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 $E[M(b)]$. 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 \( \nu_1 = 1, 2, \ldots, b, \)

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

**Proof.** From Proposition 1, an allocation \( b \) is superior to an allocation \( b' \) if

\[
E[M(b)] - E[M(b')] = \int_0^\infty 1 - \prod_{i=1}^{r} \Phi \left( \frac{y - \mu_i}{\sigma_i} \right)^{b_i} dy - \int_{-\infty}^{0} \prod_{i=1}^{r} \Phi \left( \frac{y - \mu_i}{\sigma_i} \right)^{b_i} dy \\
- \int_0^\infty 1 - \prod_{i=1}^{r} \Phi \left( \frac{y - \mu_i}{\sigma_i} \right)^{b_i} dy + \int_{-\infty}^{0} \prod_{i=1}^{r} \Phi \left( \frac{y - \mu_i}{\sigma_i} \right)^{b_i} dy \\
= \int_{-\infty}^{\infty} \prod_{i=1}^{r} \Phi \left( \frac{y - \mu_i}{\sigma_i} \right)^{b_i} - \prod_{i=1}^{r} \Phi \left( \frac{y - \mu_i}{\sigma_i} \right)^{b_i} dy > 0. \tag{B.2} 
\]

Then for two parent pairs, the result follows since

\[
E[M(b_1 = \nu_1 - 1)] - E[M(b_1 = \nu_1)] \\
= \int_{-\infty}^{\infty} \Phi \left( \frac{y - \mu_1}{\sigma_1} \right)^{\nu_1 - 1} \Phi \left( \frac{y - \mu_2}{\sigma_2} \right)^{b - \nu_1 - 1} dy. \tag{B.2} 
\]

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 \( i, k \in \{1, \ldots, r\}, \ i \neq k \), 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 \( b_i^* = 0 \) for all \( b \geq 2 \); (ii) if \( \mu_1 = \mu_2 \) and \( \sigma_1 < \sigma_2 \), then \( b_i^* = 0 \) 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 \( \nu_1 \) 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
\begin{align*}
y = 0, \Phi \left( \frac{y}{\sigma_1} \right) &< \Phi \left( \frac{y}{\sigma_2} \right) \text{ for all } y < 0, \text{ and } \Phi \left( \frac{y}{\sigma_1} \right) > \Phi \left( \frac{y}{\sigma_2} \right) \text{ for all } y > 0. \text{ From (B.1),} \\
E[M(b_1 = \nu_1 - 1)] - 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_1} \right) - \Phi \left( \frac{y}{\sigma_2} \right) \right) dy. \tag{B.3}
\end{align*}

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 \( \Phi \left( \frac{c}{\sigma_1} \right) - \Phi \left( \frac{c}{\sigma_2} \right) = \Phi \left( -\frac{c}{\sigma_1} \right) - \Phi \left( -\frac{c}{\sigma_2} \right). \) Therefore the difference between the two integrals in (B.3), if any, will be the result of the multiplier \( \Phi \left( \frac{y}{\sigma_1} \right)^{\nu_1 - 1} \Phi \left( \frac{y}{\sigma_2} \right)^{b - \nu_1}. \) Since \( \Phi \left( \frac{y}{\sigma_1} \right)^{\nu_1 - 1} \Phi \left( \frac{y}{\sigma_2} \right)^{b - \nu_1} \) is a cdf composed of the product of strictly-increasing cdfs,

\( \Phi \left( \frac{c}{\sigma_1} \right)^{\nu_1 - 1} \Phi \left( \frac{c}{\sigma_2} \right)^{b - \nu_1} > \Phi \left( -\frac{c}{\sigma_1} \right)^{\nu_1 - 1} \Phi \left( -\frac{c}{\sigma_2} \right)^{b - \nu_1}. \)

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 \( i = 1, 2, \ldots, r, \) each with unknown primary objective value \( \mu_i \) and secondary objective value \( \sigma_i^2, \) 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 \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_r) \) 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 \( i = 1, \ldots, r. \) 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 \( \alpha \) 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 $\S C.4$.

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 $j \in \mathcal{P}$ 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 $(\hat{Y}_j, \hat{\sigma}_j^2) \in \cup_{c \in 0,...,p}\{(\hat{Y}, \hat{\sigma}^2) : (\hat{Y} \geq \hat{Y}_c) \cap (\hat{\sigma}^2 \geq \hat{\sigma}_c^2)\}$, 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 $j^* \in \mathcal{P}$. (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 $\mathcal{P} \neq \mathcal{P}$, which implies $MC$. \hfill \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

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

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

$$- \lim_{n \to \infty} \frac{1}{n} \log P\{MCE\} = \min_{i \in \mathcal{P}} \min_{k \leq r, k \neq i} \left( - \lim_{n \to \infty} \frac{1}{n} \log P\{\hat{Y}_i \leq \hat{Y}_k\} - \lim_{n \to \infty} \frac{1}{n} \log P\{\hat{\sigma}_i^2 \leq \hat{\sigma}_k^2\} \right). \quad (C.2)$$
By Glynn and Juneja (2004), under Assumption 1, for all \( i \in \mathcal{P}, k \leq r \) such that \( \mu_k < \mu_i \),

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

(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 \( \hat{\sigma}_i^2 \leq \hat{\sigma}_k^2 \) in Proposition 2.

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

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

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 \( \mu_i \) and variance \( \sigma_i^2 \) for all \( m = 1, 2, \ldots, n \), then

\[
n_{\alpha_i} \hat{\sigma}_i^2 = \sum_{m=1}^{n_{\alpha_i}} (Y_{im} - \frac{1}{n_{\alpha_i}} \sum_{m=1}^{n_{\alpha_i}} Y_{im})^2 = \sigma_i^2 \sum_{m=1}^{n_{\alpha_i}} (Z_m - \frac{1}{n_{\alpha_i}} \sum_{m=1}^{n_{\alpha_i}} Z_m)^2 = \sigma_i^2 \sum_{m=1}^{n_{\alpha_i}} U_{m}^2,
\]

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 \( \hat{\sigma}_i^2 \) and \( \hat{\sigma}_k^2 \) 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 \((\hat{\sigma}_i^2, \hat{\sigma}_k^2)\),

\[
A(\hat{\sigma}_i^2, \hat{\sigma}_k^2)(t_i, t_k) = \log E\left[ e^{t_i \hat{\sigma}_i^2 / \alpha_i} \sum_{m=1}^{n_{\alpha_i}} U_{m}^{2} + t_k \hat{\sigma}_k^2 / \alpha_k \sum_{m=1}^{n_{\alpha_k}} U_{m}^{2} \right].
\]

By the independence systems \( i \) and \( k \),

\[
A(\hat{\sigma}_i^2, \hat{\sigma}_k^2)(t_i, t_k) = \log E\left[ e^{t_i \hat{\sigma}_i^2 / \alpha_i} \sum_{m=1}^{n_{\alpha_i}} U_{m}^{2} \right] + \log E\left[ e^{t_k \hat{\sigma}_k^2 / \alpha_k} \sum_{m=1}^{n_{\alpha_k}} U_{m}^{2} \right] = (n_{\alpha_i} - 1) \log E\left[ e^{t_i \sigma_i^2 / \alpha_i U_i^2} \right] + (n_{\alpha_k} - 1) \log E\left[ e^{t_k \sigma_k^2 / \alpha_k U_k^2} \right].
\]

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

\[
\lim_{n \to \infty} \frac{1}{n} A(\hat{\sigma}_i^2, \hat{\sigma}_k^2)(nt_i, nt_k) = \alpha_i \log E\left[ e^{t_i \sigma_i^2 / \alpha_i U_i^2} \right] + \alpha_k \log E\left[ e^{t_k \sigma_k^2 / \alpha_k U_k^2} \right]
\]

\[
= -\frac{\alpha_i}{2} \log(1 - 2\sigma_i^2 / \alpha_i) - \frac{\alpha_k}{2} \log(1 - 2\sigma_k^2 / \alpha_k)
\]

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

\[
J(x_i, x_k) = \sup_{t_i, t_k} \{ t_i x_i + t_k x_k + \frac{\alpha_i}{2} \log(1 - 2 \sigma_i^2 / \alpha_i) + \frac{\alpha_k}{2} \log(1 - 2 \sigma_k^2 / \alpha_k) \}
\]

\[
= \alpha_i \sup_{t_i / \alpha_i} \{ \frac{t_i}{\alpha_i} x_i + \frac{1}{2} \log(1 - 2 \sigma_i^2 / \alpha_i) \} + \alpha_k \sup_{t_k / \alpha_k} \{ \frac{t_k}{\alpha_k} x_k + \frac{1}{2} \log(1 - 2 \sigma_k^2 / \alpha_k) \}
\]

\[
= \alpha_i \left[ \frac{x_i}{\sigma_i^2} - \log \left( \frac{x_i}{\sigma_i^2} \right) - 1 \right] + \alpha_k \left[ \frac{x_k}{\sigma_k^2} - \log \left( \frac{x_k}{\sigma_k^2} \right) - 1 \right]
\]

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

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

After some algebra,

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

which implies the result.

Combining the result of Proposition 2 with the results in equations (C.2) and (C.3), the overall rate of decay of \( P\{MCE\} \) 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 \( P\{MCE\} \) is

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

Lemma C.1 states that the rate of decay of \( P\{MCE\} \) 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 \( P\{MCI_{ph}\} \); instead, we consider the overall rate of \( P\{MC\} \), since certain terms in the rate of decay of \( P\{MCI_{ph}\} \) will never be the unique minimum in the overall rate of decay of \( P\{MC\} \).

First, we require additional notation. Recall that we labeled the Pareto systems from 1, 2, \ldots, \( p \). Let the ordered list \( \mathcal{O}_{\mathcal{P}} = \{(1, 1), (2, 2), \ldots, (p, p)\} \) 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

\[ MCI_{ph}^* := \bigcup_{\ell \neq ph} \bigcup_{\ell = 0}^p (\hat{Y}_\ell \leq \hat{Y}_j) \cap (\hat{\sigma}^2_\ell \leq \hat{\sigma}^2_j). \]

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

**Lemma C.2.** The rate of decay of \( P\{MC\} \) is

\[
- \lim_{n \to \infty} \frac{1}{n} \log P\{MC\} = \min \left( - \lim_{n \to \infty} \frac{1}{n} \log P\{MCE\}, - \lim_{n \to \infty} \frac{1}{n} \log P\{MCI_{ph}^* \cap (\hat{O}[p] = O_P)\} \right).
\]

**(Proof.** We derive upper and lower bounds on the rate of decay of the \( P\{MC\} \) by considering bounds on the rate of decay of \( P\{MCI_{ph}\} \). By the law of total probability, it follows that

\[
P\{MCI_{ph}\} = P\{MCI_{ph} \cap (\hat{O}[p] = O_P)\} + \sum_{\text{all } S \neq O_P} P\{MCI_{ph} \cap (\hat{O}[p] = S)\},
\]

where \( S \) is an ordered set of \( p \) elements of the form \( \{(x_1, y_1), (x_2, y_2), \ldots, (x_p, y_p)\} \) where the \( x \) coordinates and \( y \) coordinates are separately drawn without replacement from the set \( \{1, 2, \ldots, p\} \).

Consider the term \( P\{MCI_{ph} \cap (\hat{O}[p] = O_P)\} \). Since \( \hat{O}[p] = O_P \) occurs, we may rewrite \( MCI_{ph} \) without order statistics so that

\[
P\{MCI_{ph}\} = P\{MCI_{ph}^* \cap (\hat{O}[p] = O_P)\} + \sum_{S \neq O_P} P\{MCI_{ph} \cap (\hat{O}[p] = S)\}. \quad \text{(C.5)}
\]

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

\[
P\{MCI_{ph}\} \leq P\{MCI_{ph}^* \cap (\hat{O}[p] = O_P)\} + \sum_{S \neq O_P} P\{\hat{O}[p] = S\}.
\]

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),

\[
- \lim_{n \to \infty} \frac{1}{n} \log P\{MCI_{ph}\} \geq - \lim_{n \to \infty} \frac{1}{n} \log \left( P\{MCI_{ph}^* \cap (\hat{O}[p] = O_P)\} + \sum_{S \neq O_P} P\{\hat{O}[p] = S\} \right)
\]

\[
= \min \left( - \lim_{n \to \infty} \frac{1}{n} \log P\{MCI_{ph}^* \cap (\hat{O}[p] = O_P)\}, \min_{S \neq O_P} \lim_{n \to \infty} \frac{1}{n} \log P\{\hat{O}[p] = S\} \right).
\]

From (C.1), it follows that

\[
- \lim_{n \to \infty} \frac{1}{n} \log P\{MC\} \geq \min \left( - \lim_{n \to \infty} \frac{1}{n} \log P\{MCE\}, - \lim_{n \to \infty} \frac{1}{n} \log P\{MCI_{ph}^* \cap (\hat{O}[p] = O_P)\}, \min_{S \neq O_P} \lim_{n \to \infty} \frac{1}{n} \log P\{\hat{O}[p] = S\} \right).
\]
Consider the term \( \min_{S \neq \mathcal{O}_p} - \lim_{n \to \infty} \frac{1}{n} \log P(\hat{O}_p = S) \). It is clear that any value of \( \hat{O}_p \) 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 \( \hat{O}_p \) 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., \( \hat{O}_p = \{(2, 2), (1, 1), (3, 3), \ldots, (p, p)\} \), where Pareto systems 1 and 2 have exchanged positions. In this case, it is sufficient to consider only instances of \( \hat{O}_p \) that contain pairwise exchanges, and further, it is sufficient to consider instances of \( \hat{O}_p \) in which there is exactly one pair exchange. For any \( \ell_1, \ell_2 \) 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 \( \hat{O}^{(\ell_1, \ell_2)}_p \) denote an ordering with exactly one pair exchange where \( \ell_1 \) and \( \ell_2 \), \( \ell_1 < \ell_2 \), are the Pareto systems whose places have been exchanged. Then

\[
- \lim_{n \to \infty} \frac{1}{n} \log P(\hat{O}^{(\ell_1, \ell_2)}_p) \geq - \lim_{n \to \infty} \frac{1}{n} \log P((\hat{Y}_{\ell_1} \geq \hat{Y}_{\ell_2}) \cap (\hat{\sigma}_{\ell_1}^2 \leq \hat{\sigma}_{\ell_2}^2))
\]

\[
\geq - \lim_{n \to \infty} \frac{1}{n} \log P(\hat{\sigma}_{\ell_1}^2 \leq \hat{\sigma}_{\ell_2}^2) \geq - \lim_{n \to \infty} \frac{1}{n} \log P(MCE).
\]

Thus the rate function for \( \min_{S \neq \mathcal{O}_p} - \lim_{n \to \infty} \frac{1}{n} \log P(\hat{O}_p = S) \) will never be the unique minimum, and

\[
- \lim_{n \to \infty} \frac{1}{n} \log P(MC) \geq \min \left( - \lim_{n \to \infty} \frac{1}{n} \log P(MCE), - \lim_{n \to \infty} \frac{1}{n} \log P(MCI_{ph} \cap (\hat{O}_p = \mathcal{O}_p)) \right).
\]

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

\[
- \lim_{n \to \infty} \frac{1}{n} \log P(MCI_{ph}) \leq - \lim_{n \to \infty} \frac{1}{n} \log P(MCI_{ph} \cap (\hat{O}_p = \mathcal{O}_p))
\]

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

\[
- \lim_{n \to \infty} \frac{1}{n} \log P(MC) \leq \min \left( - \lim_{n \to \infty} \frac{1}{n} \log P(MCE), - \lim_{n \to \infty} \frac{1}{n} \log P(MCI_{ph} \cap (\hat{O}_p = \mathcal{O}_p)) \right).
\]

The rate of decay of \( P(MC) \) follows from the upper and lower bounds.

Theorem 3 states the overall rate of decay of \( P(MC) \), where once again, we do not directly derive the rate of decay of \( P(MCI_{ph} \cap (\hat{O}_p = \mathcal{O}_p)) \). 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 \( P(MCE) \).
Theorem 3. Let $\alpha_0 := 1$ and $\alpha_{p+1} := 1$. Under Assumption 1, the rate function of $P\{MC\}$ is

\begin{equation}
- \lim_{n \to \infty} \frac{1}{n} \log P\{MC\} = \min_{i \in P} \min_{k \leq r, k \neq i} \left( G(\alpha_i, \alpha_k) I_{\{\mu_k < \mu_i\}} + H(\alpha_i, \alpha_k) I_{\{\sigma_k^2 < \sigma_i^2\}} \right),
\end{equation}

\begin{equation}
- \lim_{n \to \infty} \frac{1}{n} \log P\{MC\} = \min_{j \not\in P} \min_{\ell = 0, \ldots, p} \left( G(\alpha_\ell, \alpha_j) I_{\{\mu_j < \mu_\ell\}} + H(\alpha_\ell, \alpha_j) I_{\{\sigma_\ell^2 < \sigma_j^2\}} \right).
\end{equation}

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 $P\{MCI_{ph} \cap (\hat{O}[\rho] = \rho)\}$ in line (C.7). Expanding the term corresponding to $\hat{O}[\rho] = \rho$, we have

\begin{equation}
P\{MCI_{ph} \cap (\hat{O}[\rho] = \rho)\} = P\left\{ \bigcup_{j \not\in P} \bigcup_{\ell = 0, \ldots, p} \left( \hat{Y}_\ell \leq \hat{Y}_j \right) \cap (\hat{\sigma}_\ell^2 + 1 \leq \hat{\sigma}_j^2) \cap \left[ \bigcap_{\ell' = 0}^p (\hat{Y}_{\ell'} \leq \hat{Y}_{\ell' + 1}) \right] \right\},
\end{equation}

which, assuming the limits exist, for

\begin{equation}
\ell b_3 = \max_j \max_{\ell = 0, \ldots, p} P\{ (\hat{Y}_\ell \leq \hat{Y}_j) \cap (\hat{\sigma}_\ell^2 + 1 \leq \hat{\sigma}_j^2) \cap \left[ \bigcap_{\ell' = 0}^p (\hat{Y}_{\ell'} \leq \hat{Y}_{\ell' + 1}) \right] \},
\end{equation}

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

\begin{equation}
- \lim_{n \to \infty} \frac{1}{n} \log P\{MCI_{ph} \cap (\hat{O}[\rho] = \rho)\}
= \min_{j \not\in P} \min_{\ell = 0, \ldots, p} \left( - \lim_{n \to \infty} \frac{1}{n} \log P\{ (\hat{Y}_\ell \leq \hat{Y}_j) \cap (\hat{\sigma}_\ell^2 + 1 \leq \hat{\sigma}_j^2) \cap \left[ \bigcap_{\ell' = 0}^p (\hat{Y}_{\ell'} \leq \hat{Y}_{\ell' + 1}) \right] \} \right).
\end{equation}

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

\begin{equation}
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\},
\end{equation}

for any $j \not\in P, \ell \in \{0, 1, \ldots, p\}$, from equation (C.8) we have

\begin{equation}
- \lim_{n \to \infty} \frac{1}{n} \log P\{ (\hat{Y}_\ell \leq \hat{Y}_j) \cap (\hat{\sigma}_\ell^2 + 1 \leq \hat{\sigma}_j^2) \cap \left[ \bigcap_{\ell' = 0}^p (\hat{Y}_{\ell'} \leq \hat{Y}_{\ell' + 1}) \right] \} = - \inf_{y_1 \leq y_j} \left[ \alpha_j I_j(y_j) + \alpha_1 I_1(y_1) + \ldots + \alpha_\ell I_\ell(y_\ell) + \ldots + \alpha_p I_p(y_p) \right] I_{\{\ell \not= 0\}}
\end{equation}

\begin{equation}
+ \left[ \inf_{z_1 \leq z_j} \alpha_j J_j(z_j) + \alpha_1 J_1(z_1) + \ldots + \alpha_\ell J_\ell(z_\ell) + \ldots + \alpha_p J_p(z_p) \right] I_{\{\ell \not= p\}},
\end{equation}

\begin{equation} \tag{9}
\end{equation}
When rate function in (C.11) is equal to $G_y$ and $\sigma_y$ denoted as $j$, since otherwise $\ell'y = \ell'p$ for all $\ell' \in \{0, 1, \ldots, p-1\}$, and line (C.10) equals zero. However notice that $\mu_j < \mu_\ell$ or $\sigma_j^2 < \sigma_\ell^2$, 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), $y = y_j$ and $z_{\ell+1} = z_j$; thus we let $y = y_j$ be denoted as $y$ and $z_{\ell+1} = z_j$ be denoted as $z$. Further, since $\mu_1 \leq \ldots \leq \mu_{\ell-1} \leq \mu_\ell \leq \mu_{\ell+1} \leq \ldots \leq \mu_p$ and $\sigma_1^2 \geq \ldots \geq \sigma_\ell^2 \geq \sigma_{\ell+2}^2 \geq \ldots \sigma_p^2$, and $\mu_j < \mu_\ell$ or $\sigma_j^2 < \sigma_\ell^2$, it is sufficient to let $y_{\ell+1}, \ldots, y_p$ equal $\mu_{\ell+1}, \ldots, \mu_p$, respectively; $z_1, \ldots, z_{\ell+2}$ equal $\sigma_1^2, \ldots, \sigma_{\ell+2}^2$, 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),

$$-\lim_{n \to \infty} \frac{1}{n} \log P\{\hat{Y}_\ell \leq \hat{Y}_j \cap (\hat{\sigma}_{\ell+1}^2 \leq \hat{\sigma}_j^2) \cap \left[ \bigcap_{\ell' = 0}^p (\hat{Y}_{\ell'} \leq \hat{Y}_{\ell+1}) \cap (\hat{\sigma}_{\ell'}^2 \geq \hat{\sigma}_{\ell+1}^2) \right]\}$$

$$= \left[ \inf_{\mu_j < y \leq \mu_{\ell-1}} \alpha_j I_j(y) + \alpha_{\ell-1} I_{\ell-1}(y) + \sum_{i=1}^{\ell-2} \alpha_i I_i(y_i) \mathbb{I}_{\{\mu_j < \mu_{\ell-1}\}} \right]$$

$$+ \left[ \inf_{\mu_{\ell-1} < z < \sigma_{\ell+1}^2} \alpha_j J_j(z) + \alpha_{\ell+1} J_{\ell+1}(z) + \sum_{i=\ell+2}^p J_i(z_i) \mathbb{I}_{\{\sigma_j^2 < \sigma_{\ell+1}^2\}} \right].$$

In (C.11) and (C.12), note that the values of $y_1, \ldots, y_{\ell-1}$ and $z_p, \ldots, z_{\ell+2}$ 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 min in (C.11) and $z^*$ denote the arg min 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

(C.11) $$\geq \inf_{y < \mu_{\ell-1}} \alpha_j I_j(y) + \alpha_{\ell-1} I_{\ell-1}(y) + \alpha_{\ell} I_{\ell}(y) + \sum_{i=1}^{\ell-2} \alpha_i I_i(y_i)$$

$$\geq \inf_{y < \mu_{\ell-1}} \alpha_{\ell-1} I_{\ell-1}(y) + \alpha_{\ell} I_{\ell}(y)$$

$$\geq \inf_{\mu_{\ell-1} < y < \mu_\ell} \alpha_{\ell-1} I_{\ell-1}(y) + \alpha_{\ell} I_{\ell}(y)$$

$$= G(\alpha_{\ell-1}, \alpha_\ell)\mathbb{I}_{\{\mu_j < \mu_{\ell-1}\}} \geq - \lim_{n \to \infty} \frac{1}{n} \log P\{\text{MCE}\}.$$
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.

\[\Box\]

### 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

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

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

\[
-\lim_{n \to \infty} \frac{1}{n} \log E[NMC] = \min_{i \in \hat{P}} \left( -\lim_{n \to \infty} \frac{1}{n} \log P\{i \notin \hat{P}\} \right), \min_{j \notin \hat{P}} \left( -\lim_{n \to \infty} \frac{1}{n} \log P\{j \in \hat{P}\} \right)
\]

Consider that for \( i \in \hat{P} \),

\[
P\{i \notin \hat{P}\} = P\{ \bigcup_{k \leq r, k \neq i} (\hat{Y}_i \leq \hat{Y}_k) \cap (\hat{\sigma}_i^2 \leq \hat{\sigma}_k^2) \},
\]

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

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

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

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

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

where \( \ell \) 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 \( P\{j \in \hat{P}\} = P\{\text{MC}(j)\} = P\{\text{MC}_{ph}(j)\} \). Now the contribution of this probability to the overall rate of decay of \( E[NMC] \) 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 $\mu$ and $\sigma$ 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.