45** 699–715. doi:10.1080/0740817X.2012.733580.
- Lynch, M., B. Walsh. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Assoc, Sunderland, MA.
- Mak, W.-K., D. P. Morton, R. K. Wood. 1999. Monte carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters* **24** 47–56.
- McConnel, M. B., D. T. Galligan. 2004. The use of integer programming to select bulls across breeding companies with volume price discounts. *J Dairy Sci* **87** 3542–3549.
- McMullen, M. D., et al. 2009. Genetic properties of the maize nested association mapping population. *Science* **325** 737–740.
- Nagaraj, K., R. Pasupathy. 2013. R-SPLINE for local integer-ordered simulation optimization problems with stochastic constraints. R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, M. E. Kuhl, eds., *Proc. of the 2013 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers, Inc., Piscataway, NJ, 846–855. URL <http://informatics-sim.org/wsc13papers/includes/files/073.pdf>.
- Park, C., S.-H. Kim. 2011. Handling stochastic constraints in discrete optimization via simulation. S. Jain, R. R. Creasey, J. Himmelspace, K. P. White, M. Fu, eds., *Proc. of the 2011 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers, Inc., Piscataway, NJ, 4217–4226. doi:10.1109/WSC.2011.6148109.
- Pasupathy, R., S. R. Hunter, N. A. Pujowidianto, L. H. Lee, C.-H. Chen. 2014. Stochastically constrained ranking and selection via SCORE. *ACM Transactions on Modeling and Computer Simulation* **25** 1–26. doi:10.1145/2630066.

- Purdue University. 2013. Purdue moves the world forward. URL www.purdue.edu/purduemoves/research.html.
- Ryu, J., S. Kim, H. Wan. 2009. Pareto front approximation with adaptive weighted sum method in multi objective simulation optimization. M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, R. G. Ingalls, eds., *Proc. of the 2009 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers, Inc., 623–633. doi:10.1109/WSC.2009.5429562.
- Schnell, F. W. 1983. Probleme der elternwahl - ein überblick. *Arbeitstagung der Arbeitsgemeinschaft der Saatzuchleiter in Gumpenstein* 1–11.
- Simmonds, N. W. 1979. *Principles of Crop Improvement*. Longman, London.
- Smith, S. P., K. Hammond. 1987. Portfolio theory, utility theory and mate selection. *Génétique, Sélection, Évolution* **19** 321–336.
- Snape, J. W. 1982. Predicting the frequencies of transgressive segregants for yield and yield components in wheat. *Theoretical and Applied Genetics* **62** 127–134.
- Streeter, M. J., S. F. Smith. 2006a. An asymptotically optimal algorithm for the max k -armed bandit problem. *Proc. of the Twenty-First National Conference on Artificial Intelligence*. American Assoc. for Artificial Intelligence.
- Streeter, M. J., S. F. Smith. 2006b. A simple distribution-free approach to the max k -armed bandit problem. *Proc. of the Twelfth International Conference on Principles and Practice of Constraint Programming*.
- Szechtman, R., E. Yücesan. 2008. A new perspective on feasibility determination. S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler, eds., *Proc. of the 2008 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers, Inc., Piscataway, NJ, 273–280. doi:10.1109/WSC.2008.4736078.
- Zhong, S., J. L. Jannink. 2007. Using quantitative trait loci results to discriminate among crosses on the basis of their progeny mean and variance. *Genetics* **177** 567–576.

Author Biographies

Susan R. Hunter is an assistant professor in the School of Industrial Engineering at Purdue University. She received her Ph.D. in Industrial and Systems Engineering from Virginia Tech in 2011 and spent two years as a postdoctoral associate in the School of Operations Research and Information Engineering at Cornell University in Ithaca, NY. Her research interests include Monte Carlo methods and simulation optimization.

Benjamin McClosky contributed to this work as a research scientist at Nature Source Genetics in Ithaca, NY. He received his Ph.D. in Computational and Applied Mathematics from Rice University in 2006. His research focus is on scientific computing and optimization.

Online Supplement for Maximizing Quantitative Traits in the Mating Design Problem via Simulation-Based Pareto Estimation

Susan R. Hunter

School of Industrial Engineering, Purdue University, West Lafayette, IN 47907, USA,
susanhunter@purdue.edu

Benjamin McClosky

Nature Source Genetics, 33 Thornwood Drive, Ithaca, NY 14850, USA
bmcclosky@naturesourcegenetics.com

A. Simulation of quantitative traits

The computational results presented in this paper are based on experiments run on simulated F_2 populations derived from a bi-parental inbred cross (Simmonds, 1979). The mating design problem in this context asks which F_2 individuals to self in order to produce the best F_3 individual. A detailed description of the trait simulator can be found in McClosky and Tanksley (2013).

Our simulations used the standard maize genetic map (1350 cM; 10 chromosomes) (McMullen et al., 2009). Maize is considered a model organism for plants (Strable and Scanlon, 2009), analogous to mice or flies in animal breeding and genetics. This suggests that our conclusions are relevant to other plant species as well. Here we briefly review the simulation design.

To produce a population, the simulator randomly draws the number of Quantitative Trait Loci (QTL) uniformly between 5 and 100. The QTL can be thought of as casual genetic factors controlling the trait. The magnitude of the single allele substitution effect of each QTL was drawn from a gamma distribution with shape parameter 1.45, defined to be consistent with empirical evidence in maize (Buckler et al., 2009). To simulate dominance, we sampled a random subset of the QTL and introduced a heterozygous deviation from the midpoint of the homozygous classes. The magnitude of the d/a deviation was drawn from a normal distribution with mean zero and standard deviation 0.3. We scaled all genetic scores to ensure that exactly half of the trait variance was attributable to the QTL. The remainder of the trait variance modeled environmental factors.

We implemented the simulator and the branch and bound code in C⁺⁺. The SO was run using MATLAB. Numerical integration during branch and bound was performed using the composite Simpson’s Rule (Atkinson, 1989).

B. Proof of Theorem 1

Let b_i^* for $i \leq r$ be the integer allocations that maximize $E[M(\mathbf{b})]$. Further, let $\mu_r = \max_{i \leq r} \mu_i$. We consider the case of pairwise comparisons between parent pairs to derive an allocation result that will help us reduce the set of parent pairs in contention. Considering two parent pairs ($r=2$) sharing a total of b samples, Lemma B.1 states that the first parent pair will receive zero samples

and the second parent pair will receive b samples if taking one sample away from parent pair 1 and giving it to parent pair 2 always results in a higher expected value of the maximum.

Lemma B.1. *For any two parent pairs and finite $b \geq 2$, parent pair 1 will receive zero samples and parent pair 2 will receive b samples if for all $\nu_1 = 1, 2, \dots, b$,*

$$\begin{aligned} & E[M(b_1 = \nu_1 - 1)] - E[M(b_1 = \nu_1)] \\ &= \int_{-\infty}^{\infty} \Phi\left(\frac{y - \mu_1}{\sigma_1}\right)^{\nu_1 - 1} \Phi\left(\frac{y - \mu_2}{\sigma_2}\right)^{b - \nu_1} \left(\Phi\left(\frac{y - \mu_1}{\sigma_1}\right) - \Phi\left(\frac{y - \mu_2}{\sigma_2}\right)\right) dy > 0. \end{aligned} \quad (\text{B.1})$$

Proof. From Proposition 1, an allocation \mathbf{b} is superior to an allocation \mathbf{b}' if

$$\begin{aligned} E[M(\mathbf{b})] - E[M(\mathbf{b}')] &= \int_0^{\infty} 1 - \prod_{i=1}^r \Phi\left(\frac{y - \mu_i}{\sigma_i}\right)^{b_i} dy - \int_{-\infty}^0 \prod_{i=1}^r \Phi\left(\frac{y - \mu_i}{\sigma_i}\right)^{b_i} dy \\ &\quad - \int_0^{\infty} 1 - \prod_{i=1}^r \Phi\left(\frac{y - \mu_i}{\sigma_i}\right)^{b'_i} dy + \int_{-\infty}^0 \prod_{i=1}^r \Phi\left(\frac{y - \mu_i}{\sigma_i}\right)^{b'_i} dy \\ &= \int_{-\infty}^{\infty} \prod_{i=1}^r \Phi\left(\frac{y - \mu_i}{\sigma_i}\right)^{b'_i} - \prod_{i=1}^r \Phi\left(\frac{y - \mu_i}{\sigma_i}\right)^{b_i} dy > 0. \end{aligned} \quad (\text{B.2})$$

Then for two parent pairs, the result follows since

$$\begin{aligned} & E[M(b_1 = \nu_1 - 1)] - E[M(b_1 = \nu_1)] \\ &= \int_{-\infty}^{\infty} \Phi\left(\frac{y - \mu_1}{\sigma_1}\right)^{\nu_1} \Phi\left(\frac{y - \mu_2}{\sigma_2}\right)^{b - \nu_1} - \Phi\left(\frac{y - \mu_1}{\sigma_1}\right)^{\nu_1 - 1} \Phi\left(\frac{y - \mu_2}{\sigma_2}\right)^{b - (\nu_1 - 1)} dy. \end{aligned} \quad \square$$

Still in the context of two parent pairs, we use Lemma B.1 to prove the following Lemma B.2, which states that none of the breeding budget will be allocated to any parent pair dominated by another parent pair whose trait distribution has larger mean and variance. Theorem 1 follows as an immediate generalization of Lemma B.2.

Lemma B.2. *If $\mu_i \leq \mu_k$ and $\sigma_i \leq \sigma_k$ with at least one inequality holding strictly for some $i, k \in \{1, \dots, r\}$, $i \neq k$, then $b_i^* = 0$ for all $b \geq 2$.*

Proof. Consider the optimal allocation of a total of b children between two parent pairs, parent pair 1 and parent pair 2. We prove the result in two parts: (i) if $\mu_1 < \mu_2$ and $\sigma_1 = \sigma_2$, then $b_1^* = 0$ for all $b \geq 2$; (ii) if $\mu_1 = \mu_2$ and $\sigma_1 < \sigma_2$, then $b_1^* = 0$ for all $b \geq 2$; the proof of the main result follows from (i) and (ii). We use Lemma B.1 to show that for any allocation ν_1 to parent pair 1 with lower mean in (i) or with lower variance in (ii), taking a sample away from parent pair 1 and giving it to parent pair 2 increases the expected maximum observed characteristic in the children.

(i). Without loss of generality, let $\sigma_1 = \sigma_2 = 1$. Since $\mu_1 < \mu_2$, then $\Phi(y - \mu_1) > \Phi(y - \mu_2)$ for all y . Further, the cdfs $\Phi(y - \mu_1)^{\nu_1 - 1}$ and $\Phi(y - \mu_2)^{b - \nu_1}$ are both positive for all y . Therefore the integrand in (B.1) is always positive, and $E[M(b_1 = \nu_1 - 1)] > E[M(b_1 = \nu_1)]$.

(ii). Without loss of generality, let $\mu_1 = \mu_2 = 0$. Since $\sigma_1 < \sigma_2$, then $\Phi\left(\frac{y}{\sigma_1}\right) = \Phi\left(\frac{y}{\sigma_2}\right)$ at

$y = 0$, $\Phi\left(\frac{y}{\sigma_1}\right) < \Phi\left(\frac{y}{\sigma_2}\right)$ for all $y < 0$, and $\Phi\left(\frac{y}{\sigma_1}\right) > \Phi\left(\frac{y}{\sigma_2}\right)$ for all $y > 0$. From (B.1),

$$\begin{aligned} & \mathbb{E}[M(b_1 = \nu_1 - 1)] - \mathbb{E}[M(b_1 = \nu_1)] \\ &= \int_0^\infty \Phi\left(\frac{y}{\sigma_1}\right)^{\nu_1-1} \Phi\left(\frac{y}{\sigma_2}\right)^{b-\nu_1} \left(\Phi\left(\frac{y}{\sigma_1}\right) - \Phi\left(\frac{y}{\sigma_2}\right)\right) dy \\ & \quad - \int_{-\infty}^0 \Phi\left(\frac{y}{\sigma_1}\right)^{\nu_1-1} \Phi\left(\frac{y}{\sigma_2}\right)^{b-\nu_1} \left(\Phi\left(\frac{y}{\sigma_2}\right) - \Phi\left(\frac{y}{\sigma_1}\right)\right) dy. \end{aligned} \quad (\text{B.3})$$

where each integral in (B.3) is non-negative. Let $c \in \mathbb{R}, c > 0$. Since the normal distribution is symmetric about zero, then $\Phi(c/\sigma_1) - \Phi(c/\sigma_2) = \Phi(-c/\sigma_2) - \Phi(-c/\sigma_1)$. Therefore the difference between the two integrals in (B.3), if any, will be the result of the multiplier $\Phi(y/\sigma_1)^{\nu_1-1} \Phi(y/\sigma_2)^{b-\nu_1}$. Since $\Phi(y/\sigma_1)^{\nu_1-1} \Phi(y/\sigma_2)^{b-\nu_1}$ is a cdf composed of the product of strictly-increasing cdfs,

$$\Phi(c/\sigma_1)^{\nu_1-1} \Phi(c/\sigma_2)^{b-\nu_1} > \Phi(-c/\sigma_1)^{\nu_1-1} \Phi(-c/\sigma_2)^{b-\nu_1}.$$

Therefore equation (B.3) is positive. \square

C. Derivation of the Asymptotically Optimal Sampling Allocation

In this section, we provide the technical details and proofs for the asymptotically optimal sampling allocation, assuming all parameters of the normal distributions are known. For brevity, we refer to the parent pairs as “systems,” consistent with the literature on SO on finite sets.

C.1 Problem Statement

Formally, we consider the following problem statement:

Given a finite set of systems $i = 1, 2, \dots, r$, each with unknown primary objective value μ_i and secondary objective value σ_i^2 , and each of which can be estimated as output from a Monte Carlo simulation, we wish to identify all systems that belong to the Pareto set \mathcal{P} ,

$$\mathcal{P} = \{\text{systems } i : (\mu_i > \mu_k) \cup (\sigma_i^2 > \sigma_k^2) \text{ for all } k \leq r; k \neq i\}.$$

That is, \mathcal{P} is the set of systems non-dominated in terms of mean and variance, and \mathcal{P} is characterized by increasing means and decreasing variances.

Let $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_r)$ be a vector denoting the portion of the total sampling budget allocated to each system, where $\sum_{i=1}^r \alpha_i = 1, \alpha_i \geq 0$ for all $i = 1, \dots, r$. Let $\hat{\mathcal{P}}$ be the estimator of the Pareto set \mathcal{P} after the simulation budget has been expended. If $\hat{\mathcal{P}} \neq \mathcal{P}$, then at least one system has been *misclassified*, that is, a Pareto system has been falsely estimated as non-Pareto, or a non-Pareto system has been falsely estimated as Pareto. As the sampling budget tends to infinity, the probability of misclassification tends to zero. Then we ask, what sampling budget α maximizes the rate of decay of the probability of misclassification?

Remark 1. While we focus on allocating the sample to maximize the rate of decay of the probability of misclassification, one could also allocate to minimize the expected number of misclassifications. We show that these two objectives result in identical asymptotic allocations in §C.4.

C.2 MC Event Reformulation: Proof of Theorem 2

Theorem 2. Let $MC_{ph} := MCE \cup MCI_{ph}$. Then $P\{MC\} = P\{MC_{ph}\}$.

Proof. First, note that

$$MC = MCE \cup (MCI \cap MCE^c) \quad \text{and} \quad MC_{ph} = MCE \cup (MCI_{ph} \cap MCE^c).$$

(\Rightarrow) It is sufficient to show that $MCI \cap MCE^c$ implies MCI_{ph} . Suppose $MCI \cap MCE^c$. Let $j \notin \mathcal{P}$ and $j \in \hat{\mathcal{P}}$ be a non-Pareto system falsely estimated as Pareto. Then for each $i \in \mathcal{P}$, $\hat{Y}_j \geq \hat{Y}_i$ or $\hat{\sigma}_j^2 \geq \hat{\sigma}_i^2$. Thus $(\hat{Y}_j, \hat{\sigma}_j^2) \in \cap_{i \in \mathcal{P}} \{(\hat{Y}, \hat{\sigma}^2) : (\hat{Y} \geq \hat{Y}_i) \cup (\hat{\sigma}^2 \geq \hat{\sigma}_i^2)\}$. Since MCE^c , no Pareto systems dominate other Pareto systems. Therefore $(\hat{Y}_j, \hat{\sigma}_j^2) \in \cup_{\ell=0, \dots, p} \{(\hat{Y}, \hat{\sigma}^2) : (\hat{Y} \geq \hat{Y}_{[\ell]}) \cap (\hat{\sigma}^2 \geq \hat{\sigma}_{[\ell+1]}^2)\}$, that is, j lies in the union of the northeast quadrants defined by origins at the estimated phantom systems. Therefore MCI_{ph} occurs.

(\Leftarrow) It is sufficient to show that $MCI_{ph} \cap MCE^c$ implies MC . Suppose $MCI_{ph} \cap MCE^c$. Then all Pareto systems are estimated as Pareto. From the set of all $j \notin \mathcal{P}$ dominating some estimated phantom Pareto system, there exists $j^* \notin \mathcal{P}$ such that $j^* \in \hat{\mathcal{P}}$. (Otherwise, if there exists no such j^* , then each $j \notin \mathcal{P}$ is dominated by some $i \in \mathcal{P}$, and MCI_{ph} does not occur.) Therefore $\hat{\mathcal{P}} \neq \mathcal{P}$, which implies MC . \square

C.3 Rate Function Derivation: Proofs of Proposition 2 and Theorem 3

To efficiently identify the Pareto set, we wish to identify the sampling allocation vector that maximizes the rate of decay of the $P\{MC\}$. For $\ell b = \max(P\{MCE\}, P\{MCI_{ph}\})$, we have $\ell b \leq P\{MC\} \leq 2\ell b$, which, assuming the limits exist, implies

$$-\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{MC\} = \min \left(-\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{MCE\}, -\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{MCI_{ph}\} \right). \quad (C.1)$$

Since the rate function corresponding to the MCE term in equation (C.1) is most straightforward, we analyze it first. Since the estimated mean and estimated variance for a normal distribution are independent, \hat{Y}_i and $\hat{\sigma}_i^2$ are independent for all $i \leq r$. Bound the $P\{MCE\}$ by noting that for

$$\ell b_2 = \max_{i \in \mathcal{P}} \max_{k \leq r, k \neq i} P\{\hat{Y}_i \leq \hat{Y}_k\} P\{\hat{\sigma}_i^2 \leq \hat{\sigma}_k^2\},$$

we have $\ell b_2 \leq P\{MCE\} \leq r^2 \times \ell b_2$. Thus

$$-\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{MCE\} = \min_{i \in \mathcal{P}} \min_{\substack{k \leq r \\ k \neq i}} \left(-\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\hat{Y}_i \leq \hat{Y}_k\} - \lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\hat{\sigma}_i^2 \leq \hat{\sigma}_k^2\} \right). \quad (C.2)$$

By Glynn and Juneja (2004), under Assumption 1, for all $i \in \mathcal{P}, k \leq r$ such that $\mu_k < \mu_i$,

$$G(\alpha_i, \alpha_k) := -\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\widehat{Y}_i \leq \widehat{Y}_k\} = \frac{(\mu_i - \mu_k)^2}{2(\sigma_i^2/\alpha_i + \sigma_k^2/\alpha_k)}, \quad (\text{C.3})$$

where $\alpha_i, \alpha_k > 0$. If $\alpha_i = 0$ or $\alpha_k = 0$, then $G(\alpha_i, \alpha_k) := 0$. We derive an expression for the rate function corresponding to the event $\widehat{\sigma}_i^2 \leq \widehat{\sigma}_k^2$ in Proposition 2.

Proposition 2. *Under Assumption 1, for all $i \in \mathcal{P}, k \leq r$ such that $\sigma_k^2 < \sigma_i^2$,*

$$H(\alpha_i, \alpha_k) := -\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\widehat{\sigma}_i^2 \leq \widehat{\sigma}_k^2\} = \frac{\alpha_i}{2} \log \left(\frac{\sigma_i^2/\alpha_i + \sigma_k^2/\alpha_k}{\sigma_k^2(1/\alpha_i + 1/\alpha_k)} \right) + \frac{\alpha_k}{2} \log \left(\frac{\sigma_i^2/\alpha_i + \sigma_k^2/\alpha_k}{\sigma_i^2(1/\alpha_i + 1/\alpha_k)} \right),$$

where $\alpha_i, \alpha_k > 0$. If $\alpha_i = 0$ or $\alpha_k = 0$, then $H(\alpha_i, \alpha_k) := 0$.

Proof. Since Y_{im} are iid normal random variables with mean μ_i and variance σ_i^2 for all $m = 1, 2, \dots, n\alpha_i$, then

$$n\alpha_i \widehat{\sigma}_i^2 = \sum_{m=1}^{n\alpha_i} (Y_{im} - \frac{1}{n\alpha_i} \sum_{m=1}^{n\alpha_i} Y_{im})^2 = \sigma_i^2 \sum_{m=1}^{n\alpha_i} (Z_m - \frac{1}{n\alpha_i} \sum_{m=1}^{n\alpha_i} Z_m)^2 = \sigma_i^2 \sum_{m=1}^{n\alpha_i-1} U_m^2,$$

where the U_m 's arise from Helmert's transformation (Stuart and Ord, 1994, p. 379), and are iid normal random variables with zero mean and unit variance (here, we assume $n\alpha_i$ is an integer).

Note that since we were able to write the scaled $\widehat{\sigma}_i^2$ and $\widehat{\sigma}_k^2$ as sums of iid random variables, the rate function will be of the type specified in Glynn and Juneja (2004). To see this, consider the joint cumulant generating function of $(\widehat{\sigma}_i^2, \widehat{\sigma}_k^2)$,

$$\Lambda_{(\widehat{\sigma}_i^2, \widehat{\sigma}_k^2)}(t_i, t_k) = \log E \left[e^{t_i \frac{\sigma_i^2}{n\alpha_i} \sum_{m=1}^{n\alpha_i-1} U_m^2 + t_k \frac{\sigma_k^2}{n\alpha_k} \sum_{m=1}^{n\alpha_k-1} U_m^2} \right].$$

By the independence systems i and k ,

$$\begin{aligned} \Lambda_{(\widehat{\sigma}_i^2, \widehat{\sigma}_k^2)}(t_i, t_k) &= \log E \left[e^{t_i \frac{\sigma_i^2}{n\alpha_i} \sum_{m=1}^{n\alpha_i-1} U_m^2} \right] + \log E \left[e^{t_k \frac{\sigma_k^2}{n\alpha_k} \sum_{m=1}^{n\alpha_k-1} U_m^2} \right] \\ &= (n\alpha_i - 1) \log E \left[e^{t_i \frac{\sigma_i^2}{n\alpha_i} U_1^2} \right] + (n\alpha_k - 1) \log E \left[e^{t_k \frac{\sigma_k^2}{n\alpha_k} U_1^2} \right]. \end{aligned}$$

Thus noting that U_1^2 is a chi-square random variable with one degree of freedom, it follows that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \Lambda_{(\widehat{\sigma}_i^2, \widehat{\sigma}_k^2)}(nt_i, nt_k) &= \alpha_i \log E \left[e^{\sigma_i^2 \frac{t_i}{\alpha_i} U_1^2} \right] + \alpha_k \log E \left[e^{\sigma_k^2 \frac{t_k}{\alpha_k} U_1^2} \right] \\ &= -\frac{\alpha_i}{2} \log(1 - 2\sigma_i^2 \frac{t_i}{\alpha_i}) - \frac{\alpha_k}{2} \log(1 - 2\sigma_k^2 \frac{t_k}{\alpha_k}) \end{aligned}$$

for $t_i < \alpha_i/2\sigma_i^2$, $t_k < \alpha_k/2\sigma_k^2$. By the Gärtner-Ellis theorem (Dembo and Zeitouni, 1998), the large

deviations principle holds with good rate function

$$\begin{aligned}
J(x_i, x_k) &= \sup_{t_i, t_k} \left\{ t_i x_i + t_k x_k + \frac{\alpha_i}{2} \log(1 - 2\sigma_i^2 \frac{t_i}{\alpha_i}) + \frac{\alpha_k}{2} \log(1 - 2\sigma_k^2 \frac{t_k}{\alpha_k}) \right\} \\
&= \alpha_i \sup_{t_i/\alpha_i} \left\{ \frac{t_i}{\alpha_i} x_i + \frac{1}{2} \log(1 - 2\sigma_i^2 \frac{t_i}{\alpha_i}) \right\} + \alpha_k \sup_{t_k/\alpha_k} \left\{ \frac{t_k}{\alpha_k} x_k + \frac{1}{2} \log(1 - 2\sigma_k^2 \frac{t_k}{\alpha_k}) \right\} \\
&= \frac{\alpha_i}{2} \left[\frac{x_i}{\sigma_i^2} - \log\left(\frac{x_i}{\sigma_i^2}\right) - 1 \right] + \frac{\alpha_k}{2} \left[\frac{x_k}{\sigma_k^2} - \log\left(\frac{x_k}{\sigma_k^2}\right) - 1 \right]
\end{aligned}$$

for $x_i > 0, x_k > 0$. By the same simplification as in Glynn and Juneja (2004), it follows that

$$-\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\hat{\sigma}_i^2 \leq \hat{\sigma}_k^2\} = \inf_x \left(\frac{\alpha_i}{2} \left[\frac{x}{\sigma_i^2} - \log\left(\frac{x}{\sigma_i^2}\right) - 1 \right] + \frac{\alpha_k}{2} \left[\frac{x}{\sigma_k^2} - \log\left(\frac{x}{\sigma_k^2}\right) - 1 \right] \right).$$

After some algebra,

$$x(\alpha_i, \alpha_k) = \frac{\alpha_i + \alpha_k}{\alpha_i/\sigma_i^2 + \alpha_k/\sigma_k^2} = \frac{\sigma_k^2(\sigma_i^2/\alpha_i) + \sigma_i^2(\sigma_k^2/\alpha_k)}{\sigma_i^2/\alpha_i + \sigma_k^2/\alpha_k},$$

which implies the result. \square

Combining the result of Proposition 2 with the results in equations (C.2) and (C.3), the overall rate of decay of $P\{\text{MCE}\}$ is given by the following Lemma C.1.

Lemma C.1. *Let $\mathbb{I}_{\{\cdot\}}$ be the indicator function. Under Assumption 1, the rate function of $P\{\text{MCE}\}$ is*

$$-\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\text{MCE}\} = \min_{i \in \mathcal{P}} \min_{k \leq r, k \neq i} \left(G(\alpha_i, \alpha_k) \mathbb{I}_{\{\mu_k < \mu_i\}} + H(\alpha_i, \alpha_k) \mathbb{I}_{\{\sigma_k^2 < \sigma_i^2\}} \right).$$

Lemma C.1 states that the rate of decay of $P\{\text{MCE}\}$ is determined by the slowest rate function for the probability that a Pareto system is falsely dominated by some other system. For competing systems $k \in \mathcal{P}$, exactly one of the rate functions in Lemma C.1 will be nonzero at a time. For competing systems $k \notin \mathcal{P}$, at least one of the rate functions will be nonzero.

Let us turn our attention to the term corresponding to MCI_{ph} in (C.1). We do not directly derive the rate function for $P\{\text{MCI}_{ph}\}$; instead, we consider the overall rate of $P\{\text{MC}\}$, since certain terms in the rate of decay of $P\{\text{MCI}_{ph}\}$ will never be the unique minimum in the overall rate of decay of $P\{\text{MC}\}$.

First, we require additional notation. Recall that we labeled the Pareto systems from $1, 2, \dots, p$. Let the ordered list $\mathcal{O}_{\mathcal{P}} = \{(1, 1), (2, 2), \dots, (p, p)\}$ denote the positions of the true Pareto set on each objective. That is, Pareto system 1 is in “position 1” on the mean objective (smallest) and “position 1” on the variance objective (largest), corresponding to $(1, 1)$. Pareto system 2 is in position 2 on the mean objective (2nd smallest) and position 2 on the variance objective (2nd largest), corresponding to $(2, 2)$, and so on. Let $\hat{\mathcal{O}}_{[\mathcal{P}]}$ denote the ordered list of estimated positions of the true Pareto set. That is, if Pareto system 1 is estimated as being in position 1 on the mean objective and position 5 on the variance objective, the first element in the list $\hat{\mathcal{O}}_{[\mathcal{P}]}$ will be $(1, 5)$.

Further, define MCI_{ph} *without* order statistics as

$$\text{MCI}_{ph}^* := \bigcup_{j \notin \mathcal{P}} \bigcup_{\ell=0}^p (\hat{Y}_\ell \leq \hat{Y}_j) \cap (\hat{\sigma}_{\ell+1}^2 \leq \hat{\sigma}_j^2).$$

We present Lemma C.2 as an intermediary step to deriving the overall rate of decay of $P\{\text{MC}\}$.

Lemma C.2. *The rate of decay of $P\{\text{MC}\}$ is*

$$-\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\text{MC}\} = \min \left(-\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\text{MCE}\}, -\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\text{MCI}_{ph}^* \cap (\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{O}_{\mathcal{P}})\} \right). \quad (\text{C.4})$$

Proof. We derive upper and lower bounds on the rate of decay of the $P\{\text{MC}\}$ by considering bounds on the rate of decay of $P\{\text{MCI}_{ph}\}$. By the law of total probability, it follows that

$$P\{\text{MCI}_{ph}\} = P\{\text{MCI}_{ph} \cap (\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{O}_{\mathcal{P}})\} + \sum_{\mathcal{S} \neq \mathcal{O}_{\mathcal{P}}} P\{\text{MCI}_{ph} \cap (\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{S})\},$$

where \mathcal{S} is an ordered set of p elements of the form $\{(x_1, y_1), (x_2, y_2), \dots, (x_p, y_p)\}$ where the x coordinates and y coordinates are separately drawn without replacement from the set $\{1, 2, \dots, p\}$. Consider the term $P\{\text{MCI}_{ph} \cap (\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{O}_{\mathcal{P}})\}$. Since $\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{O}_{\mathcal{P}}$ occurs, we may rewrite MCI_{ph} without order statistics so that

$$P\{\text{MCI}_{ph}\} = P\{\text{MCI}_{ph}^* \cap (\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{O}_{\mathcal{P}})\} + \sum_{\mathcal{S} \neq \mathcal{O}_{\mathcal{P}}} P\{\text{MCI}_{ph} \cap (\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{S})\}. \quad (\text{C.5})$$

(*Lower bound on rate of decay of $P\{\text{MC}\}$.*) Consider an upper bound on the $P\{\text{MCI}_{ph}\}$. From equation (C.5), an upper bound is

$$P\{\text{MCI}_{ph}\} \leq P\{\text{MCI}_{ph}^* \cap (\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{O}_{\mathcal{P}})\} + \sum_{\mathcal{S} \neq \mathcal{O}_{\mathcal{P}}} P\{\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{S}\}.$$

We now derive the rate function corresponding to the upper bound. Assuming the limits exist and applying the principle of the slowest term (Ganesh et al., 2004, Lemma 2.1),

$$\begin{aligned} -\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\text{MCI}_{ph}\} &\geq -\lim_{n \rightarrow \infty} \frac{1}{n} \log \left(P\{\text{MCI}_{ph}^* \cap (\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{O}_{\mathcal{P}})\} + \sum_{\mathcal{S} \neq \mathcal{O}_{\mathcal{P}}} P\{\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{S}\} \right) \\ &= \min \left(-\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\text{MCI}_{ph}^* \cap (\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{O}_{\mathcal{P}})\}, \min_{\mathcal{S} \neq \mathcal{O}_{\mathcal{P}}} -\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{S}\} \right). \end{aligned}$$

From (C.1), it follows that

$$\begin{aligned} -\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\text{MC}\} &\geq \min \left(-\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\text{MCE}\}, -\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\text{MCI}_{ph}^* \cap (\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{O}_{\mathcal{P}})\}, \right. \\ &\quad \left. \min_{\mathcal{S} \neq \mathcal{O}_{\mathcal{P}}} -\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{S}\} \right). \end{aligned}$$

Consider the term $\min_{\mathcal{S} \neq \mathcal{O}_{\mathcal{P}}} - \lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{S}\}$. It is clear that any value of $\hat{\mathcal{O}}_{[\mathcal{P}]}$ that results in an MCE event will have a rate function that is greater than or equal to the corresponding rate function for MCE. Now consider values of $\hat{\mathcal{O}}_{[\mathcal{P}]}$ that do *not* result in MCE, such that all Pareto systems are estimated as Pareto, but may be estimated in the wrong *order*, e.g., $\hat{\mathcal{O}}_{[\mathcal{P}]} = \{(2, 2), (1, 1), (3, 3), \dots, (p, p)\}$, where Pareto systems 1 and 2 have exchanged positions. In this case, it is sufficient to consider only instances of $\hat{\mathcal{O}}_{[\mathcal{P}]}$ that contain pairwise exchanges, and further, it is sufficient to consider instances of $\hat{\mathcal{O}}_{[\mathcal{P}]}$ in which there is exactly one pair exchange. For any ℓ_1, ℓ_2 indexing the Pareto set such that $\ell_1 < \ell_2$, a pair exchange occurs if $(\hat{Y}_{\ell_1} \geq \hat{Y}_{\ell_2}) \cap (\hat{\sigma}_{\ell_1}^2 \leq \hat{\sigma}_{\ell_2}^2)$. Let $\hat{\mathcal{O}}_{[\mathcal{P}]}^{(\ell_1, \ell_2)}$ denote an ordering with exactly one pair exchange where ℓ_1 and ℓ_2 , $\ell_1 < \ell_2$, are the Pareto systems whose places have been exchanged. Then

$$\begin{aligned} - \lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\hat{\mathcal{O}}_{[\mathcal{P}]}^{(\ell_1, \ell_2)}\} &\geq - \lim_{n \rightarrow \infty} \frac{1}{n} \log P\{(\hat{Y}_{\ell_1} \geq \hat{Y}_{\ell_2}) \cap (\hat{\sigma}_{\ell_1}^2 \leq \hat{\sigma}_{\ell_2}^2)\} \\ &\geq - \lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\hat{\sigma}_{\ell_1}^2 \leq \hat{\sigma}_{\ell_2}^2\} \geq - \lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\text{MCE}\}. \end{aligned}$$

Thus the rate function for $\min_{\mathcal{S} \neq \mathcal{O}_{\mathcal{P}}} - \lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{S}\}$ will never be the unique minimum, and

$$- \lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\text{MC}\} \geq \min \left(- \lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\text{MCE}\}, - \lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\text{MCI}_{ph}^* \cap (\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{O}_{\mathcal{P}})\} \right).$$

(Upper bound on rate of decay of $P\{\text{MC}\}$.) We now seek a lower bound on the $P\{\text{MCI}_{ph}\}$ which leads to an upper bound on the rate of decay of the $P\{\text{MC}\}$. From equation (C.5), it is straightforward to see that $P\{\text{MCI}_{ph}\} \geq P\{\text{MCI}_{ph}^* \cap (\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{O}_{\mathcal{P}})\}$, and hence

$$- \lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\text{MCI}_{ph}\} \leq - \lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\text{MCI}_{ph}^* \cap (\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{O}_{\mathcal{P}})\}$$

which, together with equation (C.1), implies

$$- \lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\text{MC}\} \leq \min \left(- \lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\text{MCE}\}, - \lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\text{MCI}_{ph}^* \cap (\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{O}_{\mathcal{P}})\} \right).$$

The rate of decay of $P\{\text{MC}\}$ follows from the upper and lower bounds. \square

Theorem 3 states the overall rate of decay of $P\{\text{MC}\}$, where once again, we do not directly derive the rate of decay of $P\{\text{MCI}_{ph}^* \cap (\hat{\mathcal{O}}_{[\mathcal{P}]} = \mathcal{O}_{\mathcal{P}})\}$. Instead, we show that the rate is either equal to the rate function in the following line (C.7), or it is greater than or equal to the rate function corresponding to $P\{\text{MCE}\}$.

Theorem 3. Let $\alpha_0 := 1$ and $\alpha_{p+1} := 1$. Under Assumption 1, the rate function of $P\{MC\}$ is

$$-\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{MC\} = \min \left(\min_{i \in \mathcal{P}} \min_{k \leq r, k \neq i} \left(G(\alpha_i, \alpha_k) \mathbb{I}_{\{\mu_k < \mu_i\}} + H(\alpha_i, \alpha_k) \mathbb{I}_{\{\sigma_k^2 < \sigma_i^2\}} \right), \quad (C.6) \right.$$

$$\left. \min_{j \notin \mathcal{P}} \min_{\ell=0, \dots, p} \left(G(\alpha_\ell, \alpha_j) \mathbb{I}_{\{\mu_j < \mu_\ell\}} + H(\alpha_{\ell+1}, \alpha_j) \mathbb{I}_{\{\sigma_j^2 < \sigma_{\ell+1}^2\}} \right) \right). \quad (C.7)$$

Proof. From Lemmas C.1 and C.2, the substitution in line (C.6) corresponding to the rate of decay of $P\{MCE\}$ is straightforward. From equation (C.4), we focus on the quantity corresponding to $P\{MCI_{ph}^* \cap (\hat{\mathcal{O}}_{[p]} = \mathcal{O}_{\mathcal{P}})\}$ in line (C.7). Expanding the term corresponding to $\hat{\mathcal{O}}_{[p]} = \mathcal{O}_{\mathcal{P}}$, we have

$$\begin{aligned} P\{MCI_{ph}^* \cap (\hat{\mathcal{O}}_{[p]} = \mathcal{O}_{\mathcal{P}})\} &= P\left\{ \left[\bigcup_{j \notin \mathcal{P}} \bigcup_{\ell=0}^p (\hat{Y}_\ell \leq \hat{Y}_j) \cap (\hat{\sigma}_{\ell+1}^2 \leq \hat{\sigma}_j^2) \right] \cap \left[\bigcap_{\ell=0}^p (\hat{Y}_\ell \leq \hat{Y}_{\ell+1}) \cap (\hat{\sigma}_\ell^2 \geq \hat{\sigma}_{\ell+1}^2) \right] \right\} \\ &= P\left\{ \bigcup_{j \notin \mathcal{P}} \bigcup_{\ell=0}^p \left[(\hat{Y}_\ell \leq \hat{Y}_j) \cap (\hat{\sigma}_{\ell+1}^2 \leq \hat{\sigma}_j^2) \cap \left[\bigcap_{\ell'=0}^p (\hat{Y}_{\ell'} \leq \hat{Y}_{\ell'+1}) \cap (\hat{\sigma}_{\ell'}^2 \geq \hat{\sigma}_{\ell'+1}^2) \right] \right] \right\}, \end{aligned}$$

which, assuming the limits exist, for

$$\ell b_3 = \max_{j \notin \mathcal{P}} \max_{\ell=0, \dots, p} P\{(\hat{Y}_\ell \leq \hat{Y}_j) \cap (\hat{\sigma}_{\ell+1}^2 \leq \hat{\sigma}_j^2) \cap \left[\bigcap_{\ell'=0}^p (\hat{Y}_{\ell'} \leq \hat{Y}_{\ell'+1}) \cap (\hat{\sigma}_{\ell'}^2 \geq \hat{\sigma}_{\ell'+1}^2) \right]\},$$

we have $\ell b_3 \leq P\{(\hat{Y}_\ell \leq \hat{Y}_j) \cap (\hat{\sigma}_{\ell+1}^2 \leq \hat{\sigma}_j^2) \cap \left[\bigcap_{\ell'=0}^p (\hat{Y}_{\ell'} \leq \hat{Y}_{\ell'+1}) \cap (\hat{\sigma}_{\ell'}^2 \geq \hat{\sigma}_{\ell'+1}^2) \right]\} \leq r^2 \ell b_3$. Then

$$\begin{aligned} &-\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{MCI_{ph}^* \cap (\hat{\mathcal{O}}_{[p]} = \mathcal{O}_{\mathcal{P}})\} \\ &= \min_{j \notin \mathcal{P}} \min_{\ell=0, \dots, p} \left(-\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{(\hat{Y}_\ell \leq \hat{Y}_j) \cap (\hat{\sigma}_{\ell+1}^2 \leq \hat{\sigma}_j^2) \cap \left[\bigcap_{\ell'=0}^p (\hat{Y}_{\ell'} \leq \hat{Y}_{\ell'+1}) \cap (\hat{\sigma}_{\ell'}^2 \geq \hat{\sigma}_{\ell'+1}^2) \right]\} \right). \end{aligned} \quad (C.8)$$

Letting $\alpha_0 := 1, \alpha_{p+1} := 1, \sigma_{p+2}^2 := 0$, and

$$I_k(y) = \frac{(y - \mu_k)^2}{2\sigma_k^2} \quad \text{and} \quad J_k(z) = \frac{1}{2} \left[\frac{z}{\sigma_k^2} - \log \left(\frac{z}{\sigma_k^2} \right) - 1 \right],$$

for any $j \notin \mathcal{P}, \ell \in \{0, 1, \dots, p\}$, from equation (C.8) we have

$$\begin{aligned} &-\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{(\hat{Y}_\ell \leq \hat{Y}_j) \cap (\hat{\sigma}_{\ell+1}^2 \leq \hat{\sigma}_j^2) \cap \left[\bigcap_{\ell'=0}^p (\hat{Y}_{\ell'} \leq \hat{Y}_{\ell'+1}) \cap (\hat{\sigma}_{\ell'}^2 \geq \hat{\sigma}_{\ell'+1}^2) \right]\} \\ &= -\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{(\hat{Y}_\ell \leq \hat{Y}_j) \cap \left[\bigcap_{\ell'=0}^p (\hat{Y}_{\ell'} \leq \hat{Y}_{\ell'+1}) \right]\} - \lim_{n \rightarrow \infty} \frac{1}{n} \log P\{(\hat{\sigma}_{\ell+1}^2 \leq \hat{\sigma}_j^2) \cap \left[\bigcap_{\ell'=0}^p (\hat{\sigma}_{\ell'}^2 \geq \hat{\sigma}_{\ell'+1}^2) \right]\} \\ &= \left[\inf_{\substack{y_\ell \leq y_j \\ y_1 \leq \dots \leq y_\ell \leq \dots \leq y_p}} \alpha_j I_j(y_j) + \alpha_1 I_1(y_1) + \dots + \alpha_\ell I_\ell(y_\ell) + \dots + \alpha_p I_p(y_p) \right] \mathbb{I}_{\{\ell \neq 0\}} \end{aligned} \quad (C.9)$$

$$\begin{aligned} &+ \left[\inf_{\substack{z_{\ell+1} \leq z_j \\ z_1 \geq \dots \geq z_{\ell+1} \geq \dots \geq z_p}} \alpha_j J_j(z_j) + \alpha_1 J_1(z_1) + \dots + \alpha_{\ell+1} J_{\ell+1}(z_{\ell+1}) + \dots + \alpha_p J_p(z_p) \right] \mathbb{I}_{\{\ell \neq p\}}. \end{aligned} \quad (C.10)$$

(In line (C.9), if $\mu_\ell < \mu_j$, then the infimum occurs at $y_j = \mu_j$, $y_{\ell'} = \mu_{\ell'}$ for all $\ell' \in \{1, 2, \dots, p\}$, and line (C.9) equals zero. In line (C.10), if $\sigma_{\ell+1}^2 < \sigma_j^2$, then the infimum occurs at $z_j = \sigma_j^2$, $z_{\ell'} = \sigma_{\ell'}^2$ for all $\ell' \in \{0, 1, \dots, p-1\}$, and line (C.10) equals zero. However notice that $\mu_j < \mu_\ell$ or $\sigma_j^2 < \sigma_{\ell+1}^2$, since otherwise $j \in \mathcal{P}$. Thus the rate function in lines (C.9) and (C.10) is nonzero.)

Note that at optimality in lines (C.9) and (C.10), $y_\ell = y_j$ and $z_{\ell+1} = z_j$; thus we let $y_\ell = y_j$ be denoted as y and $z_{\ell+1} = z_j$ be denoted as z . Further, since $\mu_1 \leq \dots \leq \mu_{\ell-1} \leq \mu_\ell \leq \mu_{\ell+1} \leq \dots \leq \mu_p$ and $\sigma_1^2 \geq \dots \geq \sigma_\ell^2 \geq \sigma_{\ell+1}^2 \geq \sigma_{\ell+2}^2 \geq \dots \geq \sigma_p^2$, and $\mu_j < \mu_\ell$ or $\sigma_j^2 < \sigma_{\ell+1}^2$, it is sufficient to let $y_{\ell+1}, \dots, y_p$ equal $\mu_{\ell+1}, \dots, \mu_p$, respectively; z_1, \dots, z_ℓ equal $\sigma_1^2, \dots, \sigma_\ell^2$, respectively; and search for y in $[\mu_j, \mu_\ell]$ and z in $[\sigma_j^2, \sigma_{\ell+1}^2]$. Then from (C.9) and (C.10),

$$\begin{aligned} & - \lim_{n \rightarrow \infty} \frac{1}{n} \log P\{(\hat{Y}_\ell \leq \hat{Y}_j) \cap (\hat{\sigma}_{\ell+1}^2 \leq \hat{\sigma}_j^2) \cap [\bigcap_{\ell'=0}^p (\hat{Y}_{\ell'} \leq \hat{Y}_{\ell'+1}) \cap (\hat{\sigma}_{\ell'}^2 \geq \hat{\sigma}_{\ell'+1}^2)]\} \\ & = \left[\inf_{\substack{\mu_j < y < \mu_\ell \\ y_1 \leq \dots \leq y_{\ell-1} \leq y \leq \mu_\ell}} \alpha_j I_j(y) + \alpha_\ell I_\ell(y) + \sum_{i=1}^{\ell-1} \alpha_i I_i(y_i) \right] \mathbb{I}_{\{\mu_j < \mu_\ell\}} \end{aligned} \quad (\text{C.11})$$

$$+ \left[\inf_{\substack{\sigma_j^2 < z < \sigma_{\ell+1}^2 \\ z_p \leq \dots \leq z_{\ell+2} \leq z \leq \sigma_{\ell+1}^2}} \alpha_j J_j(z) + \alpha_{\ell+1} J_{\ell+1}(z) + \sum_{i=\ell+2}^p J_i(z_i) \right] \mathbb{I}_{\{\sigma_j^2 < \sigma_{\ell+1}^2\}} \quad (\text{C.12})$$

In (C.11) and (C.12), note that the values of $y_1, \dots, y_{\ell-1}$ and $z_p, \dots, z_{\ell+2}$ are either equal to their respective mean values, so that the rate functions evaluate to zero, or they are equal to y or z , respectively. (This observation follows from the KKT conditions, which are necessary and sufficient for global optimality in this convex minimization problem.) Then each infimum in (C.11) and (C.12) reduces to a one-dimensional problem with an unknown number of nonzero rate functions in each of the summation terms. Let y^* denote the $\arg \inf$ in (C.11) and z^* denote the $\arg \inf$ in (C.12), and let

$$y^{**} := \arg \min_y \alpha_j I_j(y) + \alpha_\ell I_\ell(y); \quad z^{**} := \arg \min_z \alpha_j J_j(z) + \alpha_{\ell+1} J_{\ell+1}(z).$$

Suppose $\mu_j < \mu_\ell$ and consider only line (C.11), since a similar argument holds for line (C.12). When $y^* = y^{**}$, which occurs when $\mu_{\ell-1} < \mu_j$ or when $\mu_j < \mu_{\ell-1}$ and $y^* \in [\mu_{\ell-1}, \mu_\ell]$, the rate function in (C.11) is equal to $G(\alpha_\ell, \alpha_j) \mathbb{I}_{\{\mu_j < \mu_\ell\}}$. Now suppose that $y^* \neq y^{**}$, which implies $\mu_j < \mu_{\ell-1}$ and $y^* \in [\mu_j, \mu_{\ell-1}]$. Then

$$\begin{aligned} (\text{C.11}) & = \inf_{\substack{\mu_j < y < \mu_{\ell-1} \\ y_1 \leq \dots \leq y_{\ell-2} \leq y \leq \mu_{\ell-1}}} \alpha_j I_j(y) + \alpha_{\ell-1} I_{\ell-1}(y) + \alpha_\ell I_\ell(y) + \sum_{i=1}^{\ell-2} \alpha_i I_i(y_i) \\ & \geq \inf_{y < \mu_{\ell-1}} \alpha_j I_j(y) + \alpha_{\ell-1} I_{\ell-1}(y) + \alpha_\ell I_\ell(y) + \sum_{i=1}^{\ell-2} \alpha_i I_i(y_i) \\ & \geq \inf_{y < \mu_{\ell-1}} \alpha_{\ell-1} I_{\ell-1}(y) + \alpha_\ell I_\ell(y) \\ & \geq \inf_{\mu_{\ell-1} < y < \mu_\ell} \alpha_{\ell-1} I_{\ell-1}(y) + \alpha_\ell I_\ell(y) \\ & = G(\alpha_{\ell-1}, \alpha_\ell) \mathbb{I}_{\{\mu_{\ell-1} < \mu_\ell\}} \geq - \lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\text{MCE}\}. \end{aligned}$$

Therefore if $y^* \neq y^{**}$, the rate function is not the unique minimum.

As mentioned previously, a similar argument holds for the case in line (C.12), such that $z^* \neq z^{**}$ implies the rate function in line (C.12) is not the unique minimum. Therefore the only rate functions that may be the unique minimum in equation (C.4) require $y^* = y^{**}$ and $z^* = z^{**}$. Therefore the result holds. \square

C.4 The Rate of Decay of the Expected Number of Misclassifications

Let $\mathbb{I}_{\{\cdot\}}$ be the indicator function, and consider that the expected number of misclassifications (NMC) is given by

$$\mathbb{E}[\text{NMC}] = E \left[\sum_{i \in \mathcal{P}} \mathbb{I}_{\{i \notin \hat{\mathcal{P}}\}} + \sum_{j \notin \mathcal{P}} \mathbb{I}_{\{j \in \hat{\mathcal{P}}\}} \right] = \sum_{i \in \mathcal{P}} P\{i \notin \hat{\mathcal{P}}\} + \sum_{j \notin \mathcal{P}} P\{j \in \hat{\mathcal{P}}\}.$$

Then assuming the limits exist (see Ganesh et al., 2004, Lemma 2.1),

$$-\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}[\text{NMC}] = \min \left(\min_{i \in \mathcal{P}} \left(-\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{i \notin \hat{\mathcal{P}}\} \right), \min_{j \notin \mathcal{P}} \left(-\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{j \in \hat{\mathcal{P}}\} \right) \right).$$

Consider that for $i \in \mathcal{P}$,

$$P\{i \notin \hat{\mathcal{P}}\} = P\left\{ \bigcup_{k \leq r, k \neq i} (\hat{Y}_i \leq \hat{Y}_k) \cap (\hat{\sigma}_i^2 \leq \hat{\sigma}_k^2) \right\},$$

which has lower bound $\ell b_4 = \max_{k \leq r, k \neq i} P\{\hat{Y}_i \leq \hat{Y}_k\} P\{\hat{\sigma}_i^2 \leq \hat{\sigma}_k^2\}$ and upper bound $r \times \ell b_4$. Then

$$\min_{i \in \mathcal{P}} \left(-\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{i \notin \hat{\mathcal{P}}\} \right) = \min_{i \in \mathcal{P}} \left(\min_{\substack{k \leq r \\ k \neq i}} \left(-\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\hat{Y}_i \leq \hat{Y}_k\} - \lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\hat{\sigma}_i^2 \leq \hat{\sigma}_k^2\} \right) \right),$$

which is equivalent to the expression for $-\lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\text{MCE}\}$ in equation (C.2).

Now consider that for $j \notin \mathcal{P}$, $P\{j \in \hat{\mathcal{P}}\} = P\{\cap_{k \leq r, k \neq j} (\hat{Y}_j \geq \hat{Y}_k) \cup (\hat{\sigma}_j^2 \geq \hat{\sigma}_k^2)\}$. Define

$$\text{MC}(j) := \bigcap_{k \leq r, k \neq j} (\hat{Y}_j \geq \hat{Y}_k) \cup (\hat{\sigma}_j^2 \geq \hat{\sigma}_k^2) \quad \text{and} \quad \text{MC}_{ph}(j) := \bigcup_{\ell=0}^p (\hat{Y}_{[\ell]} \leq \hat{Y}_j) \cap (\hat{\sigma}_{[\ell+1]}^2 \leq \hat{\sigma}_j^2),$$

where ℓ is the index of the phantom Pareto systems, and notice that

$$\text{MC}(j) = (\text{MC}(j) \cap \text{MCE}) \cup (\text{MC}(j) \cap \text{MCE}^c) \quad \text{and} \quad \text{MC}_{ph}(j) = (\text{MC}_{ph}(j) \cap \text{MCE}) \cup (\text{MC}_{ph}(j) \cap \text{MCE}^c).$$

By a proof similar to that of Theorem 2, it can be shown that $P\{j \in \hat{\mathcal{P}}\} = P\{\text{MC}(j)\} = P\{\text{MC}_{ph}(j)\}$. Now the contribution of this probability to the overall rate of decay of $\mathbb{E}[\text{NMC}]$ can be analyzed in processes similar to those in Lemma C.2 and Theorem 3.

D. Parent Pair Populations Used in §5

In §D.2, we present 98 randomly-generated populations of 100 parent pairs that were used in the numerical section. Originally, a total of 101 populations were generated for use in the numerical section, all of which are presented in §D.2. The sample paths generated for three populations, populations 16, 29, and 90, were excluded from the analyses in the numerical section for reasons we discuss in §D.1.

D.1 Description of Solver

While we were able to numerically solve Problem Q on the first try in most populations, some populations resulted in ill-conditioned Problems Q . If the solver struggled to find an optimal solution, the following steps were taken: (i) adjust the tolerances of the solver if necessary; (ii) solve a relaxed version of Problem Q with only constraints corresponding to MCE; (iii) solve a relaxed version of Problem Q with only constraints corresponding to MCI_{ph} ; (iv) evaluate the rate of decay, z , of a “guess” allocation; (v) re-solve Problem Q using the tolerances from (i) and a warm start, where the warm start is the allocation producing the largest z from among the guess allocation in (iv) and any allocations for which the solver converged in steps (ii) and (iii); (vi) return the allocation producing the largest z — equal allocation, the guess allocation in (iv), or an allocation for which the solver converged in (ii), (iii), or (v).

The guess allocation divides parent pairs into three groups: non-Pareto parent pairs in the bottom 70th percentile on both objectives (nP-B70), non-Pareto parent pairs in the top 30th percentile on both objectives (nP-T30), and Pareto parent pairs. Within each group, the allocation is distributed equally, however the guess allocation gives parent pairs in nP-T30 ten times more sample than those in nP-B70, and Pareto parent pairs receive four times more sample than parent pairs in nP-T30. The guess allocation proved routinely superior to equal allocation, and at times better than solutions to the relaxed versions of Problem Q . Indeed, sometimes equal allocation was a better solution than a relaxed version of Problems Q . All populations that were not solved on the first try have a note in the title of the graph. Graphs labeled “Warm start” indicate that the best solution is the result of solving Problem Q in step (v). A graph labeled “MCE only,” “MCI only,” or “Guess” indicates that the best solution is the result of step (ii), (iii), or (iv), respectively.

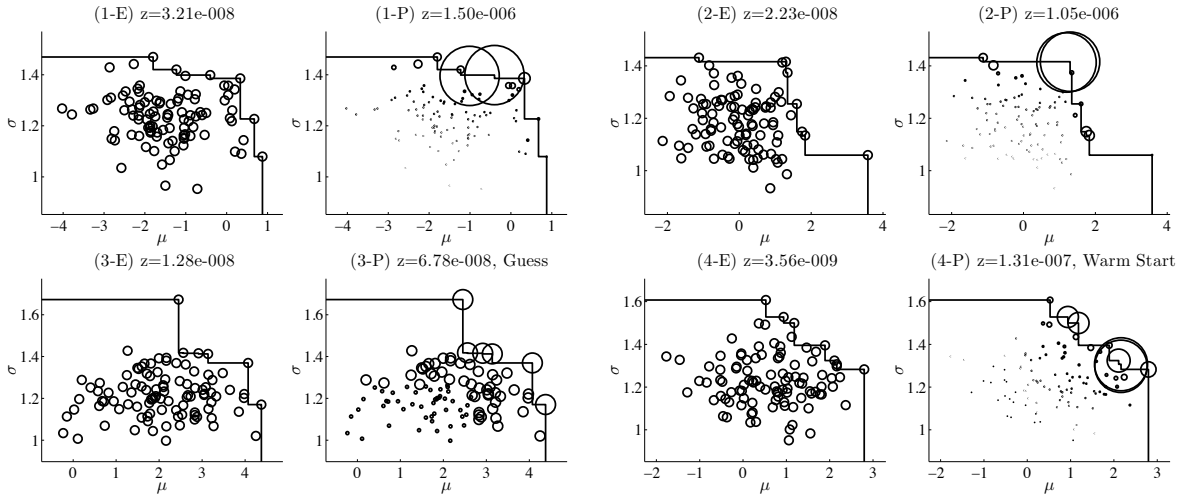
During implementation of the sequential algorithm, steps (i)–(vi) were conducted as previously described, except that a time limit of ten minutes was placed on finding a solution in any step (most solutions were found in seconds). In MATLAB, implementing the time limit required evaluating the objective function, which, on occasion, did not occur for very long periods of time as the solver searched for new directions. The sample paths for populations 16, 29, and 90 stalled in this fashion, and thus were not included in the analysis. We view this as a weakness of the solver we used and not as a weakness of the proposed algorithm — a heuristic allocation, such as the Guess allocation or equal allocation, should be applied if no solution is found in a desired time limit.

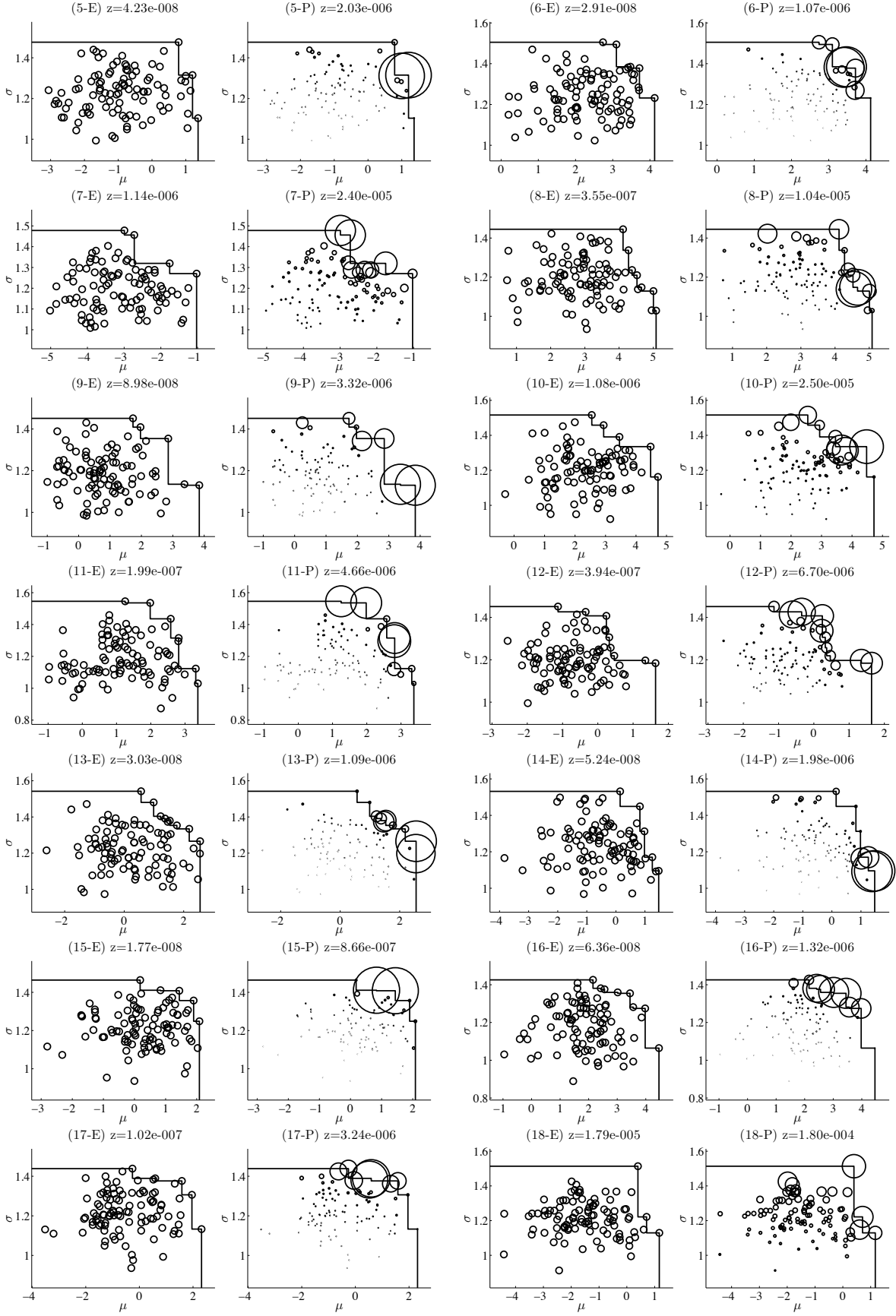
D.2 Parent Pair Populations and Allocations

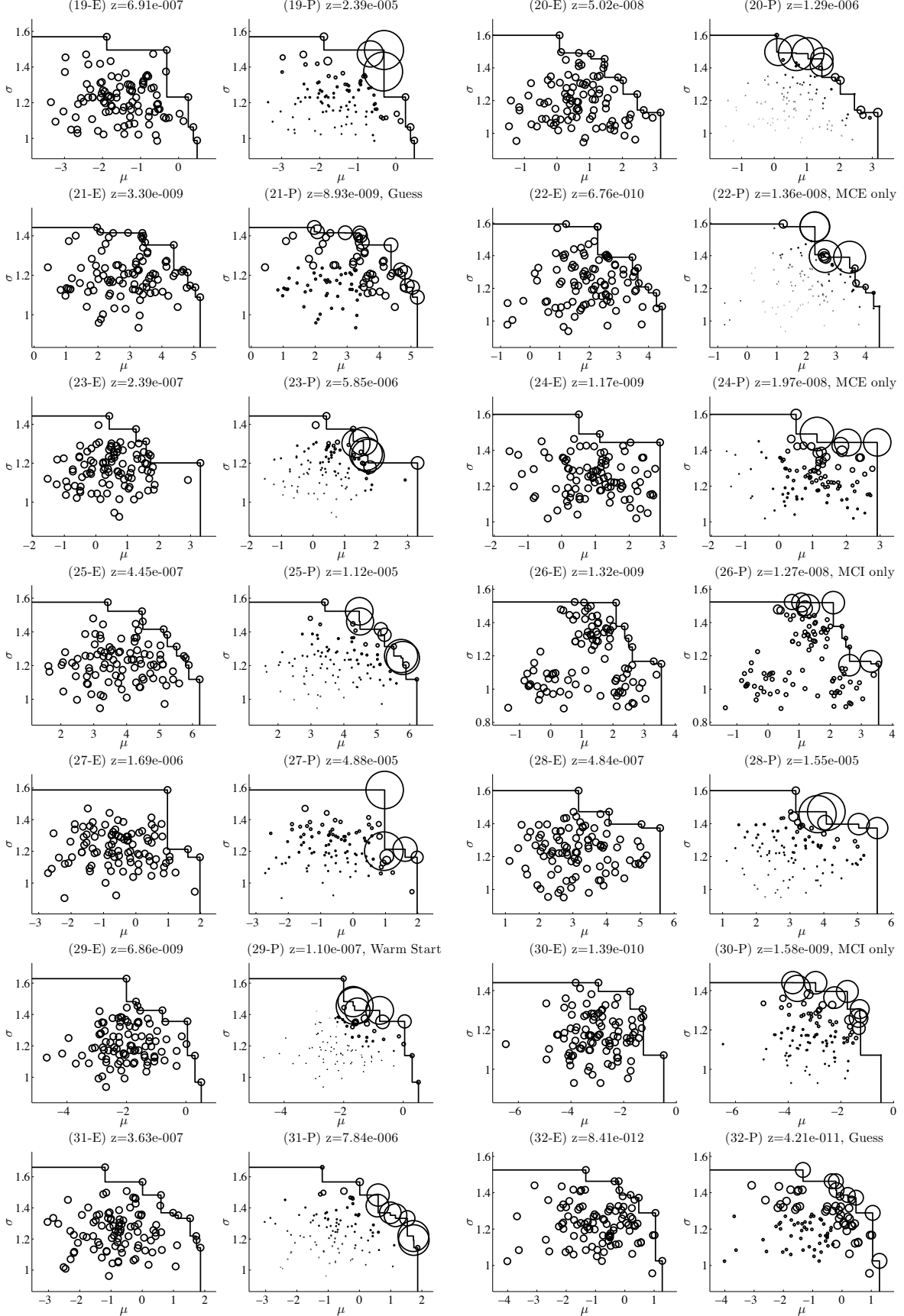
There are two graphs for each population. The first, for example labeled as (1-E), shows the amount of sample given to each parent pair under equal allocation. The second, for example (1-P), shows the amount of sample given to each parent pair under the proposed allocation, using the solver described in §D.1 and assuming the true μ and σ values are known (we solve Problem Q). In each case, the size of the circle is proportional to the amount of sampling budget n allocated to that parent pair. The rate z indicated in the title gives the true rate of decay of the probability of misclassification at the allocation presented in the graph, subject to solver tolerance.

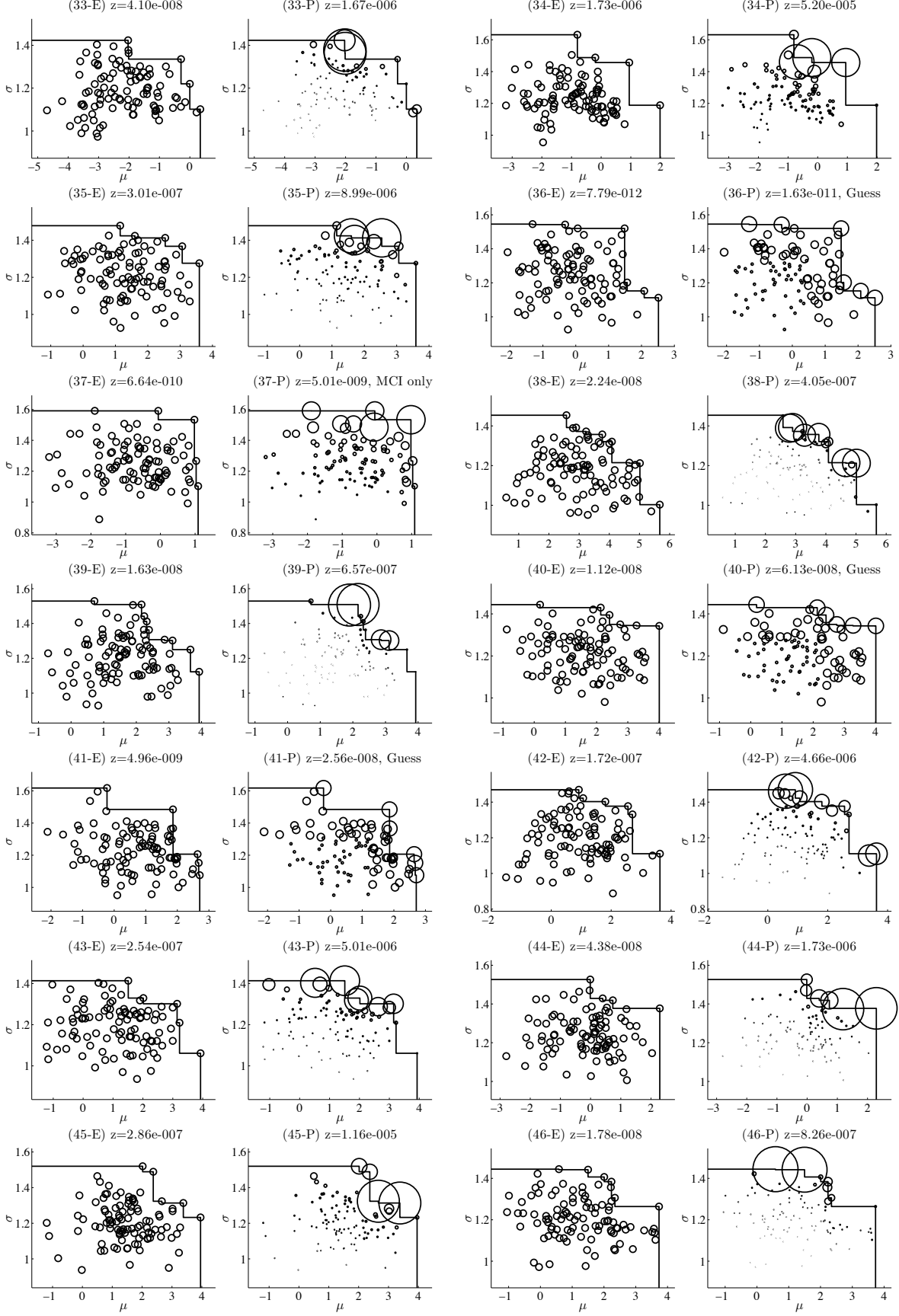
These graphical representations of the optimal allocations provide interesting insights into the nature of optimal allocations in a bi-objective context. Loosely speaking, we notice that systems in the Pareto set tend to receive more sample than systems that are “close” to the Pareto frontier, who in turn, receive more sample than systems that are “far away” from the Pareto frontier. These results are intuitive since the variances among the parent pair trait distributions are similar, and thus the relative locations of the parent pairs’ mean values primarily affect their likelihood of being falsely included in or excluded from the Pareto set.

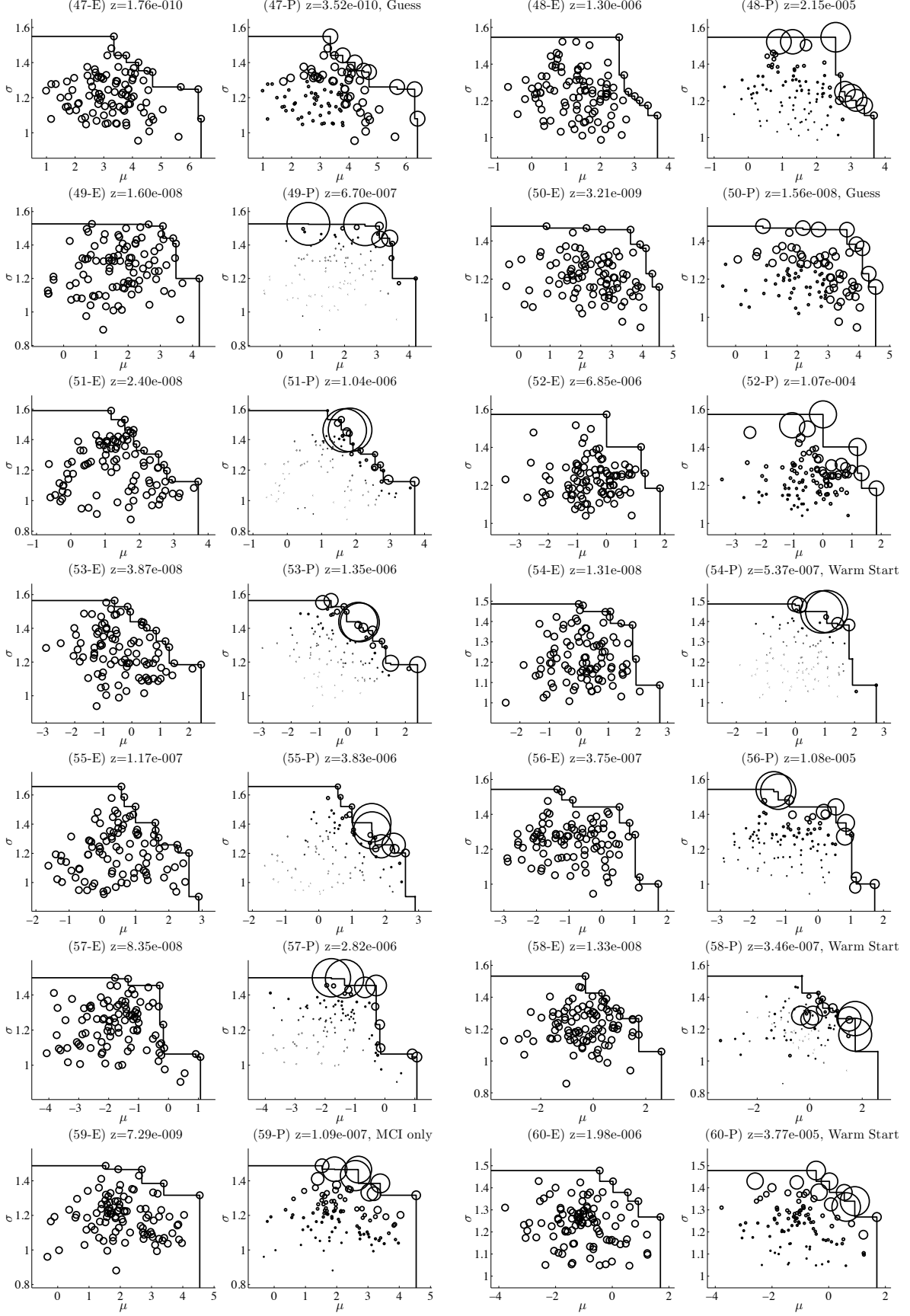
We further note that in some problem instances, such as 14-P, the sample allocation is focused on a few systems, while in other problem instances, the sample allocation is more spread out across the Pareto frontier, as in 12-P. Again loosely speaking, the problem instances in which much of the sample is concentrated on a few systems are problem instances in which there are several systems very close together requiring a large proportion of sample to determine which systems are in the Pareto set and which are not. If the systems in the Pareto set are more evenly distributed, and they have some reasonable separation from the systems they dominate, then the sample tends to be more evenly distributed.

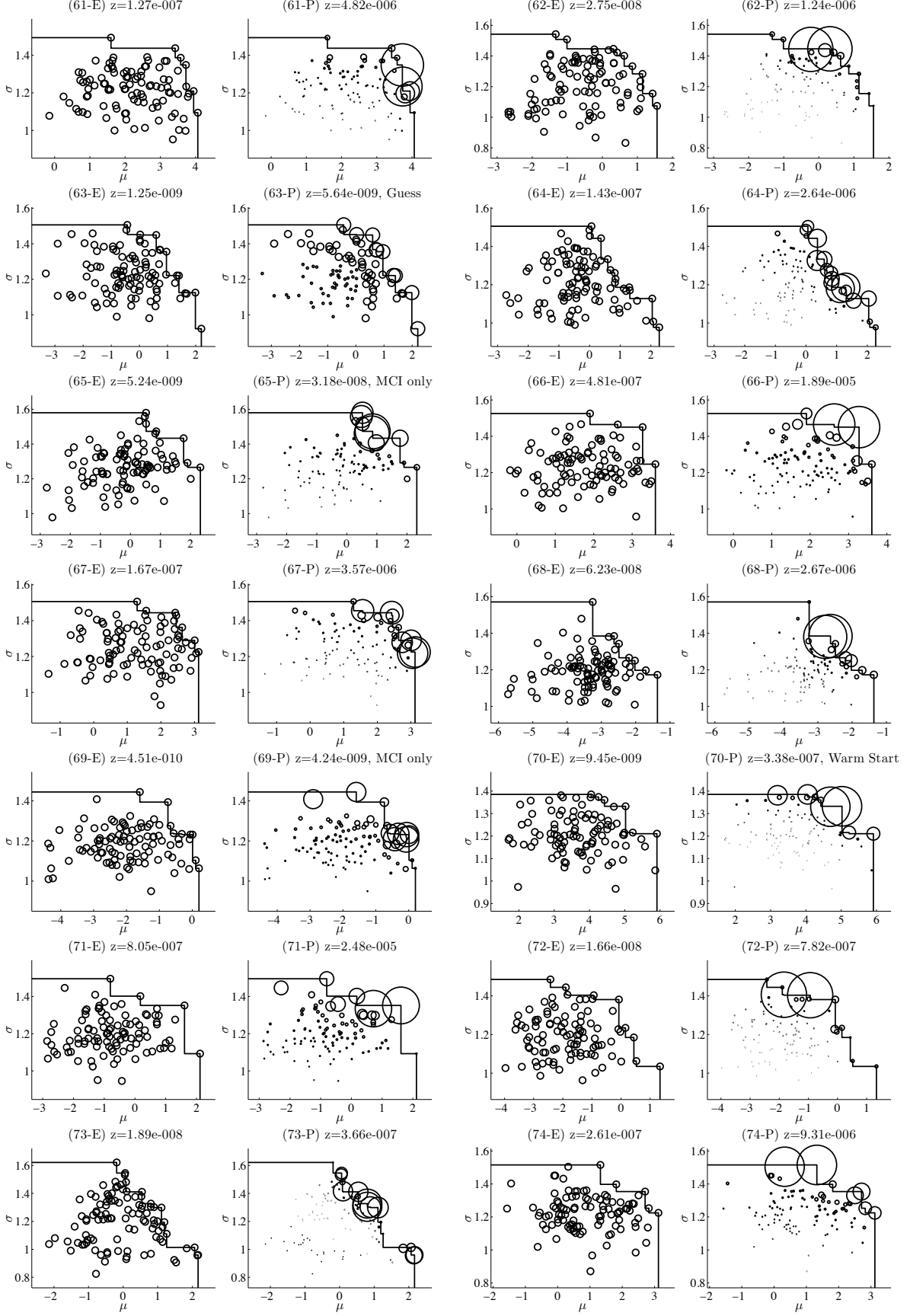


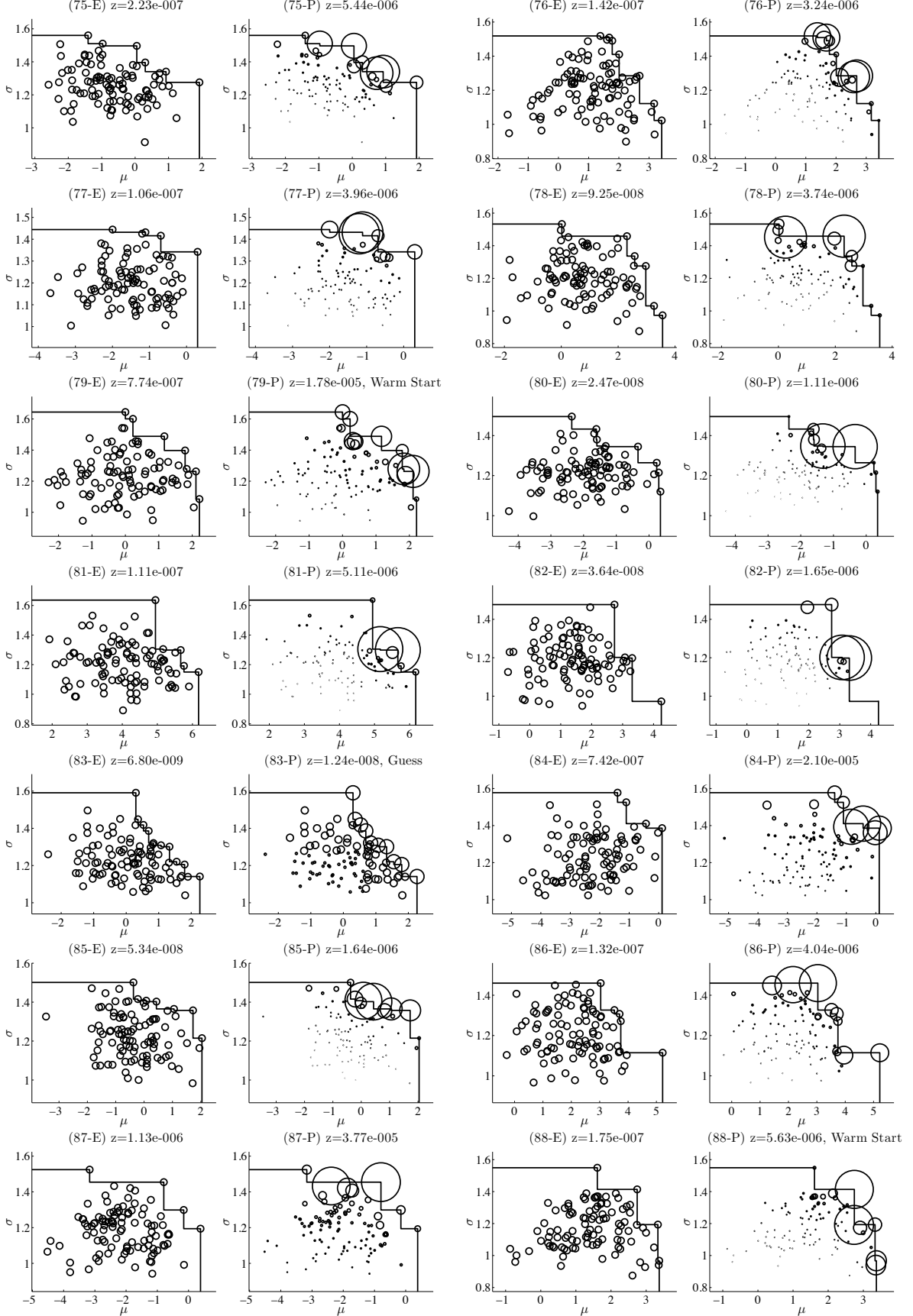


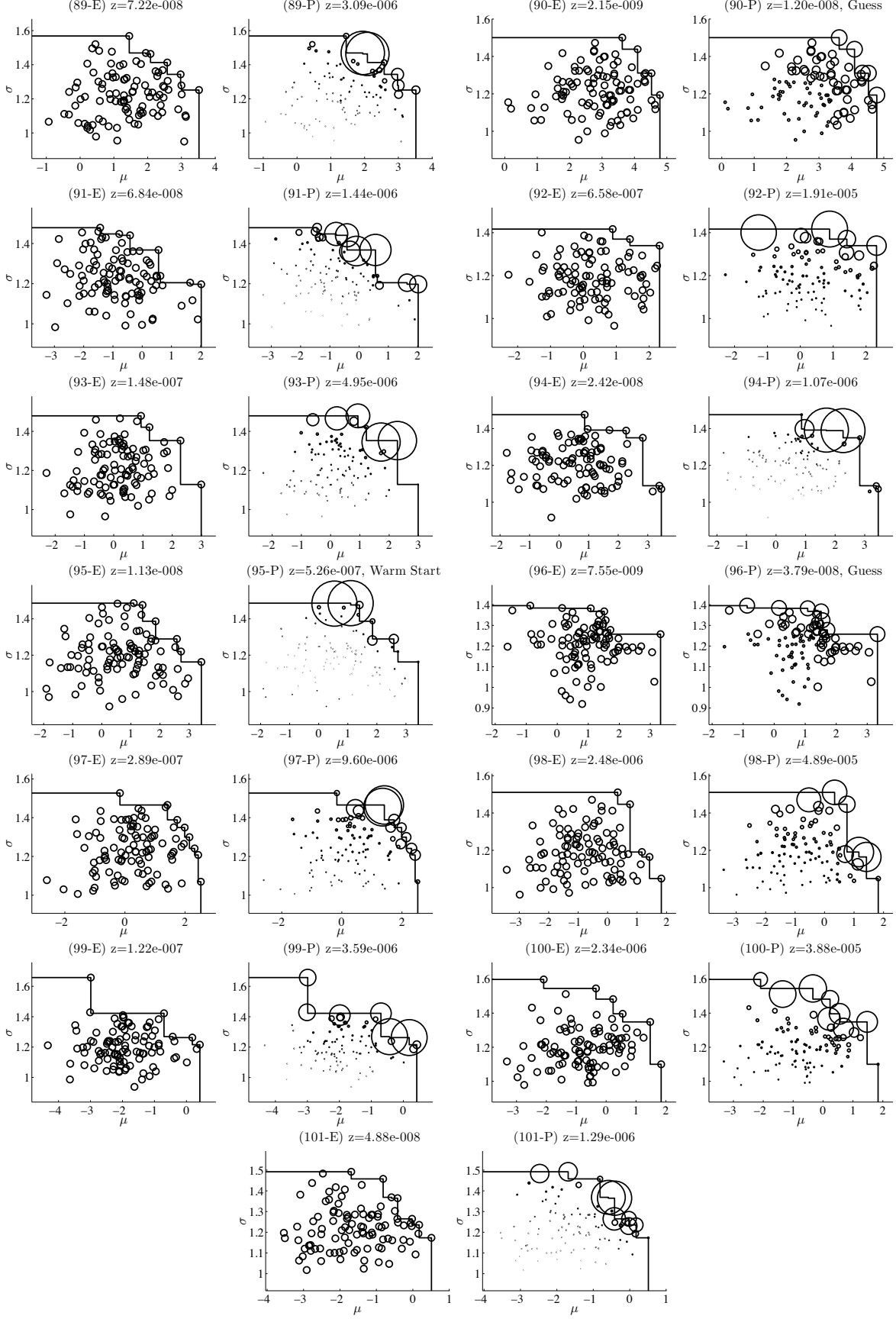












References

- Atkinson, K. E. 1989. *An Introduction to Numerical Analysis*. 2nd ed. John Wiley & Sons.
- Buckler, E. S., et al. 2009. The genetic architecture of maize flowering time. *Science* **325** 714–718.
- Dembo, A., O. Zeitouni. 1998. *Large Deviations Techniques and Applications*. 2nd ed. Springer, New York.
- Ganesh, A., N. O’Connell, D. Wischik. 2004. *Big Queues*. Lecture Notes in Mathematics, Volume 1838, Springer, New York.
- Glynn, P. W., S. Juneja. 2004. A large deviations perspective on ordinal optimization. R. G. Ingalls, M. D. Rossetti, J. S. Smith, B. A. Peters, eds., *Proc. of the 2004 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers, Inc., Piscataway, NJ, 577–585. doi:10.1109/WSC.2004.1371364.
- McClosky, B., S. D. Tanksley. 2013. The impact of recombination on short-term selection gain in plant breeding experiments. *Theoretical and Applied Genetics* **126** 2299–2312.
- McMullen, M. D., et al. 2009. Genetic properties of the maize nested association mapping population. *Science* **325** 737–740.
- Simmonds, N. W. 1979. *Principles of Crop Improvement*. Longman, London.
- Strable, J., M.J. Scanlon. 2009. Maize (*Zea mays*): A model organism for basic and applied research in plant biology. *Cold Spring Harbor Protocols* **4**.
- Stuart, A., J. K. Ord. 1994. *Kendall’s Advanced Theory of Statistics*, vol. 1 Distribution Theory. Sixth ed. Edward Arnold, London.