# Sampling Laws for Stochastically Constrained Simulation Optimization on Finite Sets

Susan R. Hunter

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Industrial & Systems Engineering

Michael R. Taaffe, Chair
Ebru K. Bish
Roberto Szechtman
G. Don Taylor

September 23, 2011
Blacksburg, Virginia

Keywords: constrained simulation optimization, optimal allocation, ranking and selection
Copyright 2011, Susan R. Hunter

# Sampling Laws for Stochastically Constrained Simulation Optimization on Finite Sets

Susan R. Hunter

(ABSTRACT)

Consider the context of selecting an optimal system from among a finite set of competing systems, based on a "stochastic" objective function and subject to multiple "stochastic" constraints. In this context, we characterize the asymptotically optimal sample allocation that maximizes the rate at which the probability of false selection tends to zero in two scenarios: first in the context of general light-tailed distributions, and second in the specific context in which the objective function and constraints may be observed together as multivariate normal random variates.

In the context of general light-tailed distributions, we present the optimal allocation as the result of a concave maximization problem for which the optimal solution is the result of solving one of two nonlinear systems of equations. The first result of its kind, the optimal allocation is particularly easy to obtain in contexts where the underlying distributions are known or can be assumed, e.g., normal, Bernoulli. A consistent estimator for the optimal allocation and a corresponding sequential algorithm for implementation are provided. Various numerical examples demonstrate where and to what extent the proposed allocation differs from competing algorithms.

In the context of multivariate normal distributions, we present an exact, asymptotically optimal allocation. This allocation is the result of a concave maximization problem in which there are at least as many constraints as there are suboptimal systems. Each constraint corresponding to a suboptimal system is a convex optimization problem. Thus the optimal allocation may easily be obtained in the context of a "small" number of systems, where the quantifier "small" depends on the available computing resources. A consistent estimator for the optimal allocation and a fully sequential algorithm, fit for implementation, are provided. The sequential algorithm performs significantly better than equal allocation in finite time across a variety of randomly generated problems.

The results presented in the general and multivariate normal context provide the first foundation of exact asymptotically optimal sampling methods in the context of "stochastically" constrained simulation optimization on finite sets. Particularly, the general optimal allocation model is likely to be most useful when correlation between the objective and constraint estimators is low, but the data are non-normal. The multivariate normal optimal allocation model is likely to be useful when the multivariate normal assumption is reasonable or the correlation is high.

# Acknowledgments

Professor Raghu Pasupathy's steadfast belief in me, particularly regarding areas in which I had interest but lacked confidence, enabled me to pursue the topics that resulted in this thesis. I am deeply grateful for his encouragement and for his constant willingness to share his knowledge. I admire his infectious curiosity, work ethic, generosity, and remarkable ability to distill and communicate complex ideas. I am fortunate to have had the privilege of working with him.

I am also indebted to my committee, Professors Michael Taaffe, Ebru Bish, Roberto Szechtman, and Don Taylor for their support, both professional and otherwise. During the creation of the chapters regarding multivariate normal distributions, a delightful collaboration arose with Nugroho Pujowidianto and Professors Chun-Hung Chen and Loo Hay Lee. Thank you for the research discussions, support, and friendship.

Further thanks are due to Professor James Thompson, whose measure theory course provided the requisite mathematical sophistication for this document, and Professor Kimberly Ellis, for her excellent and patient guidance in the beginning of my Ph.D. studies.

I have been the beneficiary of various sources of funding over the years, including Pratt, Davenport, and Walt Fellowships, departmental and GSA travel grants, and employment as a graduate teaching assistant and summer instructor. Thank you to these funding sources and those who manage them, including Professors Jaime Camellio, Pat Koelling, and Eileen Van Aken.

My graduate education has been enriched in many ways by the people I have met along the way. I am fortunate that the list of graduate students, faculty, and staff who have supported me locally and in the broader simulation community is exceedingly long. Particular thanks are due to the newly-minted Dr. Evrim Dalkiran, whose friendship and camaraderie I especially cherish.

I am also thankful for friends and family, including Patrick Withem, Eric Hunter, and my parents, Vivian Hunter and Gene Hunter, who have provided valuable tangible and intangible support during my studies.

# Contents

# List of Figures

# List of Tables

# List of Algorithms

# Chapter 1

# Introduction

In many decision-making contexts, the process underlying a physical or abstract system is governed by complex or unknown probabilistic laws and cannot be analyzed analytically. Thus constructing a computer model (simulation) of the system can be a powerful decision-making tool. A simulation model of a system enables the analysis of "what-if" scenarios in a wide variety of applications in which the underlying system cannot be expressed analytically; examples of such applications may include designing a traffic network, designing the layout of a factory, or designing an effective policy to mitigate the spread of an epidemic through a city. Given that one can use a simulation to perform such what-if analyses, it follows that a decision maker may wish to systematically construct what-if scenarios with the explicit goal of identifying "the best" design (system) with respect to a performance measure of interest. For example, from among some competing set of systems, the decision maker may wish to identify *the* traffic network system that minimizes expected travel time, *the* factory system that maximizes expected throughput, or *the* policy system that minimizes the expected disease prevalence in an epidemic.

The problem of identifying the best system with respect to a performance measure of interest by constructing what-if scenarios in a simulation is broadly called the simulation optimization (SO) problem. Since the performance measure of interest can be estimated implicitly through the simulation, the formulation of the SO problem is general and widely applicable, including in situations where the underlying systems are highly complex. In addition to the traffic network, factory, and epidemiological examples given, ready applications of the SO problem can also be found in diverse contexts such as quality control, telecommunication systems, and health care (see Henderson and Pasupathy, 2011, for a collection of contributed SO problems).

One naïve way of solving an SO problem is to fix a sampling budget, select a sample size for a fixed set of candidate systems in advance, simulate the systems, and report the system with the best estimated performance measure as the solution. For example, select five traffic network configurations, simulate each for 100 replications, and select as "best" the configuration producing the lowest estimated expected travel time. However this naïve method is inefficient in two key respects:

(i) "Structural" efficiency: We do not exploit the structure of the performance measure (objective) as a function of system design parameters. Further, if the true set of candidate systems is larger than the set of candidate systems we sample, e.g. the set is infinite, there is no method to ensure that a system "near" the best in objective value will enter our consideration. This question of structural efficiency exists in deterministic optimization and SO settings.

(ii) "Sampling" efficiency: We have made no attempt to efficiently expend the overall sampling budget among the candidate systems we sample. Often simulation runs are expensive, and it is inefficient to expend samples in such a way that clearly inferior systems are not eliminated from consideration as quickly as possible. This question of "sampling" efficiency arises only in SO settings.

These two efficiency issues have been among those considered by researchers interested in the methodological aspects of SO. The types of methods developed depend on the nature of the objective function, e.g., whether it is defined on a space classified as continuous or discrete, finite or infinite, and ordered or unordered, and the presence of secondary performance measures (constraints). For an overview of SO methods and entry points into this literature, see Andradóttir (2006); Spall (2003); Fu (2002); Barton and Meckesheimer (2006), and Ólafsson and Kim (2002).

The focus of this work is on the SO variation in which the set of candidate systems is finite and unordered, that is, no topological structure on the set of candidate systems is assumed. For example, in the case of designing a factory, the set of competing systems might be a finite set of possible machine configurations. Thus the machine configurations have no inherent "ordering." Since no topology is assumed, there is no structure in the objective function to exploit. Each system must receive some portion of the overall sampling budget, and our interest in this SO variation is exclusively on sampling efficiency: for some measure of efficiency, how do we optimally allocate the simulation budget among the systems?

Within the context of SO on unordered finite sets, two prominent SO variations arise: the unconstrained version, or selecting the best system from among a finite set of competing systems, and the "stochastically" constrained version, or selecting the best *feasible* system from among a finite set of competing systems. The unconstrained version of the SO problem on unordered finite sets has arguably seen the most development among the SO variations. Appearing broadly as ranking and selection (see Kim and Nelson, 2006, for an overview), the currently available solution methods are reliable and have stable digital implementations. In contrast, the constrained version of the problem has seen far less development, despite its usefulness in the context of multiple performance measures — for example, in the context of epidemiological policy-making, the decision maker may wish to select, from among a finite set of competing policies, that policy which minimizes the expected disease prevalence, subject to the expected cost of the intervention being less than some monetary budget. The subject of our work is this stochastically constrained version of the SO problem, about which little is currently known.

To explore this constrained SO variation in more detail, consider the following setting. Suppose there exist multiple performance measures defined on a finite set of systems, one of which is primary and called the objective function, while the others are secondary and called the constraint functions. Suppose further that the objective and constraint function values for any given system may be estimated using a stochastic simulation, and that the quality of the objective and constraint function estimators is dependent on the simulation effort expended. Then the constrained SO problem is to identify that system having the best objective function value, from among those systems whose constraint values cross a pre-specified threshold, using only the simulation output. The efficiency of a solution to this problem, which will be defined in rigorous terms later in this document, is measured in terms of the total simulation effort expended.

The broad objective of this work is to rigorously characterize the nature of the most efficient (optimal) sampling plan when solving the constrained SO problem on finite sets. Such characterization is extremely useful in that it facilitates the construction of asymptotically optimal sampling algorithms. The specific questions we ask along the way are twofold.

Q.1 Let an algorithm for solving the constrained SO problem estimate the objective and constraint functions by allocating a portion of an available simulation budget to each competing system. Suppose further that this algorithm returns to the user that system having the best estimated objective function, among the estimated-feasible systems. As the simulation budget increases, the probability that such an algorithm returns any system other than the truly best system decays to zero. Can the asymptotic behavior of this probability of false selection be characterized? Specifically, can its rate of decay be deduced as a function of the sampling proportions allocated to the various systems?

Q.2 Given a satisfactory answer to Q.1, can a method be devised to identify the sampling proportion that maximizes the rate of decay of the probability of false selection?

This work answers both of the above questions in the affirmative. We provide general results for the case in which the underlying performance measures have light-tailed distributions, as well as specific results for the case in which the underlying performance measures have multivariate normal distributions.

In each case, relying on large-deviation principles and generalizing prior work in the unconstrained context by Glynn and Juneja (2004), we fully characterize the probabilistic decay behavior of the false selection event as a function of the budget allocations. This characterization leads to the formulation of a mathematical program whose solution is the allocation that maximizes the rate of probabilistic decay. Since the constructed mathematical program is a concave maximization problem, identifying the asymptotically optimal allocation is “easy” in contexts where the underlying distributional family of the simulation estimator is known or assumed. This asymptotically optimal allocation is then used to create a sequential algorithm guaranteeing that, in the limit, the systems will be sampled in proportion to the optimal allocation – that is, in the limit, the proposed sampling strategy achieves maximum efficiency.

## 1.1 This Work in Context

In the SO literature on finite sets, there exist two broad problem statements related to ensuring the quality of the solution or ensuring efficiency in obtaining a solution. The first problem statement is to identify the best feasible system in finite-time with a pre-specified probabilistic guarantee (the finite-time guarantee problem statement). This problem statement yields finite-time results, but provides no guarantees regarding computational efficiency. The second problem statement in the literature, and the focus of this work, is to find a computing budget allocation that minimizes the probability of falsely selecting any system other than the best feasible system (the efficiency problem statement). This problem statement yields infinite-time (asymptotic) results with efficiency guarantees.

With these two problem statements in mind, prior research on selecting the best system in the unconstrained context falls broadly under one of three categories:

– traditional ranking and selection (R&S) procedures (see, e.g., Kim and Nelson, 2006, for an overview), which typically require a normality assumption and provide finite-time probabilistic guarantees on the probability of false selection or some other loss function.

– the Optimal Computing Budget Allocation (OCBA) framework (see, e.g., Chen et al., 2000), which, under the assumption of normality, provides an approximately optimal sample allocation, and

– the large-deviations (LD) approach (see, e.g., Glynn and Juneja, 2004), which provides an asymptotically optimal sample allocation in the context of general light-tailed distributions.

R&S procedures are developed in response to the finite-time guarantee problem statement, while OCBA and LD approaches are developed in response to the efficiency problem statement.

Corresponding research in the stochastically constrained context is taking an analogous route. For example, as illustrated in table 1.1, recent work by Batur and Kim (2010) and Andradóttir and Kim (2010) provide finite-time guarantees on feasibility determination and on the probability of false selection in the context of stochastically constrained SO, respectively. Both of these papers parallel traditional R&S work on the finite-time guarantee problem statement. Similarly, recent work on the efficiency problem statement by Lee et al. (2011) in the context of stochastically constrained SO parallels the previous OCBA work in the unconstrained context. This work on the efficiency problem statement, which appears in the bottom right-hand cell of table 1.1, is the first to provide a complete generalization of previous large deviations work in ordinal optimization by Glynn and Juneja (2004) and in feasibility determination by Szechtman and Yücesan (2008).

Table 1.1: Research in the area of simulation optimization on finite sets can be categorized by the nature of the result, the required distributional assumption, and the presence of objective function or constraints.

| Result Time | Required Dist'n | Optimization: only objective(s) | Feasibility: only constraint(s) | Constrained Optimization: objective(s) & constraint(s) |
|---|---|---|---|---|
| Finite | Normal | Ranking & Selection (e.g., Kim and Nelson, 2006) | Batur and Kim (2010) | Andradóttir and Kim (2010) |
| Infinite | Normal | OCBA (e.g., Chen et al., 2000) | [application of general solution][1] | OCBA-CO (Lee et al., 2011) |
| | General | Glynn and Juneja (2004) | Szechtman and Yücesan (2008) | Chapters 2 & 3 of this work |

[1] Problems lying in the infinite-time, normal row are also solved as applications of the solutions in the infinite-time, general row.

In addition to characterizing the optimal allocation strategy for stochastically constrained SO on finite sets in the context of general light-tailed distributions, we also use an LD approach to provide an exact characterization of the optimal allocation in the context of multivariate normal distributions. This characterization fully accounts for dependence between the objective and constraint functions, and has not previously been explored in the literature. In OCBA-CO (Lee et al., 2011), Bonferroni bounds dissolve the effect of dependence in the solution, and hence correlation is not explicitly taken into account in the allocation. In chapters 2 and 3 of this document, which provide results in the bottom right-hand cell of table 1.1, the objective and constraints are assumed independent. The ability to account for correlation in the optimal allocation thus remains an important open question. Towards answering this question, chapters 4 and 5 contain a series of results that use the

LD approach to precisely characterize the optimal allocation in the context of multivariate normal distributions. In chapter 4, the characterized optimal allocations are asymptotically exact and expressed explicitly as a function of the correlation between the performance measures. Chapter 5 contains a characterization of the asymptotically exact optimal allocation as the solution to a concave optimization problem.

## 1.2 Problem Statement

The following rigorous problem statement will form the context of the analyses that follow. Suppose there exists a finite set $i = 1, 2, \ldots, r$ of systems, each with an unknown objective value $h_i \in \mathbb{R}$ and unknown constraint values $g_{ij} \in \mathbb{R}$, $j = 1, 2, \ldots, s$ and $i = 1, 2, \ldots, r$. Given constants $\gamma_j \in \mathbb{R}$, we wish to select the system with the lowest objective value $h_i$, subject to the constraints $g_{ij} \leq \gamma_j$, $j = 1, 2, \ldots, s$ and $i = 1, 2, \ldots, r$. That is, consider

$$\begin{aligned} \text{Problem } P: \qquad & \arg\min_{i=1,\ldots,r} \quad h_i \\ & \text{s.t.} \quad g_{ij} \leq \gamma_j, \text{ for all } j = 1, 2, \ldots, s \text{ and } i = 1, 2, \ldots, r; \end{aligned}$$

where $h_i$ and $g_{ij}$ are expectations, estimates of $h_i$ and $g_{ij}$ are observed together through simulation as sample means, and a unique solution to problem $P$ is assumed to exist.

Let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_r)$ be a vector denoting the proportion of the total sampling budget $n$ given to each system, so that $\sum_{i=1}^{r} \alpha_i = 1$ and $\alpha_i \geq 0$ for all $i = 1, 2, \ldots, r$. Furthermore, let the system having the smallest estimated objective value among the estimated-feasible systems be selected as the estimated solution to problem $P$. Then one may ask, what vector of proportions $\boldsymbol{\alpha}$ maximizes the rate of decay of the probability that this procedure returns a suboptimal solution to problem $P$?

## 1.3 Organization

The remainder of this chapter contains a list of the contributions of this work and a description of the notation, conventions, and assumptions that hold throughout the document. In chapters 2 and 3, we consider the case of general light-tailed distributions with one constraint and multiple constraints, respectively. In chapters 4 and 5, we consider the case of multivariate normal distributions with one constraint and multiple constraints, respectively. Chapter 6 contains concluding remarks and a discussion of future research.

## 1.4 Contributions

This work addresses the question of identifying "the best" among a finite set of systems in the presence of multiple "stochastic" performance measures, one of which is used as an objective function and the rest as constraints. This question is a crucial generalization of the work on unconstrained simulation optimization on finite sets by Glynn and Juneja (2004). The following are the specific contributions of this work.

In chapters 2 and 3, we consider the problem statement in section 1.2 in the context of general light-tailed distributions with one constraint and multiple constraints, respectively. These chapters contain the following contributions.

- C.1 We present the first complete characterization of the optimal sampling plan for constrained SO on finite sets when the performance measures can be observed as simulation output. Relying on a large-deviations framework, we derive the probability law for erroneously obtaining a suboptimal solution as a function of the sampling plan. We then demonstrate that the optimal sampling plan can be identified as the solution to a strictly concave maximization problem.
- C.2 We present a consistent estimator and a corresponding algorithm toward estimating the optimal sampling plan. The algorithm is easy to implement in contexts where the underlying distributions governing the performance measures are known or assumed, e.g., the underlying distributions are normal or Bernoulli. In the absence of such distributional knowledge or assumption, the proposed framework inspires an approximate algorithm derived through an approximation of the rate function using Taylor's theorem (Rudin, 1976, p. 110).
- C.3 For the specific context involving performance measures constructed using normal random variables, numerical examples demonstrate where and to what extent the only competitor in the normal context, OCBA-CO (Lee et al., 2011), is suboptimal. There currently appear to be no competitors to the proposed framework for more general contexts.

In chapters 4 and 5, we consider the problem statement in section 1.2 in the context of multivariate normal distributions with one constraint and multiple constraints, respectively. The normal context is particularly relevant since a substantial portion of the corresponding literature in the unconstrained context makes a normality assumption. These chapters contain the following contributions.

- C.4 We present a series of results that precisely characterize the effect of correlation on the optimal allocation in the context in which the underlying performance measures are independent and identically distributed (iid) replications from a bivariate normal distribution (one constraint).

These results provide particular insight into the effect of correlation on the rate of decay of the probability of false selection.

C.5 In the multivariate normal context (multiple constraints), we present the first characterization of the asymptotically exact optimal allocation as the solution to a concave maximization problem, which is itself composed of $r-1$ convex optimization problems. Thus one may solve for the optimal allocation in the context of a "small" number of systems and constraints, where the quantifier "small" depends on available computing resources.

C.6 We present a consistent estimator and fully sequential algorithm that is fit for implementation, and demonstrate its performance on numerous randomly generated problems. The fully sequential algorithm displays significant gains in the rate of decay of the probability of false selection over an equal allocation scheme. These results demonstrate that, although the optimal allocation results are asymptotic in nature, the proposed sequential algorithm performs well in finite time.

## 1.5 Notation and Conventions

In the remainder of the document, let the following notation and conventions hold. Let the feasible system with the lowest objective value be denoted system 1. Further, let us partition the set of $r$ competing systems into the following four mutually exclusive and collectively exhaustive subsets, where $i \leq r$ and $j \leq s$ are shorthand for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, s$.

$$
\begin{aligned}
1 :=& \arg\min_i \{h_i : g_{ij} \leq \gamma_j \text{ for all } j \leq s\} \text{ is the unique best feasible system;}\\
\Gamma :=& \{i : g_{ij} \leq \gamma_j \text{ for all } j \leq s,\ i \neq 1\} \text{ is the set of suboptimal feasible systems;}\\
\mathcal{S}_b :=& \{i : h_i \leq h_1 \text{ and } g_{ij} > \gamma_j \text{ for at least one } j \leq s\} \text{ is the set of infeasible systems}\\
& \quad \text{that have } \textit{better} \text{ (lower) objective values than system 1; and}\\
\mathcal{S}_w :=& \{i : h_i > h_1 \text{ and } g_{ij} > \gamma_j \text{ for at least one } j \leq s\} \text{ is the set of infeasible systems}\\
& \quad \text{that have } \textit{worse} \text{ (higher) objective values than system 1.}
\end{aligned}
$$

The partitioning of the suboptimal systems into the sets $\Gamma$, $\mathcal{S}_b$ and $\mathcal{S}_w$ is strategic and facilitates analyzing the behavior of the false selection probability.

The following notation allows us to distinguish between constraints on which the system is classified as feasible or infeasible.

$$\mathcal{C}_F^i := \{j : g_{ij} \leq \gamma_j\} \text{ is the set of constraints satisfied by system } i\text{; and}$$

$\mathcal{C}_I^i := \{j : g_{ij} > \gamma_j\}$ is the set of constraints not satisfied by system $i$.

Let us interpret the minimum over the empty set as infinity (see, e.g., Dembo and Zeitouni, 1998, p. 5). Likewise, the union over the empty set is an event having probability zero, and the intersection over the empty set is an event having probability one. Also, a sequence of sets $\mathcal{A}_m$ converges to the set $\mathcal{A}$, denoted $\mathcal{A}_m \to \mathcal{A}$, if for large enough $m$ the symmetric difference $(\mathcal{A}_m \cap \mathcal{A}^c) \cup (\mathcal{A} \cap \mathcal{A}_m^c) = \emptyset$.

For two vectors $\boldsymbol{x} = \{x_1, x_2, \ldots, x_k\}$ and $\boldsymbol{y} = \{y_1, y_2, \ldots, y_k\}$, let the notation $\boldsymbol{x} \leq \boldsymbol{y}$ indicate that $x_i \leq y_i$ for all $i = 1, \ldots, k$. This convention holds for row vectors, column vectors, and all operators $=$, $\neq$, $\leq$, $<$, $\geq$, and $>$.

To aid readability, whenever it is reasonable to do so, the following notational conventions hold throughout this document: lower-case letters denote fixed values; upper-case letters denote random variables; upper-case Greek or script letters denote fixed sets; estimated (random) quantities are accompanied by a "hat," e.g., $\hat{H}_1$ estimates the fixed value $h_1$; optimal values have an asterisk, e.g., $x^*$, and vectors are displayed in bold, e.g., $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \ldots, \alpha_r\}$.

## 1.6 Assumptions

The following assumptions hold throughout the document. First, let us assume that the systems are simulated independent of each other.

**Assumption 1.** *The output random variables* $(H_i, \boldsymbol{G_i}) = (H_i, G_{i1}, \ldots, G_{is})$ *are mutually independent for all* $i \leq r$.

Thus correlation across systems, which may result from the use of common random numbers, is not specifically taken into account in the asymptotically optimal allocations presented here.

To ensure that each system is distinguishable from the quantity on which its potential false evaluation as the "best" system depends, and to ensure that the sets of systems maybe correctly estimated with probability one (wp1), the following assumption holds.

**Assumption 2.** *No system has the same objective value as system 1, and no system lies exactly on a constraint, that is,* $h_1 \neq h_i$ *for all* $i \leq r, i \neq 1$, *and* $g_{ij} \neq \gamma_j$ *for all* $i \leq r, j \leq s$.

Assumptions of this type also appear in Glynn and Juneja (2004) and Szechtman and Yücesan (2008).

Further assumptions will be made by chapter.

# Chapter 2

# General Light-Tailed Distributions: One Constraint

Consider the context of problem $P$ (see section 1.2) when there is only one constraint, that is, when $s = 1$. This chapter provides results that build intuition in the one-constraint case before the presentation of the multiple-constraint case in chapter 3. A version of the work in this chapter has been published in the Proceedings of the 2010 Winter Simulation Conference (Hunter and Pasupathy, 2010).

## 2.1 Chapter Organization

This chapter is organized as follows. Section 2.2 contains assumptions for the chapter. Section 2.3 contains a derivation of the rate function of the probability of false selection as a function of the computing budget allocation. Section 2.4 contains a framework for finding the optimal allocation as a conceptual algorithm in which the optimal allocation is the solution to one of two nonlinear systems of equations. Examples of the proposed algorithm are presented in section 2.5 for the case in which the random variables follow normal distributions. Section 2.5 also contains a closed-form solution for the allocation between systems other than system 1 when the assumption that $\alpha_1^* \gg \alpha_i^*$ is made.

## 2.2 Assumptions for Chapter 2

To estimate the unknown quantities $h_i$ and $g_i$, we assume we may obtain replicates of the output random variables $(H_i, G_i)$ from each system. In addition to assumptions 1 and 2, in this chapter, we make the following assumptions.

**Assumption 2.1.** *For any particular system $i$, the random variables $H_i$ and $G_i$ are independent.*

Let $\bar{H}_i(n) = n^{-1}\sum_{k=1}^{n} H_{ik}$ and $\bar{G}_i(n) = n^{-1}\sum_{k=1}^{n} G_{ik}$. We will use $\hat{H}_i \equiv \bar{H}_i(\alpha_i n)$ and $\hat{G}_i \equiv \bar{G}_i(\alpha_i n)$ as shorthand for the estimators of $h_i$ and $g_i$ after scaling the sample size by $\alpha_i$. Let

$$\Lambda_{H_i}^{(n)}(\theta) = \log \mathrm{E}\big[e^{\theta \bar{H}_i(n)}\big] \text{ and } \Lambda_{G_i}^{(n)}(\theta) = \log \mathrm{E}\big[e^{\theta \bar{G}_i(n)}\big]$$

be the cumulant generating functions of $\bar{H}_i(n)$ and $\bar{G}_i(n)$, respectively. Let the effective domain of a function $f(\cdot)$ be denoted by $\mathcal{D}_f = \{x : f(x) < \infty\}$ and its interior by $\mathcal{D}_f^\circ$. Let $f'(x)$ denote the derivative of $f$ with respect to the argument $x$. As is usual in large deviations (LD) contexts, we make the following assumption.

**Assumption 2.2.** *For each system $i \leq r$,*

*(1) the limits*

$$\Lambda_{H_i}(\theta) = \lim_{n\to\infty} \frac{1}{n}\Lambda_{H_i}^{(n)}(n\theta) = \lim_{n\to\infty}\frac{1}{n}\log E\big[e^{\theta\sum_{k=1}^{n} H_{ik}}\big]$$

*and*

$$\Lambda_{G_i}(\theta) = \lim_{n\to\infty} \frac{1}{n}\Lambda_{G_i}^{(n)}(n\theta) = \lim_{n\to\infty}\frac{1}{n}\log E\big[e^{\theta\sum_{k=1}^{n} G_{ik}}\big]$$

*exist as extended real numbers for all $\theta$;*

*(2) the origin belongs to the interior of $\mathcal{D}_{\Lambda_{H_i}}$ and $\mathcal{D}_{\Lambda_{G_i}}$, that is, $0 \in \mathcal{D}^\circ_{\Lambda_{H_i}}$ and $0 \in \mathcal{D}^\circ_{\Lambda_{G_i}}$;*

*(3) $\Lambda_{H_i}(\theta)$ and $\Lambda_{G_i}(\theta)$ are strictly convex and $C^\infty$ on $\mathcal{D}^\circ_{\Lambda_{H_i}}$ and $\mathcal{D}^\circ_{\Lambda_{G_i}}$, respectively;*

*(4) $\Lambda_{H_i}(\theta)$ and $\Lambda_{G_i}(\theta)$ are steep, that is, for any sequence $\{\theta_n\} \in \mathcal{D}_{\Lambda_{H_i}}$ that converges to a boundary point of $\mathcal{D}_{\Lambda_{H_i}}$, $\lim_{n\to\infty} |\Lambda'_{H_i}(\theta_n)| = \infty$, and likewise, for $\{\theta_n\} \in \mathcal{D}_{\Lambda_{G_i}}$ converging to a boundary point of $\mathcal{D}_{\Lambda_{G_i}}$, $\lim_{n\to\infty} |\Lambda'_{G_i}(\theta_n)| = \infty$.*

Assumption 2.2 implies that $\bar{H}_i(n) \to h_i$ wp1 and $\bar{G}_i(n) \to g_i$ wp1 (see Bucklew, 2003, remark 3.2.1). Furthermore, assumption 2.2 ensures that, by the Gärtner-Ellis theorem (Dembo and Zeitouni, 1998, p. 44), $\bar{H}_i(n)$ and $\bar{G}_i(n)$ satisfy the large deviations principle (LDP) (Dembo and Zeitouni, 1998, p. 4) with good rate functions

$$I_i(x) = \sup_{\theta\in\mathbb{R}}\{\theta x - \Lambda_{H_i}(\theta)\} \text{ and } J_i(y) = \sup_{\theta\in\mathbb{R}}\{\theta y - \Lambda_{G_i}(\theta)\}.$$

Assumption 2.2(3) is stronger than what is needed for the Gärtner-Ellis theorem to hold. However, we require $\Lambda_{H_i}(\theta)$ and $\Lambda_{G_i}(\theta)$ to be strictly convex and $C^\infty$ on $\mathcal{D}^\circ_{\Lambda_{H_i}}$ and $\mathcal{D}^\circ_{\Lambda_{G_i}}$, respectively, so that $I_i(x)$ and $J_i(y)$ are strictly convex and $C^\infty$ for $x \in \mathcal{F}^\circ_{H_i} = \text{int}\{\Lambda'_{H_i}(\theta) : \theta \in \mathcal{D}^\circ_{\Lambda_{H_i}}\}$ and $y \in \mathcal{F}^\circ_{G_i} = \text{int}\{\Lambda'_{G_i}(\theta) : \theta \in \mathcal{D}^\circ_{\Lambda_{G_i}}\}$, respectively.

Let $h_\ell = \arg\min_i\{h_i\}$ and let $h_u = \arg\max_i\{h_i\}$. We further assume,

**Assumption 2.3.** *(1) the interval* $[h_\ell, h_u] \subset \cap_{i=1}^r \mathcal{F}^\circ_{H_i}$*, and (2)* $\gamma \in \cap_{i=1}^r \mathcal{F}^\circ_{G_i}$.

Assumption 2.3 assigns nonzero probability to false selection events. As in Glynn and Juneja (2004), assumption 2.3(1) ensures that $\bar{H}_i(n)$ may take any value in the interval $[h_\ell, h_u]$. Consequently, this assumption ensures that we may falsely estimate any system $i$, $i = 2, \ldots, r$, as at least as good in objective function value as system 1. Assumption 2.3(2) ensures there is a nonzero probability that each system will be estimated feasible and a nonzero probability that each system will be estimated infeasible. Specifically, it ensures there is a nonzero probability that system 1 will be estimated infeasible and a nonzero probability that an infeasible system will be estimated feasible.

## 2.3 Rate Function of the Probability of False Selection

The false selection (FS) event is the event that the actual best feasible system, system 1, is not the estimated best feasible system. That is, FS is the event that the optimal system is estimated infeasible, or the optimal system is estimated feasible but another estimated-feasible system has a lower estimated objective value. Let $\bar{\Gamma}$ be the set of estimated-feasible systems excluding system 1, that is, $\bar{\Gamma} = \{i : \hat{G}_i \leq \gamma, i \neq 1\}$. Then the probability of false selection is

$$\begin{aligned} P\{FS\} &= P\{(\hat{G}_1 > \gamma) \cup ((\hat{G}_1 \leq \gamma) \cap (\hat{H}_1 \geq \min_{i \in \bar{\Gamma}} \hat{H}_i))\} \\ &= P\{\hat{G}_1 > \gamma\} + P\{(\hat{G}_1 \leq \gamma) \cap (\hat{H}_1 \geq \min_{i \in \bar{\Gamma}} \hat{H}_i)\} \qquad (2.1) \\ &= P\{FS_1\} + P\{FS_2\}, \end{aligned}$$

where $FS_1$ denotes the event that system 1 is estimated infeasible, and $FS_2$ denotes the event that system 1 is estimated feasible but is "beaten" in objective function value by another estimated-feasible system. We wish to obtain an expression for the rate at which the probability of false selection tends to zero with increasing sampling budget $n$. We first derive expressions for the rate functions of $P\{FS_1\}$ and $P\{FS_2\}$ appearing in equation (2.1), then we combine these results to find the rate function of $P\{FS\}$.

First, consider the rate function for $P\{FS_1\}$, the probability that system 1 is declared infeasible on any of its constraints. Theorem 2.1 establishes the asymptotic behavior of $P\{FS_1\}$ as the rate function corresponding to the constraint on system 1 times the proportion of sample given to system 1.

**Theorem 2.1.** *The rate function for $P\{FS_1\}$ is given by*

$$-\lim_{n\to\infty}\frac{1}{n}\log P\{FS_1\} = \alpha_1 J_1(\gamma).$$

*Proof.* Under assumption 2.2, by the Gärtner-Ellis theorem,

$$-\inf_{y\in(\gamma,\infty)} J_1(y) \leq \liminf_{n\to\infty}\frac{1}{n}\log P(\bar{G}_1(n)>\gamma) \leq \limsup_{n\to\infty}\frac{1}{n}\log P(\bar{G}_1(n)>\gamma) \leq -\inf_{y\in[\gamma,\infty)} J_1(y)$$

Since $J_1(y)$ is strictly convex and $C^\infty$ in $\mathcal{F}^\circ_{G_1}$ and $g_1 < \gamma \in \mathcal{F}^\circ_{G_1}$, it follows that $-\inf_{y\in(\gamma,\infty)} J_1(y) = -\inf_{y\in[\gamma,\infty)} J_1(y) = -J_1(\gamma)$. Therefore the limit exists.

Now let the cumulant generating function of $\hat{G}_1$ be denoted $\Lambda^{(\alpha_1 n)}_{G_1}(\theta)$. By assumption 2.2,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\Lambda^{(\alpha_1 n)}_{G_1}(n\theta) &= \lim_{n\to\infty}\frac{1}{n}\log \mathrm{E}[e^{\theta/\alpha_1\sum_{k=1}^{\alpha_1 n} G_{1k}}] = \lim_{n\to\infty}\frac{\alpha_1}{n}\log \mathrm{E}[e^{\theta/\alpha_1\sum_{k=1}^{n} G_{1k}}] \\ &= \alpha_1\Lambda_{G_1}(\theta/\alpha_1).\end{aligned}$$

Hence $\hat{G}_1$ also satisfies the LDP with good rate function

$$\sup_\theta\{\theta y - \alpha_1\Lambda_{G_1}(\theta/\alpha_1)\} = \alpha_1\sup_{\theta/\alpha_1}\{(\theta/\alpha_1)y - \Lambda_{G_1}(\theta/\alpha_1)\} = \alpha_1 J_1(y).$$

Therefore by similar arguments to those given above, the infimum over the set $(\gamma,\infty)$ is achieved at $\alpha_1 J_1(\gamma)$, and the result follows. □

Therefore the rate at which the probability that system 1 is estimated infeasible goes to zero is a function of the constant $J_1(\gamma)$. Under our assumptions and with logic similar to that given in the proof of theorem 2.1, it can be shown that for any infeasible system $i$, the rate function for the probability that system $i$ is estimated feasible is

$$-\lim_{n\to\infty}\frac{1}{n}\log P\{\hat{G}_i \leq \gamma\} = \alpha_i J_i(\gamma).$$

Now consider the rate function for $P\{FS_2\}$. Under the assumption of independence between the objective function and constraint (assumption 2.1), and since the probability that system 1 is estimated feasible tends to one,

$$\lim_{n\to\infty}\frac{1}{n}\log P\{(\hat{G}_1\leq\gamma)\cap(\hat{H}_1 \geq \min_{i\in\bar{\Gamma}}\hat{H}_i)\} = \lim_{n\to\infty}\frac{1}{n}\log P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}}\hat{H}_i\}. \tag{2.2}$$

As the equality in equation (2.2) always holds, in the remainder of this section, we omit the explicit statement of the event that system 1 is estimated feasible.

The rate function for $P\{FS_2\}$ is the rate function for the probability that system 1 is estimated feasible, but another estimated-feasible system has a better estimated objective value. Since the estimated set of feasible systems $\bar{\Gamma}$ may contain worse feasible systems ($i \in \Gamma$), better infeasible systems ($i \in \mathcal{S}_b$), and worse infeasible systems ($i \in \mathcal{S}_w$), we strategically consider the rate functions for the probability that system 1 is beaten by a system in $\bar{\Gamma} \cap \Gamma$, $\bar{\Gamma} \cap \mathcal{S}_b$, or $\bar{\Gamma} \cap \mathcal{S}_w$ separately. Assuming for now that the required limits exist, lemma 2.2 states that the rate function of $P\{FS_2\}$ is determined by the slowest-converging probability that system 1 will be "beaten" by an estimated-feasible system from $\Gamma, \mathcal{S}_b$, or $\mathcal{S}_w$.

**Lemma 2.2.** *The rate function for $P\{FS_2\}$ is given by the minimum rate function of the probability that system 1 is beaten by an estimated-feasible system that is (i) feasible and worse, (ii) infeasible and better, or (iii) infeasible and worse. That is,*

$$-\lim_{n\to\infty}\frac{1}{n}\log P\{FS_2\} = \min\Big(-\lim_{n\to\infty}\frac{1}{n}\log P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\Gamma}\hat{H}_i\},$$
$$-\lim_{n\to\infty}\frac{1}{n}\log P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\mathcal{S}_b}\hat{H}_i\}, -\lim_{n\to\infty}\frac{1}{n}\log P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\mathcal{S}_w}\hat{H}_i\}\Big). \tag{2.3}$$

*Proof.* From equation (2.2), the probability that system 1 is beaten by another estimated-feasible system can be written as,

$$P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}}\hat{H}_i\} = P\{(\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\Gamma}\hat{H}_i)\cup(\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\mathcal{S}_b}\hat{H}_i)\cup(\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\mathcal{S}_w}\hat{H}_i)\}$$

Therefore

$$\begin{aligned}
&\frac{1}{n}\log\max\Big(P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\Gamma}\hat{H}_i\}, P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\mathcal{S}_b}\hat{H}_i\}, P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\mathcal{S}_w}\hat{H}_i\}\Big)\\
&\leq \frac{1}{n}\log P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}}\hat{H}_i\}\\
&\leq \frac{1}{n}\log\Big(P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\Gamma}\hat{H}_i\} + P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\mathcal{S}_b}\hat{H}_i\} + P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\mathcal{S}_w}\hat{H}_i\}\Big).
\end{aligned}$$

Assuming the relevant limits exist, the conclusion is reached by noting that the limit of the left-hand and right-hand sides are equivalent by proposition A.2 and the principle of the slowest term, proposition A.1, respectively. □

Next, let us individually consider each of the terms on the right hand side of equation (2.3), and establish their respective limits.

First, consider the rate function of the probability that system 1 is "beaten" by a worse estimated-feasible system from $\Gamma$. Since $\bar{\Gamma}$ is equivalent to $\Gamma$ in the limit, and we are considering only the probability that system 1 is beaten by another truly feasible system, we expect that the rate function will be the same as in the unconstrained case presented in Glynn and Juneja (2004). Also, since system 1 can be beaten by any system in $\bar{\Gamma} \cap \Gamma$, we intuitively expect the rate function to be the minimum rate function across all systems in $\Gamma$, corresponding to the system that is "best" at crossing the optimality hurdle. Lemma 2.3 states that this is indeed the case.

**Lemma 2.3.** *The rate function for the probability that system 1 is estimated feasible and has a worse estimated objective value than an estimated-feasible system from $\Gamma$ (feasible and worse) is*

$$-\lim_{n\to\infty} \frac{1}{n} \log P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\Gamma} \hat{H}_i\} = \min_{i\in\Gamma} \left( \inf_x (\alpha_1 I_1(x) + \alpha_i I_i(x)) \right).$$

*Proof.* Let $C \subseteq \Gamma$ denote a set of systems. A lower bound for $P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\Gamma} \hat{H}_i\}$ is given by,

$$\begin{aligned} P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\Gamma} \hat{H}_i\} &= P\{\cup_{i\in\bar{\Gamma}\cap\Gamma}\, \hat{H}_1 \geq \hat{H}_i\} = \sum_C P\{(\cup_{i\in\bar{\Gamma}\cap\Gamma}\, \hat{H}_1 \geq \hat{H}_i) \cap (\bar{\Gamma}\cap\Gamma = C)\} \\ &= \sum_C P\{(\cup_{i\in C} \hat{H}_1 \geq \hat{H}_i) \cap (\bar{\Gamma}\cap\Gamma = C)\} = \sum_C P\{\cup_{i\in C} \hat{H}_1 \geq \hat{H}_i\} P\{\bar{\Gamma}\cap\Gamma = C\} \\ &\geq P\{\cup_{i\in\Gamma} \hat{H}_1 \geq \hat{H}_i\} P\{\bar{\Gamma}\cap\Gamma = \Gamma\} \geq \max_{i\in\Gamma} P\{\hat{H}_1 \geq \hat{H}_i\} P\{\bar{\Gamma}\cap\Gamma = \Gamma\}. \end{aligned}$$

An upper bound is given by

$$P\{\cup_{i\in\bar{\Gamma}\cap\Gamma}\, \hat{H}_1 \geq \hat{H}_i\} \leq P\{\cup_{i\in\Gamma}\, \hat{H}_1 \geq \hat{H}_i\} \leq |\Gamma| \max_{i\in\Gamma} P\{\hat{H}_1 \geq \hat{H}_i\},$$

so that

$$\max_{i\in\Gamma} P\{\hat{H}_1 \geq \hat{H}_i\} P\{\bar{\Gamma}\cap\Gamma = \Gamma\} \leq P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\Gamma} \hat{H}_i\} \leq |\Gamma| \max_{i\in\Gamma} P\{\hat{H}_1 \geq \hat{H}_i\}.$$

By Glynn and Juneja (2004),

$$\lim_{n\to\infty} \frac{1}{n} \log P\{\hat{H}_1 \geq \hat{H}_i\} = -\inf_x \big(\alpha_1 I_1(x) + \alpha_i I_i(x)\big).$$

Therefore by proposition A.2,

$$\lim_{n\to\infty} \frac{1}{n} \log P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\Gamma} \hat{H}_i\} = -\min_{i\in\Gamma} \left( \inf_x (\alpha_1 I_1(x) + \alpha_i I_i(x)) \right). \qquad \square$$

Intuitively, the rate functions $I_1(x)$ and $I_i(x)$ may be added because of the assumed independence between systems (assumption 1). Loosely speaking, in the prior lemma, the rate function of the probability that system 1 is beaten by a worse feasible system is determined by the worse feasible system which is "best" at being falsely estimated as the best feasible system. In this case, the rate at which feasible systems become feasible does not affect the rate function.

Now consider the rate function of the probability that system 1 is beaten by a better infeasible system. Since the only hurdle to a better, infeasible system being declared optimal is feasibility, the rate function is determined by the better infeasible system which is best at being falsely estimated as feasible. Lemma 2.4 states this result rigorously.

**Lemma 2.4.** *The rate function for the probability that system 1 is estimated-feasible and beaten by a better, infeasible, and estimated-feasible system is*

$$-\lim_{n\to\infty}\frac{1}{n}\log P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\mathcal{S}_b}\hat{H}_i\} = \min_{i\in\mathcal{S}_b}\alpha_i J_i(\gamma).$$

*Proof.* Let $C \subseteq \mathcal{S}_b$ denote a set of systems. An upper bound for $P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\mathcal{S}_b}\hat{H}_i\}$ is given by

$$\begin{aligned}
P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\mathcal{S}_b}\hat{H}_i\} &= P\{\cup_{i\in\bar{\Gamma}\cap\mathcal{S}_b}\hat{H}_1 \geq \hat{H}_i\} = \sum_C P\{(\cup_{i\in\bar{\Gamma}\cap\mathcal{S}_b}\hat{H}_1 \geq \hat{H}_i)\cap(\bar{\Gamma}\cap\mathcal{S}_b = C)\}\\
&= \sum_C P\{(\cup_{i\in C}\hat{H}_1 \geq \hat{H}_i)\cap(\bar{\Gamma}\cap\mathcal{S}_b = C)\}\\
&= \sum_C P\{\cup_{i\in C}\hat{H}_1 \geq \hat{H}_i\}P\{\bar{\Gamma}\cap\mathcal{S}_b = C\}\\
&= \sum_C P\{\cup_{i\in C}\hat{H}_1 \geq \hat{H}_i\}\prod_{i\in C}P\{i\in\bar{\Gamma}\}\prod_{i\in\mathcal{S}_b\setminus C}P\{i\notin\bar{\Gamma}\} \qquad (2.4)\\
&\leq \sum_C\prod_{i\in C}P\{i\in\bar{\Gamma}\} \leq \sum_{d=1}^{|\mathcal{S}_b|}\binom{|\mathcal{S}_b|}{d}\max_{i\in\mathcal{S}_b}P\{i\in\bar{\Gamma}\} \leq 2^{|\mathcal{S}_b|}\max_{i\in\mathcal{S}_b}P\{i\in\bar{\Gamma}\}.
\end{aligned}$$

Let $k^* = \arg\max_{i\in\mathcal{S}_b}P\{i\in\bar{\Gamma}\}$. From equation (2.4), it can be seen that $P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\mathcal{S}_b}\hat{H}_i\}$ is bounded below by $P\{\hat{H}_1 \geq \hat{H}_{k^*}\}P\{k^*\in\bar{\Gamma}\}\prod_{i\in\mathcal{S}_b\setminus\{k^*\}}P\{i\notin\bar{\Gamma}\}$. Therefore,

$$P\{\hat{H}_1 \geq \hat{H}_{k^*}\}P\{k^*\in\bar{\Gamma}\}\prod_{i\in\mathcal{S}_b\setminus\{k^*\}}P\{i\notin\bar{\Gamma}\} \leq P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\mathcal{S}_b}\hat{H}_i\} \leq 2^{|\mathcal{S}_b|}P\{k^*\in\bar{\Gamma}\}.$$

Then by proposition A.2,

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}\log P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\mathcal{S}_b}\hat{H}_i\} &= \lim_{n\to\infty}\frac{1}{n}\log P\{k^*\in\bar{\Gamma}\} = \lim_{n\to\infty}\frac{1}{n}\log\max_{i\in\mathcal{S}_b}P\{i\in\bar{\Gamma}\}\\
&= \max_{i\in\mathcal{S}_b}\lim_{n\to\infty}\frac{1}{n}\log P\{i\in\bar{\Gamma}\} = -\min_{i\in\mathcal{S}_b}\alpha_i J_i(\gamma). \qquad \square
\end{aligned}$$

Therefore the rate function is determined by the better infeasible system that is most easily falsely estimated as being in the feasible set.

Finally, we consider the rate function for the probability that system 1 is beaten by a worse infeasible system. Systems which are infeasible and worse must cross both feasibility and optimality hurdles. Therefore the rate function includes both optimality and feasibility terms, where the addition between these terms is a result of the independence between the objective function and constraints (assumption 1).

**Lemma 2.5.** *The rate function for the probability that system 1 is estimated-feasible and beaten by a worse, infeasible, and estimated-feasible system is*

$$-\lim_{n\to\infty}\frac{1}{n}\log P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\mathcal{S}_w}\hat{H}_i\} = \min_{i\in\mathcal{S}_w}\left(\inf_x(\alpha_1 I_1(x)+\alpha_i I_i(x))+\alpha_i J_i(\gamma)\right).$$

*Proof.* Let $C \subseteq \mathcal{S}_w$ denote a set of systems. An upper bound for $P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\mathcal{S}_w}\hat{H}_i\}$ is given by

$$\begin{aligned}
P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\mathcal{S}_w}\hat{H}_i\} &= P\{\cup_{i\in\bar{\Gamma}\cap\mathcal{S}_w}\hat{H}_1 \geq \hat{H}_i\} = \sum_C P\{(\cup_{i\in\bar{\Gamma}\cap\mathcal{S}_w}\hat{H}_1 \geq \hat{H}_i)\cap(\bar{\Gamma}\cap\mathcal{S}_w = C)\} \\
&= \sum_C P\{\cup_{i\in C}\hat{H}_1 \geq \hat{H}_i\}\prod_{i\in C}P\{i\in\bar{\Gamma}\}\prod_{i\in\mathcal{S}_w\setminus C}P\{i\notin\bar{\Gamma}\} \qquad (2.5) \\
&\leq \sum_C |C|\max_{i\in C}(P\{\hat{H}_1 \geq \hat{H}_i\}P\{i\in\bar{\Gamma}\}) \\
&\leq \sum_{d=1}^{|\mathcal{S}_w|}\binom{|\mathcal{S}_w|}{d}|\mathcal{S}_w|\max_{i\in\mathcal{S}_w}(P\{\hat{H}_1 \geq \hat{H}_i\}P\{i\in\bar{\Gamma}\}) \\
&\leq 2^{|\mathcal{S}_w|}|\mathcal{S}_w|\max_{i\in\mathcal{S}_w}(P\{\hat{H}_1 \geq \hat{H}_i\}P\{i\in\bar{\Gamma}\}).
\end{aligned}$$

Let $k^* = \arg\max_{i\in\mathcal{S}_w}\big(P\{\hat{H}_1 \geq \hat{H}_i\}P\{i\in\bar{\Gamma}\}\big)$. Then from equation (2.5), we find that

$$P\{\hat{H}_1 \geq \hat{H}_{k^*}\}P\{k^*\in\bar{\Gamma}\}\prod_{i\in\mathcal{S}_w\setminus\{k^*\}}P\{i\notin\bar{\Gamma}\}$$

is a lower bound for $P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\mathcal{S}_w}\hat{H}_i\}$. Then,

$$\begin{aligned}
P\{\hat{H}_1 \geq \hat{H}_{k^*}\}P\{k^*\in\bar{\Gamma}\}\prod_{i\in\mathcal{S}_w\setminus\{k^*\}}P\{i\notin\bar{\Gamma}\} &\leq P\{\hat{H}_1 \geq \min_{i\in\bar{\Gamma}\cap\mathcal{S}_w}\hat{H}_i\} \\
&\leq 2^{|\mathcal{S}_w|}|\mathcal{S}_w|P\{\hat{H}_1 \geq \hat{H}_{k^*}\}P\{k^*\in\bar{\Gamma}\}.
\end{aligned}$$

Since by proposition A.2 we have,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\log\max_{i\in\mathcal{S}_w}(P\{\hat{H}_1\geq\hat{H}_i\}P\{i\in\bar{\Gamma}\}) &= \max_{i\in\mathcal{S}_w}\lim_{n\to\infty}\frac{1}{n}\log P\{\hat{H}_1\geq\hat{H}_i\}P\{i\in\bar{\Gamma}\}\\ &= \max_{i\in\mathcal{S}_w}\lim_{n\to\infty}\left(\frac{1}{n}\log P\{\hat{H}_1\geq\hat{H}_i\}+\frac{1}{n}\log P\{i\in\bar{\Gamma}\}\right),\end{aligned}$$

then by prior results,

$$-\lim_{n\to\infty}\frac{1}{n}\log P\{\hat{H}_1\geq\min_{i\in\bar{\Gamma}\cap\mathcal{S}_w}\hat{H}_i\} = \min_{i\in\mathcal{S}_w}\Big(\inf_x(\alpha_1 I_1(x)+\alpha_i I_i(x))+\alpha_i J_i(\gamma)\Big). \qquad \square$$

As before, the rate function is determined by the worse infeasible system which is "best" at falsely being declared both feasible and optimal.

Combining lemmas 2.2 through 2.5 and applying the principle of the largest term, we arrive at the following theorem.

**Theorem 2.6.** *The rate function for the probability of false selection, that is, the probability that we return to the user a system other than system 1 is given by*

$$\lim_{n\to\infty}\frac{1}{n}\log P\{FS\} = -\min\Bigg(\overbrace{\alpha_1 J_1(\gamma),}^{\text{1 declared infeasible}}\ \overbrace{\underbrace{\min_{i\in\Gamma}\Big(\inf_x(\alpha_1 I_1(x)+\alpha_i I_i(x))\Big),}_{\text{1 beaten by worse feasible system}}\quad \underbrace{\min_{i\in\mathcal{S}_b}\alpha_i J_i(\gamma),}_{\text{1 beaten by better infeasible system}}\quad \underbrace{\min_{i\in\mathcal{S}_w}\Big(\inf_x(\alpha_1 I_1(x)+\alpha_i I_i(x))+\alpha_i J_i(\gamma)\Big)}_{\text{1 beaten by worse infeasible system}}}^{\text{1 declared feasible}}\Bigg).$$

## 2.4 Optimal Allocation Strategy

From theorem 2.6, an asymptotically optimal allocation strategy will result from maximizing the rate at which $P\{FS\}$ tends to zero with increasing $n$. Thus we wish to allocate the $\alpha_i$'s to solve the following optimization problem:

$$\begin{aligned} \max \quad & \min\Bigg(\alpha_1 J_1(\gamma),\ \min_{i\in\Gamma}\Big(\inf_x(\alpha_1 I_1(x)+\alpha_i I_i(x))\Big),\ \min_{i\in\mathcal{S}_b}\alpha_i J_i(\gamma), \\ & \min_{i\in\mathcal{S}_w}\Big(\inf_x(\alpha_1 I_1(x)+\alpha_i I_i(x))+\alpha_i J_i(\gamma)\Big)\Bigg) \quad s.t. \\ & \sum_{i=1}^{r}\alpha_i = 1, \alpha_i \geq 0. \end{aligned} \tag{2.6}$$

By Glynn and Juneja (2006), $\inf_x(\alpha_1 I_1(x)+\alpha_i I_i(x))$ is a concave, strictly increasing, $C^\infty$ function of $\alpha_1$ and $\alpha_i$. Let $x(\alpha_1,\alpha_i) = \arg\inf_x(\alpha_1 I_1(x)+\alpha_i I_i(x))$. As Glynn and Juneja (2006) demonstrate, for $\alpha_1 > 0$ and $\alpha_i > 0$, $x(\alpha_1,\alpha_i)$ is a $C^\infty$ function of $\alpha_1$ and $\alpha_i$. Likewise, the linear functions $\alpha_1 J_1(\gamma)$ and $\alpha_i J_i(\gamma)$ and the sum $\inf_x(\alpha_1 I_1(x)+\alpha_i I_i(x))+\alpha_i J_i(\gamma)$ are also concave, strictly increasing, $C^\infty$ functions of $\alpha_1$ and $\alpha_i$. Since the minimum of concave, strictly increasing functions is also concave and strictly increasing, the problem in (2.6) is a concave maximization problem. Equivalently, we may rewrite the problem in (2.6) as the following problem $Q$.

$$\begin{aligned} \text{Problem } Q: \quad \max \quad z \quad s.t.& \\ \alpha_1 J_1(\gamma) &\geq z, \\ \alpha_1 I_1(x(\alpha_1,\alpha_i))+\alpha_i I_i(x(\alpha_1,\alpha_i)) &\geq z,\ i\in\Gamma \\ \alpha_i J_i(\gamma) &\geq z,\ i\in\mathcal{S}_b \\ \alpha_1 I_1(x(\alpha_1,\alpha_i))+\alpha_i\big[I_i(x(\alpha_1,\alpha_i))+J_i(\gamma)\big] &\geq z,\ i\in\mathcal{S}_w \\ \sum_{i=1}^{r}\alpha_i &= 1, \alpha_i\geq 0. \end{aligned}$$

Slater's condition (see Boyd and Vandenberghe, 2004, p. 226) holds for problem $Q$, that is, there exists a point $\boldsymbol{\alpha} = (\alpha_1,\alpha_2,\ldots,\alpha_r)$ in the relative interior of the feasible set such that $\boldsymbol{\alpha}$ is "strictly feasible," or so that the inequality constraints hold with strict inequality. For example, a point satisfying this condition is $z = 0, \alpha_i = 1/r, i = 1,\ldots,r$. Since problem $Q$ is concave with differentiable objective function and constraints and Slater's condition holds, strong duality holds and the Karush-Kuhn Tucker conditions are necessary and sufficient for global optimality (see, e.g., Boyd and Vandenberghe, 2004).

Since problem $Q$ is a strictly concave, continuous function of $\boldsymbol{\alpha}$ on a compact set, a unique solution exists. Proposition 2.7 states this result.

**Proposition 2.7.** *There exists a unique solution* $\boldsymbol{\alpha}^* = \{\alpha_1^*,\alpha_2^*,\ldots,\alpha_r^*\}$ *to problem* $Q$ *with optimal value* $z^*$.

Let us define problem $Q^*$ by replacing the inequality constraints corresponding to systems in $\Gamma, \mathcal{S}_b$, and $\mathcal{S}_w$ with equality constraints, and forcing each $\alpha_i$ to be strictly greater than zero.

$$\begin{aligned} \text{Problem } Q^*: \quad \max \quad z \quad s.t.& \\ \alpha_1 J_1(\gamma) &\geq z, \\ \alpha_1 I_1(x(\alpha_1,\alpha_i)) + \alpha_i I_i(x(\alpha_1,\alpha_i)) &= z,\ i \in \Gamma \\ \alpha_i J_i(\gamma) &= z,\ i \in \mathcal{S}_b \\ \alpha_1 I_1(x(\alpha_1,\alpha_i)) + \alpha_i\big[I_i(x(\alpha_1,\alpha_i)) + J_i(\gamma)\big] &= z,\ i \in \mathcal{S}_w \\ \sum_{i=1}^{r} \alpha_i &= 1, \alpha_i > 0, \end{aligned}$$

The following proposition formally states the equivalence of problem $Q$ and problem $Q^*$.

**Proposition 2.8.** *Problems $Q$ and $Q^*$ are equivalent, that is, problem $Q^*$ has the unique solution $\boldsymbol{\alpha}^*$ with optimal value $z^*$.*

*Proof.* For $\alpha_i = 1/r, i = 1, \ldots, r$, it follows that $z > 0$ in problem $Q$. Therefore if $\alpha_i = 0$ for $i \in \{1\} \cup \mathcal{S}_b$, then $z = 0$, which is suboptimal. Now consider $\alpha_i = 0$ for $i \in \Gamma \cup \mathcal{S}_w$. In this case, the constraints for $i \in \Gamma \cap \mathcal{S}_w$ reduce to $\alpha_1 \inf_x I_1(x) = \alpha_1 I_1(h_1) = 0$, and hence $z = 0$. Therefore in problem $Q$, it must be the case that $\alpha_i > 0$ for all $i \leq r$. As a matter of notation, for the remainder of this proof we temporarily append the variable $z$ to the vector $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha} = (z, \alpha_1, \ldots, \alpha_r)$.

For dual variables $\nu$ and $\boldsymbol{\lambda} = (\lambda_i \geq 0 : i \leq r)$, the Lagrangian is,

$$\begin{aligned} L(\boldsymbol{\alpha}, \boldsymbol{\lambda}, \nu) &= z + \lambda_1(\alpha_1 J_1(\gamma) - z) + \sum_{i \in \Gamma} \lambda_i(\alpha_1 I_1(x(\alpha_1,\alpha_i)) + \alpha_i I_i(x(\alpha_1,\alpha_i)) - z) \\ &\quad + \sum_{i \in \mathcal{S}_b} \lambda_i(\alpha_i J_i(\gamma) - z) + \sum_{i \in \mathcal{S}_w} \lambda_i(\alpha_1 I_1(x(\alpha_1,\alpha_i)) + \alpha_i\big[I_i(x(\alpha_1,\alpha_i)) + J_i(\gamma)\big] - z) \\ &\quad + \nu\left(\sum_{i=1}^{r} \alpha_i - 1\right) \\ &= z\left(\sum_{i=1}^{r} \lambda_i - 1\right) + \nu\left(\sum_{i=1}^{r} \alpha_i - 1\right) + \alpha_1\Big(\lambda_1 J_1(\gamma) + \sum_{i \in \Gamma \cup \mathcal{S}_w} \lambda_i I_1(x(\alpha_1,\alpha_i))\Big) \\ &\quad + \sum_{i \in \Gamma} \alpha_i \lambda_i I_i(x(\alpha_1,\alpha_i)) + \sum_{i \in \mathcal{S}_b} \alpha_i(\lambda_i J_i(\gamma)) + \sum_{i \in \mathcal{S}_w} \alpha_i \lambda_i\big[I_i(x(\alpha_1,\alpha_i)) + J_i(\gamma)\big]. \end{aligned}$$

Now we solve for the first-order KKT conditions by taking taking the gradient of the Lagrangian and setting it equal to zero. Since $x(\alpha_1,\alpha_i)$ solves $\alpha_1 I_1'(x) + \alpha_i I_i'(x) = 0$, then $\frac{\partial}{\partial \alpha_1}\big(\alpha_1 I_1(x(\alpha_1,\alpha_i)) + \alpha_i I_i(x(\alpha_1,\alpha_i))\big) = I_1(x(\alpha_1,\alpha_i))$ and $\frac{\partial}{\partial \alpha_i}\big(\alpha_1 I_1(x(\alpha_1,\alpha_i)) + \alpha_i I_i(x(\alpha_1,\alpha_i))\big) = I_i(x(\alpha_1,\alpha_i))$ (see Glynn and Juneja, 2004). Then the stationarity conditions are

$$\sum_{i=1}^{r} \lambda_i = 1 \tag{2.7}$$

$$\lambda_1 J_1(\gamma) + \sum_{i \in \Gamma \cup \mathcal{S}_w} \lambda_i I_1(x(\alpha_1^*, \alpha_i^*)) = \nu \tag{2.8}$$

$$\lambda_i I_i(x(\alpha_1^*, \alpha_i^*)) = \nu, \ i \in \Gamma \tag{2.9}$$

$$\lambda_i J_i(\gamma) = \nu, \ i \in \mathcal{S}_b \tag{2.10}$$

$$\lambda_i [I_i(x(\alpha_1^*, \alpha_i^*)) + J_i(\gamma)] = \nu, \ i \in \mathcal{S}_w \tag{2.11}$$

and the complementary slackness conditions are

$$\lambda_1(\alpha_1^* J_1(\gamma) - z^*) = 0 \tag{2.12}$$

$$\lambda_i(\alpha_1^* I_1(x(\alpha_1^*, \alpha_i^*)) + \alpha_i^* I_i(x(\alpha_1^*, \alpha_i^*)) - z^*) = 0, \ i \in \Gamma \tag{2.13}$$

$$\lambda_i(\alpha_i^* J_i(\gamma) - z^*) = 0, \ i \in \mathcal{S}_b \tag{2.14}$$

$$\lambda_i(\alpha_1^* I_1(x(\alpha_1^*, \alpha_i^*)) + \alpha_i^* \big[I_i(x(\alpha_1^*, \alpha_i^*)) + J_i(\gamma)\big] - z^*) = 0, \ i \in \mathcal{S}_w. \tag{2.15}$$

Equation (2.7) implies that at least one $\lambda_i > 0$. Suppose any $\lambda_i = 0$ for $i \in \Gamma \cup \mathcal{S}_b \cup \mathcal{S}_w$. Since $\alpha_i > 0$ for all $i \leq r$, the rate functions in equations (2.9) – (2.11) are strictly greater than zero, which implies $\nu = 0$, $\lambda_i = 0$ for all $i \in \Gamma \cup \mathcal{S}_b \cup \mathcal{S}_w$, and $\lambda_1 = 1$. Then from equation (2.8), $J_1(\gamma) = 0$. However we have a contradiction since by assumption, $J_1(\gamma) > 0$. Therefore $\lambda_i > 0$ for $i \in \Gamma \cup \mathcal{S}_b \cup \mathcal{S}_w$.

Since $\lambda_i > 0$ in equations (2.13) – (2.15), then the complementary slackness conditions implies that each of these constraints is binding. Therefore the inequality constraints corresponding to $i \in \Gamma \cup \mathcal{S}_b \cup \mathcal{S}_w$ in problem $Q$ can be replaced with equality constraints as in problem $Q^*$. □

The structure of the identical problem $Q^*$ lends intuition to the structure of the optimal allocation, as noted in the following steps: (i) Solve a relaxation of problem $Q^*$ without the feasibility constraint for system 1. Let this problem be called problem $\widetilde{Q}^*$, and let $\tilde{z}^*$ be the optimal value at the optimal solution $\tilde{\boldsymbol{\alpha}}^* = (\tilde{\alpha}_1^*, \ldots, \tilde{\alpha}_r^*)$ to problem $\widetilde{Q}^*$. (ii) Check if the feasibility constraint for system 1 is satisfied by the solution $\tilde{\boldsymbol{\alpha}}^*$. If the feasibility constraint is satisfied, $\tilde{\boldsymbol{\alpha}}^*$ is the optimal solution for problem $Q^*$. Otherwise, (iii) force the feasibility constraint to be binding. The steps (i), (ii), and (iii) are equivalent to solving one of two systems of nonlinear equations, as identified by the KKT conditions of problems $Q^*$ and $\widetilde{Q}^*$. Theorem 2.9 asserts this result formally.

**Theorem 2.9.** *Let the set of suboptimal feasible systems $\Gamma$ be non-empty, and define problem $\widetilde{Q}^*$ as problem $Q^*$, but with the inequality constraint relaxed. Let $(\boldsymbol{\alpha}^*, z^*)$ and $(\tilde{\boldsymbol{\alpha}}^*, \tilde{z}^*)$ denote the* *unique optimal solution and optimal value pairs for problems $Q^*$ and $\widetilde{Q}^*$, respectively. Consider the conditions,*

**C0.** $\sum_{i=1}^{r} \alpha_i = 1, \boldsymbol{\alpha} > 0,$ *and*

$$z = \alpha_1 I_1(x(\alpha_1, \alpha_i)) + \alpha_i I_i(x(\alpha_1, \alpha_i)) = \alpha_j J_j(\gamma) = \alpha_1 I_1(x(\alpha_1, \alpha_k)) + \alpha_k [I_k(x(\alpha_1, \alpha_k)) + J_k(\gamma)]$$
*for all* $i \in \Gamma, j \in \mathcal{S}_b, k \in \mathcal{S}_w,$

**C1.** $\displaystyle\sum_{i\in\Gamma} \frac{I_1(x(\alpha_1, \alpha_i))}{I_i(x(\alpha_1, \alpha_i))} + \sum_{i\in\mathcal{S}_w} \frac{I_1(x(\alpha_1, \alpha_i))}{I_i(x(\alpha_1, \alpha_i)) + J_i(\gamma)} = 1.,$

**C2.** $\alpha_1 J_1(\gamma) = z.$

*Then (i)* $\tilde{\boldsymbol{\alpha}}^*$ *solves C0 and C1 and* $\tilde{\alpha}_1^* J_1(\gamma) \geq \tilde{z}^*$ *if and only if* $\tilde{\boldsymbol{\alpha}}^* = \boldsymbol{\alpha}^*$*; and*

*(ii)* $\boldsymbol{\alpha}^*$ *solves C0 and C2 and* $\tilde{\alpha}_1^* J_1(\gamma) < \tilde{z}^*$ *if and only if* $\boldsymbol{\alpha^*} \neq \tilde{\boldsymbol{\alpha}}^*$.

*Proof.* Due to the structure of problem $Q$, the KKT conditions are necessary and sufficient for global optimality. From prior results, recall that the solutions to problems $Q$, $Q^*$, and $\widetilde{Q}^*$ exist, and that condition C0 holds for the solutions $\boldsymbol{\alpha^*}$ and $\tilde{\boldsymbol{\alpha}}^*$.

We now simplify the KKT equations for problem $Q$. Since $\lambda_i > 0$ for all $i \in \Gamma \cup \mathcal{S}_b \cup \mathcal{S}_w$ in the proof of proposition 2.8, it follows that $\nu > 0$. Dividing (2.8) by $\nu$ and appropriately substituting in values from (2.9)–(2.11),

$$\frac{\lambda_1 J_1(\gamma)}{\nu} + \sum_{i\in\Gamma} \frac{I_1(x(\alpha_1^*, \alpha_i^*))}{I_i(x(\alpha_1^*, \alpha_i^*))} + \sum_{i\in\mathcal{S}_w} \frac{I_1(x(\alpha_1^*, \alpha_i^*))}{I_i(x(\alpha_1^*, \alpha_i^*)) + J_i(\gamma)} = 1. \tag{2.16}$$

By a similar logic to that given in the proof of proposition 2.8 and the simplification provided in equation (2.16), omitting the term with $\lambda_1$ in equation (2.16) yields condition C1 as a KKT condition for problem $\widetilde{Q}^*$. Taken together, C0 and C1 create a fully-specified system of equations that form the KKT conditions for Problem $\widetilde{Q}^*$. A solution $\boldsymbol{\alpha}$ is thus optimal to problem $\widetilde{Q}^*$ if and only if it solves C0 and C1.

Let $\mathcal{D}(Q^*)$ and $\mathcal{D}(\widetilde{Q}^*)$ denote the feasible regions of problems $Q^*$ and $\widetilde{Q}^*$, respectively.

*Proof of Claim (i).* ($\Rightarrow$) Suppose $\tilde{\boldsymbol{\alpha}}^*$ solves C0 and C1, and $\tilde{\alpha}_1^* J_1(\gamma) \geq \tilde{z}^*$. Then $\tilde{\boldsymbol{\alpha}}^* \in \mathcal{D}(Q^*)$. Since the objective functions of problems $Q^*$ and $\widetilde{Q}^*$ are identical, and $\mathcal{D}(Q^*) \subset \mathcal{D}(\widetilde{Q}^*)$, it follows that $z^* \leq \tilde{z}^*$. Therefore $\tilde{\boldsymbol{\alpha}}^* \in \mathcal{D}(Q^*)$ implies $\tilde{\boldsymbol{\alpha}}^*$ is the optimal solution to problem $Q^*$, and by the uniqueness of the optimal solution, $\tilde{\boldsymbol{\alpha}}^* = \boldsymbol{\alpha}^*$.

($\Leftarrow$) Now suppose $\tilde{\boldsymbol{\alpha}}^* = \boldsymbol{\alpha^*}$. Since $\tilde{\boldsymbol{\alpha}}^*$ is the optimal solution to problem $\widetilde{Q}^*$, then $\tilde{\boldsymbol{\alpha}}^*$ solves C0 and C1. Further, since $\boldsymbol{\alpha^*}$ is the optimal solution to problem $Q$, $\boldsymbol{\alpha^*} = \tilde{\boldsymbol{\alpha}}^* \in \mathcal{D}(Q^*)$. Therefore $\tilde{\alpha}_1^* J_1(\gamma) \geq \tilde{z}^*$.

*Proof of Claim (ii).* ($\Rightarrow$) Suppose $\boldsymbol{\alpha^*}$ solves C0 and C2, and $\tilde{\alpha}_1^* J_1(\gamma) < \tilde{z}^*$. Then $\tilde{\boldsymbol{\alpha}}^* \notin \mathcal{D}(Q^*)$, and therefore $\tilde{\boldsymbol{\alpha}}^* \neq \boldsymbol{\alpha^*}$.

($\Leftarrow$) By prior arguments, C0 holds for $\boldsymbol{\alpha^*}$ and $\tilde{\boldsymbol{\alpha}}^*$. Now suppose $\boldsymbol{\alpha^*} \neq \tilde{\boldsymbol{\alpha}}^*$, which implies $\tilde{\boldsymbol{\alpha}}^* \notin \mathcal{D}(Q^*)$. Then it must be the case that $\tilde{\alpha}_1^* J_1(\gamma) < \tilde{z}^*$. Further, since $\tilde{\boldsymbol{\alpha}}^*$ uniquely solves C0 and C1, $\boldsymbol{\alpha^*} \neq \tilde{\boldsymbol{\alpha}}^*$ implies that C1 does not hold for $\boldsymbol{\alpha^*}$. Therefore when solving problem $Q$, it follows that $\lambda_1 > 0$ in equation (2.16). By the complementary slackness condition in equation (2.12), $\alpha_1^* J_1(\gamma) = z^*$, and hence C2 holds for $\boldsymbol{\alpha^*}$. □

Theorem 2.9 implies that, since a solution to problem $Q^*$ always exists, an optimal solution to problem $Q$ can be obtained as the solution to one of the two sets of nonlinear equations C0 and C1 or C0 and C2. We state the procedure implicit in theorem 2.9 as algorithm 2.1.

**Algorithm 2.1** Conceptual algorithm to solve for the optimal allocation $\boldsymbol{\alpha^*}$ in the case general light-tailed distributions and one constraint

```
1: Solve the nonlinear system C0, C1 to obtain α̃* and z̃*.
2: if J_1(γ) ≥ z̃*/α̃_1* then
3:     return α* = α̃*.
4: else
5:     Solve the nonlinear system C0, C2 to obtain α*.
6:     return α*.
7: end if
```

Theorem 2.9 assumes that we have at least one system in $\Gamma$. In the event that $\Gamma$ is empty, conditions C0 and C1 may not form a fully-specified system of equations (e.g., $\Gamma$ and $\mathcal{S}_w$ are empty), or may not have a solution. In such a case, C0 and C2 provide the optimal allocation. When the sets $\mathcal{S}_b$ and $\mathcal{S}_w$ are empty but $\Gamma$ is nonempty, theorem 2.9 reduces to the result presented in Glynn and Juneja (2004).

### 2.4.1 Additional Mathematical Insight into the Structure of the Optimal Allocation Problem (Problem $Q$)

One can a solution for the dual variables from the KKT conditions presented in the proof of proposition 2.8 and the simplification presented in the proof of theorem 2.9 as follows.

Since $\nu > 0$ and $J_1(\gamma) > 0$, from equation (2.16),

$$\sum_{i \in \Gamma} \frac{I_1(x(\alpha_1^*, \alpha_i^*))}{I_i(x(\alpha_1^*, \alpha_i^*))} + \sum_{i \in \mathcal{S}_w} \frac{I_1(x(\alpha_1^*, \alpha_i^*))}{I_i(x(\alpha_1^*, \alpha_i^*)) + J_i(\gamma)} \leq 1.$$

Rearranging terms and substituting $1-\sum_{i=2}^{r}\lambda_i$ for $\lambda_1$,

$$\frac{J_1(\gamma)}{\nu}-J_1(\gamma)\sum_{i=2}^{r}\frac{\lambda_i}{\nu}=1-\sum_{i\in\Gamma}\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*))}-\sum_{i\in\mathcal{S}_w}\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*))+J_i(\gamma)}.$$

Further substitution yields,

$$\begin{aligned}\nu = \Bigg[&\sum_{i\in\Gamma}\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*))}+\sum_{i\in\mathcal{S}_b}\frac{1}{J_i(\gamma)}+\sum_{i\in\mathcal{S}_w}\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*))+J_i(\gamma)}\\ &+\frac{1}{J_1(\gamma)}\left(1-\sum_{i\in\Gamma}\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*))}-\sum_{i\in\mathcal{S}_w}\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*))+J_i(\gamma)}\right)\Bigg]^{-1}.\end{aligned}$$

Then $\lambda_1$ is given by

$$\begin{aligned}\lambda_1 = &\frac{1}{J_1(\gamma)}\left(1-\sum_{i\in\Gamma}\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*))}-\sum_{i\in\mathcal{S}_w}\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*))+J_i(\gamma)}\right)\\ &\times\Bigg[\sum_{i\in\Gamma}\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*))}+\sum_{i\in\mathcal{S}_b}\frac{1}{J_i(\gamma)}+\sum_{i\in\mathcal{S}_w}\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*))+J_i(\gamma)}\\ &+\frac{1}{J_1(\gamma)}\left(1-\sum_{i\in\Gamma}\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*))}-\sum_{i\in\mathcal{S}_w}\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*))+J_i(\gamma)}\right)\Bigg]^{-1}.\end{aligned}$$

The value of $z^*$ may also be found in terms of other variables from condition C0 in theorem 2.9. Dividing through by $\alpha_1^*$ and the appropriate rate functions yields

$$\begin{aligned}\frac{z^*}{\alpha_1^*}\left(\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*))}\right)&=\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*))}+\frac{\alpha_i^*}{\alpha_1^*}\ \text{ for all } i\in\Gamma\\ \frac{z^*}{\alpha_1^*}\left(\frac{1}{J_i(\gamma)}\right)&=\frac{\alpha_i^*}{\alpha_1^*}\ \text{ for all } i\in\mathcal{S}_b\\ \frac{z^*}{\alpha_1^*}\left(\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*))+J_i(\gamma)}\right)&=\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*))+J_i(\gamma)}+\frac{\alpha_i^*}{\alpha_1^*}\ \text{ for all } i\in\mathcal{S}_w.\end{aligned}$$

Summing up the left hand side and the right hand side and noticing

$$\sum_{i\in\Gamma\cup\mathcal{S}_b\cup\mathcal{S}_w}\frac{\alpha_i^*}{\alpha_1^*}=\frac{1}{\alpha_1^*}-1,$$

it follows that

$$\frac{z^*}{\alpha_1^*}\left(\sum_{i\in\Gamma}\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*))}+\sum_{i\in\mathcal{S}_b}\frac{1}{J_i(\gamma)}+\sum_{i\in\mathcal{S}_w}\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*))+J_i(\gamma)}\right)$$
$$=\sum_{i\in\Gamma}\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*))}+\sum_{i\in\mathcal{S}_w}\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*))+J_i(\gamma)}+\frac{1}{\alpha_1^*}-1.$$

Rearranging terms yields

$$z^*=\frac{1-\alpha_1^*\left(1-\sum_{i\in\Gamma}\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*))}-\sum_{i\in\mathcal{S}_w}\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*))+J_i(\gamma)}\right)}{\sum_{i\in\Gamma}\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*))}+\sum_{i\in\mathcal{S}_b}\frac{1}{J_i(\gamma)}+\sum_{i\in\mathcal{S}_w}\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*))+J_i(\gamma)}}.$$

If C0 and C1 hold, then

$$z^*=\left(\sum_{i\in\Gamma}\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*))}+\sum_{i\in\mathcal{S}_b}\frac{1}{J_i(\gamma)}+\sum_{i\in\mathcal{S}_w}\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*))+J_i(\gamma)}\right)^{-1}.$$

If C0 and C2 hold, then $\alpha_1^* J_1(\gamma)=z^*$. Hence,

$$z^*=\left[\left(\sum_{i\in\Gamma}\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*))}+\sum_{i\in\mathcal{S}_b}\frac{1}{J_i(\gamma)}+\sum_{i\in\mathcal{S}_w}\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*))+J_i(\gamma)}\right)\right.$$
$$\left.+\frac{1}{J_1(\gamma)}\left(1-\sum_{i\in\Gamma}\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*))}-\sum_{i\in\mathcal{S}_w}\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*))+J_i(\gamma)}\right)\right]^{-1}.$$

### 2.4.2 An Approximate Closed-Form Solution for Normally Distributed Performance Measures

We now consider an approximation for the case in which the random variables corresponding to both the objective and constraint have normal distributions. The relevant rate functions for the normal case are derived as follows.

Let $H_i \sim N(h_i,\sigma_{h_i}^2)$ and $G_i \sim N(g_i,\sigma_{g_i}^2)$ for all $i \le r$. Then

$$I_i(x)=\frac{(x-h_i)^2}{2\sigma_{h_i}^2}\quad\text{and}\quad J_i(\gamma)=\frac{(\gamma-g_i)^2}{2\sigma_{g_i}^2}.$$

Differentiating to find the value of $x$ at which $\inf_x(\alpha_1 I_1(x) + \alpha_i I_i(x))$ is achieved yields

$$x(\alpha_1, \alpha_i) = \frac{(\sigma_{h_i}^2/\alpha_i)h_1}{(\sigma_{h_i}^2/\alpha_i) + (\sigma_{h_1}^2/\alpha_1)} + \frac{(\sigma_{h_1}^2/\alpha_1)h_i}{(\sigma_{h_1}^2/\alpha_1) + (\sigma_{h_i}^2/\alpha_i)}.$$

Then the rate functions are,

$$\begin{aligned}
\alpha_i J_i(\gamma) &= \frac{\alpha_i(\gamma - g_i)^2}{2\sigma_{g_i}^2},\ i \in \{1\} \cup \mathcal{S}_b, \\
\alpha_1 I_1(x(\alpha_1, \alpha_i)) + \alpha_i I_i(x(\alpha_1, \alpha_i)) &= \frac{(h_1 - h_i)^2}{2(\sigma_{h_1}^2/\alpha_1 + \sigma_{h_i}^2/\alpha_i)},\ i \in \Gamma \\
\alpha_1 I_1(x(\alpha_1, \alpha_i)) + \alpha_i[I_i(x(\alpha_1, \alpha_i)) + J_i(\gamma)] &= \frac{(h_1 - h_i)^2}{2(\sigma_{h_1}^2/\alpha_1 + \sigma_{h_i}^2/\alpha_i)} + \frac{\alpha_i(\gamma - g_i)^2}{2\sigma_{g_i}^2},\ i \in \mathcal{S}_w.
\end{aligned}$$

Taking partial derivatives with respect to $\alpha_i$ yields

$$\frac{\partial}{\partial \alpha_1}\left[\alpha_1 I_1(x(\alpha_1, \alpha_i)) + \alpha_i I_i(x(\alpha_1, \alpha_i))\right] = I_1(x(\alpha_1, \alpha_i)) = \frac{\alpha_i^2 \sigma_{h_1}^2 (h_1 - h_i)^2}{2(\alpha_i \sigma_{h_1}^2 + \alpha_1 \sigma_{h_i}^2)^2}, \tag{2.17}$$

$$\frac{\partial}{\partial \alpha_i}\left[\alpha_1 I_1(x(\alpha_1, \alpha_i)) + \alpha_i I_i(x(\alpha_1, \alpha_i))\right] = I_i(x(\alpha_1, \alpha_i)) = \frac{\alpha_1^2 \sigma_{h_i}^2 (h_1 - h_i)^2}{2(\alpha_i \sigma_{h_1}^2 + \alpha_1 \sigma_{h_i}^2)^2}. \tag{2.18}$$

The general solution for the optimal allocation in the normal case does not have a closed-form expression.

In the unconstrained normal case, Chen et al. (2000) and Glynn and Juneja (2004) simplify the optimal allocation by allowing $\alpha_1^* \gg \alpha_i^*$. However it is not clear what this approximation means in terms of problem parameters. For example, to take the limit as $\alpha_1^* \to 1$ and maintain the KKT conditions for optimality, problem parameters such as $\sigma_{h_1}$ must change to maintain the optimality and feasibility of the new allocation. We circumvent this problem by removing the optimal allocation requirement between system 1 and all other systems, that is, fixing $\alpha_1$ and optimizing with respect to $\alpha_i, i \in \{2, \ldots, r\}$. From condition C0 in theorem 2.9, the following corollary follows, regarding the optimal allocation amongst all systems other than system 1.

**Corollary 2.10.** *Let $\boldsymbol{\alpha}'^* = (\alpha_1, \alpha_2^*, \ldots, \alpha_r^*)$. If $\boldsymbol{\alpha}'^* > 0, \sum_{i=2}^r \alpha_i^* = 1 - \alpha_1$ minimizes the asymptotic probability of false selection, then*

$$\begin{aligned}
&\alpha_i^* J_i(\gamma) = \alpha_1 I_1(x(\alpha_1, \alpha_j^*)) + \alpha_j^* I_j(x(\alpha_1, \alpha_j^*)) = \alpha_1 I_1(x(\alpha_1, \alpha_k^*)) + \alpha_k^*[I_k(x(\alpha_1, \alpha_k^*)) + J_k(\gamma)], \\
&\qquad \textit{for all } i \in \mathcal{S}_b, j \in \Gamma,\ k \in \mathcal{S}_w.
\end{aligned} \tag{2.19}$$

To find the allocation for $\alpha_1 \gg \alpha_i^*$ for a fixed $\alpha_1$, we fall within the purview of corollary 2.10.

Using equation (2.19),

$$\frac{\alpha_i^*}{\alpha_j^*} = \frac{\alpha_1 I_1(x(\alpha_1,\alpha_j^*))\mathbb{I}_{j\in\Gamma\cup\mathcal{S}_w} - \alpha_1 I_1(x(\alpha_1,\alpha_i^*))\mathbb{I}_{i\in\Gamma\cup\mathcal{S}_w} + \alpha_j^* I_j(x(\alpha_1,\alpha_j^*))\mathbb{I}_{j\in\Gamma\cup\mathcal{S}_w} + \alpha_j^* J_j(\gamma)\mathbb{I}_{j\in\mathcal{S}_b\cup\mathcal{S}_w}}{\alpha_j^* I_i(x(\alpha_1,\alpha_i^*))\mathbb{I}_{i\in\Gamma\cup\mathcal{S}_w} + \alpha_j^* J_i(\gamma)\mathbb{I}_{i\in\mathcal{S}_b\cup\mathcal{S}_w}}$$

$$= \frac{\dfrac{\alpha_1\alpha_j^*\sigma_{h_1}^2(h_1-h_j)^2}{(\alpha_j^*\sigma_{h_1}^2+\alpha_1\sigma_{h_j}^2)^2}\mathbb{I}_{j\in\Gamma\cup\mathcal{S}_w} + \dfrac{\alpha_1^2\sigma_{h_j}^2(h_1-h_j)^2}{(\alpha_j^*\sigma_{h_1}^2+\alpha_1\sigma_{h_j}^2)^2}\mathbb{I}_{j\in\Gamma\cup\mathcal{S}_w} + \dfrac{(\gamma-g_j)^2}{\sigma_{g_j}^2}\mathbb{I}_{j\in\mathcal{S}_b\cup\mathcal{S}_w}}{\dfrac{\alpha_1\alpha_i^*\sigma_{h_1}^2(h_1-h_i)^2}{(\alpha_i^*\sigma_{h_1}^2+\alpha_1\sigma_{h_i}^2)^2}\mathbb{I}_{i\in\Gamma\cup\mathcal{S}_w} + \dfrac{\alpha_1^2\sigma_{h_i}^2(h_1-h_i)^2}{(\alpha_i^*\sigma_{h_1}^2+\alpha_1\sigma_{h_i}^2)^2}\mathbb{I}_{i\in\Gamma\cup\mathcal{S}_w} + \dfrac{(\gamma-g_i)^2}{\sigma_{g_i}^2}\mathbb{I}_{i\in\mathcal{S}_b\cup\mathcal{S}_w}}. \tag{2.20}$$

Now let $\alpha_1 \to 1$ in equation (2.20). Since $1-\alpha_1 = \sum_{i=2}^{r}\alpha_i^*$, then $\alpha_i^* \to 0$ for all $i \in \{2\ldots,r\}$. Then

$$\frac{\alpha_i^*}{\alpha_j^*} \approx \frac{\left(\dfrac{h_1-h_j}{\sigma_{h_j}}\right)^2 \mathbb{I}_{j\in\Gamma\cup\mathcal{S}_w} + \left(\dfrac{\gamma-g_j}{\sigma_{g_j}}\right)^2 \mathbb{I}_{j\in\mathcal{S}_b\cup\mathcal{S}_w}}{\left(\dfrac{h_1-h_i}{\sigma_{h_i}}\right)^2 \mathbb{I}_{i\in\Gamma\cup\mathcal{S}_w} + \left(\dfrac{\gamma-g_i}{\sigma_{g_i}}\right)^2 \mathbb{I}_{i\in\mathcal{S}_b\cup\mathcal{S}_w}}.$$

where equality holds for all $i,j \in \mathcal{S}_b$. When considering only feasible systems, these results reduce to those presented by Glynn and Juneja (2004).

## 2.5 Normal Examples

To illustrate the proposed algorithm, we consider several examples where the underlying performance measures have normal distributions. (See section 2.4.2 for the rate functions in the normal case.) Specifically, we present cases in which $\alpha_1 J_1(\gamma)$ is nonbinding (only C0 and C1 hold), binding with $\lambda_1 > 0$ (only C0 and C2 hold), and binding with $\lambda_1 = 0$ (C0, C1, and C2 hold). To demonstrate the effect of $g_1$ on the allocation to system 1, we vary $g_1$ while holding all other parameters of the examples fixed.

Let there be $r = 3$ systems, each with normally distributed objective function and constraint such that $\gamma = 0$ and with means and variances given in table 2.1.

Table 2.1: Means and variances for normal examples illustrating the proposed conceptual algorithm in the case of general light-tailed distributions and one constraint

| System ($i$) | $h_i$ | $\sigma_{h_i}^2$ | $g_i$ | $\sigma_{g_i}^2$ |
|---|---|---|---|---|
| 1 | 0 | 1.0 | $g_1 < 0$ varies | 1.0 |
| 2 | 2.0 | 1.0 | $-1.0$ | 1.0 |
| 3 | 2.0 | 1.0 | $-2.0$ | 1.0 |

Thus in the examples that follow, $\Gamma = \{1,2,3\}$, $\mathcal{S}_b = \emptyset$, and $\mathcal{S}_w = \emptyset$. Since the basis for the proposed allocation to systems in $\Gamma$ regards their "scaled distance" from system 1, and systems 2 and 3 are equal in this respect, one can intuitively expect that they will receive equal allocation.

**Example 2.11** ($\alpha_1 J_1(\gamma)$ nonbinding)**.** In this example, let $g_1 = -2$. Then problem $\widetilde{Q}^*$ is

$$\begin{aligned} \max \quad z \quad & s.t. \\ \frac{(h_1-h_2)^2}{2(\sigma_{h_1}^2/\alpha_1 + \sigma_{h_2}^2/\alpha_2)} &= z \\ \frac{(h_1-h_3)^2}{2(\sigma_{h_1}^2/\alpha_1 + \sigma_{h_3}^2/\alpha_3)} &= z \\ \alpha_1 + \alpha_2 + \alpha_3 &= 1 \\ \alpha_i &> 0, \end{aligned}$$

which is equivalent to,

$$\begin{aligned} \max \quad z \quad & s.t. \\ \frac{2}{1/\alpha_1 + 1/\alpha_2} &= z \\ \alpha_2 - \alpha_3 &= 0 \\ \alpha_1 + \alpha_2 + \alpha_3 &= 1 \\ \alpha_i &> 0. \end{aligned}$$

Solving for $\alpha_1$ yields the equation

$$\frac{2\alpha_1(1-\alpha_1)}{\alpha_1+1} = z. \tag{2.21}$$

Taking the derivative and setting it equal to zero yields the quadratic equation

$$\alpha_1^2 + 2\alpha_1 - 1 = 0,$$

with solution

$$\tilde{\alpha}_1^* = \sqrt{2} - 1 \approx 0.4142, \qquad \tilde{\alpha}_2^* = 1 - \frac{1}{\sqrt{2}} \approx 0.2929, \qquad \tilde{\alpha}_3^* = 1 - \frac{1}{\sqrt{2}} \approx 0.2929. \tag{2.22}$$

Then $\tilde{z}^* = 6 - 4\sqrt{2} \approx 0.3431$.

Now let us check the value of $\tilde{\alpha}_1^* J_1(\gamma)$. We have $\tilde{\alpha}_1^* J_1(\gamma) = (\sqrt{2}-1)(0+2)^2/2 \approx 0.8284 > \tilde{z}^*$. Therefore $\tilde{\boldsymbol{\alpha}}^*$ solves Problem $Q^*$, and the allocation given in (2.22) is optimal. □

**Example 2.12** ($\alpha_1 J_1(\gamma)$ binding, $\lambda_1 > 0$)**.** Now let $g_1 = -1$. Thus this example is identical to the previous example, except that the mean of the distribution on $G_1$ has changed. Therefore the solution to problem $\widetilde{Q}^*$, $\tilde{\boldsymbol{\alpha}}^*$, is given by equation (2.22). However when upon checking the value of $\tilde{\alpha}_1^* J_1(\gamma)$, we find

$$\tilde{\alpha}_1^* J_1(\gamma) = (\sqrt{2}-1)\frac{(0+1)^2}{2(1)} \approx 0.2071 < \tilde{z}^*.$$

Therefore $\tilde{\boldsymbol{\alpha}}^*$ does not solve $Q^*$, and the allocation to system 1 must increase due to the slow rate function for $P\{FS_1\}$. Now let the constraint $\alpha_1 J_1(\gamma)$ be binding such that $\alpha_1 J_1(\gamma) = z$. Then from equation (2.21), it follows that

$$\frac{2\alpha_1(1-\alpha_1)}{\alpha_1+1} = \frac{\alpha_1}{2} = z.$$

Solving for $\alpha_1$ in the above equation results in $\alpha_1 = 0.6$. Since the problem is fully specified, the optimal allocation is,

$$\alpha_1^* = 0.6, \qquad \alpha_2^* = 0.2, \qquad \alpha_3^* = 0.2,$$

and the overall rate achieved is $z^* = 0.3$. □

**Example 2.13** ($\alpha_1 J_1(\gamma)$ binding, $\lambda_1 = 0$)**.** Now let

$$g_1 = -\sqrt{\frac{2(6-4\sqrt{2})}{\sqrt{2}-1}} \approx -1.2872.$$

This example is identical to the previous examples, except that once again the mean of the distribution on $G_1$ has changed. Solving the relaxed problem $\tilde{Q}^*$, we find $\tilde{\boldsymbol{\alpha}}^*$ is given by equation (2.22). Now checking $\tilde{\alpha}_1^* J_1(\gamma)$ results in $\tilde{\alpha}_1^* J_1(\gamma) = 6 - 4\sqrt{2} = z^*$. For the allocation $\tilde{\boldsymbol{\alpha}}^*$,

$$\sum_{i\in\Gamma} \frac{I_1(x(\alpha_1,\alpha_i))}{I_i(x(\alpha_1,\alpha_i))} = \left(\frac{\alpha_2^*}{\alpha_1^*}\right)^2 + \left(\frac{\alpha_3^*}{\alpha_1^*}\right)^2 = 1,$$

and hence $\lambda_1 = 0$. Thus this example was constructed so that the value of $\alpha_1 J_1(\gamma)$ sits exactly at the optimal allocation that results from solving problem $\tilde{Q}^*$. □

The values of the optimal allocations and the achieved rate of decay of $P\{FS\}$ can be plotted as a function of the value $g_1$. Solving for the optimal allocation as a function of $g_1$ yields the allocations displayed in figure 2.1 and the rate $z^*$ displayed in figure 2.2.

Figure 2.1 shows that as $g_1$ becomes farther from $\gamma = 0$, system 1 requires a smaller portion of the sample to determine its feasibility. For values of $g_1$ smaller than $g_1 = -1.2872$, the feasibility of system 1 is not binding in this example. Therefore the optimal allocation as a function of $g_1$ does

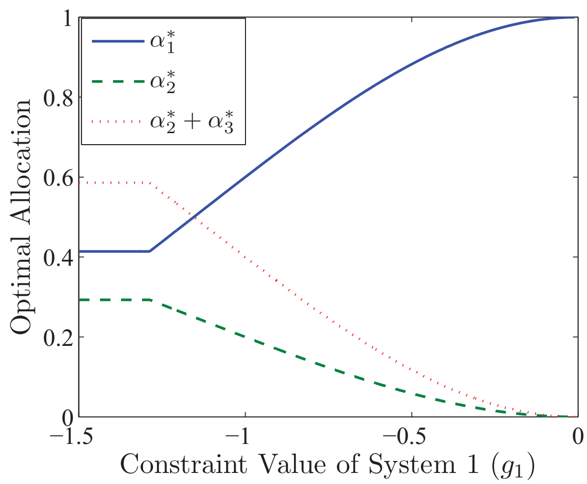


Figure 2.1: The graph of the constraint value of system 1 ($g_1$) versus the optimal allocation for the systems in table 2.1 shows that when the system 1 is sufficiently "close" to the constraint, the allocation to system 1 increases. When system 1 is sufficiently "far" from the constraint, the allocation to systems as a function of the position of system 1 does not change.

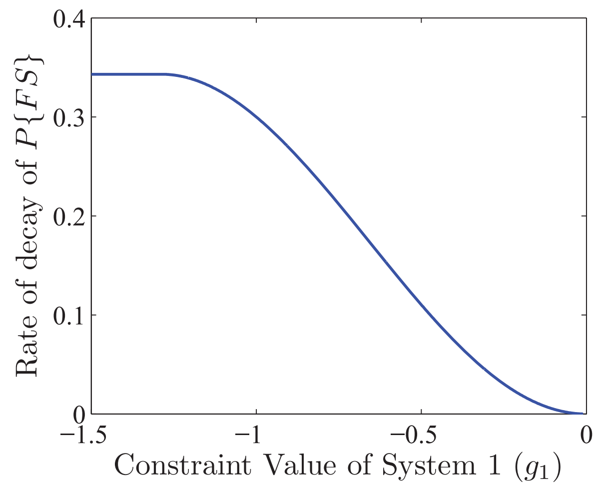


Figure 2.2: The graph of the constraint value of system 1 ($g_1$) versus the rate of decay of of the probability of false selection ($P\{FS\}$) for the systems in table 2.1 shows that the rate of decay of the $P\{FS\}$ decreases as system 1 becomes closer to the constraint.

not change for $g_1 < -1.2872$. Likewise, in figure 2.2, the rate of decay of $P\{FS\}$ remains constant at $z^* = 0.3431$ for $g_1 < -1.2872$. For $g_1 > -1.2873$, the rate decreases to zero as a function of increasing $g_1$.

# Chapter 3

# General Light-Tailed Distributions: Multiple Constraints

Consider the context of problem $P$ fully, with multiple constraints (see section 1.2). In this chapter, we define a relaxation of the assumption of independence between the objective function and constraint random variables. Before presenting results analogous to those presented in chapter 2 for the multiple-constraint case, we discuss the structure of permissible dependence. Since many of the proofs in this chapter are similar to those presented in chapter 2, full proofs appear in appendix B. A version of the work in this chapter is presented in the journal article by Hunter and Pasupathy (2011).

## 3.1 Chapter Organization

Section 3.2 contains an exploration of the structure of permissible dependence for the results presented in this chapter. Assumptions for the chapter are described in section 3.3. Section 3.4 contains an expression for the rate function of the probability of false selection as a function of the computing budget allocation. Section 3.5 contains a general sampling framework and a conceptual algorithm to solve for the optimal allocation. A consistent estimator and an implementable sequential algorithm for the optimal allocation is provided in section 3.6. Section 3.7 contains numerical illustrations for the normal case and a comparison with OCBA-CO (Lee et al., 2011).

## 3.2 Structure of Permissible Dependence

In this chapter, we replace the assumption regarding independence between the objective function and constraint for each system with an assumption allowing certain types of dependence that "wash out" asymptotically. For independent events $A_n$ and $B_n$, when the relevant limits exist,

$$\lim_{n\to\infty} \frac{1}{n} \log P\{A_n \cap B_n\} = \lim_{n\to\infty} \frac{1}{n} \log P\{A_n\} + \lim_{n\to\infty} \frac{1}{n} \log P\{B_n\}.$$

That is, when computing the rate function of the intersection of two independent events, the rate functions of each of the individual events simply add up. Intuitively, because $A_n$ and $B_n$ are independent, $P\{A_n \cap B_n\} = P\{A_n\}P\{B_n\}$, and hence the summation results from taking the log of both sides. While independence is a sufficient condition for the rate function of the intersection to be the sum of the rate functions of each event, independence is not a necessary condition.

The events considered in large deviations contexts tend to be dichotomously categorized as having probabilities that tend to zero or one. For example, the event that system 1 is deemed feasible tends to one, and the event that system 1 is deemed infeasible tends to zero. Therefore in the following discussion of permissible dependence, we only consider events $A_n$ and $B_n$ such that $P\{A_n\} \to 0$, either $P\{B_n\} \to 0$ or $P\{B_n\} \to 1$, and $P\{A_n \cap B_n\} \to 0$.

Let us define

$$c_n = \frac{P\{A_n \cap B_n\}}{P\{A_n\}P\{B_n\}}.$$

Then by definition $P\{A_n \cap B_n\} = c_n P\{A_n\}P\{B_n\}$, and it follows that

$$\frac{1}{n} \log P\{A_n \cap B_n\} = \frac{1}{n} \log c_n + \frac{1}{n} \log P\{A_n\} + \frac{1}{n} \log P\{B_n\}.$$

Now suppose all relevant limits exist. Then

$$\lim_{n\to\infty} \frac{1}{n} \log P\{A_n \cap B_n\} = \lim_{n\to\infty} \frac{1}{n} \log c_n + \lim_{n\to\infty} \frac{1}{n} \log P\{A_n\} + \lim_{n\to\infty} \frac{1}{n} \log P\{B_n\}.$$

When $P\{B_n\} \to 1$, the above equation simplifies to

$$\lim_{n\to\infty} \frac{1}{n} \log P\{A_n \cap B_n\} = \lim_{n\to\infty} \frac{1}{n} \log c_n + \lim_{n\to\infty} \frac{1}{n} \log P\{A_n\}.$$

From this argument, the conditions for asymptotic independence become clear: we require

$$\lim_{n\to\infty} \frac{1}{n} \log c_n = 0.$$

The following definition states this requirement formally.

**Definition 3.1.** Let $A_n$ and $B_n$ be events with $P\{A_n\} \to 0$ and either $P\{B_n\} \to 0$ or $P\{B_n\} \to 1$.

Then $A_n$ and $B_n$ exhibit *limiting independence* if

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{P\{A_n \cap B_n\}}{P\{A_n\}P\{B_n\}} = 0.$$

Equivalently, we require

$$\lim_{n\to\infty} \left( \frac{P\{A_n \cap B_n\}}{P\{A_n\}P\{B_n\}} \right)^{1/n} = 1.$$

This definition does not allow certain types of extreme dependence, such as the case where events $A_n$ and $B_n$ are disjoint and hence $P\{A_n \cap B_n\} = 0$, or the case where $P\{B_n\} \to 0$, $B_n \subseteq A_n$, and hence $P\{A_n \cap B_n\} = P\{B_n\}$. In addition, the following example shows that when $P\{A_n\} \to 0$ and $P\{B_n\} \to 1$, limiting independence does not hold if $B_n$ acts in such a way that it increases the rate at which $P\{A_n \cap B_n\}$ tends to zero.

**Example 3.2.** Let $A_n$ and $B_n$ be events such that $P\{A_n\} = e^{-2n} + e^{-3n}$, $P\{B_n\} = 1 - e^{-2n}$, and $P\{A_n \cap B_n\} = P\{B_n\} - P\{A_n^c\} = e^{-3n}$. One can imagine these measures on a $[0, 1]$ number line where $B_n = [0, 1 - e^{-2n}]$, $A_n = [1 - e^{-2n} - e^{-3n}, 1]$, and $A_n \cap B_n = [1 - e^{-2n} - e^{-3n}, 1 - e^{-2n}]$. Then $\lim_{n\to\infty} \frac{1}{n} \log P\{A_n \cap B_n\} = -3$, and by the principle of the largest term, $\lim_{n\to\infty} \frac{1}{n} \log P\{A_n\} = \lim_{n\to\infty} \frac{1}{n} \log \left(e^{-2n} + e^{-3n}\right) = -2$. Therefore $P\{A_n \cap B_n\} \to 0$ faster than $P\{A_n\} \to 0$. □

Among the types of allowable dependence are those where $c_n$ is bounded away from 0 and $\infty$, and where $c_n$ tends to $\infty$ slowly enough that the overall term $\lim_{n\to\infty} \frac{1}{n} \log c_n$ still tends to zero. The following example explores one type of permissible dependence for bivariate normal distributions.

**Example 3.3.** For $i = 1, \ldots, n$, let

$$\begin{bmatrix} X_{1i} \\ X_{2i} \end{bmatrix} \overset{iid}{\sim} BN\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right),$$

where BN indicates the bivariate normal distribution, correlation $\rho = E[(X_1 - \mu_1)(X_2 - \mu_2)]/\sigma_1\sigma_2$, and $|\rho| < 1$. Then for $\bar{X}_1 = \frac{1}{n}\sum_{i=1}^{n} X_{1i}$ and $\bar{X}_2 = \frac{1}{n}\sum_{i=1}^{n} X_{2i}$,

$$\begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \end{bmatrix} \sim BN\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \frac{1}{n} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right).$$

Consider the events $P\{\bar{X}_1 > a\}$ and $P\{\bar{X}_2 > b\}$ for some $a > 0$ and $b < 0$, where $a, b < \infty$, and $\rho \geq 0$. Then

$$P\{\bar{X}_1 > a, \bar{X}_2 > b\} \geq P\{\bar{X}_1 > a\}P\{\bar{X}_2 > b\}, \tag{3.1}$$

and hence $c_n \geq 1$. If $b < 0$, then $P\{\bar{X}_2 > b\} \to 1$, and since

$$c_n = \underbrace{\frac{P\{\bar{X}_1 > a, \bar{X}_2 > b\}}{P\{\bar{X}_1 > a\}}}_{\leq 1} \underbrace{\frac{1}{P\{\bar{X}_2 > b\}}}_{\to 1}, \tag{3.2}$$

then $\limsup c_n \leq 1$. Since $c_n \geq 1$ and $\limsup c_n \leq 1$, it follows that $\lim_{n\to\infty} c_n = 1$ and hence $\lim_{n\to\infty} \frac{1}{n} \log c_n = 0$. □

## 3.3 Assumptions for Chapter 3

To estimate the unknown quantities $h_i$ and $g_{ij}$, we assume we may obtain replicates of the output random variables $(H_i, G_{i1}, \ldots, G_{is})$ from each system. In addition to assumptions 1 and 2, let the following assumptions hold in this chapter.

We assume limiting independence across events regarding the classification of systems on optimality and feasibility.

**Assumption 3.1.** *For any particular system $i$, all events involving the output random variables $H_i, G_{i1}, \ldots, G_{is}$ exhibit mutual limiting independence.*

The assumption that for any particular system $i$, the output random variables $H_i, G_{i1}, \ldots, G_{is}$ are mutually independent is a more strict version of Assumption 3.1.

Let us define $\bar{H}_i(n) = n^{-1} \sum_{k=1}^{n} H_{ik}$ and $\bar{G}_{ij}(n) = n^{-1} \sum_{k=1}^{n} G_{ijk}$. We will use $\hat{H}_i \equiv \bar{H}_i(\alpha_i n)$ and $\hat{G}_{ij} \equiv \bar{G}_{ij}(\alpha_i n)$ as shorthand for the estimators of $h_i$ and $g_{ij}$ after scaling the sample size by $\alpha_i > 0$, the proportion of the total sample $n$ which is allocated to system $i$.

Let

$$\Lambda_{H_i}^{(n)}(\theta) = \log \mathrm{E}\big[e^{\theta \bar{H}_i(n)}\big] \text{ and } \Lambda_{G_{ij}}^{(n)}(\theta) = \log \mathrm{E}\big[e^{\theta \bar{G}_{ij}(n)}\big]$$

be the cumulant generating functions of $\bar{H}_i(n)$ and $\bar{G}_{ij}(n)$, respectively. Let the effective domain of a function $f(\cdot)$ be denoted by $\mathcal{D}_f = \{x : f(x) < \infty\}$ and its interior by $\mathcal{D}_f^{\circ}$. Let $f'(x)$ denote the derivative of $f$ with respect to the argument $x$. As is usual in LD contexts, we make the following assumption.

**Assumption 3.2.** *For each system $i \leq r$ and constraint $j \leq s$,*

*(1) the limits*

$$\Lambda_{H_i}(\theta) = \lim_{n\to\infty} \frac{1}{n} \Lambda_{H_i}^{(n)}(n\theta) \text{ and } \Lambda_{G_{ij}}(\theta) = \lim_{n\to\infty} \frac{1}{n} \Lambda_{G_{ij}}^{(n)}(n\theta)$$

*exist as extended real numbers for all $\theta$;*

*(2) the origin belongs to the interior of $\mathcal{D}_{\Lambda_{H_i}}$ and $\mathcal{D}_{\Lambda_{G_{ij}}}$, that is, $0 \in \mathcal{D}^{\circ}_{\Lambda_{H_i}}$ and $0 \in \mathcal{D}^{\circ}_{\Lambda_{G_{ij}}}$;*

*(3) $\Lambda_{H_i}(\theta)$ and $\Lambda_{G_{ij}}(\theta)$ are strictly convex and $C^{\infty}$ on $\mathcal{D}^{\circ}_{\Lambda_{H_i}}$ and $\mathcal{D}^{\circ}_{\Lambda_{G_{ij}}}$, respectively;*

*(4) $\Lambda_{H_i}(\theta)$ and $\Lambda_{G_{ij}}(\theta)$ are steep, that is, for any sequence $\{\theta_n\} \in \mathcal{D}_{\Lambda_{H_i}}$ that converges to a boundary point of $\mathcal{D}_{\Lambda_{H_i}}$, $\lim_{n\to\infty} |\Lambda'_{H_i}(\theta_n)| = \infty$, and likewise, for $\{\theta_n\} \in \mathcal{D}_{\Lambda_{G_{ij}}}$ converging to a boundary point of $\mathcal{D}_{\Lambda_{G_{ij}}}$, $\lim_{n\to\infty} |\Lambda'_{G_{ij}}(\theta_n)| = \infty$.*

Assumption 3.2 implies that $\bar{H}_i \to h_i$ wp1 and $\bar{G}_{ij} \to g_{ij}$ wp1 (see Bucklew, 2003, remark 3.2.1). Furthermore, assumption 3.2 ensures that, by the Gärtner-Ellis theorem (Dembo and Zeitouni, 1998, p. 44), $\bar{H}_i$ and $\bar{G}_{ij}$ satisfy the large deviations principle (LDP) (Dembo and Zeitouni, 1998, p. 4) with good rate functions $I_i(x) = \sup_{\theta\in\mathbb{R}}\{\theta x - \Lambda_{H_i}(\theta)\}$ and $J_{ij}(y) = \sup_{\theta\in\mathbb{R}}\{\theta y - \Lambda_{G_{ij}}(\theta)\}$. Assumption 3.2(3) is stronger than what is needed for the Gärtner-Ellis theorem to hold. However, we require $\Lambda_{H_i}(\theta)$ and $\Lambda_{G_{ij}}(\theta)$ to be strictly convex and $C^{\infty}$ on $\mathcal{D}^{\circ}_{\Lambda_{H_i}}$ and $\mathcal{D}^{\circ}_{\Lambda_{G_{ij}}}$, respectively, so that $I_i(x)$ and $J_{ij}(y)$ are strictly convex and $C^{\infty}$ for $x \in \mathcal{F}^{\circ}_{H_i} = \text{int}\{\Lambda'_{H_i}(\theta) : \theta \in \mathcal{D}^{\circ}_{\Lambda_{H_i}}\}$ and $y \in \mathcal{F}^{\circ}_{G_{ij}} = \text{int}\{\Lambda'_{G_{ij}}(\theta) : \theta \in \mathcal{D}^{\circ}_{\Lambda_{G_{ij}}}\}$, respectively.

Let $h_\ell = \arg\min_i\{h_i\}$ and let $h_u = \arg\max_i\{h_i\}$. We further assume,

**Assumption 3.3.** *(1) the interval $[h_\ell, h_u] \subset \cap_{i=1}^{r} \mathcal{F}^{\circ}_{H_i}$, and (2) $\gamma_j \in \cap_{i=1}^{r} \mathcal{F}^{\circ}_{G_{ij}}$ for all $j \le s$.*

Assumption 3.3 assigns nonzero probability to false selection events. As in Glynn and Juneja (2004), assumption 3.3(1) ensures that $\bar{H}_i$ may take any value in the interval $[h_\ell, h_u]$. Consequently, this assumption ensures that we may falsely estimate any system $i$, $i = 2, \ldots, r$, as at least as good in objective function value as system 1. Assumption 2.3(2) ensures there is a nonzero probability that each system will be estimated feasible and a nonzero probability that each system will be estimated infeasible. Specifically, it ensures there is a nonzero probability that system 1 will be estimated infeasible and a nonzero probability that an infeasible system will be estimated feasible.

## 3.4 Rate Function of Probability of False Selection

The false selection (FS) event is the event that the actual best feasible system, system 1, is not the estimated best feasible system. More specifically, FS is the event that system 1 is incorrectly estimated infeasible on *any* of its constraints, or that system 1 is estimated feasible on all of its constraints but another system, also estimated feasible on all of its constraints, has the best estimated-objective value. Let $\bar{\Gamma}$ be the set of estimated-feasible systems, excluding system 1, that is, $\bar{\Gamma} = \{i : \hat{G}_{ij} \leq \gamma_j \text{ for all } j \leq s, i \neq 1\}$. Then formally, the probability of false selection is

$$\begin{aligned} P\{FS\} &= P\{(\cup_j \hat{G}_{1j} > \gamma_j) \cup ((\cap_j \hat{G}_{1j} \leq \gamma_j) \cap (\hat{H}_1 \geq \min_{i\in\bar{\Gamma}} \hat{H}_i))\} \\ &= P\{\cup_j \hat{G}_{1j} > \gamma_j\} + P\{(\cap_j \hat{G}_{1j} \leq \gamma_j) \cap (\cup_{i\in\bar{\Gamma}} \hat{H}_1 \geq \hat{H}_i)\} \qquad (3.3) \\ &= P\{FS_1\} + P\{FS_2\}. \end{aligned}$$

The following theorems 3.4 and 3.5 provide derivations of the rate functions for $P\{FS_1\}$ and $P\{FS_2\}$, respectively, whose expressions appear in equation (3.3). In theorem 3.5, through the strategic division of systems into the sets $\Gamma, \mathcal{S}_b$, and $\mathcal{S}_w$, the rate function of $P\{FS_2\}$ splits into the minimum of three separate terms corresponding to the rate functions for systems in each of these three sets. Finally, in theorem 3.6, these results are combined to produce the overall rate function for $P\{FS\}$.

First consider the rate function for $P\{FS_1\}$, the probability that system 1 is declared infeasible on any of its constraints. Theorem 3.4 establishes the asymptotic behavior of $P\{FS_1\}$ as the rate function corresponding to the constraint on system 1 that is most likely to be declared unsatisfied.

**Theorem 3.4.** *The rate function for $P\{FS_1\}$ is given by*

$$-\lim_{n\to\infty} \frac{1}{n} \log P\{FS_1\} = \min_{j\in\mathcal{C}_F^1} \alpha_1 J_{1j}(\gamma_j).$$

*Proof.* The following upper and lower bounds hold for $P\{FS_1\}$:

$$\max_{j\in\mathcal{C}_F^1} P\{\hat{G}_{1j} > \gamma_j\} \leq P\{\cup_{j\in\mathcal{C}_F^1} \hat{G}_{1j} > \gamma_j\} \leq s \max_{j\in\mathcal{C}_F^1} P\{\hat{G}_{1j} > \gamma_j\}.$$

Assuming the relevant limits exist and by application of proposition A.2 (see appendix A) the rate function for $\max_{j\in\mathcal{C}_F^1} P\{\hat{G}_{1j} > \gamma_j\}$ is given by

$$\lim_{n\to\infty} \frac{1}{n} \log \max_{j\in\mathcal{C}_F^1} P\{\hat{G}_{1j} > \gamma_j\} = \max_{j\in\mathcal{C}_F^1} \lim_{n\to\infty} \frac{1}{n} \log P\{\hat{G}_{1j} > \gamma_j\}.$$

Under assumption 3.2, by the Gärtner-Ellis theorem, we have

$$-\inf_{y\in(\gamma_j,\infty)} J_{1j}(y) \leq \liminf_{n\to\infty} \frac{1}{n} \log P(\bar{G}_{1j}(n) > \gamma_j) \leq \limsup_{n\to\infty} \frac{1}{n} \log P(\bar{G}_{1j}(n) > \gamma_j) \leq -\inf_{y\in[\gamma_j,\infty)} J_{1j}(y)$$

Since by assumption 3.3, $J_{1j}(y)$ is strictly convex and $C^\infty$ in $\mathcal{F}^\circ_{G_{1j}}$ and $g_{1j} < \gamma_j \in \mathcal{F}^\circ_{G_{1j}}$, it follows that $-\inf_{y\in(\gamma_j,\infty)} J_{1j}(y) = -\inf_{y\in[\gamma_j,\infty)} J_{1j}(y) = -J_{1j}(\gamma_j)$. Therefore the limit exists.

Now let the cumulant generating function of $\hat{G}_{1j}$ be denoted $\Lambda^{(\alpha_1 n)}_{G_{1j}}(\theta)$. By assumption 3.2,

$$\lim_{n\to\infty}\frac{1}{n}\Lambda^{(\alpha_1 n)}_{G_{1j}}(n\theta) = \lim_{n\to\infty}\frac{1}{n}\log \mathrm{E}[e^{\theta/\alpha_1 \sum_{k=1}^{\alpha_1 n} G_{1jk}}] = \lim_{n\to\infty}\frac{\alpha_1}{n}\log \mathrm{E}[e^{\theta/\alpha_1 \sum_{k=1}^{n} G_{1jk}}]$$
$$= \alpha_1 \Lambda_{G_{1j}}(\theta/\alpha_1).$$

Hence $\hat{G}_{1j}$ also satisfies the LDP with good rate function

$$\sup_\theta\{\theta y - \alpha_1\Lambda_{G_{1j}}(\theta/\alpha_1)\} = \alpha_1 \sup_{\theta/\alpha_1}\{(\theta/\alpha_1)y - \Lambda_{G_{1j}}(\theta/\alpha_1)\} = \alpha_1 J_{1j}(y).$$

Therefore by similar arguments to those given above, the infimum over the set $(\gamma,\infty)$ is achieved at $\alpha_1 J_{1j}(\gamma)$. Therefore the limit exists and

$$\lim_{n\to\infty}\frac{1}{n}\log P\{FS_1\} = \max_{j\in\mathcal{C}_F^1}\lim_{n\to\infty}\frac{1}{n}\log P\{\hat{G}_{1j} > \gamma_j\} = -\min_{j\in\mathcal{C}_F^1}\alpha_1 J_{1j}(\gamma_j). \qquad \square$$

Theorem 3.4 implies that the rate function for $P\{FS_1\}$ is determined by the constraint that is most likely to qualify system 1 as infeasible. Under our assumptions and with logic similar that given in the proof of theorem 3.4, it can be shown that for any system $i$ that does not satisfy constraint $j$, the rate function for the probability that system $i$ is incorrectly estimated feasible on constraint $j$ is

$$\lim_{n\to\infty}\frac{1}{n}\log P\{\hat{G}_{ij}\le\gamma_j\} = -\alpha_i J_{ij}(\gamma_j).$$

Now consider $P\{FS_2\}$. Since the probability that system 1 is estimated feasible tends to one and under our assumptions regarding independence (assumptions 1 and 3.1), it holds that

$$\lim_{n\to\infty}\frac{1}{n}\log P\{(\cap_j\ \hat{G}_{1j}\ge\gamma_j)\cap(\cup_{i\in\bar{\Gamma}}\ \hat{H}_1\ge\hat{H}_i)\} = \lim_{n\to\infty}\frac{1}{n}\log P\{\cup_{i\in\bar{\Gamma}}\ \hat{H}_1\ge\hat{H}_i\}. \qquad (3.4)$$

Therefore the rate function of $P\{FS_2\}$ is governed by the rate at which the probability that system 1 is "beaten" by another estimated-feasible system tends to zero. Since the equality in equation (3.4) always holds, in the remainder of the chapter, we omit the explicit statement of the event that system 1 is estimated feasible. Since the estimated set of feasible systems $\bar{\Gamma}$ may contain worse feasible systems ($i\in\Gamma$), better infeasible systems ($i\in\mathcal{S}_b$), and worse infeasible systems ($i\in\mathcal{S}_w$), we strategically consider the rate functions for the probability that system 1 is beaten by a system in $\Gamma,\mathcal{S}_b$, or $\mathcal{S}_w$ separately. Theorem 3.5 states that the rate function of $P\{FS_2\}$ is determined by the slowest-converging probability that system 1 will be "beaten" by an estimated-feasible system from $\Gamma,\mathcal{S}_b$, or $\mathcal{S}_w$.

**Theorem 3.5.** *The rate function for $P\{FS_2\}$ is given by the minimum rate function of the probability that system 1 is beaten by an estimated-feasible system that is (i) feasible and worse, (ii) infeasible and better, or (iii) infeasible and worse. That is,*

$$-\lim_{n\to\infty}\frac{1}{n}\log P\{FS_2\} = \min\Bigg(\overbrace{\min_{i\in\Gamma}(\inf_x(\alpha_1 I_1(x)+\alpha_i I_i(x))),}^{\substack{\textit{system 1 beaten by a}\\ \textit{feasible and worse system}}}\quad \overbrace{\min_{i\in\mathcal{S}_b}\alpha_i\sum_{j\in\mathcal{C}_I^i}J_{ij}(\gamma_j),}^{\substack{\textit{system 1 beaten by an}\\ \textit{infeasible and better system}}}$$
$$\underbrace{\min_{i\in\mathcal{S}_w}(\inf_x(\alpha_1 I_1(x)+\alpha_i I_i(x))+\alpha_i\sum_{j\in\mathcal{C}_I^i}J_{ij}(\gamma_j))}_{\substack{\textit{system 1 beaten by an}\\ \textit{infeasible and worse system}}}\Bigg).$$

*Proof.* See appendix B, section B.1. □

Like the intuition behind theorem 3.4, that the rate function of $P\{FS_1\}$ is determined by the constraint most likely to disqualify system 1, in theorem 3.5, the rate function of $P\{FS_2\}$ is determined by the system most likely to "beat" system 1. However systems in $\Gamma, \mathcal{S}_b$, and $\mathcal{S}_w$ must overcome different obstacles to be falsely estimated as the best feasible system. Since systems in $\Gamma$ are truly feasible, they must overcome one obstacle: optimality. The rate function for systems in $\Gamma$ is thus identical to the unconstrained optimization case presented in Glynn and Juneja (2004) and is determined by the system in $\Gamma$ most likely to be falsely estimated as optimal. Systems in $\mathcal{S}_b$ are truly better than system 1, but are infeasible. They also have one obstacle to overcome to be selected as best: feasibility. The rate function for systems in $\mathcal{S}_b$ is thus determined by the system in $\mathcal{S}_b$ that is most likely to be falsely estimated as feasible. Since an infeasible system in $\mathcal{S}_b$ must falsely be estimated as feasible on *all* of its infeasible constraints, the rate functions for the infeasible constraints simply add up inside the overall rate function for each system in $\mathcal{S}_b$. Systems in $\mathcal{S}_w$ are worse and infeasible, so two obstacles must be overcome: optimality and feasibility. The rate function for systems in $\mathcal{S}_w$ is thus determined by the system that is most likely to be falsely estimated as optimal and feasible, and there are two terms added in the rate function corresponding to optimality and feasibility.

Theorem 3.6 presents the rate function for the $P\{FS\}$ that results from combining the rate function results for $P\{FS_1\}$ and $P\{FS_2\}$. Recalling from equation (3.3) that $P\{FS\} = P\{FS_1\} + P\{FS_2\}$, the overall rate function for the probability of false selection is determined by the minimum of the rate functions for $P\{FS_1\}$ and $P\{FS_2\}$.

**Theorem 3.6.** *The rate function for the probability of false selection, that is, the probability that we return to the user a system other than system 1 is given by*

$$-\lim_{n\to\infty}\frac{1}{n}\log P\{FS\} = \min\Bigg(\overbrace{\min_j \alpha_1 J_{1j}(\gamma_j)}^{\substack{\text{system 1}\\ \text{estimated infeasible}}},$$

$$\overbrace{\underbrace{\min_{i\in\Gamma}(\inf_x(\alpha_1 I_1(x)+\alpha_i I_i(x)))}_{\substack{\text{system 1 beaten by}\\ \text{feasible and worse system}}},\ \underbrace{\min_{i\in\mathcal{S}_b}\alpha_i\sum_{j\in\mathcal{C}_I^i}J_{ij}(\gamma_j)}_{\substack{\text{system 1 beaten}\\ \text{by infeasible and}\\ \text{better system}}},\ \underbrace{\min_{i\in\mathcal{S}_w}(\inf_x(\alpha_1 I_1(x)+\alpha_i I_i(x))+\alpha_i\sum_{j\in\mathcal{C}_I^i}J_{ij}(\gamma_j))}_{\substack{\text{system 1 beaten by}\\ \text{infeasible and worse system}}}}^{\text{system 1 estimated feasible}}\Bigg).$$

Theorem 3.6 asserts that the overall rate function of the probability of false selection is determined by the most likely of the following four events: (i) system 1 is incorrectly declared infeasible on one of its constraints; (ii) a feasible and worse system is correctly declared feasible, but incorrectly declared best; (iii) an infeasible and better system is correctly declared better, but incorrectly declared feasible; (iv) an infeasible and worse system is incorrectly declared feasible and best. This result is intuitive since we expect an unlikely event to happen in the most likely way.

## 3.5 Optimal Allocation Strategy

In this section, we derive an optimal allocation strategy that asymptotically minimizes the probability of false selection. An asymptotically optimal allocation strategy will result from maximizing the rate at which the $P\{FS\}$ tends to zero as a function of $\boldsymbol{\alpha}$. Thus we wish to allocate the $\alpha_i$'s to solve the following optimization problem:

$$\begin{aligned}\max\quad & \min\Bigg(\min_j \alpha_1 J_{1j}(\gamma_j),\ \min_{i\in\Gamma}\Big(\inf_x(\alpha_1 I_1(x)+\alpha_i I_i(x))\Big),\ \min_{i\in\mathcal{S}_b}\alpha_i\sum_{j\in\mathcal{C}_I^i}J_{ij}(\gamma_j),\\ &\qquad \min_{i\in\mathcal{S}_w}\Bigg(\inf_x(\alpha_1 I_1(x)+\alpha_i I_i(x))+\alpha_i\sum_{j\in\mathcal{C}_I^i}J_{ij}(\gamma_j)\Bigg)\Bigg)\quad s.t.\\ & \sum_{i=1}^{r}\alpha_i = 1,\quad \alpha_i\ge 0.\end{aligned}\tag{3.5}$$

By Glynn and Juneja (2006), $\inf_x(\alpha_1 I_1(x)+\alpha_i I_i(x))$ is a concave, strictly increasing, $C^\infty$ function of $\alpha_1$ and $\alpha_i$. Let $x(\alpha_1,\alpha_i)=\arg\inf_x(\alpha_1 I_1(x)+\alpha_i I_i(x))$. As Glynn and Juneja (2006) demonstrate, for $\alpha_1>0$ and $\alpha_i>0$, $x(\alpha_1,\alpha_i)$ is a $C^\infty$ function of $\alpha_1$ and $\alpha_i$. Likewise, the linear functions $\alpha_1 J_{1j}(\gamma_j)$

and $\alpha_i \sum_{j\in\mathcal{C}_I^i} J_{ij}(\gamma_j)$ and the sum $\inf_x(\alpha_1 I_1(x)+\alpha_i I_i(x))+\alpha_i \sum_{j\in\mathcal{C}_I^i} J_i(\gamma_j)$ are also concave, strictly increasing $C^\infty$ functions of $\alpha_1$ and $\alpha_i$. Since the minimum of concave, strictly increasing functions is also concave and strictly increasing, the problem in (3.5) is a concave maximization problem. Equivalently, the problem in (3.5) may be written as the following problem $Q$.

$$\text{Problem } Q: \qquad \max \quad z \quad s.t.$$
$$\alpha_1 J_{1j}(\gamma_j) \geq z,\ j \in \mathcal{C}_F^1$$
$$\alpha_1 I_1(x(\alpha_1,\alpha_i)) + \alpha_i I_i(x(\alpha_1,\alpha_i)) \geq z,\ i \in \Gamma$$
$$\alpha_i \sum_{j\in\mathcal{C}_I^i} J_{ij}(\gamma_j) \geq z,\ i \in \mathcal{S}_b$$
$$\alpha_1 I_1(x(\alpha_1,\alpha_i)) + \alpha_i I_i(x(\alpha_1,\alpha_i)) + \alpha_i \sum_{j\in\mathcal{C}_I^i} J_{ij}(\gamma_j) \geq z,\ i \in \mathcal{S}_w$$
$$\sum_{i=1}^{r} \alpha_i = 1,\ \alpha_i \geq 0.$$

Since problem $Q$ is a strictly concave, continuous function of $\boldsymbol{\alpha}$ on a compact set, a unique solution exists. Proposition 3.7 states this result.

**Proposition 3.7.** *There exists a unique solution* $\boldsymbol{\alpha}^* = \{\alpha_1^*, \alpha_2^*, \ldots, \alpha_r^*\}$ *to problem* $Q$*, with optimal value* $z^*$*.*

*Proof.* Under assumptions 2 and 3.3, the rate function of $P\{FS\}$ as a function of $\boldsymbol{\alpha}$ is a map from a compact set onto the set of positive real numbers. By assumption 3.2, the rate function of $P\{FS\}$ is also a continuous function of $\boldsymbol{\alpha}$, and hence attains its maximum on its domain. Therefore a solution $\boldsymbol{\alpha}^*$ exists as a non-negative real number for problem $Q$. Since $P\{FS\}$ is a strictly increasing concave function of $\boldsymbol{\alpha}$, the solution $\boldsymbol{\alpha}^*$ is unique. □

Slater's condition (see, e.g. Boyd and Vandenberghe, 2004, p. 244) holds for problem $Q$, that is, there exists a point $\boldsymbol{\alpha}= (\alpha_1, \alpha_2, \ldots, \alpha_r)$ in the relative interior of the feasible set of problem $Q$ for which the inequality constraints are not binding. For example, a point satisfying this condition is $z = 0, \alpha_i = 1/r, i \leq r$. Since problem $Q$ is concave with differentiable objective function and constraints and Slater's condition holds, the Karush-Kuhn Tucker (KKT) conditions are necessary and sufficient for global optimality (see, e.g., Boyd and Vandenberghe, 2004). Therefore the solution to the KKT conditions for problem $Q$ will yield a unique optimal solution $\boldsymbol{\alpha}^*$. By solving the KKT conditions for problem $Q$, we encounter a more conveniently-expressed problem $Q^*$ which has an identical optimal solution and optimal value to that of problem $Q$. Let us define problem $Q^*$ by replacing the inequality constraints corresponding to systems in $\Gamma, \mathcal{S}_b$, and $\mathcal{S}_w$ with equality constraints, and forcing each $\alpha_i$ to be strictly greater than zero.

$$\begin{aligned}
\text{Problem } Q^*: \qquad \max \quad z \quad s.t.& \\
\alpha_1 J_{1j}(\gamma_j) \geq z,\ j \in \mathcal{C}_F^1& \\
\alpha_1 I_1(x(\alpha_1,\alpha_i)) + \alpha_i I_i(x(\alpha_1,\alpha_i)) = z,\ i \in \Gamma& \\
\alpha_i \sum_{j\in\mathcal{C}_I^i} J_{ij}(\gamma_j) = z,\ i \in \mathcal{S}_b& \\
\alpha_1 I_1(x(\alpha_1,\alpha_i)) + \alpha_i I_i(x(\alpha_1,\alpha_i)) + \alpha_i \sum_{j\in\mathcal{C}_I^i} J_{ij}(\gamma_j) = z,\ i \in \mathcal{S}_w& \\
\sum_{i=1}^{r} \alpha_i = 1,\ \alpha_i > 0.&
\end{aligned}$$

The following proposition states the equivalence of problem $Q$ and problem $Q^*$.

**Proposition 3.8.** *Problems $Q$ and $Q^*$ are equivalent, that is, problem $Q^*$ has the unique solution $\boldsymbol{\alpha}^*$ with optimal value $z^*$.*

*Proof.* For $\alpha_i = 1/r, i \leq r$, it follows that $z > 0$ in $Q$. Therefore $\alpha_i = 0$ for $i \in \{1\} \cup \mathcal{S}_b$ is suboptimal since $z = 0$. Now consider $\alpha_i = 0$ for $i \in \Gamma \cup \mathcal{S}_w$. In this case, the constraints for $i \in \Gamma \cup \mathcal{S}_w$ reduce to $\alpha_1 \inf_x I_1(x) = \alpha_1 I_1(h_1) = 0$, and hence $z = 0$. Therefore in problem $Q$, we must have $\alpha_i^* > 0$ for all $i \leq r$. As a matter of notation, for the remainder of this proof we temporarily append the variable $z$ to the vector $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha} = (z, \alpha_1, \ldots, \alpha_r)$.

Denoting the dual variables $\nu$ and $\boldsymbol{\lambda} = (\lambda_j^1 \geq 0, \lambda_i \geq 0 : j = 1, \ldots, |\mathcal{C}_F^1|, i = 2, \ldots, r)$, the Lagrangian dual may be written as,

$$\begin{aligned}
L(\boldsymbol{\alpha},\boldsymbol{\lambda},\nu) = z &+ \sum_{j\in\mathcal{C}_F^1} \lambda_j^1(\alpha_1 J_{1j}(\gamma_j) - z) \\
&+ \sum_{i\in\Gamma} \lambda_i\Big(\alpha_1 I_1(x(\alpha_1,\alpha_i)) + \alpha_i I_i(x(\alpha_1,\alpha_i)) - z\Big) + \sum_{i\in\mathcal{S}_b} \lambda_i\Big(\alpha_i \sum_{i\in\mathcal{C}_I^i} J_{ij}(\gamma_j) - z\Big) \\
&+ \sum_{i\in\mathcal{S}_w} \lambda_i\Big(\alpha_1 I_1(x(\alpha_1,\alpha_i)) + \alpha_i\big[I_i(x(\alpha_1,\alpha_i)) + \sum_{j\in\mathcal{C}_I^i} J_{ij}(\gamma_j)\big] - z\Big) + \nu\Big(\sum_{i=1}^{r} \alpha_i - 1\Big)
\end{aligned}$$

$$L(\boldsymbol{\alpha}, \boldsymbol{\lambda}, \nu) = z\Big(1 - \sum_{j \in \mathcal{C}_F^1} \lambda_j^1 - \sum_{i=2}^{r} \lambda_i\Big) + \nu\Big(\sum_{i=1}^{r} \alpha_i - 1\Big)$$
$$+ \alpha_1\Big(\sum_{j \in \mathcal{C}_F^1} \lambda_j^1 J_{1j}(\gamma_j) + \sum_{i \in \Gamma \cup \mathcal{S}_w} \lambda_i I_1(x(\alpha_1, \alpha_i))\Big) + \sum_{i \in \Gamma} \alpha_i \lambda_i I_i(x(\alpha_1, \alpha_i))$$
$$+ \sum_{i \in \mathcal{S}_w} \alpha_i \lambda_i \Big(I_i(x(\alpha_1, \alpha_i)) + \sum_{j \in \mathcal{C}_I^i} J_{ij}(\gamma_j)\Big) + \sum_{i \in \mathcal{S}_b} \alpha_i \lambda_i \sum_{j \in \mathcal{C}_I^i} J_{ij}(\gamma_j).$$

Now we take the gradient of the Lagrangian and set it equal to zero to solve for the stationarity KKT conditions. Since $x(\alpha_1, \alpha_i)$ solves $\alpha_1 I_1'(x) + \alpha_i I_i'(x) = 0$, it holds that $\frac{\partial}{\partial \alpha_1}\big(\alpha_1 I_1(x(\alpha_1, \alpha_i)) + \alpha_i I_i(x(\alpha_1, \alpha_i))\big) = I_1(x(\alpha_1, \alpha_i))$ and $\frac{\partial}{\partial \alpha_i}\big(\alpha_1 I_1(x(\alpha_1, \alpha_i)) + \alpha_i I_i(x(\alpha_1, \alpha_i))\big) = I_i(x(\alpha_1, \alpha_i))$ (see Glynn and Juneja, 2004). Then the stationarity conditions are,

$$\sum_{j=1}^{|\mathcal{C}_F^1|} \lambda_j^1 + \sum_{i=2}^{r} \lambda_i = 1 \tag{3.6}$$
$$\sum_{j \in \mathcal{C}_F^1} \lambda_j^1 J_{1j}(\gamma_j) + \sum_{i \in \Gamma \cup \mathcal{S}_w} \lambda_i I_1(x(\alpha_1^*, \alpha_i^*)) = \nu \tag{3.7}$$
$$\lambda_i I_i(x(\alpha_1^*, \alpha_i^*)) = \nu, \ i \in \Gamma \tag{3.8}$$
$$\lambda_i \sum_{j \in \mathcal{C}_I^i} J_{ij}(\gamma_j) = \nu, \ i \in \mathcal{S}_b \tag{3.9}$$
$$\lambda_i \big[I_i(x(\alpha_1^*, \alpha_i^*)) + \sum_{j \in \mathcal{C}_I^i} J_{ij}(\gamma_j)\big] = \nu, \ i \in \mathcal{S}_w, \tag{3.10}$$

and the complementary slackness conditions are,

$$\lambda_j^1 \big[\alpha_1^* J_{1j}(\gamma_j) - z\big] = 0, \ j \in \mathcal{C}_F^1 \tag{3.11}$$
$$\lambda_i \big[\alpha_1^* I_1(x(\alpha_1^*, \alpha_i^*)) + \alpha_i^* I_i(x(\alpha_1^*, \alpha_i^*)) - z\big] = 0, \ i \in \Gamma \tag{3.12}$$
$$\lambda_i \Big[\alpha_i^* \sum_{j \in \mathcal{C}_I^i} J_{ij}(\gamma_j) - z\Big] = 0, \ i \in \mathcal{S}_b \tag{3.13}$$
$$\lambda_i \Big[\alpha_1^* I_1(x(\alpha_1^*, \alpha_i^*)) + \alpha_i^* I_i(x(\alpha_1^*, \alpha_i^*)) + \alpha_i^* \sum_{j \in \mathcal{C}_I^i} J_{ij}(\gamma_j) - z\Big] = 0, \ i \in \mathcal{S}_w. \tag{3.14}$$

Equation (3.6) implies that at least one $\lambda_i > 0$. Suppose $\lambda_i = 0$ for some $i \in \Gamma \cup \mathcal{S}_b \cup \mathcal{S}_w$. Since $\alpha_i > 0$ for all $i \leq r$, the rate functions in equations (3.8)–(3.10) are strictly greater than zero, which implies $\nu = 0, \lambda_i = 0$ for all $i \in \Gamma \cup \mathcal{S}_b \cup \mathcal{S}_w$, and $\sum_{j=1}^{|\mathcal{C}_F^1|} \lambda_j^1 = 1$. Therefore at least one $\lambda_j^1 > 0$.

Then in equation (3.7), it must be the case that for $\lambda_j^1 > 0$, the corresponding $J_{1j}(\gamma_j) = 0$. However we have a contradiction since by assumption, $J_{1j}(\gamma_j) > 0$ for all $j \in |\mathcal{C}_F^1|$. Therefore $\lambda_i > 0$ for all $i \in \Gamma \cup \mathcal{S}_b \cup \mathcal{S}_w$.

Since $\lambda_i > 0$ in equations (3.12)–(3.14), then complementary slackness implies each of these constraints is binding. Therefore the inequality constraints corresponding to $i \in \Gamma \cup \mathcal{S}_b \cup \mathcal{S}_w$ in problem $Q$ may be replaced with equality constraints in problem $Q^*$. □

The structure of the identical problem $Q^*$ lends intuition to the structure of the optimal allocation, as noted in the following steps: (i) Solve a relaxation of problem $Q^*$ without the feasibility constraint for system 1. Let this problem be called problem $\widetilde{Q}^*$, and let $\tilde{z}^*$ be the optimal value at the optimal solution $\tilde{\boldsymbol{\alpha}}^* = (\tilde{\alpha}_1^*, \ldots, \tilde{\alpha}_r^*)$ to Problem $\widetilde{Q}^*$. (ii) Check if the feasibility constraint for system 1 is satisfied by the solution $\tilde{\boldsymbol{\alpha}}^*$. If the feasibility constraint is satisfied, $\tilde{\boldsymbol{\alpha}}^*$ is the optimal solution for problem $Q^*$. Otherwise, (iii) force the feasibility constraint to be binding. The steps (i), (ii), and (iii) are equivalent to solving one of two systems of nonlinear equations, as identified by the KKT conditions of problems $Q^*$ and $\tilde{Q}^*$. Theorem 3.9 asserts this result formally.

**Theorem 3.9.** *Let the set of suboptimal feasible systems $\Gamma$ be non-empty, and define problem $\tilde{Q}^*$ as problem $Q^*$ but with the inequality constraint on the feasibility of system 1 relaxed. Let $(\boldsymbol{\alpha}^*, z^*)$ and $(\tilde{\boldsymbol{\alpha}}^*, \tilde{z}^*)$ denote the unique optimal solution and optimal value pairs for problems $Q^*$ and $\widetilde{Q}^*$, respectively. Consider the conditions,*

*C0.* $\sum_{i=1}^r \alpha_i = 1$, $\boldsymbol{\alpha} > 0$, *and*

$$z = \alpha_1 I_1(x(\alpha_1, \alpha_i)) + \alpha_i I_i(x(\alpha_1, \alpha_i)) = \alpha_k \textstyle\sum_{j \in \mathcal{C}_I^k} J_{kj}(\gamma_j)$$

$$= \alpha_1 I_1(x(\alpha_1, \alpha_\ell)) + \alpha_\ell \big[I_\ell(x(\alpha_1, \alpha_\ell)) + \textstyle\sum_{j \in \mathcal{C}_I^\ell} J_{\ell j}(\gamma_j)\big], \textit{ for all } i \in \Gamma, k \in \mathcal{S}_b,\ \ell \in \mathcal{S}_w,$$

*C1.* $$\sum_{i \in \Gamma} \frac{I_1(x(\alpha_1, \alpha_i))}{I_i(x(\alpha_1, \alpha_i))} + \sum_{i \in \mathcal{S}_w} \frac{I_1(x(\alpha_1, \alpha_i))}{I_i(x(\alpha_1, \alpha_i)) + \sum_{j \in \mathcal{C}_I^i} J_{ij}(\gamma_j)} = 1,$$

*C2.* $\min_{j \in \mathcal{C}_F^1} \alpha_1 J_{1j}(\gamma_j) = z$.

*Then (i) $\tilde{\boldsymbol{\alpha}}^*$ solves C0 and C1 and $\min_{j \in \mathcal{C}_F^1} \tilde{\alpha}_1^* J_{1j}(\gamma_j) \geq \tilde{z}^*$ if and only if $\tilde{\boldsymbol{\alpha}}^* = \boldsymbol{\alpha}^*$; and*

*(ii) $\boldsymbol{\alpha}^*$ solves C0 and C2 and $\min_{j \in \mathcal{C}_F^1} \tilde{\alpha}_1^* J_{1j}(\gamma_j) < \tilde{z}^*$ if and only if $\boldsymbol{\alpha}^* \neq \tilde{\boldsymbol{\alpha}}^*$.*

*Proof.* Due to the structure of problem $Q$, the KKT conditions are necessary and sufficient for global optimality. From prior results, we recall that the solutions to problems $Q$, $Q^*$, and $\widetilde{Q}^*$ exist, and that condition C0 holds for the solutions $\boldsymbol{\alpha}^*$ and $\tilde{\boldsymbol{\alpha}}^*$.

We simplify the KKT equations for problem $Q$ for use in the remainder of the proof. Since $\lambda_i > 0$ for all $i \in \Gamma \cup \mathcal{S}_b \cup \mathcal{S}_w$ in the proof of proposition 3.8, it follows that $\nu > 0$. Dividing (3.7) by

$\nu$ and appropriately substituting in values from (3.8)–(3.10), we find

$$\frac{\sum_{j\in\mathcal{C}_F^1}\lambda_j^1 J_{1j}(\gamma_j)}{\nu}+\sum_{i\in\Gamma}\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*))}+\sum_{i\in\mathcal{S}_w}\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*))+\sum_{j\in\mathcal{C}_I^i}J_{ij}(\gamma_j)}=1. \tag{3.15}$$

By a similar logic to that given in the proof of proposition 3.8 and the simplification provided in (3.15), omitting terms with $\lambda_j^1$ in equation (3.15) yields condition C1 as a KKT condition for problem $\widetilde{Q}^*$. Taken together, C0 and C1 create a fully-specified system of equations that form the KKT conditions for problem $\widetilde{Q}^*$. A solution $\boldsymbol{\alpha}$ is thus optimal to problem $\widetilde{Q}^*$ if and only if it solves C0 and C1.

Let $\mathcal{D}(Q^*)$ and $\mathcal{D}(\widetilde{Q}^*)$ denote the feasible regions of problems $Q^*$ and $\widetilde{Q}^*$, respectively.

*Proof of Claim (i).* ($\Rightarrow$) Suppose $\tilde{\boldsymbol{\alpha}}^*$ solves C0 and C1, and $\min_{j\in\mathcal{C}_F^1}\tilde{\alpha}_1^* J_{1j}(\gamma_j)\geq\tilde{z}^*$. Then $\tilde{\boldsymbol{\alpha}}^*\in\mathcal{D}(Q^*)$. Since the objective functions of problems $Q^*$ and $\widetilde{Q}^*$ are identical, and $\mathcal{D}(Q^*)\subset\mathcal{D}(\widetilde{Q}^*)$, it follows that $z^*\leq\tilde{z}^*$. Therefore $\tilde{\boldsymbol{\alpha}}^*\in\mathcal{D}(Q^*)$ implies $\tilde{\boldsymbol{\alpha}}^*$ is the optimal solution to problem $Q^*$, and by the uniqueness of the optimal solution, $\tilde{\boldsymbol{\alpha}}^*=\boldsymbol{\alpha}^*$.

($\Leftarrow$) Now suppose $\tilde{\boldsymbol{\alpha}}^*=\boldsymbol{\alpha}^*$. Since $\tilde{\boldsymbol{\alpha}}^*$ is the optimal solution to problem $\widetilde{Q}^*$, then $\tilde{\boldsymbol{\alpha}}^*$ solves C0 and C1. Further, since $\boldsymbol{\alpha}^*$ is the optimal solution to problem $Q$, $\boldsymbol{\alpha}^*=\tilde{\boldsymbol{\alpha}}^*\in\mathcal{D}(Q^*)$. Therefore $\min_{j\in\mathcal{C}_F^1}\tilde{\alpha}_1^* J_{1j}(\gamma_j)\geq\tilde{z}^*$.

*Proof of Claim (ii).* ($\Rightarrow$) Let us suppose that $\boldsymbol{\alpha}^*$ solves C0 and C2, and $\min_{j\in\mathcal{C}_F^1}\tilde{\alpha}_1^* J_{1j}(\gamma_j)<\tilde{z}^*$. Then $\tilde{\boldsymbol{\alpha}}^*\notin\mathcal{D}(Q^*)$, and therefore $\tilde{\boldsymbol{\alpha}}^*\neq\boldsymbol{\alpha}^*$.

($\Leftarrow$) By prior arguments, C0 holds for $\boldsymbol{\alpha}^*$ and $\tilde{\boldsymbol{\alpha}}^*$. Now suppose $\boldsymbol{\alpha}^*\neq\tilde{\boldsymbol{\alpha}}^*$, which implies $\tilde{\boldsymbol{\alpha}}^*\notin\mathcal{D}(Q^*)$. Then it must be the case that $\min_{j\in\mathcal{C}_F^1}\tilde{\alpha}_1^* J_{1j}(\gamma_j)<\tilde{z}^*$. Further, since $\tilde{\boldsymbol{\alpha}}^*$ uniquely solves C0 and C1, $\boldsymbol{\alpha}^*\neq\tilde{\boldsymbol{\alpha}}^*$ implies that C1 does not hold for $\boldsymbol{\alpha}^*$. Therefore when solving problem $Q$, it must be the case that $\lambda_j^1>0$ for at least one $j\in\mathcal{C}_F^1$ in equation (3.15). By the complementary slackness conditions in equation (3.11), $\min_{j\in\mathcal{C}_F^1}\alpha_1^* J_{1j}(\gamma_j)=\tilde{z}^*$, and hence C2 holds for $\boldsymbol{\alpha}^*$. □

Theorem 3.9 implies that, since a solution to problem $Q^*$ always exists, an optimal solution to problem $Q$ can be obtained as the solution to one of the two sets of nonlinear equations C0 and C1 or C0 and C2. The procedure implicit in theorem 3.9 is presented as algorithm 3.2. Theorem 3.9 also assumes that we have at least one system in $\Gamma$. In the event that $\Gamma$ is empty, conditions C0 and C1 may not form a fully-specified system of equations (e.g., $\Gamma$ and $\mathcal{S}_w$ are empty), or may not have a solution. In such a case, C0 and C2 provide the optimal allocation. When the sets $\mathcal{S}_b$ and $\mathcal{S}_w$ are empty but $\Gamma$ is nonempty, theorem 3.9 reduces to the result presented in Glynn and Juneja (2004).

**Algorithm 3.2** Conceptual algorithm to solve for the optimal allocation $\boldsymbol{\alpha}^*$ in the case of general light-tailed distributions and multiple constraints

Solve the nonlinear system C0, C1 to obtain $\tilde{\boldsymbol{\alpha}}^*$ and $\tilde{z}^*$.<br>
**if** $\min_j \tilde{\alpha}_1^* J_{1j}(\gamma_j) \geq \tilde{z}^*$ **then**<br>
    **return** $\boldsymbol{\alpha}^* = \tilde{\boldsymbol{\alpha}}^*$.<br>
**else**<br>
    Solve the nonlinear system C0, C2 to obtain $\boldsymbol{\alpha}^*$.<br>
    **return** $\boldsymbol{\alpha}^*$.<br>
**end if**

### 3.5.1 Additional Mathematical Insight into the Structure of the Optimal Allocation Problem (Problem $Q$)

From the proof of theorem 3.9, recall equation (3.15), which is duplicated here.

$$\frac{\sum_{j \in \mathcal{C}_F^1} \lambda_j^1 J_{1j}(\gamma_j)}{\nu} + \sum_{i \in \Gamma} \frac{I_1(x(\alpha_1^*, \alpha_i^*))}{I_i(x(\alpha_1^*, \alpha_i^*))} + \sum_{i \in \mathcal{S}_w} \frac{I_1(x(\alpha_1^*, \alpha_i^*))}{I_i(x(\alpha_1^*, \alpha_i^*)) + \sum_{j \in \mathcal{C}_I^i} J_{ij}(\gamma_j)} = 1. \tag{3.16}$$

Since $\lambda_j^1 = 1 - \sum_{i=2}^r \lambda_i - \sum_{k \in \mathcal{C}_F^1,\, k \neq j} \lambda_k^1$, then expanding the first term in the prior equation, a bit of algebra yields

$$\frac{\sum_{j \in \mathcal{C}_F^1} \lambda_j J_{1j}(\gamma_j)}{\nu} = \frac{1}{\nu}\Bigg[\sum_{j \in \mathcal{C}_F^1} J_{1j}(\gamma_j) - \sum_{j \in \mathcal{C}_F^1}\Bigg(J_{1j}(\gamma_j) \sum_{\substack{k \in \mathcal{C}_F^1 \\ k \neq j}} \lambda_k\Bigg)\Bigg] - \sum_{j \in \mathcal{C}_F^1} J_{1j}(\gamma_j)\left(\sum_{i=2}^r \frac{\lambda_i}{\nu}\right).$$

Substituting values from the stationarity KKT conditions for $\sum_{i=2}^r \lambda_i/\nu$ above and plugging into equation (3.16), further algebra yields,

$$\begin{aligned}
\nu = {} & \left[1 - \frac{\sum_{j \in \mathcal{C}_F^1}\left(J_{1j}(\gamma_j) \sum_{k \in \mathcal{C}_F^1,\, k \neq j} \lambda_k\right)}{\sum_{j \in \mathcal{C}_F^1} J_{1j}(\gamma_j)}\right] \\
& \times \Bigg[\sum_{i \in \Gamma} \frac{1}{I_i(x(\alpha_1^*, \alpha_i^*))} + \sum_{i \in \mathcal{S}_b} \frac{1}{\sum_{j \in \mathcal{C}_I^i} J_{ij}(\gamma_j)} + \sum_{i \in \mathcal{S}_w} \frac{1}{I_i(x(\alpha_1^*, \alpha_i^*)) + \sum_{j \in \mathcal{C}_I^i} J_{ij}(\gamma_j)} \\
& \qquad + \frac{1}{\sum_{j \in \mathcal{C}_F^1} J_{1j}(\gamma_j)}\Bigg(1 - \sum_{i \in \Gamma} \frac{I_1(x(\alpha_1^*, \alpha_i^*))}{I_i(x(\alpha_1^*, \alpha_i^*))} - \sum_{i \in \mathcal{S}_w} \frac{I_1(x(\alpha_1^*, \alpha_i^*))}{I_i(x(\alpha_1^*, \alpha_i^*)) + \sum_{j \in \mathcal{C}_I^i} J_{ij}(\gamma_j)}\Bigg)\Bigg]^{-1}
\end{aligned}$$

Suppose that all constraints for feasibility of system 1 are non-binding, that is, $\min_j \alpha_1^* J_{1j}(\gamma_j) > z^*$. Then by complimentary slackness, $\lambda_j^1 = 0$ for all $j = 1 \ldots |\mathcal{C}_F^1|$ and the numerator of $\nu$ equals 1. If condition C2 holds such that at least one constraint for the feasibility of system 1 is binding, no particular insight is gained by further simplifying the expression for $\nu$.

Further insight into the optimal rate of decay of the probability of false selection, $z^*$, can be obtained as follows. From condition C0, dividing through by $\alpha_1^*$ and the appropriate rate functions, we have,

$$\frac{z^*}{\alpha_1^*}\left(\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*))}\right) = \frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*))} + \frac{\alpha_i^*}{\alpha_1^*} \text{ for all } i \in \Gamma \tag{3.17}$$

$$\frac{z^*}{\alpha_1^*}\left(\frac{1}{\sum_{j\in\mathcal{C}_I^i} J_{ij}(\gamma_j)}\right) = \frac{\alpha_i^*}{\alpha_1^*} \text{ for all } i \in \mathcal{S}_b \tag{3.18}$$

$$\frac{z^*}{\alpha_1^*}\left(\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*)) + \sum_{j\in\mathcal{C}_I^i} J_{ij}(\gamma_j)}\right) = \frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*)) + \sum_{j\in\mathcal{C}_I^i} J_{ij}(\gamma_j)} + \frac{\alpha_i^*}{\alpha_1^*} \text{ for all } i \in \mathcal{S}_w \tag{3.19}$$

Summing up the left hand side and the right hand side of equations (3.17)–(3.19), noticing that

$$\sum_{i\in\Gamma\cup\mathcal{S}_b\cup\mathcal{S}_w} \frac{\alpha_i^*}{\alpha_1^*} = \frac{1}{\alpha_1^*} - 1,$$

and rearranging terms,

$$z^* = \frac{1-\alpha_1^*\left(1-\sum_{i\in\Gamma}\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*))} - \sum_{i\in\mathcal{S}_w}\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*)) + \sum_{j\in\mathcal{C}_I^i} J_{ij}(\gamma_j)}\right)}{\left(\sum_{i\in\Gamma}\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*))} + \sum_{i\in\mathcal{S}_b}\frac{1}{\sum_{j\in\mathcal{C}_I^i} J_{ij}(\gamma_j)} + \sum_{i\in\mathcal{S}_w}\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*)) + \sum_{j\in\mathcal{C}_I^i} J_{ij}(\gamma_j)}\right)}.$$

If C0 and C1 hold, then

$$z^* = \left(\sum_{i\in\Gamma}\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*))} + \sum_{i\in\mathcal{S}_b}\frac{1}{\sum_{j\in\mathcal{C}_I^i} J_{ij}(\gamma_j)} + \sum_{i\in\mathcal{S}_w}\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*)) + \sum_{j\in\mathcal{C}_I^i} J_{ij}(\gamma_j)}\right)^{-1}.$$

If C0 and C2 hold, then $\alpha_1^* \min_j J_{1j}(\gamma_j) = z^*$, and hence

$$z^* = \left[\sum_{i\in\Gamma}\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*))} + \sum_{i\in\mathcal{S}_b}\frac{1}{\sum_{j\in\mathcal{C}_I^i} J_{ij}(\gamma_j)} + \sum_{i\in\mathcal{S}_w}\frac{1}{I_i(x(\alpha_1^*,\alpha_i^*)) + \sum_{j\in\mathcal{C}_I^i} J_{ij}(\gamma_j)} \right.$$
$$\left. + \frac{1}{\min_j J_{1j}(\gamma_j)}\left(1-\sum_{i\in\Gamma}\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*))} - \sum_{i\in\mathcal{S}_w}\frac{I_1(x(\alpha_1^*,\alpha_i^*))}{I_i(x(\alpha_1^*,\alpha_i^*)) + \sum_{j\in\mathcal{C}_I^i} J_{ij}(\gamma_j)}\right)\right]^{-1}.$$

## 3.6 Consistency and Implementation

In practice, the rate functions in algorithm 3.2 are unavailable and must be estimated. Therefore with a view toward implementation, we address consistency of estimators in this section. Specifically, we first show that the important sets, $\{1\}, \Gamma, \mathcal{S}_b, \mathcal{S}_w, \mathcal{C}_F^i$ and $\mathcal{C}_I^i$, can be estimated consistently, that is, they can be identified correctly as simulation effort tends to infinity. Next, we demonstrate that the optimal allocation estimator, identified by using estimated rate functions in algorithm 3.2, is a consistent estimator of the true optimal allocation $\boldsymbol{\alpha}^*$. These generic consistency results inspire the sequential algorithm presented in section 3.6.2, which is easily implementable at least in contexts where the distribution families underlying the rate functions are known or assumed.

### 3.6.1 Generic Consistency Results

To simplify notation, let each system be allocated $m$ samples, where we explicitly denote the dependence of the estimators on $m$ in this section. Suppose we have at our disposal consistent estimators $\hat{I}_i^m(x), \hat{J}_{ij}^m(y), i \leq r, j \leq s$ of the corresponding rate functions $I_i(x), J_{ij}(y), i \leq r, j \leq s$. Such consistent estimators are easy to construct when the distributional families underlying the true rate functions $I_i(x), J_{ij}(y), i \leq r, j \leq s$ are known or assumed. For example, suppose $H_{ik},\ k = 1, 2, \ldots, m$ are simulation observations of the objective function of the $i$th system, assumed to be resulting from a normal distribution with unknown mean $h_i$ and unknown variance $\sigma_{h_i}^2$. The obvious consistent estimator for the rate function $I_i(x) = \frac{(x-h_i)^2}{2\sigma_{h_i}^2}$ is then $\hat{I}_i^m(x) = \frac{(x-\bar{H}_i)^2}{2\hat{\sigma}_i^2}$, where $\bar{H}_i$ and $\hat{\sigma}_{h_i}$ are the sample mean and sample standard deviation of $H_{ik}, k = 1, 2, \ldots, m$ respectively. In the more general case where the distributional family is unknown or not assumed, the rate function may be estimated as the Legendre-Fenchel transform (see, e.g., Dembo and Zeitouni, 1998, p. 26) of the cumulant generating function estimator

$$\hat{I}_i^m(x) = \sup_\theta(\theta x - \hat{\Lambda}_{H_i}^m(\theta)), \tag{3.20}$$

where $\hat{\Lambda}_{H_i}^m(\theta) = \log m^{-1} \sum_{k=1}^m \exp(\theta H_{ik})$. In what follows, to preserve generality, the discussion pertains to estimators of the type displayed in equation (3.20). By arguments analogous to those in Glynn and Juneja (2004) and under our assumptions, the estimator in equation (3.20) is consistent.

Let $(\bar{H}_i(m),\ \bar{G}_{i1}(m),\ \ldots,\ \bar{G}_{is}(m)) = \left(\frac{1}{m}\sum_{k=1}^m H_{ik},\ \frac{1}{m}\sum_{k=1}^m G_{i1k},\ \ldots,\ \frac{1}{m}\sum_{k=1}^m G_{isk}\right)$ denote the estimators of $(h_i, g_{i1}, \ldots, g_{is})$. We also define the following set estimators.

$\hat{1}(m) := \arg\min_i \{\bar{H}_i(m) : \bar{G}_{ij}(m) \leq \gamma_j$ for all $j \leq s\}$ is the estimated best feasible system;

$\widehat{\Gamma}(m) := \{i : \bar{G}_{ij}(m) \leq \gamma_j \text{ for all } j \leq s, i \neq \hat{1}(m)\}$ is the estimated set of suboptimal feasible systems;

$\widehat{\mathcal{S}}_b(m) := \{i : \bar{H}_{\hat{1}(m)}(m) \geq \bar{H}_i(m) \text{ and } \bar{G}_{ij}(m) > \gamma_j \text{ for some } j \leq s\}$ is the estimated set of infeasible, better systems;

$\widehat{\mathcal{S}}_w(m) := \{i : \bar{H}_{\hat{1}(m)}(m) < \bar{H}_i(m) \text{ and } \bar{G}_{ij}(m) > \gamma_j \text{ for some } j \leq s\}$ is the estimated set of infeasible, worse systems;

$\widehat{\mathcal{C}}^i_F(m) := \{j : \bar{G}_{ij}(m) \leq \gamma_j\}$ is the set of constraints on which system $i$ is estimated feasible;

$\widehat{\mathcal{C}}^i_I(m) := \{j : \bar{G}_{ij}(m) > \gamma_j\}$ is the set of constraints on which system $i$ is estimated infeasible.

Notice $\bar{\Gamma}$ (defined in section 3.4) excludes system 1 while $\hat{\Gamma}(m)$ excludes the *estimated* system 1.

Since assumption 3.2 implies $\bar{H}_i(m) \to h_i$ wp1 and $\bar{G}_{ij}(m) \to g_{ij}$ wp1 for all $i \leq r$ and $j \leq s$, and the numbers of systems and constraints are finite, all estimated sets converge to their true counterparts wp1 as $m \to \infty$. (See section 1.5 for a rigorous definition of the convergence of sets.) Proposition 3.10 formally states this result.

**Proposition 3.10.** *Under assumption 3.2, system $\hat{1}(m) \to$ system 1, $\hat{\Gamma}(m) \to \Gamma$, $\widehat{\mathcal{S}}_b(m) \to \mathcal{S}_b$, $\widehat{\mathcal{S}}_w(m) \to \mathcal{S}_w$, $\widehat{\mathcal{C}}^i_F(m) \to \mathcal{C}^i_F$, and $\widehat{\mathcal{C}}^i_I(m) \to \mathcal{C}^i_I$ wp1 as $m \to \infty$.*

*Proof.* We only prove that $\widehat{\mathcal{C}}^i_F \to \mathcal{C}^i_F$ wp1 as $m \to \infty$. The proofs for the other parts of the theorem follow in a very similar fashion.

By assumption 3.2, $\bar{G}_{ij}(m) \to g_{ij}$ wp1 for all $i \leq r$ and $j \leq s$. By definition, $g_{ij} < \gamma_j$ for each $j \in \mathcal{C}^i_F$. Since $|\mathcal{C}^i_F| < \infty$, for large enough $m$, $\bar{G}_{ij}(m) < \gamma_j$ uniformly in $j \in \mathcal{C}^i_F$ wp1, and hence the assertion holds. □

Let $\hat{\boldsymbol{\alpha}}^*(m)$ denote the estimator of the optimal allocation vector $\boldsymbol{\alpha}^*$ obtained by replacing the rate functions $I_i(x), J_{ij}(x), i \leq r, j \leq s$ appearing in conditions C0, C1, and C2 with their corresponding estimators $\hat{I}^m_i(x), \hat{J}^m_{ij}(x), i \leq r, j \leq s$ obtained through sampling, and then using algorithm 3.2. Since the search space $\{\boldsymbol{\alpha} : \alpha_i \geq 0, \sum_{i=1}^r \alpha_i = 1\}$ is a compact set, and the estimated (consistent) rate functions can be shown to converge uniformly over the search space, it is no surprise that $\hat{\boldsymbol{\alpha}}^*(m)$ converges to the optimal allocation vector $\boldsymbol{\alpha}^*$ as $m \to \infty$ wp1. Theorem 3.13 formally asserts this result, with a proof that is a direct application of results found in the stochastic root-finding literature (see, e.g., Pasupathy and Kim, 2011, theorem 5.7).

Before the statement of theorem 3.13, two additional lemmas are required. The proof of the following lemma 3.11 follows closely along the lines of the proofs in Glynn and Juneja (2004).

**Lemma 3.11.** *Suppose assumption 3.3 holds. Then there exists* $\epsilon > 0$ *such that* $\hat{I}_i^m(x) \to I_i(x)$ *as* $m \to \infty$ *uniformly in* $x \in [h_\ell - \epsilon, h_u + \epsilon]$ *wp1, for all* $i \in \{1\} \cup \Gamma \cup \mathcal{S}_w$.

*Proof.* By assumption 3.3, $[h_\ell, h_u] \subset \cap_{i=1}^r \mathcal{F}_{H_i}^\circ$, recalling that $\mathcal{F}_{H_i}^\circ = \text{int}\{\Lambda'_{H_i}(\theta) : \theta \in \mathcal{D}_{\Lambda_{H_i}}^\circ\}$. Then there exists an $\epsilon > 0$ such that $[h_\ell - \epsilon, h_u + \epsilon] \subset \cap_{i=1}^r \mathcal{F}_{H_i}^\circ$. Glynn and Juneja (2004) show that $\hat{I}_i(x) \to I_i(x)$ pointwise wp1 on $[h_\ell, h_u]$, however by the exact same arguments, we may show that $\hat{I}_i(x) \to I_i(x)$ pointwise wp1 on $[h_\ell - \epsilon, h_u + \epsilon]$.

Suppose that $\hat{I}_i^m(x)$ does not converge uniformly in $x \in [h_\ell - \epsilon, h_u + \epsilon]$ to $I_i(x)$ as $m \to \infty$ wp1. Then for some $\epsilon_0 > 0$, there exists a sequence $\{x_k\} \to x^*$, $\{x_k\}$ and $x^*$ in $[h_\ell - \epsilon, h_u + \epsilon]$, and a subsequence $\hat{I}_{i,m_k}(x)$ of $\hat{I}_i(x)$ such that $|\hat{I}_{i,m_k}(x_k) - I_i(x_k)| > \epsilon_0$ for all $k$. This statement will be contradicted.

Consider

$$\begin{aligned} |\hat{I}_{i,m_k}(x_k) - I_i(x_k)| &= |\hat{I}_{i,m_k}(x_k) - \hat{I}_{i,m_k}(x^*) + \hat{I}_{i,m_k}(x^*) - I_i(x^*) + I_i(x^*) - I_i(x_k)| \\ &\leq |\hat{I}_{i,m_k}(x_k) - \hat{I}_{i,m_k}(x^*)| + |\hat{I}_{i,m_k}(x^*) - I_i(x^*)| + |I_i(x_k) - I_i(x^*)|. \qquad (3.21) \end{aligned}$$

From the last term in equation (3.21), by the continuity of $I_i(\cdot)$ at $x^*$, $I_i(x_k) \to I_i(x^*)$. Therefore there exists $K_a(\epsilon_0/6)$ such that for all $k \geq K_a(\epsilon_0/6)$, $|I_i(x_k) - I_i(x^*)| < \epsilon_0/6$.

From the middle term in equation (3.21), by the pointwise convergence of $\hat{I}_{i,m_k}(x^*)$ to $I_i(x^*)$, there exists $M_a(\epsilon_0/6)$ such that for all $m_k \geq M_a(\epsilon_0/6)$, $|\hat{I}_{i,m_k}(x^*) - I_i(x^*)| < \epsilon_0/6$.

From the first term in equation (3.21), since $\hat{I}_{i,m_k}(\cdot)$ is convex for all $m$,

$$|\hat{I}_{i,m_k}(x_k) - \hat{I}_{i,m_k}(x^*)| \leq \max\{|\hat{I}'_{i,m_k}(h_1 - \epsilon)|, |\hat{I}'_{i,m_k}(h_r + \epsilon)|\}|x_k - x^*|. \qquad (3.22)$$

By arguments analogous to those in Glynn and Juneja (2004), $\hat{I}'_i(x) \to I'_i(x)$ pointwise in $x$ wp1 for $x \in [h_\ell - \epsilon, h_u + \epsilon]$. Therefore there exists $M_{b_1}(\epsilon_1)$ and $M_{b_2}(\epsilon_1)$ such that wp1, for all $m_k > \max\{M_{b_1}(\epsilon_1), M_{b_2}(\epsilon_1)\}$, $|\hat{I}'_{i,m_k}(h_r - \epsilon) - I'_{i,m_k}(h_r - \epsilon)| < \epsilon_1$ and $|\hat{I}'_{i,m_k}(h_r + \epsilon) - I'_{i,m_k}(h_r + \epsilon)| < \epsilon_1$. Since $[h_1 - \epsilon, h_r + \epsilon]$ is a compact set in $\cap_{i=1}^r \mathcal{F}_{H_i}^\circ$, then $I'_i(x)$ is bounded. Hence for all $m_k > \max\{M_{b_1}(\epsilon_1), M_{b_2}(\epsilon_1)\}$, $|\hat{I}'_{i,m_k}(h_r - \epsilon)|$ and $|\hat{I}'_{i,m_k}(h_r + \epsilon)|$ are uniformly bounded wp1; let this bound be $B$. Since $\{x_k\} \to x^*$, there exists $K_b(\epsilon_0/6B)$ such that for all $k \geq K_b(\epsilon_0/6B)$, $|x_k - x| < \epsilon_0/6B$. Then from equation (3.22), for all $k \geq K_b(\epsilon_0/6B)$ and $m_k > \max\{M_{b_1}(\epsilon_1), M_{b_2}(\epsilon_1)\}$,

$$\max\{|\hat{I}'_{i,m_k}(h_1 - \epsilon)|, |\hat{I}'_{i,m_k}(h_r + \epsilon)|\}|x_k - x| \leq B|x_k - x| < \frac{\epsilon_0}{6}.$$

Now let $k > \max\{K_a(\epsilon_0/6), K_b(\epsilon_0/6B)\}$ and $m_k > \max\{M_a(\epsilon_0/6), M_{b_1}(\epsilon_1), M_{b_2}(\epsilon_1)\}$. Then from equation (3.21),

$$|\hat{I}_{i,m_k}(x_k) - I_i(x_k)| \le \frac{\epsilon_0}{6} + \frac{\epsilon_0}{6} + \frac{\epsilon_0}{6} = \frac{\epsilon_0}{2},$$

and the conclusion follows by contradiction. $\square$

**Lemma 3.12.** *Let the system of equations C0 and C1 be denoted $f_1(\boldsymbol{\alpha}) = 0$, and let the system of equations C0 and C2 be denoted by $f_2(\boldsymbol{\alpha}) = 0$, where $f_1$ and $f_2$ are vector-valued functions with compact support $\sum_{i=1}^{r} \alpha_i = 1, \boldsymbol{\alpha} \ge 0$. Let the estimators $\hat{F}_1^m(\boldsymbol{\alpha})$ and $\hat{F}_2^m(\boldsymbol{\alpha})$ be the same set of equations as $f_1(\boldsymbol{\alpha})$ and $f_2(\boldsymbol{\alpha})$, respectively, except with all unknown rate functions replaced by their corresponding estimated quantities. If assumption 3.3 holds, then the functional sequences $\hat{F}_1^m(\boldsymbol{\alpha}) \to f_1(\boldsymbol{\alpha})$ and $\hat{F}_2^m(\boldsymbol{\alpha}) \to f_2(\boldsymbol{\alpha})$ uniformly in $\boldsymbol{\alpha}$ as $m \to \infty$ wp1.*

*Proof.* The proof of the theorem proceeds in two steps. We first show that $\alpha_1 \hat{I}_1^m(\hat{x}_m(\alpha_1, \alpha_i) + \alpha_i \hat{I}_i^m(\hat{x}_m(\alpha_1, \alpha_i))$ converges uniformly in $\boldsymbol{\alpha}$ as $m \to \infty$ wp1 for all $i \in \Gamma \cup \mathcal{S}_w$, where $\hat{x}_m(\alpha_1, \alpha_i) = \arg\inf_x(\alpha_1 \hat{I}_1^m(x) + \alpha_i \hat{I}_i^m(x))$. Next we show that the quantities $\alpha_i \sum_{j \in \mathcal{C}_I^i} \hat{J}_{ij}^m(\gamma_j), i \in \mathcal{S}_b \cup \mathcal{S}_w, j \le s$ and $\alpha_1 \hat{J}_{1j}^m(\gamma_j), j \in \mathcal{C}_F^1$ converge uniformly in $\boldsymbol{\alpha}$ as $m \to \infty$ wp1. These assertions, together with the observation that we search only in the set $\{\boldsymbol{\alpha} : \sum_{i=1}^{r} \alpha_i = 1, \alpha_i > 0\}$ and hence $I_i(x(\alpha_1, \alpha_i)) > \delta > 0$, which implies for large enough $m$, $\hat{I}_i^m(\hat{x}_m(\alpha_1, \alpha_i)) > \delta$, proves the theorem.

By lemma 3.11, $\hat{I}_i^m(x) \to I_i(x)$ uniformly in $x$ on $[h_\ell - \epsilon, h_u + \epsilon]$ wp1 for some $\epsilon > 0$. By Glynn and Juneja (2004), $\hat{x}_m(\alpha_1, \alpha_i) \to x(\alpha_1, \alpha_i)$ wp1, where $x(\alpha_1, \alpha_i) = \arg\inf_x(\alpha_1 I_1(x) + \alpha_i I_i(x)) \in [h_\ell, h_u]$. Therefore for $m$ large enough and for all feasible $\alpha_1, \alpha_i$, we have $\hat{x}_m(\alpha_1, \alpha_i) \in [h_\ell - \epsilon/2, h_u + \epsilon/2]$ wp1 for all $i \in \{1\} \cup \Gamma \cup \mathcal{S}_w$. It then follows that $\alpha_1 \hat{I}_1^m(\hat{x}_m(\alpha_1, \alpha_i) + \alpha_i \hat{I}_i^m(\hat{x}_m(\alpha_1, \alpha_i))$ converges uniformly in $\boldsymbol{\alpha}$ as $m \to \infty$ wp1, for all $i \in \Gamma \cup \mathcal{S}_w$.

Under assumption 3.3, it follows from analogous arguments to those in Glynn and Juneja (2004) that $\hat{J}_{ij}^m(\gamma_j) \to J_{ij}(\gamma_j)$ as $m \to \infty$ wp1, for all $i \in \mathcal{S}_b \cup \mathcal{S}_w$ and $j \le s$. Therefore the terms $\alpha_i \sum_{j \in \mathcal{C}_I^i} \hat{J}_{ij}^m(\gamma_j)$ converge uniformly in $\boldsymbol{\alpha}$ as $m \to \infty$ wp1. Likewise, for all $j \in \mathcal{C}_F^1$, $\alpha_1 \hat{J}_{1j}^m(\gamma_j)$ converges uniformly in $\boldsymbol{\alpha}$ as $m \to \infty$ wp1. $\square$

**Theorem 3.13.** *Let the postulates of lemma 3.12 hold, and assume $\Gamma$ is nonempty. Then the empirical estimate of the optimal allocation is consistent, that is, $\hat{\boldsymbol{\alpha}}^*(m) \to \boldsymbol{\alpha}^*$ as $m \to \infty$ wp1.*

*Proof.* As argued previously, $f_1(\boldsymbol{\alpha})$ and $f_2(\boldsymbol{\alpha})$ are continuous functions of $\boldsymbol{\alpha}$ on a compact set. Further, the solutions $f_1(\boldsymbol{\alpha}) = 0$ and $f_2(\boldsymbol{\alpha}) = 0$ exist. If each rate function in problem $Q$ is replaced with estimated rate functions, these new problems remain continuous, concave maximization problems on a compact set, which attain their maximums. Therefore the systems $\hat{F}_1^m(\boldsymbol{\alpha}) = 0$ and $\hat{F}_2^m(\boldsymbol{\alpha}) = 0$ have a solution for large enough $m$ wp1. By lemma 3.12, it also holds that $\hat{F}_1^m(\boldsymbol{\alpha}) \to f_1(\boldsymbol{\alpha})$ and $\hat{F}_2^m(\boldsymbol{\alpha}) \to f_2(\boldsymbol{\alpha})$ uniformly in $\boldsymbol{\alpha}$ as $m \to \infty$ wp1. Thus all the requirements are satisfied for convergence of the sample-path solution $\hat{\boldsymbol{\alpha}}^*(m)$ to its true counterpart $\boldsymbol{\alpha}^*$ as $m \to \infty$ wp1 (see Pasupathy and Kim, 2011, theorem 5.7). $\square$

### 3.6.2 A Sequential Algorithm for Implementation

This section concludes with a sequential algorithm that naturally stems from the conceptual algorithm (algorithm 3.2) outlined in section 3.5 and the consistent estimator that was discussed in the previous section. Algorithm 3.3 formally outlines this procedure, where $n$ denotes the total simulation budget, and $n_i$ denotes the total sample expended at system $i$.

**Algorithm 3.3** Sequential algorithm for the case of general light-tailed distributions with multiple constraints

**Require:** Number of pilot samples $\delta_0 > 0$; number of samples between allocation vector updates $\delta > 0$; and a minimum-sample vector $\boldsymbol{\varepsilon} > 0$.

1: Initialize: collect $\delta_0$ samples from each system $i \leq r$.
2: Initialize: $n = r\delta_0, n_i = \delta_0$. {Initialize total simulation effort and effort for each system.}
3: Update rate function estimators $\hat{I}_i^{n_i}(x), \hat{J}_{ij}^{n_i}(x), i \leq r, j \leq s$.
4: **if** no systems are estimated feasible **then**
5:    Set $\hat{\boldsymbol{\alpha}}^*(n) = (1/r, 1/r, \ldots, 1/r)$.
6: **else**
7:    Solve the system C0, C1 using the updated rate function estimators to obtain $\widehat{\tilde{\boldsymbol{\alpha}}}^*(n)$ and $\widehat{\tilde{z}}^*(n)$.
8:    **if** $\min_j \widehat{\tilde{\alpha}}_1^*(n)\hat{J}_{1j}^{n_1}(\gamma_j) \geq \widehat{\tilde{z}}^*(n)$ **then**
9:       $\hat{\boldsymbol{\alpha}}^*(n) = \widehat{\tilde{\boldsymbol{\alpha}}}^*(n)$.
10:    **else**
11:       Solve the system C0, C2 using the updated rate function estimators to obtain $\hat{\boldsymbol{\alpha}}^*(n)$.
12:    **end if**
13: **end if**
14: Collect one sample at each of the systems $X_k, k = 1, 2, \ldots, \delta$, where the $X_k$'s are iid random variates having probability mass function $\hat{\boldsymbol{\alpha}}^*(n)$ and support $\{1, 2, \ldots, r\}$, and update $n_{X_k} = n_{X_k} + 1$.
15: Set $n = n + \delta$ update $\bar{\boldsymbol{\alpha}}(n) = \{n_1/n, n_2/n, \ldots, n_r/n\}$.
16: **if** $\bar{\boldsymbol{\alpha}}(n) > \boldsymbol{\varepsilon}$ **then**
17:    Set $\delta^+ = 0$.
18: **else**
19:    Collect one sample from each system in the set of systems receiving insufficient sample $\mathcal{I}_n$.
20:    Update $n_i = n_i + 1$ for all $i \in \mathcal{I}_n$. Let $\delta^+ = |\mathcal{I}_n|$.
21: **end if**
22: Set $n = n + \delta^+$ and go to step 3.

The essential idea in algorithm 3.3 is straightforward. At the end of each iteration, the optimal allocation vector is estimated using rate function estimators constructed from samples already gathered from the various systems. Systems are chosen for sampling at the subsequent iteration by using the estimated optimal allocation vector as the sampling distribution. Since the true optimal allocation $\boldsymbol{\alpha}^* > 0$, it follows that the sequential algorithm should sample from each system infinitely often. To ensure systems with small allocations continue to be sampled, we assume knowledge of an "indifference zone" vector $\boldsymbol{\varepsilon} > 0$ such that if the actual proportion of sample expended at each system in algorithm 5.4, defined as $\bar{\boldsymbol{\alpha}}(n) = \{n_1/n, n_2/n, \ldots, n_r/n\}$, falls below $\boldsymbol{\varepsilon}$, we sample once from each system receiving insufficient sample. All elements of $\boldsymbol{\varepsilon}$ should be "small" relative to $1/r$. Therefore the algorithm guarantees that we spend a minimal amount of sample at each system in the limit.

In a context where the distributional family underlying the simulation observations is known or assumed, the rate function estimators should be estimated in step 3 accordingly — by simply estimating the distributional parameters appearing within the expression for the rate function. Also, algorithm 3.3 provides flexibility on how often the optimal allocation vector is re-estimated through the algorithm parameter $\delta$. The choice of the parameter $\delta$ will depend on the particular problem, and specifically, on how expensive the simulation execution is relative to solving the nonlinear systems in steps 4 and 7. Lastly, as is clear from the algorithm listing, algorithm 3.3 relies on fully sequential and simultaneous observation of the objective and constraint functions. Deviation from these assumptions, while interesting, renders the present context inapplicable.

## 3.7 Comparison with OCBA-CO

Now consider the case in which the random variables corresponding to both the objective and constraint have normal distributions. The relevant rate functions for the normal case are presented in section 2.4.2. In what follows, the actual rate functions governing the simulation estimators have been used for analysis instead of the sequential estimator outlined in algorithm 3.3. This choice helps display the asymptotic allocation proposed by theory and highlights its deviation from the asymptotic solution proposed by OCBA-CO without introducing simulation noise. Owing to their routine nature, results from numerical tests demonstrating that the sequential estimator in algorithm 3.3 indeed converges to the optimal allocation vector identified by theory have been omitted.

### 3.7.1 Theoretical Comparison

Lee et al. (2011) describe an OCBA framework for an asymptotic simulation budget allocation for constrained simulation optimization on finite sets (OCBA-CO). The work by Lee et al. (2011)

provides the only other asymptotic sample allocation result for constrained simulation optimization on finite sets in the literature.

For suboptimal systems, Lee et al. (2011) divide the systems into a "feasibility dominance" set and an "optimality dominance" set. Formally, these sets are defined as

$\mathcal{S}_F$ : the feasibility dominance set, $\mathcal{S}_F = \{i : P\{\hat{G}_i \leq \gamma\} < P\{\hat{H}_1 > \hat{H}_i\}, i \neq 1\}$,
$\mathcal{S}_O$ : the optimality dominance set, $\mathcal{S}_O = \{i : P\{\hat{G}_i \leq \gamma\} \geq P\{\hat{H}_1 > \hat{H}_i\}, i \neq 1\}$.

Lee et al. (2011) make the assumption $\alpha_1 \gg \alpha_{i\in\mathcal{S}_O}$ and then note that systems in $\mathcal{S}_w$ are in set $\mathcal{S}_O$ if $\left(\frac{h_1-h_i}{\sigma_{h_i}}\right)^2 \geq \left(\frac{\gamma-g_i}{\sigma_{g_i}}\right)^2$, and in set $\mathcal{S}_F$ otherwise. The assumption $\alpha_1 \gg \alpha_{i\in\mathcal{S}_O}$, along with an approximation to the probability of correct selection, allows Lee et al. (2011) to solve explicitly for each $\alpha_i$. This closed-form approximation provided by Lee et al. (2011) can be written as, for systems $i = 2, \ldots, r$,

$$\begin{aligned}\alpha_i &= \frac{\left(\frac{\sigma_{h_i}}{h_1-h_i}\right)^2 \mathbb{I}_{i\in\mathcal{S}_O} + \left(\frac{\sigma_{g_i}}{\gamma-g_i}\right)^2 \mathbb{I}_{i\in\{1\}\cup\mathcal{S}_F}}{\sum_{i\in\mathcal{S}_O}\left(\frac{\sigma_{h_i}}{h_1-h_i}\right)^2 + \sum_{i\in\{1\}\cup\mathcal{S}_F}\left(\frac{\sigma_{g_i}}{\gamma-g_i}\right)^2} \quad \text{if } \frac{(\gamma-g_1)^2}{2\sigma^2_{g_1}} \leq \frac{1}{2\sigma_{h_1}\sqrt{\sum_{i\in\mathcal{S}_O}\sigma^2_{h_i}/(h_1-h_i)^4}}, \\ &= \frac{\left(\frac{\sigma_{h_i}}{h_1-h_i}\right)^2 \mathbb{I}_{i\in\mathcal{S}_O} + \left(\frac{\sigma_{g_i}}{\gamma-g_i}\right)^2 \mathbb{I}_{i\in\mathcal{S}_F}}{\sum_{i\in\mathcal{S}_O}\left(\frac{\sigma_{h_i}}{h_1-h_i}\right)^2 + \sum_{i\in\mathcal{S}_F}\left(\frac{\sigma_{g_i}}{\gamma-g_i}\right)^2 + \sigma_{h_1}\sqrt{\sum_{i\in\mathcal{S}_O}\sigma^2_{h_i}/(h_1-h_i)^4}} \quad \text{otherwise.}\end{aligned} \tag{3.23}$$

The allocation for system 1 is equal to the allocation given in equation (3.23), provided the condition in equation (3.23) holds. If this condition does not hold, then $\alpha_1 = \sigma_{h_1}\sqrt{\sum_{i\in\mathcal{S}_O}(\alpha_i^2/\sigma^2_{h_i})}$.

Like the LD-based allocation discussed in this paper, if rate of decay of the probability of false selection due to the infeasibility of system 1 is "fast," as defined by the criteria in equation (3.23), Lee et al. (2011) allocate sample to system 1 in a way that is similar to the unconstrained case (see Chen et al., 2000). Otherwise, the allocation to system 1 is calculated in a way that is similar to the allocation of other systems.

Particular insight into the allocation provided by Lee et al. (2011) is given by writing the proportional allocation for suboptimal sets under OCBA-CO as,

$$\frac{\alpha_i}{\alpha_k} = \frac{\left(\frac{h_1-h_k}{\sigma_{h_k}}\right)^2 \mathbb{I}_{k\in\mathcal{S}_O} + \left(\frac{\gamma-g_k}{\sigma_{g_k}}\right)^2 \mathbb{I}_{k\in\mathcal{S}_F}}{\left(\frac{h_1-h_i}{\sigma_{h_i}}\right)^2 \mathbb{I}_{i\in\mathcal{S}_O} + \left(\frac{\gamma-g_i}{\sigma_{g_i}}\right)^2 \mathbb{I}_{i\in\mathcal{S}_F}} \quad \text{for all } i, k = 2, \ldots, r. \tag{3.24}$$

The allocation is written in this format for comparison with the corresponding expression recommended by the proposed allocation for $\alpha_1 \gg \alpha_i^*$, which is given in section 2.4.2 as,

$$\frac{\alpha_i}{\alpha_k} \approx \frac{\left(\dfrac{h_1 - h_k}{\sigma_{h_k}}\right)^2 \mathbb{I}_{k \in \Gamma \cup \mathcal{S}_w} + \left(\dfrac{\gamma - g_k}{\sigma_{g_k}}\right)^2 \mathbb{I}_{k \in \mathcal{S}_b \cup \mathcal{S}_w}}{\left(\dfrac{h_1 - h_i}{\sigma_{h_i}}\right)^2 \mathbb{I}_{i \in \Gamma \cup \mathcal{S}_w} + \left(\dfrac{\gamma - g_i}{\sigma_{g_i}}\right)^2 \mathbb{I}_{i \in \mathcal{S}_b \cup \mathcal{S}_w}}, \tag{3.25}$$

where equality holds for all $i, k \in \mathcal{S}_b$.

As can be seen from equation (3.24), in OCBA-CO, only one term in each of the numerator and denominator is active at a time. This artifact of the set definitions and the assumptions used in Lee et al. (2011) can sometimes lead to severely suboptimal allocations for infeasible and worse systems. The example we present next is designed to highlight this issue and the consequent inefficiency incurred in the form of a decreased convergence rate of false selection.

### 3.7.2 Numerical Comparison

The following numerical examples demonstrate the differences between the proposed allocation and the allocation provided by OCBA-CO (Lee et al., 2011).

**Example 3.14.** Suppose there are two systems and one constraint such that each $H_i$ and $G_i$ are normally distributed. Let the means and variances be as given in table 3.1, and let $\gamma = 0$.

Table 3.1: Means and variances for numerical example 3.14 illustrating the differences in optimal allocation between the proposed allocation and OCBA-CO (Lee et al., 2011)

| System ($i$) | $h_i$ | $\sigma^2_{h_i}$ | $g_i$ | $\sigma^2_{g_i}$ |
|---|---|---|---|---|
| 1 | 0 | 2.0 | $-10.0$ | 1.0 |
| 2 | 2.0 | 1.0 | $g_2 \in (0, 1.9]$ | 1.0 |

Note the following features of this example: (i) Since system 2 belongs to $\mathcal{S}_O$ for large enough $n$ and $g_2 \in (0, 1.9]$, the OCBA-CO allocation to system 2 does not depend on $g_2$; (ii) For all values of $g_2$, system 2 is an element of $\mathcal{S}_w$, and hence the proposed allocation *will* change as a function of $g_2$; (iii) System 1 is decidedly feasible ($g_1 = -10$ and $\sigma_{g_1} = 1$), and so does not require much sample for detecting its feasibility. Solving for the optimal allocation as a function of $g_2$ yields the allocations displayed in figure 3.1 and the overall rate of decay of $P\{FS\}$ displayed in figure 3.2.

From the proposed optimal allocation in figure 3.1, it can be seen that the allocation to system 2 should not remain constant as a function of $g_1$, as proposed by Lee et al. (2011). In fact, for certain values of $g_1$, nearly all the sample should be allocated to system 2. □

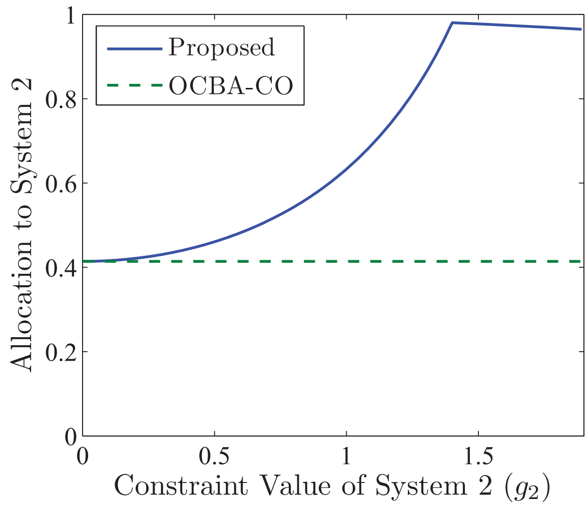


Figure 3.1: The graph of the constraint value of the second system ($g_2$) versus optimal allocation ($\boldsymbol{\alpha}^*$) for the systems in example 3.14 shows the differences between the proposed allocation and OCBA-CO.

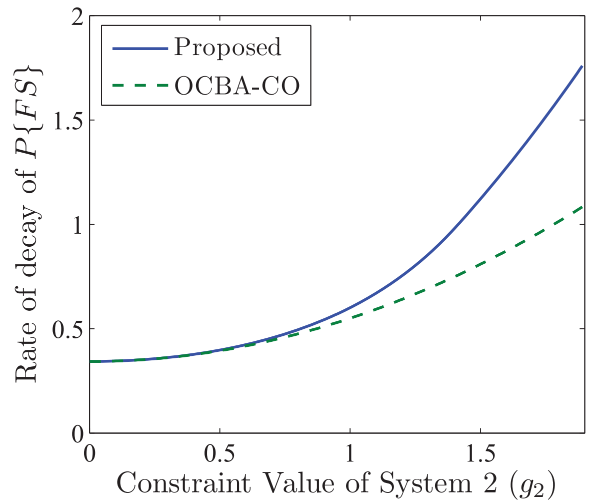


Figure 3.2: The graph of the constraint value of the second system ($g_2$) versus the rate of decay of the probability of false selection ($P\{FS\}$) for the systems in example 3.14 shows the differences in the rate of decay of $P\{FS\}$ for the proposed allocation and OCBA-CO.

Now suppose the constraint value for system $g_2$ is fixed, and explore the allocation to system 1 as a function of $\sigma_{h_1}$. As a result of the $\alpha_1 \gg \alpha_i$ assumption, the OCBA-CO allocation to $\alpha_1$ increases as a function $\sigma_{h_1}$, the variance of the objective value of system 1. The next example in this section is designed to show how this allocation policy can be severely suboptimal.

**Example 3.15.** Retain the two systems and their values from example 3.14, except fix $g_2 = 1.6$, and vary $\sigma^2_{h_1}$ in the interval $[0.2, 4]$. Table 3.2 displays the means and variances for this example.

Table 3.2: Means and variances for numerical example 3.15 illustrating the differences in optimal allocation between the proposed allocation and OCBA-CO (Lee et al., 2011)

| System ($i$) | $h_i$ | $\sigma^2_{h_i}$ | $g_i$ | $\sigma^2_{g_i}$ |
|---|---|---|---|---|
| 1 | 0 | $\sigma^2_{h_1} \in [0.2, 4]$ | $-10.0$ | 1.0 |
| 2 | 2.0 | 1.0 | 1.6 | 1.0 |

Solving for the optimal allocation as a function of $\sigma^2_{h_1}$ yields the allocations displayed in figure 3.3 and the achieved rate of decay of $P\{FS\}$ displayed in figure 3.4.

From figure 3.3, it can be seen that the proposed allocation to system 1 increases slightly at first, and then decreases to a very low, steady allocation from approximately $\sigma^2_{h_1} = 1.5$ onwards. The steady allocation occurs because system 1 requires only a minimal proportion of the total sample to determine its feasibility.

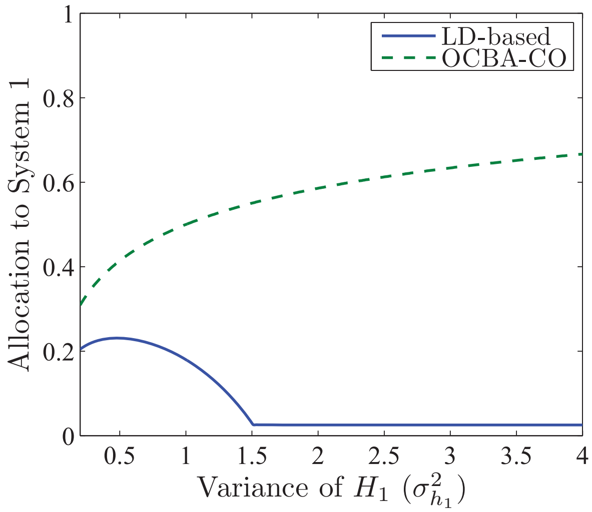


Figure 3.3: The graph of variance of the best feasible system ($\sigma^2_{h_1}$) versus optimal allocation allocation ($\boldsymbol{\alpha}^*$) for the systems in example 3.15 shows the differences between the proposed allocation and OCBA-CO.

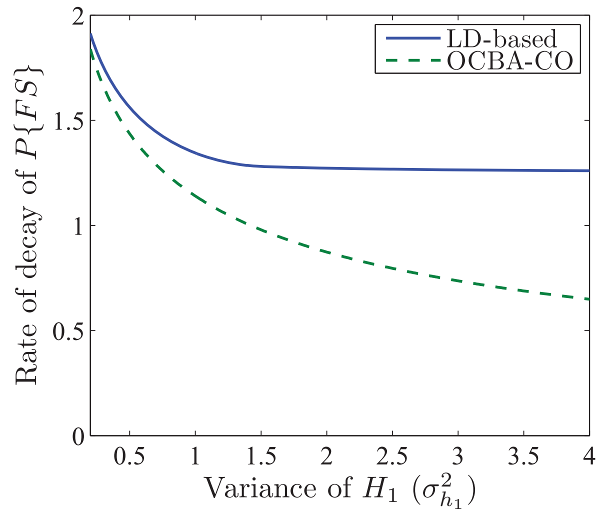


Figure 3.4: The graph of variance of the best feasible system ($\sigma^2_{h_1}$) versus the rate of decay of the probability of false selection ($P\{FS\}$) for the systems in example 3.15 shows the differences in the rate of decay of $P\{FS\}$ for the proposed allocation and OCBA-CO.

Contrasting this allocation is the OCBA-CO allocation, which constantly increases as $\sigma^2_{h_1}$ increases. The OCBA-CO allocation does not exploit the fact that system 1 can be correctly selected by allocating more sample to system 2 to disqualify it faster. In figure 3.4, while the proposed allocation achieves a rate of decay that remains constant as $\sigma^2_{h_1}$ increases beyond approximately $\sigma^2_{h_1} = 1.5$, the rate of decay of $P\{FS\}$ for the OCBA-CO allocation continues to decrease as a function of $\sigma^2_{h_1}$. $\square$

# Chapter 4

# Bivariate Normal: One Constraint

Consider the special case of problem $P$ (see section 1.2) in which there is only one constraint ($s = 1$), and the underlying random variables follow a bivariate normal distribution. We derive results analogous to those presented in chapters 2 and 3 as an explicit function of correlation. Because the bounds on the probability of false selection can be formulated in multiple ways, the formulation of the $P\{FS\}$ that was used in previous chapters is not the simplest version. Therefore we derive rate functions for the two $P\{FS\}$ formulations (the $P\{FS\}$ as written in previous chapters, and a simpler version) and show that the rate functions are equivalent. The simpler formulation leads to the proposed optimal allocation strategy. As with chapter 2, this chapter is used to build intuition before considering the multiple constraint case in chapter 5. A version of the work in this chapter appears in the Proceedings of the 2011 Winter Simulation Conference (Hunter et al., 2011).

## 4.1 Chapter Organization

This chapter is organized as follows. Section 4.2 contains assumptions for this chapter. Section 4.3 contains a derivation of the rate function of the $P\{FS\}$. Section 4.4, contains an optimal allocation strategy and a closed-form solution for the allocation between systems other than system 1 under the assumption that $\alpha_1^* \gg \alpha_i^*$.

## 4.2 Assumptions for Chapter 4

Let $(H_i, G_i)$ be a random output vector from the simulation. In addition to assumptions 1 and 2, we assume the following throughout the chapter.

**Assumption 4.1.** *We assume we may obtain independent and identically distributed (iid) replicates of the bivariate normal random vector $(H_i, G_i)$, with correlation $\rho_i$, $|\rho_i| < 1$. That is,*

$$\begin{bmatrix} H_i \\ G_i \end{bmatrix} \sim BVN\left( \begin{bmatrix} h_i \\ g_i \end{bmatrix}, \Sigma_i \right),$$

*where $\Sigma_i$ is the variance-covariance matrix*

$$\Sigma_i = \begin{pmatrix} Var(H_i) & Cov(H_i, G_i) \\ Cov(G_i, H_i) & Var(G_i) \end{pmatrix} = \begin{pmatrix} \sigma_{h_i}^2 & \rho_i\, \sigma_{h_i}\sigma_{g_i} \\ \rho_i\, \sigma_{h_i}\sigma_{g_i} & \sigma_{g_i}^2 \end{pmatrix},$$

*and $\sigma_{h_i}^2 < \infty, \sigma_{g_i}^2 < \infty$. We denote the vector of correlations as $\rho = (\rho_1, \rho_2, \ldots, \rho_r)$.*

Let us define $(\bar{H}_i(n), \bar{G}_i(n)) = (n^{-1}\sum_{k=1}^{n} H_{ik}, n^{-1}\sum_{k=1}^{n} G_{ik})$, and define the vector $(\hat{H}_i, \hat{G}_i) \equiv (\bar{H}_i(\alpha_i n), \bar{G}_i(\alpha_i n))$ as shorthand notation for the estimator of $(h_i, g_i)$ after scaling the sample size by $\alpha_i$. Under assumption 4.1, by the Gärtner-Ellis theorem (see, e.g., Dembo and Zeitouni, 1998, p. 44), $(\bar{H}_i, \bar{G}_i)$ satisfies the large deviations principle (LDP) with good rate function

$$\begin{aligned} I_i(x, y) &= \frac{1}{2} \begin{bmatrix} x - h_i & y - g_i \end{bmatrix} \Sigma_i^{-1} \begin{bmatrix} x - h_i \\ y - g_i \end{bmatrix} \\ &= \frac{1}{2(1 - \rho_i^2)} \left( \frac{(x - h_i)^2}{\sigma_{h_i}^2} - \frac{2\rho_i(x - h_i)(y - g_i)}{\sigma_{h_i}\sigma_{g_i}} + \frac{(y - g_i)^2}{\sigma_{g_i}^2} \right). \end{aligned}$$

Since the marginal distributions for each $H_i$ and $G_i$ are also normal, $\bar{H}_i$ and $\bar{G}_i$ likewise satisfy the LDP with good rate functions

$$I_i(x) = \frac{(x - h_i)^2}{2\sigma_{h_i}^2} \quad \text{and} \quad J_i(y) = \frac{(y - g_i)^2}{2\sigma_{g_i}^2}.$$

## 4.3 Rate Function of the Probability of False Selection

We present two formulations of the $P\{FS\}$ and their resulting rate functions. We then reconcile these formulations, and proceed using the simplest.

### 4.3.1 Probability of False Selection: First Formulation

Consistent with the formulation of the probability of false selection in previous chapters, let us formulate the $P\{FS\}$ as the probability that system 1 is declared infeasible or that system 1 estimated feasible and is beaten in estimated objective value by another estimated-feasible system. That is,

$$P\{FS\} = P\{(\underbrace{\hat{G}_1 > \gamma}_{\substack{\text{system 1}\\ \text{estimated}\\ \text{infeasible}}}) \cup (\cup_{i=2}^{r} \overbrace{(\underbrace{\hat{G}_i \leq \gamma}_{\substack{\text{system i}\\ \text{estimated}\\ \text{feasible}}}) \cap (\underbrace{\hat{H}_i \leq \hat{H}_1}_{\substack{\text{system i}\\ \text{“beats”}\\ \text{system 1}}}) \cap (\underbrace{\hat{G}_1 \leq \gamma}_{\substack{\text{system 1}\\ \text{estimated}\\ \text{feasible}}})}^{\substack{\text{systems 1 and i estimated feasible}\\ \text{and system i “beats” system 1}}})\}. \tag{4.1}$$

We now derive the rate function for the $P\{FS\}$ in equation (4.1). The $P\{FS\}$ in equation (4.1) has the lower bound

$$\max\left(P\{\hat{G}_1 > \gamma\}, \max_{2\leq i\leq r}\left(P\{(\hat{G}_i \leq \gamma) \cap (\hat{H}_i \leq \hat{H}_1) \cap (\hat{G}_1 \leq \gamma)\}\right)\right)$$

and the upper bound

$$r \times \max\left(P\{\hat{G}_1 > \gamma\}, \max_{2\leq i\leq r}\left(P\{(\hat{G}_i \leq \gamma) \cap (\hat{H}_i \leq \hat{H}_1) \cap (\hat{G}_1 \leq \gamma)\}\right)\right)$$

such that

$$\lim_{n\to\infty}\frac{1}{n}\log P\{FS\} = \lim_{n\to\infty}\frac{1}{n}\log\max\left(P\{\hat{G}_1 > \gamma\}, \max_{2\leq i\leq r}\left(P\{(\hat{G}_i \leq \gamma) \cap (\hat{H}_i \leq \hat{H}_1) \cap (\hat{G}_1 \leq \gamma)\}\right)\right). \tag{4.2}$$

As in chapter 3, there are two main terms: one for the feasibility of system 1, and a second term representing the event that another estimated-feasible system “beats” the estimated-feasible system 1 in estimated objective value. The following proposition states that the overall rate function is the minimum rate function of the probability of each of these two events.

**Proposition 4.1.** *The rate function for $P\{FS\}$ is*

$$-\lim_{n\to\infty}\frac{1}{n}\log P\{FS\} = \min\left(\alpha_1 J_1(\gamma), \min_{2\leq i\leq r}\left(\inf_{\substack{x_i\leq x_1\\ y_1\leq\gamma\\ y_i\leq\gamma}}\left(\alpha_1 I_1(x_1,y_1) + \alpha_i I_i(x_i,y_i)\right)\right)\right).$$

*Proof.* From equation (4.2) and proposition A.2, assuming the relevant limits exist, it follows that

$$\begin{aligned}-\lim_{n\to\infty}\frac{1}{n}\log P\{FS\} = \min\Bigg(&-\lim_{n\to\infty}\frac{1}{n}\log P\{\hat{G}_1 > \gamma\},\\ &\min_{2\leq i\leq r}\left(-\lim_{n\to\infty}\frac{1}{n}\log P\{(\hat{G}_i \leq \gamma) \cap (\hat{H}_i \leq \hat{H}_1) \cap (\hat{G}_1 \leq \gamma)\}\right)\Bigg).\end{aligned}$$

Since assumption 4.1 satisfies assumptions 2.2 and 2.3, it follows from theorem 2.1 that

$$-\lim_{n\to\infty}\frac{1}{n}\log P\{\hat{G}_1>\gamma\}=\alpha_1 J_1(\gamma).$$

Let us consider $-\lim_{n\to\infty}\frac{1}{n}\log P\{(\hat{G}_i\le\gamma)\cap(\hat{H}_i\le\hat{H}_1)\cap(\hat{G}_1\le\gamma)\}$ for a system $i\neq 1$. Let

$$\Lambda^{(\alpha_1 n,\alpha_i n)}_{(H_1,G_1,H_i,G_i)}(\boldsymbol{\theta}_1,\boldsymbol{\theta_i})=\log \mathrm{E}\big[e^{\langle(\boldsymbol{\theta}_1,\boldsymbol{\theta}_i),(\hat{H}_1,\hat{G}_1,\hat{H}_i,\hat{G}_i)\rangle}\big]$$

be the cumulant generating function of the random vector $(\hat{H}_1,\hat{G}_1,\hat{H}_i,\hat{G}_i)$. By the independence of system 1 and system $i$ (assumption 1),

$$\Lambda^{(\alpha_1 n,\alpha_i n)}_{(H_1,G_1,H_i,G_i)}(\boldsymbol{\theta}_1,\boldsymbol{\theta_i})=\Lambda^{(\alpha_1 n)}_{(H_1,G_1)}(\boldsymbol{\theta}_1)+\Lambda^{(\alpha_i n)}_{(H_i,G_i)}(\boldsymbol{\theta}_i).$$

Under assumption 4.1,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\Lambda^{(\alpha_1 n,\alpha_i n)}_{(H_1,G_1,H_i,G_i)}(n\boldsymbol{\theta}_1,n\boldsymbol{\theta_i})&=\lim_{n\to\infty}\frac{1}{n}\Lambda^{(\alpha_1 n)}_{(H_1,G_1)}(n\boldsymbol{\theta}_1)+\lim_{n\to\infty}\frac{1}{n}\Lambda^{(\alpha_i n)}_{(H_i,G_i)}(n\boldsymbol{\theta}_i)\\&=\alpha_1\log\mathrm{E}[e^{\langle\boldsymbol{\theta}_1/\alpha_1,(H_{11},G_{11})\rangle}]+\alpha_i\log\mathrm{E}[e^{\langle\boldsymbol{\theta}_i/\alpha_i,(H_{i1},G_{i1})\rangle}].\end{aligned}$$

Then by the Gärtner-Ellis theorem, $(\hat{H}_1,\hat{G}_1,\hat{H}_i,\hat{G}_i)$ satisfies an LDP with good rate function

$$I_i(x_1,y_1,x_i,y_i)=\alpha_1 I_1(x_1,y_1)+\alpha_i I_i(x_i,y_i),$$

and hence

$$-\lim_{n\to\infty}\frac{1}{n}\log P\{(\hat{G}_i\le\gamma)\cap(\hat{H}_i\le\hat{H}_1)\cap(\hat{G}_1\le\gamma)\}=\inf_{\substack{x_i\le x_1\\ y_1\le\gamma\\ y_i\le\gamma}}\Big(\alpha_1 I_1(x_1,y_1)+\alpha_i I_i(x_i,y_i)\Big).\qquad\square$$

For notational simplicity, let the rate function $\inf_{x_i\le x_1,\,y_1\le\gamma,\,y_i\le\gamma}(\alpha_1 I_1(x_1,y_1)+\alpha_i I_i(x_i,y_i))$ be denoted by $K_i(\alpha_1,\alpha_i)$. Then expanding the bivariate normal rate functions,

$$\begin{aligned}K_i(\alpha_1,\alpha_i)=\inf_{\substack{x_i\le x_1\\ y_1\le\gamma\\ y_i\le\gamma}}\Bigg(&\frac{\alpha_1}{2(1-\rho_1^2)}\left(\frac{(x_1-h_1)^2}{\sigma_{h_1}^2}-\frac{2\rho_1(x_1-h_1)(y_1-g_1)}{\sigma_{h_1}\sigma_{g_1}}+\frac{(y_1-g_1)^2}{\sigma_{g_1}^2}\right)\\&+\frac{\alpha_i}{2(1-\rho_i^2)}\left(\frac{(x_i-h_i)^2}{\sigma_{h_i}^2}-\frac{2\rho_i(x_i-h_i)(y_i-g_i)}{\sigma_{h_i}\sigma_{g_i}}+\frac{(y_i-g_i)^2}{\sigma_{g_i}^2}\right)\Bigg).\end{aligned}$$

The following proposition provides the location of the infimum in $K_i(\alpha_1,\alpha_i)$.

**Proposition 4.2.** *Let*

$$\Gamma_\ell(\rho,\alpha) = \Bigg\{ i : h_i > h_1,\ g_1 \le \gamma - \rho_1 \frac{\sigma_{g_1}}{\sigma_{h_1}} \frac{(\sigma_{h_1}^2/\alpha_1)}{(\sigma_{h_1}^2/\alpha_1)+(\sigma_{h_i}^2/\alpha_i)}(h_i - h_1),$$
$$g_i \le \gamma + \rho_i \frac{\sigma_{g_i}}{\sigma_{h_i}} \frac{(\sigma_{h_i}^2/\alpha_i)}{(\sigma_{h_1}^2/\alpha_1)+(\sigma_{h_i}^2/\alpha_i)}(h_i - h_1) \Bigg\},$$
$$\Gamma_u(\rho,\alpha) = \Bigg\{ i : h_i > h_1 + \rho_1\sigma_{h_1}\frac{(\gamma - g_1)}{\sigma_{g_1}},\ \gamma - \rho_1 \frac{\sigma_{g_1}}{\sigma_{h_1}} \frac{(\sigma_{h_1}^2/\alpha_1)}{(\sigma_{h_1}^2/\alpha_1)+(\sigma_{h_i}^2/\alpha_i)}(h_i - h_1) < g_1 \le \gamma,$$
$$g_i \le \gamma + \rho_i \frac{\sigma_{g_i}}{\sigma_{h_i}} \frac{(\sigma_{h_i}^2/\alpha_i)}{((1-\rho_1^2)\sigma_{h_1}^2/\alpha_1)+(\sigma_{h_i}^2/\alpha_i)} \left( h_i - \big(h_1 + \rho_1\sigma_{h_1}\frac{(\gamma-g_1)}{\sigma_{g_1}}) \right) \Bigg\},$$
$$\mathcal{S}_b(\rho) = \Bigg\{ i : h_i - \rho_i\sigma_{h_i}\frac{(g_i-\gamma)}{\sigma_{g_i}} \le h_1 \ \ \textit{and}\ \ g_i > \gamma \Bigg\},$$
$$\mathcal{S}_{w,\ell}(\rho,\alpha) = \Bigg\{ i : h_i - \rho_i\sigma_{h_i}\frac{(g_i-\gamma)}{\sigma_{g_i}} > h_1,$$
$$g_1 \le \gamma - \rho_1 \frac{\sigma_{g_1}}{\sigma_{h_1}} \frac{(\sigma_{h_1}^2/\alpha_1)}{(\sigma_{h_1}^2/\alpha_1)+((1-\rho_i^2)\sigma_{h_i}^2/\alpha_i)} \left( \big(h_i - \rho_i\sigma_{h_i}\frac{(g_i-\gamma)}{\sigma_{g_i}}) - h_1 \right),$$
$$g_i > \gamma + \rho_i \frac{\sigma_{g_i}}{\sigma_{h_i}} \frac{(\sigma_{h_i}^2/\alpha_i)}{(\sigma_{h_1}^2/\alpha_1)+(\sigma_{h_i}^2/\alpha_i)}(h_i - h_1) \Bigg\},$$
$$\mathcal{S}_{w,u}(\rho,\alpha) = \Bigg\{ i : h_i - \rho_i\sigma_{h_i}\frac{(g_i-\gamma)}{\sigma_{g_i}} > h_1 + \rho_1\sigma_{h_1}\frac{(\gamma-g_1)}{\sigma_{g_1}},$$
$$\gamma - \rho_1 \frac{\sigma_{g_1}}{\sigma_{h_1}} \frac{(\sigma_{h_1}^2/\alpha_1)}{(\sigma_{h_1}^2/\alpha_1)+((1-\rho_i^2)\sigma_{h_i}^2/\alpha_i)} \left( \big(h_i - \rho_i\sigma_{h_i}\frac{(g_i-\gamma)}{\sigma_{g_i}}) - h_1 \right) < g_1 \le \gamma,$$
$$g_i > \gamma + \rho_i \frac{\sigma_{g_i}}{\sigma_{h_i}} \frac{\sigma_{h_i}^2/\alpha_i}{((1-\rho_1^2)\sigma_{h_1}^2/\alpha_1)+(\sigma_{h_i}^2/\alpha_i)} \left( h_i - \big(h_1 + \rho_1\sigma_{h_1}\frac{(\gamma-g_1)}{\sigma_{g_1}}) \right) \Bigg\}.$$

*Then the infimum in $K_i(\alpha_1,\alpha_i)$ is achieved at the following locations.*

*For $i \in \Gamma_\ell(\rho,\alpha)$,*

$$x_1^* = x_i^* = \frac{(\sigma_{h_i}^2/\alpha_i)h_1 + (\sigma_{h_1}^2/\alpha_1)h_i}{(\sigma_{h_1}^2/\alpha_1)+(\sigma_{h_i}^2/\alpha_i)};\ y_1^* = g_1 + \rho_1 \frac{\sigma_{g_1}}{\sigma_{h_1}} \frac{(\sigma_{h_1}^2/\alpha_1)}{(\sigma_{h_1}^2/\alpha_1)+(\sigma_{h_i}^2/\alpha_i)}(h_i - h_1);$$
$$y_i^* = g_i - \rho_i \frac{\sigma_{g_i}}{\sigma_{h_i}} \frac{(\sigma_{h_i}^2/\alpha_i)}{(\sigma_{h_1}^2/\alpha_1)+(\sigma_{h_i}^2/\alpha_i)}(h_i - h_1). \tag{4.3}$$

*For $i \in \Gamma_u(\rho,\alpha)$,*

$$x_1^* = x_i^* = \frac{(\sigma_{h_i}^2/\alpha_i)(h_1 + \rho_1\sigma_{h_1}\frac{(\gamma-g_1)}{\sigma_{g_1}}) + ((1-\rho_1^2)\sigma_{h_1}^2/\alpha_1)h_i}{(1-\rho_1^2)\sigma_{h_1}^2/\alpha_1 + \sigma_{h_i}^2/\alpha_i};\ y_1^* = \gamma; \tag{4.4}$$

$$y_i^* = g_i - \rho_i \frac{\sigma_{g_i}}{\sigma_{h_i}} \frac{\sigma_{h_i}^2/\alpha_i}{((1-\rho_1^2)\sigma_{h_1}^2/\alpha_1 + \sigma_{h_i}^2/\alpha_i)} \left( h_i - \left(h_1 + \rho_1 \sigma_{h_1} \frac{(\gamma - g_1)}{\sigma_{g_1}}\right) \right). \tag{4.5}$$

*For* $i \in \mathcal{S}_b(\rho)$,

$$x_1^* = h_1;\ x_i^* = h_i - \rho_i \sigma_{h_i} \frac{(g_i - \gamma)}{\sigma_{g_i}};\ y_1^* = g_1;\ y_i^* = \gamma. \tag{4.6}$$

*For* $i \in \mathcal{S}_{w,\ell}(\rho, \alpha)$,

$$x_1^* = x_i^* = \frac{(1-\rho_i^2)\sigma_{h_i}^2/\alpha_i(h_1) + (\sigma_{h_1}^2/\alpha_1)(h_i - \rho_i \sigma_{h_i} \frac{(g_i-\gamma)}{\sigma_{g_i}})}{\sigma_{h_1}^2/\alpha_1 + (1-\rho_i^2)\sigma_{h_i}^2/\alpha_i}; \tag{4.7}$$

$$y_1^* = g_1 + \rho_1 \frac{\sigma_{g_1}}{\sigma_{h_1}} \frac{(\sigma_{h_1}^2/\alpha_1)}{(\sigma_{h_1}^2/\alpha_1 + (1-\rho_i^2)\sigma_{h_i}^2/\alpha_i)} \left( (h_i - \rho_i \sigma_{h_i} \frac{(g_i - \gamma)}{\sigma_{g_i}}) - h_1 \right);\ y_i^* = \gamma. \tag{4.8}$$

*For* $i \in \mathcal{S}_{w,u}(\rho, \alpha)$,

$$x_1^* = x_i^* = \frac{(1-\rho_i^2)\sigma_{h_i}^2/\alpha_i \left(h_1 + \rho_1 \sigma_{h_1} \frac{(\gamma-g_1)}{\sigma_{g_1}}\right) + (1-\rho_1^2)\sigma_{h_1}^2/\alpha_1 \left(h_i - \rho_i \sigma_{h_i} \frac{(g_i-\gamma)}{\sigma_{g_i}}\right)}{(1-\rho_1^2)\sigma_{h_1}^2/\alpha_1 + (1-\rho_i^2)\sigma_{h_i}^2/\alpha_i}; \tag{4.9}$$

$$y_1^* = \gamma;\ y_i^* = \gamma. \tag{4.10}$$

*Proof.* $K_i(\alpha_1, \alpha_i)$ is a convex function of $(x_1, x_i, y_1, y_i)$. Let $\lambda_x \geq 0, \lambda_{y_1} \geq 0$, and $\lambda_{y_i} \geq 0$ be the Lagrange multipliers such that

$$\begin{aligned} L(x_1, x_i, y_1, y_i, \lambda_x, \lambda_{y_1}, \lambda_{y_i}) = {} & \frac{\alpha_1}{2(1-\rho_1^2)} \left( \frac{(x_1 - h_1)^2}{\sigma_{h_1}^2} - \frac{2\rho_1(x_1 - h_1)(y_1 - g_1)}{\sigma_{h_1}\sigma_{g_1}} + \frac{(y_1 - g_1)^2}{\sigma_{g_1}^2} \right) \\ & + \frac{\alpha_i}{2(1-\rho_i^2)} \left( \frac{(x_i - h_i)^2}{\sigma_{h_i}^2} - \frac{2\rho_i(x_i - h_i)(y_i - g_i)}{\sigma_{h_i}\sigma_{g_i}} + \frac{(y_i - g_i)^2}{\sigma_{g_i}^2} \right) \\ & + \lambda_x(x_i - x_1) + \lambda_{y_1}(y_1 - \gamma) + \lambda_{y_i}(y_i - \gamma). \end{aligned}$$

Under assumption 4.1, $K_i(\alpha_1, \alpha_i)$ is a continuous, differentiable, strictly convex function of $(x_1, x_i, y_1, y_i)$ which attains its unconstrained minimum at $(h_1, h_i, g_1, g_i)$. Further, Slater's condition (Boyd and Vandenberghe, 2004, p. 244) holds, that is, there exists a "strictly feasible" point in the interior of the space $x_i \leq x_1, y_1 \leq \gamma, y_i \leq \gamma$. Therefore the Karush-Kuhn-Tucker (KKT) conditions (see, e.g., Boyd and Vandenberghe, 2004) are necessary and sufficient for a global minimum, and by the strict convexity of $K_i(\alpha_1, \alpha_i)$, this minimum is unique. Then the KKT conditions are,

$$\frac{\alpha_1}{(1-\rho_1^2)}\left(\frac{(x_1^*-h_1)}{\sigma_{h_1}^2}-\frac{\rho_1(y_1^*-g_1)}{\sigma_{h_1}\sigma_{g_1}}\right)=\frac{-\alpha_i}{(1-\rho_i^2)}\left(\frac{(x_i^*-h_i)}{\sigma_{h_i}^2}-\frac{\rho_i(y_i^*-g_i)}{\sigma_{h_i}\sigma_{g_i}}\right)=\lambda_x \tag{4.11}$$

$$\frac{-\alpha_1}{(1-\rho_1^2)}\left(-\frac{\rho_1(x_1^*-h_1)}{\sigma_{h_1}\sigma_{g_1}}+\frac{(y_1^*-g_1)}{\sigma_{g_1}^2}\right)=\lambda_{y_1} \tag{4.12}$$

$$\frac{-\alpha_i}{(1-\rho_i^2)}\left(-\frac{\rho_i(x_i^*-h_i)}{\sigma_{h_i}\sigma_{g_i}}+\frac{(y_i^*-g_i)}{\sigma_{g_i}^2}\right)=\lambda_{y_i} \tag{4.13}$$

$$\lambda_x(x_i^*-x_1^*)=0,\ \lambda_{y_1}(y_1^*-\gamma)=0,\ \lambda_{y_i}(y_i^*-\gamma)=0 \tag{4.14}$$

$$(x_i^*-x_1^*)\le 0,\ (y_1^*-\gamma)\le 0,\ (y_i^*-\gamma)\le 0. \tag{4.15}$$

*Case:* $\lambda_x = 0$. From equation (4.11), $\lambda_x = 0$ implies

$$x_1^*=h_1+\rho_1\sigma_{h_1}\frac{(y_1^*-g_1)}{\sigma_{g_1}}\ \text{ and }\ x_i^*=h_i+\rho_i\sigma_{h_i}\frac{(y_i^*-g_i)}{\sigma_{g_i}}. \tag{4.16}$$

Since $\lambda_x = 0$, conditions in equations (4.14) and (4.15) imply $x_i^* \le x_1^*$. Hence

$$\left(h_i+\rho_i\sigma_{h_i}\frac{(y_i^*-g_i)}{\sigma_{g_i}}\right)\le\left(h_1+\rho_1\sigma_{h_1}\frac{(y_1^*-g_1)}{\sigma_{g_1}}\right).$$

*Case:* $\lambda_x > 0$. The complementary slackness (CS) conditions in equation (4.14) imply that if $\lambda_x > 0$, then $x_1^* = x_i^*$. Letting $x_1^* = x_i^*$ in equation (4.11),

$$x_1^*=x_i^*=\frac{\frac{\alpha_1}{(1-\rho_1^2)\sigma_{h_1}^2}\left(h_1+\rho_1\sigma_{h_1}\frac{(y_1^*-g_1)}{\sigma_{g_1}}\right)+\frac{\alpha_i}{(1-\rho_i^2)\sigma_{h_i}^2}\left(h_i+\rho_i\sigma_{h_i}\frac{(y_i^*-g_i)}{\sigma_{g_i}}\right)}{\frac{\alpha_1}{(1-\rho_1^2)\sigma_{h_1}^2}+\frac{\alpha_i}{(1-\rho_i^2)\sigma_{h_i}^2}} \tag{4.17}$$

Equivalently,

$$x_1^*=x_i^*=\frac{(1-\rho_1^2)\sigma_{h_1}^2/\alpha_1\left(h_i+\rho_i\sigma_{h_i}\frac{(y_i^*-g_i)}{\sigma_{g_i}}\right)+(1-\rho_i^2)\sigma_{h_i}^2/\alpha_i\left(h_1+\rho_1\sigma_{h_1}\frac{(y_1^*-g_1)}{\sigma_{g_1}}\right)}{(1-\rho_1^2)\sigma_{h_1}^2/\alpha_1+(1-\rho_i^2)\sigma_{h_i}^2/\alpha_i}. \tag{4.18}$$

From equation (4.11), $\lambda_x > 0$ and equation (4.17) imply

$$\left(h_i+\rho_i\sigma_{h_i}\frac{(y_i^*-g_i)}{\sigma_{g_i}}\right)>\left(h_1+\rho_1\sigma_{h_1}\frac{(y_1^*-g_1)}{\sigma_{g_1}}\right)$$

*Case:* $\lambda_{y_1} = 0$. From equation (4.12), $\lambda_{y_1} = 0$ and $y_1^* \leq \gamma$ imply

$$y_1^* = g_1 + \rho_1 \sigma_{g_1} \frac{(x_1^* - h_1)}{\sigma_{h_1}} \text{ and } g_1 \leq \gamma - \rho_1 \sigma_{g_1} \frac{(x_1^* - h_1)}{\sigma_{h_1}}. \tag{4.19}$$

*Case:* $\lambda_{y_1} > 0$. From the CS conditions in equation (4.14), $\lambda_{y_1} > 0$ implies $y_1^* = \gamma$. Further, from equation (4.12), it follows that $g_1 > \gamma - \rho_1 \sigma_{g_1}(x_1^* - h_1)/\sigma_{h_1}$.

*Case:* $\lambda_{y_i} = 0$. From equation (4.13), $\lambda_{y_i} = 0$ and $y_i^* \leq \gamma$ imply

$$y_i^* = g_i + \rho_i \sigma_{g_i} \frac{(x_i^* - h_i)}{\sigma_{h_i}} \text{ and } g_i \leq \gamma - \rho_i \sigma_{g_i} \frac{(x_i^* - h_i)}{\sigma_{h_i}}. \tag{4.20}$$

*Case:* $\lambda_{y_i} > 0$. From the CS conditions in equation (4.14), $\lambda_{y_i} > 0$ implies $y_i^* = \gamma$. Further, from equation (4.13), it follows that $g_i > \gamma - \rho_i \sigma_{g_i}(x_i^* - h_i)/\sigma_{h_i}$.

Considering all possible combinations of the cases and solving for the relevant $(x_1^*, x_i^*, y_1^*, y_i^*)$ values yields the results. Table 4.1 shows which $\lambda$ values correspond to which sets. Table 4.2 displays the infeasible solutions to the KKT conditions.

Table 4.1: Feasible solutions to the KKT conditions for finding the infimum in the rate function of the probability of false selection $P\{FS\}$ for the bivariate normal case

| $\lambda_x$ | $\lambda_{y_1}$ | $\lambda_{y_i}$ | Set |
|---|---|---|---|
| $> 0$ | $0$ | $0$ | $\Gamma_\ell(\rho, \alpha)$ |
| $> 0$ | $> 0$ | $0$ | $\Gamma_u(\rho, \alpha)$ |
| $0$ | $0$ | $> 0$ | $\mathcal{S}_b(\rho)$ |
| $> 0$ | $0$ | $> 0$ | $\mathcal{S}_{w,\ell}(\rho, \alpha)$ |
| $> 0$ | $> 0$ | $> 0$ | $\mathcal{S}_{w,u}(\rho, \alpha)$ |

Table 4.2: Infeasible solutions to the KKT conditions for finding the infimum in the rate function of the probability of false selection $P\{FS\}$ for the bivariate normal case

| $\lambda_x$ | $\lambda_{y_1}$ | $\lambda_{y_i}$ | Solution |
|---|---|---|---|
| $0$ | $0$ | $0$ | $x_1^* = h_1;\ x_i^* = h_i;\ y_1^* = g_1;\ y_i^* = g_i$<br>$h_i \leq h_1;\ g_1 \leq \gamma;\ g_i \leq \gamma.$ |
| $0$ | $> 0$ | $0$ | $x_1^* = h_1 - \rho_1 \sigma_{h_1} \frac{(g_1 - \gamma)}{\sigma_{g_1}};\ x_i^* = h_i;\ y_1^* = \gamma;\ y_i^* = g_i$<br>$h_i \leq h_1 - \rho_1 \sigma_{h_1} \frac{(g_1 - \gamma)}{\sigma_{g_1}};\ g_1 > \gamma;\ g_i \leq \gamma$ |
| $0$ | $> 0$ | $> 0$ | $x_1^* = h_1 - \rho_1 \sigma_{h_1} \frac{(g_1 - \gamma)}{\sigma_{g_1}};\ x_i^* = h_i - \rho_i \sigma_{h_i} \frac{(g_i - \gamma)}{\sigma_{g_i}};\ y_1^* = \gamma;\ y_i^* = \gamma$<br>$h_i - \rho_i \sigma_{h_i} \frac{(g_i - \gamma)}{\sigma_{g_i}} \leq h_1 - \rho_1 \sigma_{h_1} \frac{(g_1 - \gamma)}{\sigma_{g_1}};\ g_1 > \gamma;\ g_i > \gamma$ |

$\square$

Using the infima from proposition 4.2 in the rate function expression, one can derive the rate functions presented in theorem 4.3.

**Theorem 4.3.** *The relevant rate functions are*

$$
\begin{aligned}
K_{i\in\Gamma_\ell(\rho,\alpha)}(\alpha_1,\alpha_i) =& \frac{(h_i-h_1)^2}{2(\sigma^2_{h_1}/\alpha_1+\sigma^2_{h_i}/\alpha_i)},\\
K_{i\in\Gamma_u(\rho,\alpha)}(\alpha_1,\alpha_i) =& \frac{\left((h_i-h_1)-\rho_1\sigma_{h_1}\dfrac{(\gamma-g_1)}{\sigma_{g_1}}\right)^2}{2((1-\rho_1^2)\sigma^2_{h_1}/\alpha_1+\sigma^2_{h_i}/\alpha_i)}+\frac{\alpha_1(\gamma-g_1^2)}{2\sigma^2_{g_1}},\\
K_{i\in\mathcal{S}_b(\rho)}(\alpha_1,\alpha_i) =& \frac{\alpha_i(\gamma-g_i)^2}{2\sigma^2_{g_i}},\\
K_{i\in\mathcal{S}_{w,\ell}(\rho,\alpha)}(\alpha_1,\alpha_i) =& \frac{\left((h_i-h_1)-\rho_i\sigma_{h_i}\dfrac{(g_i-\gamma)}{\sigma_{g_i}}\right)^2}{2(\sigma^2_{h_1}/\alpha_1+(1-\rho_i^2)\sigma^2_{h_i}/\alpha_i)}+\frac{\alpha_i(g_i-\gamma)^2}{2\sigma^2_{g_i}},\\
K_{i\in\mathcal{S}_{w,u}(\rho,\alpha)}(\alpha_1,\alpha_i) =& \frac{\left((h_i-h_1)-\rho_1\sigma_{h_1}\dfrac{(\gamma-g_1)}{\sigma_{g_1}}-\rho_i\sigma_{h_i}\dfrac{g_i-\gamma}{\sigma_{g_i}}\right)^2}{2((1-\rho_1^2)\sigma^2_{h_1}/\alpha_1+(1-\rho_i^2)\sigma^2_{h_i}/\alpha_i)}+\frac{\alpha_1(\gamma-g_1)^2}{2\sigma^2_{g_1}}+\frac{\alpha_i(g_i-\gamma)^2}{2\sigma^2_{g_i}}.
\end{aligned}
$$

*Proof.* Mathematica code using results from proposition 4.2 to find the rate function expressions is provided in appendix C.1. The rate functions for $K_{i\in\Gamma_u(\rho,\alpha)}(\alpha_1,\alpha_i), K_{i\in\mathcal{S}_{w,\ell}(\rho,\alpha)}(\alpha_1,\alpha_i)$, and $K_{i\in\mathcal{S}_{w,u}(\rho,\alpha)}(\alpha_1,\alpha_i)$ can be found by completing the square in the following expanded formulations:

$$
\begin{aligned}
K_{i\in\Gamma_u(\rho,\alpha)}(\alpha_1,\alpha_i) =& \frac{(h_i-h_1)^2-2\rho_1\sigma_{h_1}(h_i-h_1)\dfrac{(\gamma-g_1)}{\sigma_{g_1}}+\dfrac{\alpha_1(\gamma-g_1)^2}{\sigma^2_{g_1}}\left(\sigma^2_{h_1}/\alpha_1+\sigma^2_{h_i}/\alpha_i\right)}{2((1-\rho_1^2)\sigma^2_{h_1}/\alpha_1+\sigma^2_{h_i}/\alpha_i)},\\
K_{i\in\mathcal{S}_{w,\ell}(\rho,\alpha)}(\alpha_1,\alpha_i) =& \frac{(h_i-h_1)^2-2\rho_i\sigma_{h_i}(h_i-h_1)\dfrac{(g_i-\gamma)}{\sigma_{g_i}}+\dfrac{\alpha_i(g_i-\gamma)^2}{\sigma^2_{g_i}}\left(\sigma^2_{h_1}/\alpha_1+\sigma^2_{h_i}/\alpha_i\right)}{2(\sigma^2_{h_1}/\alpha_1+(1-\rho_i^2)\sigma^2_{h_i}/\alpha_i)},\\
K_{i\in\mathcal{S}_{w,u}(\rho,\alpha)}(\alpha_1,\alpha_i) =& \Bigg[(h_i-h_1)^2-2(h_i-h_1)\left(\rho_1\sigma_{h_1}\frac{(\gamma-g_1)}{\sigma_{g_1}}+\rho_i\sigma_{h_i}\frac{(g_i-\gamma)}{\sigma_{g_i}}\right)\\
&+\ 2\rho_1\sigma_{h_1}\frac{(\gamma-g_1)}{\sigma_{g_1}}\rho_i\sigma_{h_i}\frac{(g_i-\gamma)}{\sigma_{g_i}}+\frac{\alpha_1(\gamma-g_1)^2}{\sigma^2_{g_1}}\left(\sigma^2_{h_1}/\alpha_1+(1-\rho_i^2)\sigma^2_{h_i}/\alpha_i\right)\\
&+\ \frac{\alpha_i(g_i-\gamma)^2}{\sigma^2_{g_i}}\left((1-\rho_1^2)\sigma^2_{h_1}/\alpha_1+\sigma^2_{h_i}/\alpha_i\right)\Bigg]\\
&\times\left[2((1-\rho_1^2)\sigma^2_{h_1}/\alpha_1+(1-\rho_i^2)\sigma^2_{h_i}/\alpha_i)\right]^{-1}.
\end{aligned}
$$

□

The following proposition states that when $\rho_1 \le 0$, $\Gamma_u(\rho,\alpha)$ and $\mathcal{S}_{w,u}(\rho,\alpha)$ are empty. This result depends on the formulation of problem $P$ in section 1.2.

**Proposition 4.4.** *Given the formulation of Problem P in Section 1.2 and under Assumption 4.1, if $\rho_1 < 0$, then $\Gamma_u(\rho, \alpha)$ and $\mathcal{S}_{w,u}(\rho, \alpha)$ are empty.*

*Proof.* The following proof is based on the notion of limiting independence (see definition 3.1). See appendix B.2 for an alternate, algebraic proof.

In the right-hand side of equation (4.2), the rate function is determined by the system that has the maximum probability of being declared feasible and beating system 1 when system 1 is also declared feasible. To derive a rate function for this "duel" between system 1 and some other system $i$, we derive a rate function for the probability

$$P\{(\hat{G}_i \leq \gamma) \cap (\hat{H}_i \leq \hat{H}_1) \cap (\hat{G}_1 \leq \gamma)\}. \tag{4.21}$$

Since $\rho_1 < 0$, it follows that

$$\begin{aligned} P\{(\hat{G}_i \leq \gamma) \cap (\hat{H}_i \leq \hat{H}_1)\} P\{\hat{G}_1 \leq \gamma\} &\leq P\{(\hat{G}_i \leq \gamma) \cap (\hat{H}_i \leq \hat{H}_1) \cap (\hat{G}_1 \leq \gamma)\} \\ &\leq P\{(\hat{G}_i \leq \gamma) \cap (\hat{H}_i \leq \hat{H}_1)\} \end{aligned}$$

Since $P\{\hat{G}_1 \leq \gamma\} \to 1$, then the rate function we derive for the probability in equation (4.21) considers only the random vector $(\hat{H}_1, \hat{H}_i, \hat{G}_i)$, and hence the correlation between $\hat{H}_1$ and $\hat{G}_1$ does not appear. Thus when $\rho_1 < 0$, these events display limiting independence, and loosely speaking, the effect of correlation is washed away by the logarithmic limit. $\square$

Thus $\boldsymbol{\alpha}$ plays a different role in the correlated bivariate normal than it does in the independent case: changing $\boldsymbol{\alpha}$ can affect which terms enter the rate function by changing the set assignment of that system. The following example provides insight into the structure of the five sets defined in proposition 4.2.

**Example 4.5.** To better understand the sets defined in proposition 4.2, the following two figures show the set boundaries for the case of equal allocation and variances across sets, where system 1 is located at $(h_1, g_1) = (0, -2)$ with $\sigma^2_{h_1} = \sigma^2_{g_1} = \sigma^2_{h_i} = \sigma^2_{g_i} = 1$, and the constraint is $\gamma = 0$.

In figure 4.1, the correlations are identical between system 1 and system i, that is, $\rho_1 = \rho_i$. As predicted in proposition 4.4, for negative correlation $\rho_1$, systems fall into one of $\Gamma_\ell(\rho, \alpha), \mathcal{S}_b(\rho)$, or $\mathcal{S}_{w,\ell}(\rho, \alpha)$. For very negative correlation, set $\mathcal{S}_{w,\ell}$ is the "largest," but as the negative correlation approaches zero, $\mathcal{S}_{w,\ell}$ shrinks to approach the "true" sets, $\Gamma, \mathcal{S}_b$, and $\mathcal{S}_w$. For positive correlations, systems fall into one of each of the five sets, with $\mathcal{S}_{w,\ell}$ and $\mathcal{S}_{w,u}$ shrinking as correlation increases. For high, positive correlation, the sets $\Gamma_u(\rho, \alpha)$ and $\mathcal{S}_b$ seem to encompass most systems.

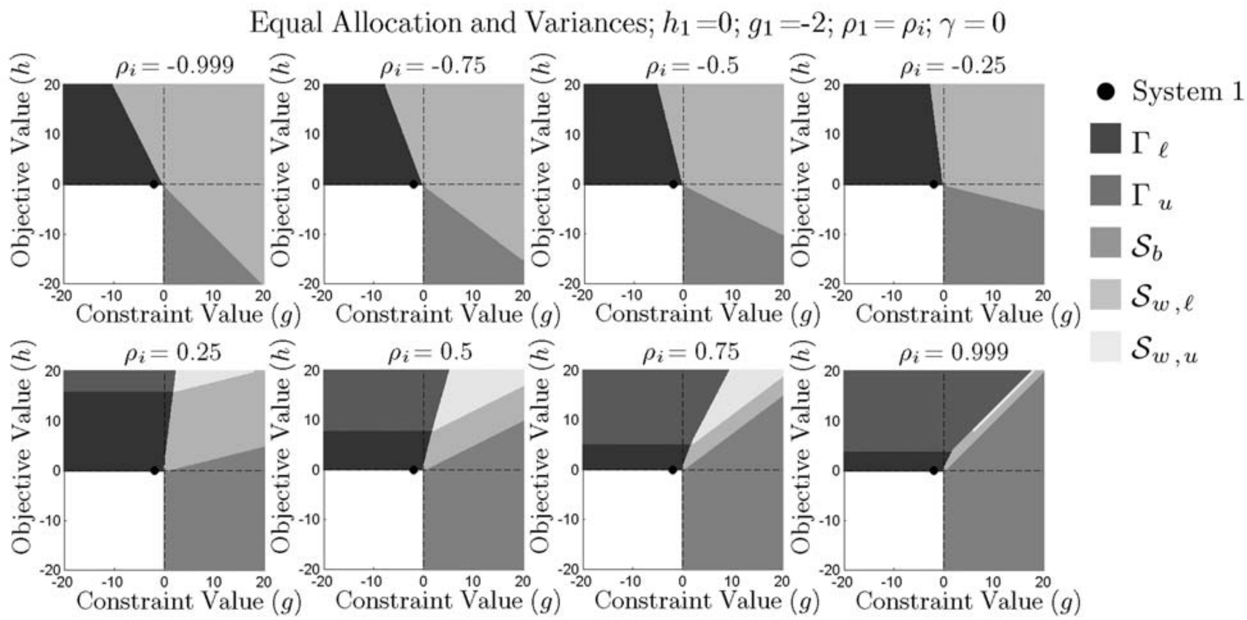


Figure 4.1: Graphs of sets $\Gamma_\ell(\rho, \alpha), \Gamma_u(\rho, \alpha), \mathcal{S}_b(\rho), \mathcal{S}_{w,\ell}(\rho, \alpha)$, and $\mathcal{S}_{w,u}(\rho, \alpha)$ for equal allocations, variances, and correlations between system 1 and system $i$ in example 4.5.

Now let us retain the same example, but consider opposite correlations between system 1 and system $i$ such that $\rho_1 = -\rho_i$. Although this scenario seems unlikely to occur in practice, it provides interesting insights into the possible shapes of the sets. Figure 4.2 shows that systems are classified by all five sets for negative correlations ($\rho_1 > 0$), but as predicted by proposition 4.4, positive correlations ($\rho_1 < 0$) implies systems are classified by only three sets $\Gamma_\ell(\rho, \alpha), \mathcal{S}_b(\rho)$, or $\mathcal{S}_{w,\ell}(\rho, \alpha)$. For very negative correlations, set $\mathcal{S}_{w,u}$ is the largest, with sets $\Gamma_u$ and $\mathcal{S}_b$ encompassing most other systems. As correlation becomes increasingly positive, sets $\mathcal{S}_{w,u}$ and $\mathcal{S}_{w,\ell}$ together shrink to the space occupied by the "true" set $\mathcal{S}_w$. In the case of positive correlation, as correlation increases, the sets $\Gamma_\ell(\rho, \alpha)$ and $\mathcal{S}_b(\rho)$ progress from the "true" set definitions inwards until they encompass most systems and $\mathcal{S}_{w,\ell}(\rho, \alpha)$ has shrunk significantly. $\square$

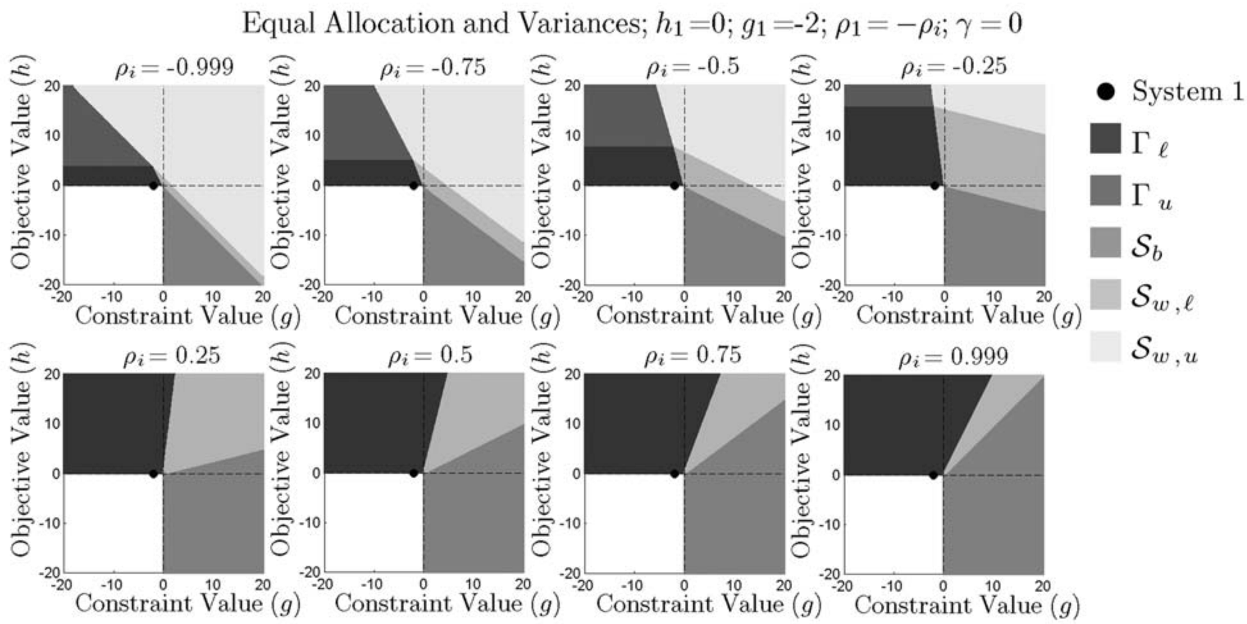


Figure 4.2: Graphs of sets $\Gamma_\ell(\rho,\alpha), \Gamma_u(\rho,\alpha), \mathcal{S}_b(\rho), \mathcal{S}_{w,\ell}(\rho,\alpha)$, and $\mathcal{S}_{w,u}(\rho,\alpha)$ for equal allocations, variances, and equal-but-opposite-sign correlations between system 1 and system $i$ in example 4.5.

The following theorem states the probability of false selection for the bivariate normal case as a function of the correlation $\rho$.

**Theorem 4.6.** *Under assumption 4.1, the rate function of the probability of false selection is*

$$-\lim_{n\to\infty}\frac{1}{n}\log P\{FS\} = \min\Bigg(\frac{\alpha_1(\gamma-g_1)^2}{2\sigma_{g_1}^2},\ \min_{i\in\Gamma_\ell(\rho,\alpha)}\frac{(h_i-h_1)^2}{2(\sigma_{h_1}^2/\alpha_1+\sigma_{h_i}^2/\alpha_i)},\ \min_{i\in\mathcal{S}_b(\rho)}\frac{\alpha_i(g_i-\gamma)^2}{2\sigma_{g_i}^2},$$
$$\min_{i\in\Gamma_u(\rho,\alpha)}\frac{\left((h_i-h_1)-\rho_1\sigma_{h_1}\dfrac{(\gamma-g_1)}{\sigma_{g_1}}\right)^2}{2((1-\rho_1^2)\sigma_{h_1}^2/\alpha_1+\sigma_{h_i}^2/\alpha_i)}+\frac{\alpha_1(\gamma-g_1^2)}{2\sigma_{g_1}^2},$$
$$\min_{i\in\mathcal{S}_{w,\ell}(\rho,\alpha)}\frac{\left((h_i-h_1)-\rho_i\sigma_{h_i}\dfrac{(g_i-\gamma)}{\sigma_{g_i}}\right)^2}{2(\sigma_{h_1}^2/\alpha_1+(1-\rho_i^2)\sigma_{h_i}^2/\alpha_i)}+\frac{\alpha_i(g_i-\gamma)^2}{2\sigma_{g_i}^2},$$
$$\min_{i\in\mathcal{S}_{w,u}(\rho,\alpha)}\frac{\left((h_i-h_1)-\rho_1\sigma_{h_1}\dfrac{(\gamma-g_1)}{\sigma_{g_1}}-\rho_i\sigma_{h_i}\dfrac{g_i-\gamma}{\sigma_{g_i}}\right)^2}{2((1-\rho_1^2)\sigma_{h_1}^2/\alpha_1+(1-\rho_i^2)\sigma_{h_i}^2/\alpha_i)}+\frac{\alpha_1(\gamma-g_1)^2}{2\sigma_{g_1}^2}+\frac{\alpha_i(g_i-\gamma)^2}{2\sigma_{g_i}^2}\Bigg).$$

Thus the overall rate function for the probability of false selection is the minimum rate function across five sets of systems, four of which are themselves defined by functions of the budget allocation.

### 4.3.2 Probability of False Selection: Second Formulation

From equation (4.1), recall that in the previous section the $P\{FS\}$ was stated as,

$$P\{FS\} = P\{(\hat{G}_1 > \gamma) \cup (\cup_{i=2}^{r} (\hat{G}_i \leq \gamma) \cap (\hat{H}_i \leq \hat{H}_1) \cap (\hat{G}_1 \leq \gamma))\},$$

which is the probability that system 1 is estimated infeasible or system 1 is estimated feasible and is beaten by another estimated-feasible system. Equivalently,

$$\begin{aligned} P\{FS\} &= P\{(\hat{G}_1 > \gamma) \cup (\cup_{i=2}^{r} (\hat{G}_i \leq \gamma) \cap (\hat{H}_i \leq \hat{H}_1) \cap (\hat{G}_1 \leq \gamma))\}, \\ &= P\{(\hat{G}_1 > \gamma) \cup (\cup_{i=2}^{r} (\hat{G}_i \leq \gamma) \cap (\hat{H}_i \leq \hat{H}_1))\}, \end{aligned} \tag{4.22}$$

where system 1 is no longer restricted to be estimated-feasible when it is beaten by another estimated-feasible system in equation (4.22). Both formulations are correct, but lead to slightly different analyses of the rate function of the $P\{FS\}$. In this section, we formulate the probability of false selection as in equation (4.22), that is, as the probability that system 1 is declared infeasible or that another system $i$ is declared feasible and beats system 1 in estimated objective value.

The $P\{FS\}$ in equation (4.22) has the lower bound

$$\max\left(P\{\hat{G}_1 > \gamma\}, \max_{2 \leq i \leq r}\left(P\{(\hat{G}_i \leq \gamma) \cap (\hat{H}_i \leq \hat{H}_1)\}\right)\right)$$

and the upper bound

$$r \times \max\left(P\{\hat{G}_1 > \gamma\}, \max_{2 \leq i \leq r}\left(P\{(\hat{G}_i \leq \gamma) \cap (\hat{H}_i \leq \hat{H}_1)\}\right)\right)$$

such that

$$\lim_{n\to\infty} \frac{1}{n} \log P\{FS\} = \lim_{n\to\infty} \frac{1}{n} \log \max\left(P\{\hat{G}_1 > \gamma\}, \max_{2 \leq i \leq r}\left(P\{(\hat{G}_i \leq \gamma) \cap (\hat{H}_i \leq \hat{H}_1)\}\right)\right). \tag{4.23}$$

As in chapter 3, we retain two main terms: one for the feasibility of system 1, and a second term representing the event that another estimated-feasible system "beats" the estimated-feasible system 1 in estimated objective value. The following proposition states that the overall rate function is the minimum rate function of the probability of each of these two events.

**Proposition 4.7.** *The rate function for $P\{FS\}$ is*

$$-\lim_{n\to\infty} \frac{1}{n} \log P\{FS\} = \min\left(\alpha_1 J_1(\gamma), \min_{2 \leq i \leq r}\left(\inf_{x_i \leq x_1,\, y_i \leq \gamma} (\alpha_1 I_1(x_1) + \alpha_i I_i(x_i, y_i))\right)\right).$$

*Proof.* From equation (4.23) and proposition A.2, it follows that

$$-\lim_{n\to\infty}\frac{1}{n}\log P\{FS\} = \min\Big(-\lim_{n\to\infty}\frac{1}{n}\log P\{\hat{G}_1 > \gamma\},\\ \min_{2\le i\le r}\big(-\lim_{n\to\infty}\frac{1}{n}\log P\{(\hat{G}_i \le \gamma)\cap(\hat{H}_i \le \hat{H}_1)\}\big)\Big).$$

Since assumption 4.1 satisfies assumptions 2.2 and 2.3, it follows from theorem 2.1 that

$$-\lim_{n\to\infty}\frac{1}{n}\log P\{\hat{G}_1 > \gamma\} = \alpha_1 J_1(\gamma).$$

Let us consider $-\lim_{n\to\infty}\frac{1}{n}\log P\{(\hat{G}_i \le \gamma)\cap(\hat{H}_i \le \hat{H}_1)\}$ for a system $i \neq 1$. Let

$$\Lambda^{(\alpha_1 n, \alpha_i n)}_{(H_1, H_i, G_i)}(\theta_1, \boldsymbol{\theta_i}) = \log \mathrm{E}\big[e^{\langle(\theta_1, \boldsymbol{\theta}_i), (\hat{H}_1, \hat{H}_i, \hat{G}_i)\rangle}\big]$$

be the cumulant generating function of the random vector $(\hat{H}_1, \hat{H}_i, \hat{G}_i)$. By the independence of system 1 and system $i$ (assumption 1),

$$\Lambda^{(\alpha_1 n, \alpha_i n)}_{(H_1, H_i, G_i)}(\theta_1, \boldsymbol{\theta_i}) = \Lambda^{(\alpha_1 n)}_{H_1}(\theta_1) + \Lambda^{(\alpha_i n)}_{(H_i, G_i)}(\boldsymbol{\theta}_i).$$

Under assumption 4.1,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\Lambda^{(\alpha_1 n, \alpha_i n)}_{(H_1, H_i, G_i)}(n\theta_1, n\boldsymbol{\theta_i}) &= \lim_{n\to\infty}\frac{1}{n}\Lambda^{(\alpha_1 n)}_{H_1}(n\theta_1) + \lim_{n\to\infty}\frac{1}{n}\Lambda^{(\alpha_i n)}_{(H_i, G_i)}(n\boldsymbol{\theta}_i)\\ &= \alpha_1 \log \mathrm{E}[e^{(\theta_1/\alpha_1)H_{11}}] + \alpha_i \log \mathrm{E}[e^{\langle \boldsymbol{\theta}_i/\alpha_i, (H_{i1}, G_{i1})\rangle}].\end{aligned}$$

Then by the Gärtner-Ellis theorem, $(\hat{H}_1, \hat{H}_i, \hat{G}_i)$ satisfies an LDP with good rate function

$$I_i(x_1, x_i, y_i) = \alpha_1 I_1(x_1) + \alpha_i I_i(x_i, y_i),$$

and hence

$$-\lim_{n\to\infty}\frac{1}{n}\log P\{(\hat{G}_i \le \gamma)\cap(\hat{H}_i \le \hat{H}_1)\} = \inf_{\substack{x_i \le x_1\\ y_i \le \gamma}}\Big(\alpha_1 I_1(x_1) + \alpha_i I_i(x_i, y_i)\Big). \qquad \square$$

For notational simplicity, let the rate function $\inf_{x_i \le x_1,\, y_i \le \gamma}(\alpha_1 I_1(x_1) + \alpha_i I_i(x_i, y_i))$ be denoted by $K_i(\alpha_1, \alpha_i)$. Then expanding the normal and bivariate normal rate functions, we have

$$K_i(\alpha_1, \alpha_i) = \inf_{\substack{x_i \le x_1\\ y_i \le \gamma}}\left(\frac{\alpha_1(x_1 - h_1)^2}{2\sigma_{h_1}^2} + \frac{\alpha_i}{2(1-\rho_i^2)}\left(\frac{(x_i - h_i)^2}{\sigma_{h_i}^2} - \frac{2\rho_i(x_i - h_i)(y_i - g_i)}{\sigma_{h_i}\sigma_{g_i}} + \frac{(y_i - g_i)^2}{\sigma_{g_i}^2}\right)\right).$$

The following lemma provides the location of the infimum in $K_i(\alpha_1, \alpha_i)$.

**Lemma 4.8.** *Under assumption 4.1, the infimum in $K_i(\alpha_1, \alpha_i)$ is achieved at*

$$x_1^* = x_i^* = \frac{(\alpha_1/\sigma_{h_1}^2)h_1 + (\alpha_i/\sigma_{h_i}^2)h_i}{(\alpha_1/\sigma_{h_1}^2) + (\alpha_i/\sigma_{h_i}^2)},\ y_i^* = g_i + \frac{\rho_i\sigma_{g_i}(\alpha_1/\sigma_{h_1}^2)}{(\alpha_1/\sigma_{h_1}^2) + (\alpha_i/\sigma_{h_i}^2)}\frac{(h_1 - h_i)}{\sigma_{h_i}},\ i \in \Gamma(\rho, \alpha) \tag{4.24}$$

$$x_1^* = h_1,\ x_i^* = h_i + \rho_i\sigma_{h_i}\frac{(\gamma - g_i)}{\sigma_{g_i}},\ y_i^* = \gamma,\ i \in \mathcal{S}_b(\rho) \tag{4.25}$$

$$x_1^* = x_i^* = \frac{\dfrac{\alpha_1 h_1}{\sigma_{h_1}^2} + \dfrac{\alpha_i}{\sigma_{h_i}^2(1-\rho_i^2)}\left(h_i + \rho_i\sigma_{h_i}\dfrac{(\gamma - g_i)}{\sigma_{g_i}}\right)}{\left(\dfrac{\alpha_1}{\sigma_{h_1}^2} + \dfrac{\alpha_i}{\sigma_{h_i}^2(1-\rho_i^2)}\right)},\ y_i^* = \gamma,\ i \in \mathcal{S}_w(\rho, \alpha) \tag{4.26}$$

*where*

$$\Gamma(\rho, \alpha) = \left\{ i : h_i > h_1 \ \ \text{and} \ \ g_i \leq \gamma + \rho_i\sigma_{g_i}\frac{(\sigma_{h_i}^2/\alpha_i)}{(\sigma_{h_1}^2/\alpha_1) + (\sigma_{h_i}^2/\alpha_i)}\frac{(h_i - h_1)}{\sigma_{h_i}} \right\},$$

$$\mathcal{S}_b(\rho) = \left\{ i : h_i \leq h_1 + \rho_i\sigma_{h_i}\frac{(g_i - \gamma)}{\sigma_{g_i}} \ \ \text{and} \ \ g_i > \gamma \right\},$$

$$\mathcal{S}_w(\rho, \alpha) = \left\{ i : h_i > h_1 + \rho_i\sigma_{h_i}\frac{(g_i - \gamma)}{\sigma_{g_i}} \ \ \text{and} \ \ g_i > \gamma + \rho_i\sigma_{g_i}\frac{(\sigma_{h_i}^2/\alpha_i)}{(\sigma_{h_1}^2/\alpha_1) + (\sigma_{h_i}^2/\alpha_i)}\frac{(h_i - h_1)}{\sigma_{h_i}} \right\}.$$

*Proof.* Finding the infimum in $K_i(\alpha_1, \alpha_i)$ is a convex minimization problem. Letting $\lambda_x \geq 0$ and $\lambda_y \geq 0$ be Lagrange multipliers, the KKT conditions are

$$\frac{\alpha_1(x_1^* - h_1)}{\sigma_{h_1}^2} = -\frac{\alpha_i}{1-\rho^2}\left(\frac{x_i^* - h_i}{\sigma_{h_i}^2} - \frac{\rho(y_i^* - g_i)}{\sigma_{h_i}\sigma_{g_i}}\right) = \lambda_x \tag{4.27}$$

$$\frac{\alpha_i}{1-\rho^2}\left(\frac{y_i^* - g_i}{\sigma_{g_i}^2} - \frac{\rho(x_i^* - h_i)}{\sigma_{h_i}\sigma_{g_i}}\right) = \lambda_y \tag{4.28}$$

$$\lambda_x(x_i^* - x_1^*) = 0 \tag{4.29}$$

$$\lambda_y(\gamma - y_i^*) = 0 \tag{4.30}$$

$$x_i^* - x_1^* \leq 0,\ \gamma - y_i^* \leq 0$$

*Case 1: $x_1^* > x_i^*$ and $y_i^* > \gamma$.* Then $\lambda_x = \lambda_y = 0$, which implies $x_1^* = h_1, x_i^* = h_i$, and $y_i^* = g_i$, which is not a feasible point.

*Case 2: $x_1^* = x_i^*$ and $y_i^* > \gamma$.* Then $\lambda_x > 0, \lambda_y = 0$. From equation (4.28),

$$y_i^* = g_i + \rho\sigma_{g_i}\frac{(x_i^* - h_i)}{\sigma_{h_i}}. \tag{4.31}$$

Using this value for $y_i^*$ in equation (4.27) and solving for $x_i^*$ in yields the value of $x_1^* = x_i^*$ given in equation (4.24). Substituting this value of $x_i^*$ into equation (4.31) yields the $y_i^*$ value given in equation (4.24). Since $\lambda_x > 0$ implies $x_i^* > h_1$, then $h_1 < h_i$. The second condition in set $\Gamma(\rho, \alpha_1, \alpha_i)$ results from the fact that $y_i^* > \gamma$. When $\rho = 0$, $\Gamma(\rho, \alpha_1, \alpha_i) = \Gamma$.

*Case 3:* $x_1^* > x_i^*$ *and* $y_i^* = \gamma$. Then $\lambda_x = 0, \lambda_y > 0$, and from equation (4.27), $x_1^* = h_1$. From equation (4.30), $y^* = \gamma$. Solving for $x_i^*$ in (4.27) results in the value of $x_i^*$ given in equation (4.25). Since $x_1^* = h_1 > x_i^*$, then

$$\frac{h_1 - h_i}{\sigma_{h_i}} > \frac{\rho(\gamma - g_i)}{\sigma_{g_i}}$$

Since $\lambda_y > 0$ and $\rho^2 < 1$, equation (4.28) results in the condition $g_i < \gamma$ for $\mathcal{S}_b(\rho)$. When $\rho = 0$, $\mathcal{S}_b(\rho) = \mathcal{S}_b$.

*Case 4:* $x_1^* = x_i^*$ *and* $y_i^* = \gamma$. Then $\lambda_x > 0, \lambda_y > 0$, and from equations (4.29) – (4.30), we have $x_1^* = x_i^*$ and $y_i^* = \gamma$. Solving for the value of $x_i^*$ in (4.27) yields the result in equation (4.26).

Since $\lambda_x > 0, \lambda_y > 0$, from equations (4.27)–(4.28), we have

$$x_1^* = x_i^* > h_1 \quad \text{and} \quad \frac{x_i^* - h_i}{\sigma_{h_i}} < \frac{\rho(\gamma - g_i)}{\sigma_{g_i}} \quad \text{and} \quad \frac{\gamma - g_i}{\sigma_{g_i}} > \frac{\rho(x_i^* - h_i)}{\sigma_{h_i}}.$$

Plugging in $x_i^*$ yields the conditions for $\mathcal{S}_w(\rho, \alpha_1, \alpha_i)$. When $\rho = 0$, $\mathcal{S}_w(\rho, \alpha_1, \alpha_i) = \mathcal{S}_w$. □

Recall the definitions of the sets for zero correlation as feasible and worse ($\Gamma$), infeasible and better ($\mathcal{S}_b$), and infeasible and worse ($\mathcal{S}_w$). Formally,

$$\begin{aligned} \Gamma &= \{i : h_i > h_1 \ \text{ and } \ g_i \leq \gamma\}, \\ \mathcal{S}_b &= \{i : h_i \leq h_1 \ \text{ and } \ g_i > \gamma\}, \\ \mathcal{S}_w &= \{i : h_i > h_1 \ \text{ and } \ g_i > \gamma\}. \end{aligned}$$

In the set definitions given in lemma 4.8, for positive $\rho_i$, $\Gamma(\rho, \alpha) \supseteq \Gamma$ and $\mathcal{S}_b(\rho) \supseteq \mathcal{S}_b$, while $\mathcal{S}_w(\rho, \alpha) \subseteq \mathcal{S}_w$. This result is intuitive: since the performance measures are positively correlated, a system from $\mathcal{S}_w$ that is "close enough" to $h_1$ in objective value or "close enough" to $\gamma$ in constraint value and succeeds in crossing one of these barriers will likely succeed in crossing the other. Therefore some systems in $\mathcal{S}_w$ may have rate functions similar to other systems in either $\Gamma$ or $\mathcal{S}_b$ when there is positive correlation. When there is negative correlation, the opposite occurs. Thus $\Gamma(\rho, \alpha) \subseteq \Gamma$ and $\mathcal{S}_b(\rho) \subseteq \mathcal{S}_b$, while $\mathcal{S}_w(\rho, \alpha) \supseteq \mathcal{S}_w$. Now systems in $\Gamma$ and $\mathcal{S}_b$ have rate functions similar to other systems in $\mathcal{S}_w$. For example, due to negative correlation, a system in $\Gamma$ that is "close enough" to the constraint might be declared infeasible when it "beats" system 1 in objective value. Of the two types of correlation, intuitively it seems that negative correlation is better since negative correlation may result in an increased rate of decay of the probability of false selection. However this result depends on the formulation of the problem in section 1.2.

For each set $\Gamma(\rho,\alpha)$, $\mathcal{S}_b(\rho)$, and $\mathcal{S}_w(\rho,\alpha)$, given the result from lemma 4.8, one can solve for the rate functions. These rate functions are given in the following proposition.

**Proposition 4.9.** *Under assumption 4.1, the relevant rate functions are*

$$\begin{aligned}
K_{i\in\Gamma(\rho,\alpha)}(\alpha_1,\alpha_i) =& \frac{(h_i-h_1)^2}{2(\sigma^2_{h_1}/\alpha_1+\sigma^2_{h_i}/\alpha_i)},\\
K_{i\in\mathcal{S}_b(\rho)}(\alpha_1,\alpha_i) =& \frac{\alpha_i(g_i-\gamma)^2}{2\sigma^2_{g_i}},\\
K_{i\in\mathcal{S}_w(\rho,\alpha)}(\alpha_1,\alpha_i) =& \frac{\left((h_i-h_1)-\rho_i\sigma_{h_i}\dfrac{(g_i-\gamma)}{\sigma_{g_i}}\right)^2}{2(\sigma^2_{h_1}/\alpha_1+(1-\rho_i^2)\sigma^2_{h_i}/\alpha_i)}+\frac{\alpha_i(g_i-\gamma)^2}{2\sigma^2_{g_i}}.
\end{aligned}$$

The rate functions for sets $\Gamma(\rho,\alpha)$ and $\mathcal{S}_b(\rho)$ are identical to the rate functions of $\Gamma$ and $\mathcal{S}_b$ in the independent case. That is, while the sets depend on $\rho$, rate functions for systems in these set do not depend on $\rho$. In the independent case, the rate function for systems in $\mathcal{S}_w$ comprises two added terms: one for "optimality" that is identical to the rate function for systems in $\Gamma$, and one for "feasibility" that is identical to the rate function for systems in $\mathcal{S}_b$. In the correlated case, the rate function for systems in $\mathcal{S}_w(\rho,\alpha)$ comprises two added terms: one for "optimality," adjusted for the correlation with the constraint, and one for "feasibility" that is identical to the rate function of systems in $\mathcal{S}_b(\rho)$.

The following theorem states the probability of false selection for the bivariate normal case as a function of the correlation $\rho$.

**Theorem 4.10.** *Under assumption 4.1, the rate function of the probability of false selection is*

$$\begin{aligned}
-\lim_{n\to\infty}\frac{1}{n}\log P\{FS\} = \min\Bigg(&\frac{\alpha_1(\gamma-g_1)^2}{2\sigma^2_{g_1}},\ \min_{i\in\Gamma(\rho,\alpha)}\frac{(h_i-h_1)^2}{2(\sigma^2_{h_1}/\alpha_1+\sigma^2_{h_i}/\alpha_i)},\ \min_{i\in\mathcal{S}_b(\rho)}\frac{\alpha_i(g_i-\gamma)^2}{2\sigma^2_{g_i}},\\
&\min_{i\in\mathcal{S}_w(\rho,\alpha)}\frac{\left((h_i-h_1)-\rho_i\sigma_{h_i}\dfrac{(g_i-\gamma)}{\sigma_{g_i}}\right)^2}{2(\sigma^2_{h_1}/\alpha_1+(1-\rho_i^2)\sigma^2_{h_i}/\alpha_i)}+\frac{\alpha_i(g_i-\gamma)^2}{2\sigma^2_{g_i}}\Bigg).
\end{aligned}$$

As in chapter 3, the overall rate function will be determined by the most likely of four events. In the correlated case, these events are: (i) system 1 is incorrectly estimated as infeasible; (ii) a system from $\Gamma(\rho,\alpha)$ is incorrectly estimated as optimal; (iii) a system from $\mathcal{S}_b(\rho)$ is incorrectly estimated as feasible; (iv) a system from $\mathcal{S}_w(\rho,\alpha)$ is incorrectly estimated as to be optimal and feasible.

### 4.3.3 Reconciliation of First and Second Probability of False Selection Formulations

In this section, we reconcile the two $P\{FS\}$ formulations and show that the overall rate of decay of the $P\{FS\}$ is identical in both cases. Therefore for the sake of simplicity, we proceed with the $P\{FS\}$ formulation that yields the simplest analysis,

$$P\{FS\} = P\{(\hat{G}_1 > \gamma) \cup (\cup_{i=2}^{r} (\hat{G}_i \leq \gamma) \cap (\hat{H}_i \leq \hat{H}_1))\}.$$

The following proposition states that the first and second formulations yield identical rate functions.

**Proposition 4.11.** *For every $\boldsymbol{\alpha}$, the rate functions from the $P\{FS\}$ formulations presented in theorem 4.6 and theorem 4.10 are equivalent.*

*Proof.* Recall that in addition to the rate function for the feasibility of system 1, the first formulation produces sets $\Gamma_\ell(\rho, \alpha), \Gamma_u(\rho, \alpha), \mathcal{S}_b(\rho), \mathcal{S}_{w,\ell}(\rho, \alpha)$, and $\mathcal{S}_{w,u}(\rho, \alpha)$, while the second formulation produces sets $\Gamma(\rho, \alpha), \mathcal{S}_b(\rho)$, and $\mathcal{S}_w(\rho, \alpha)$. Further, the formulations and rate functions for each $\mathcal{S}_b(\rho)$ are identical, while $\Gamma_\ell(\rho, \alpha) \subseteq \Gamma(\rho, \alpha)$ and $\mathcal{S}_{w,\ell}(\rho, \alpha) \subseteq \mathcal{S}_w(\rho, \alpha)$.

By proposition 4.4 and by the requirement that $g_1 < \gamma$, if $\rho_1 \leq 0$, $\Gamma_u(\rho, \alpha)$ and $\mathcal{S}_{w,u}(\rho, \alpha)$ are empty. Thus when $\rho_1 \leq 0$, the rate functions for the first and second formulation are identical.

Now let us suppose $\rho_1 > 0$ and that $\Gamma_u(\rho, \alpha)$ and $\mathcal{S}_{w,u}(\rho, \alpha)$ are nonempty. Then the rate function for systems in $\Gamma_u(\rho, \alpha)$ minus the rate function for the feasibility of system 1 is,

$$\frac{\left((h_i - h_1) - \rho_1 \sigma_{h_1} \dfrac{(\gamma - g_1)}{\sigma_{g_1}}\right)^2}{2((1-\rho_1^2)\sigma_{h_1}^2/\alpha_1 + \sigma_{h_i}^2/\alpha_i)} > 0,$$

and the rate function for systems in $\mathcal{S}_{w,u}(\rho, \alpha)$ minus the rate function for the feasibility of system 1 is,

$$\frac{\left((h_i - h_1) - \rho_1 \sigma_{h_1} \dfrac{(\gamma - g_1)}{\sigma_{g_1}} - \rho_i \sigma_{h_i} \dfrac{g_i - \gamma}{\sigma_{g_i}}\right)^2}{2((1-\rho_1^2)\sigma_{h_1}^2/\alpha_1 + (1-\rho_i^2)\sigma_{h_i}^2/\alpha_i)} + \frac{\alpha_i (g_i - \gamma)^2}{2\sigma_{g_i}^2} > 0.$$

Thus whenever a system falls in $\Gamma_u(\rho, \alpha)$ or $\mathcal{S}_{w,u}(\rho, \alpha)$, the rate function of that system is larger than the rate function of the feasibility of system 1. Thus for any $\boldsymbol{\alpha}$, no system in $\Gamma_u(\rho, \alpha)$ or $\mathcal{S}_{w,u}(\rho, \alpha)$ will determine the overall rate function of the $P\{FS\}$. Hence the overall rate of decay of the $P\{FS\}$ for the two formulations are identical. $\square$

Further, when we wish to optimally allocate among the systems, we wish to maximize the minimum rate function. Thus the optimal allocation, which we will derive in the next section, would never place any systems in $\Gamma_u(\rho, \alpha)$ or $\mathcal{S}_{w,u}(\rho, \alpha)$.

## 4.4 Optimal Allocation Strategy

Proceeding with the results of the second formulation of the $P\{FS\}$ presented in section 4.3.2, we wish to allocate the $\alpha_i$'s to solve the following optimization problem:

$$\begin{aligned} \max \quad & \min\left(\alpha_1 J_1(\gamma),\ \min_{2\le i\le r} K_i(\alpha_1,\alpha_i)\right) \quad s.t. \\ & \sum_{i=1}^{r} \alpha_i = 1, \quad \alpha_i \ge 0. \end{aligned} \tag{4.32}$$

This problem can be re-expressed as

$$\begin{aligned} \text{Problem } Q: \qquad \max \quad z \quad & s.t. \\ \alpha_1 J_1(\gamma) &\ge z, \\ K_i(\alpha_1,\alpha_i) &\ge z, \quad 2 \le i \le r, \\ \sum_{i=1}^{r} \alpha_i = 1,&\ \alpha_i \ge 0. \end{aligned}$$

**Proposition 4.12.** *The rate function $K_i(\alpha_1, \alpha_i)$ is a concave function of $(\alpha_1, \alpha_i)$.*

*Proof.* Let $\boldsymbol{\alpha} = (\alpha_1, \alpha_i)$ and $\bar{\boldsymbol{\alpha}} = (\bar{\alpha}_1, \bar{\alpha}_i)$, such that $\boldsymbol{\alpha} \neq \bar{\boldsymbol{\alpha}}$. Further, let $\lambda \in (0, 1)$. Then

$$\begin{aligned} K_i(\lambda\alpha_1 + &(1-\lambda)\bar{\alpha}_1, \lambda\alpha_i + (1-\lambda)\bar{\alpha}_i) \\ &= \inf_{x_i\le x_1, y_i\le\gamma} [(\lambda\alpha_1 + (1-\lambda)\bar{\alpha}_1)I_1(x_1) + (\lambda\alpha_i + (1-\lambda)\bar{\alpha}_i)I_i(x_i,y_i)] \\ &= \inf_{x_i\le x_1, y_i\le\gamma} [\lambda(\alpha_1 I_1(x_1) + \alpha_i I_i(x_i,y_i)) + (1-\lambda)(\bar{\alpha}_1 I_1(x_1) + \bar{\alpha}_i I_i(x_i,y_i))] \\ &\ge \inf_{x_i\le x_1, y_i\le\gamma} \lambda[\alpha_1 I_1(x_1) + \alpha_i I_i(x_i,y_i)] + \inf_{x_i\le x_1, y_i\le\gamma} (1-\lambda)[\bar{\alpha}_1 I_1(x_1) + \bar{\alpha}_i I_i(x_i,y_i)] \\ &= \lambda K_i(\alpha_1,\alpha_i) + (1-\lambda)K_i(\bar{\alpha}_1,\bar{\alpha}_i). \end{aligned}$$

□

Changing values of $(\alpha_1, \alpha_i)$ can change the set assignment for a particular function, and hence change the rate function. So the question may arise, are the rate functions still concave across these set boundary points? The answer is yes, since changing $(\alpha_1, \alpha_i)$ amounts to changing the location of the infimum over $x_i \le x_1, y_i \le \gamma$ in the expression of $K_i(\alpha_1, \alpha_i)$. Since the concavity of $K_i(\alpha_1, \alpha_i)$

as a function of $(\alpha_1, \alpha_i)$ does not depend on where this infimum occurs, and since $\alpha_1 J_1(\gamma)$ is a concave function of $\alpha_1$, problem $Q$ is a concave maximization problem.

Let problem $Q^*_{BVN}$ be as problem $Q$ except with the inequality constraints on $K_i(\alpha_1, \alpha_i)$ replaced by equality constraints and forcing $\alpha_i > 0$. Following a proof similar to that given in proposition 3.8, it can be shown that problems $Q$ and $Q^*_{BVN}$ are equivalent, that is, they have an identical optimal solution and optimal value. Problem $Q^*_{BVN}$ is written as

$$\begin{aligned}
\text{Problem } Q^*_{BVN}: \qquad & \max \quad z \quad s.t. \\
& \frac{\alpha_1(\gamma - g_1)^2}{2\sigma^2_{g_1}} \geq z, \\
& \frac{(h_1 - h_i)^2}{2(\sigma^2_{h_1}/\alpha_1 + \sigma^2_{h_i}/\alpha_i)} = z, \ i \in \Gamma(\rho, \alpha) \\
& \frac{\alpha_i(\gamma - g_i)^2}{2\sigma^2_{g_i}} = z, \ i \in \mathcal{S}_b(\rho) \\
& \frac{\left((h_i - h_1) - \rho_i \sigma_{h_i} \dfrac{(g_i - \gamma)}{\sigma_{g_i}}\right)^2}{2(\sigma^2_{h_1}/\alpha_1 + (1 - \rho_i^2)\sigma^2_{h_i}/\alpha_i)} + \frac{\alpha_i(g_i - \gamma)^2}{2\sigma^2_{g_i}} = z, \ i \in \mathcal{S}_w(\rho, \alpha) \\
& \sum_{i=1}^{r} \alpha_i = 1, \ \alpha_i > 0.
\end{aligned}$$

We propose that the solution to problem $Q^*_{BVN}$ be obtained using a solver.

### 4.4.1 An Approximate Closed-Form Solution

This section presents a closed-form solution to problem $Q^*_{BVN}$ under the assumption that $\alpha^*_1 \gg \alpha^*_i$. Thus the solutions in this section are only appropriate when such an assumption is reasonable. For example, such a scenario may arise when the number of systems in $\Gamma(\rho, \alpha)$ and $\mathcal{S}_w(\rho, \alpha)$ is large.

The $\alpha^*_1 \gg \alpha^*_i$ assumption appears often in previous OCBA literature (see, e.g., Lee et al., 2011). Notationally, we remove the "stars" from $\alpha$ in this section to emphasize that the $\alpha$ value presented here is an approximation to the true optimal solution. First, let us use the $\alpha_1 \gg \alpha_i$ assumption to simplify the characterization of the sets $\Gamma(\rho, \alpha)$ and $\mathcal{S}_w(\rho, \alpha)$. Define the sets $\Gamma^{\alpha_1 \gg \alpha_i}(\rho)$ and $\mathcal{S}_w^{\alpha_1 \gg \alpha_i}(\rho)$ as

$$\begin{aligned}
\Gamma^{\alpha_1 \gg \alpha_i}(\rho) &= \left\{ i : h_i > h_1 \ \text{ and } \ g_i < \gamma + \rho_i \sigma_{g_i} \frac{(h_i - h_1)}{\sigma_{h_i}} \right\}, \\
\mathcal{S}_w^{\alpha_1 \gg \alpha_i}(\rho) &= \left\{ i : h_i > h_1 + \rho_i \sigma_{h_i} \frac{(g_i - \gamma)}{\sigma_{g_i}} \ \text{ and } \ g_i > \gamma + \rho_i \sigma_{g_i} \frac{(h_i - h_1)}{\sigma_{h_i}} \right\}.
\end{aligned}$$

Set $\mathcal{S}_b(\rho)$ needs no modification, as it does not depend on $(\alpha_1, \alpha_i)$. The corresponding rate functions are

$$K_{i\in\Gamma^{\alpha_1\gg\alpha_i}(\rho)}(\alpha_1,\alpha_i) = \frac{\alpha_i(h_1-h_i)^2}{2\sigma_{h_i}^2} = \alpha_i I_i(h_1),$$
$$K_{i\in\mathcal{S}_b(\rho)}(\alpha_1,\alpha_i) = \frac{\alpha_i(\gamma-g_i)^2}{2\sigma_{g_i}^2} = \alpha_i J_i(\gamma),$$
$$K_{i\in\mathcal{S}_w^{\alpha_1\gg\alpha_i}(\rho)}(\alpha_1,\alpha_i) = \frac{\alpha_i}{2(1-\rho_i^2)}\left(\frac{(h_1-h_i)^2}{\sigma_{h_i}^2} - 2\rho_i\frac{(h_1-h_i)}{\sigma_{h_i}}\frac{(\gamma-g_i)}{\sigma_{g_i}} + \frac{(\gamma-g_i)^2}{\sigma_{g_i}^2}\right) = \alpha_i I_i(h_1,\gamma).$$

Let $\mathbb{I}_{(\cdot)}$ denote the indicator function. Under assumption 4.1 and the assumption that $\alpha_1^* \gg \alpha_i^*$, it follows from the equality constraints in problem $Q_{BVN}^*$ that the sample allocated to systems other than system 1 should follow the proportion

$$\frac{\alpha_i}{\alpha_k} = \frac{I_k(h_1)\mathbb{I}_{(k\in\Gamma^{\alpha_1\gg\alpha_i}(\rho))} + J_k(\gamma)\mathbb{I}_{(k\in\mathcal{S}_b(\rho))} + I_k(h_1,\gamma)\mathbb{I}_{(k\in\mathcal{S}_w^{\alpha_1\gg\alpha_i}(\rho))}}{I_i(h_1)\mathbb{I}_{(i\in\Gamma^{\alpha_1\gg\alpha_i}(\rho))} + J_i(\gamma)\mathbb{I}_{(i\in\mathcal{S}_b(\rho))} + I_i(h_1,\gamma)\mathbb{I}_{(i\in\mathcal{S}_w^{\alpha_1\gg\alpha_i}(\rho))}}.$$

# Chapter 5

# Multivariate Normal: Multiple Constraints

Consider the context of problem $P$ fully (see section 1.2), with multiple constraints. In this chapter, we allow correlation between the objective function and constraints, but we require that the underlying random variables follow a multivariate normal distribution. We provide a characterization of the rate of decay of the probability of false selection and an optimal allocation strategy that is the result of a concave maximization problem. We also provide an implementable sequential algorithm, as well as numerical results displaying its performance.

## 5.1 Chapter Organization

Section 5.2 provides initial assumptions for the chapter. Section 5.3 contains the general rate function of the probability of false selection and section 5.4 contains an optimal allocation strategy that maximizes the rate of decay of the probability of false selection. In Section 5.5 we make the multivariate assumption which allows the derivation of the consistent estimator of the optimal allocation and the implementable fully sequential algorithm in Section 5.6. Numerical results are presented in Section 5.7.

## 5.2 Assumptions for Chapter 5

While the main focus of this chapter is on the case in which the underlying random variables $(H_i, G_{i1}, \ldots, G_{is}), i \leq r$, follow a multivariate normal distribution, the results we present in sections 5.3 and 5.4 hold in a much more general context. Therefore we postpone making the multivariate normal assumption until our results require it. For now, in addition to assumptions 1 and 2, we make two additional assumptions as is typical in large deviations contexts.

Let $(H_i, \boldsymbol{G}_i) = (H_i, G_{i1}, \ldots, G_{is})$ be random variables with mean $(h_i, \boldsymbol{g}_i) = (h_i, g_{i1}, \ldots, g_{is})$. Let us define

$$(\bar{H}_i(n), \bar{\boldsymbol{G}}_i(n)) = (\bar{H}_i(n), \bar{G}_{i1}(n), \ldots, \bar{G}_{is}(n)) = \left( \frac{1}{n}\sum_{k=1}^{n} H_{ik}, \frac{1}{n}\sum_{k=1}^{n} G_{i1k}, \ldots, \frac{1}{n}\sum_{k=1}^{n} G_{isk} \right).$$

We will use $(\hat{H}_i, \hat{\boldsymbol{G}}_i) \equiv (\bar{H}_i(\alpha_i n), \bar{\boldsymbol{G}}_i(\alpha_i n))$ as shorthand for the estimator of $(h_i, \boldsymbol{g}_i)$ after scaling the sample size by $\alpha_i > 0$, the proportion of the total sample $n$ which is allocated to system $i$.

Let $\Lambda^{(n)}_{H_i}(\theta) = \log \mathrm{E}\big[e^{\theta \bar{H}_i(n)}\big]$, $\Lambda^{(n)}_{G_{ij}}(\theta) = \log \mathrm{E}\big[e^{\theta \bar{G}_{ij}(n)}\big]$, and

$$\Lambda^{(n)}_{(H_i, \boldsymbol{G}_i)}(\boldsymbol{\theta}) = \log \mathrm{E}\big[e^{\langle \boldsymbol{\theta}, (\bar{H}_i(n), \bar{\boldsymbol{G}}_i(n))\rangle}\big]$$

be the cumulant generating functions of $\bar{H}_i(n)$, $\bar{G}_{ij}(n)$, and $(\bar{H}_i(n), \bar{\boldsymbol{G}}_i(n))$, respectively, where $\theta \in \mathbb{R}$, $\boldsymbol{\theta} \in \mathbb{R}^{s+1}$, and $\langle \cdot, \cdot \rangle$ denotes the dot product. Let the effective domain of a function $f(\cdot)$ be denoted by $\mathcal{D}_f = \{x : f(x) < \infty\}$ and its interior by $\mathcal{D}^\circ_f$. Let $f'(x)$ denote the derivative of $f$ with respect to the argument $x$, and $\nabla f(\boldsymbol{x})$ denote the gradient of $f$.

As is usual in LD contexts, we make the following assumption.

**Assumption 5.1.** *Let the following hold for each $i \leq r$ and $j \leq s$:*

*(1) the limit $\Lambda_{(H_i, \boldsymbol{G}_i)}(\boldsymbol{\theta}) = \lim_{n\to\infty} \frac{1}{n}\Lambda^{(n)}_{(H_i, \boldsymbol{G}_i)}(n\boldsymbol{\theta})$ exists as an extended real number for all $\boldsymbol{\theta} \in \mathbb{R}^{s+1}$, where we denote $\Lambda_{H_i}(\theta) = \lim_{n\to\infty} \frac{1}{n}\Lambda^{(n)}_{H_i}(n\theta)$ and $\Lambda_{G_{ij}}(\theta) = \lim_{n\to\infty} \frac{1}{n}\Lambda^{(n)}_{G_{ij}}(n\theta)$ for all $\theta \in \mathbb{R}$;*

*(2) the origin belongs to the interior of $\mathcal{D}_{\Lambda_{(H_i, \boldsymbol{G_i})}}$, that is, $0 \in \mathcal{D}^\circ_{\Lambda_{(H_i, \boldsymbol{G_i})}}$;*

*(3) $\Lambda_{(H_i, \boldsymbol{G}_i)}(\boldsymbol{\theta})$ is strictly convex and $C^\infty$ on $\mathcal{D}^\circ_{\Lambda_{(H_i, \boldsymbol{G_i})}}$;*

*(4) $\Lambda_{(H_i, \boldsymbol{G}_i)}(\boldsymbol{\theta})$ is steep, that is, for any sequence $\{\boldsymbol{\theta}(n)\} \in \mathcal{D}_{\Lambda_{(H_i, \boldsymbol{G_i})}}$ converging to a boundary point of $\mathcal{D}_{\Lambda_{(H_i, \boldsymbol{G_i})}}$, then it holds that $\lim_{n\to\infty} |\nabla \Lambda_{(H_i, \boldsymbol{G}_i)}(\boldsymbol{\theta}(n))| = \infty$.*

Assumption 5.1 implies that $(\bar{H}_i(n), \bar{\boldsymbol{G}}_i(n)) \to (h_i, \boldsymbol{g}_i)$ wp1 (see Bucklew, 2003, remark 3.2.1). Assumption 5.1 also ensures that $\bar{H}_i(n)$, $\bar{G}_{ij}(n)$, and $(\bar{H}_i(n), \bar{\boldsymbol{G}}_i(n))$ satisfy the large deviations principle (LDP) with good rate functions

$$I_i(x) = \sup_{\theta \in \mathbb{R}} \{\theta x - \Lambda_{H_i}(\theta)\}, \; J_{ij}(y) = \sup_{\theta \in \mathbb{R}} \{\theta y - \Lambda_{G_{ij}}(\theta)\},$$

and

$$I_i(x, \boldsymbol{y}) = \sup_{\boldsymbol{\theta} \in \mathbb{R}^{s+1}} \{\langle \boldsymbol{\theta}, (x, \boldsymbol{y}) \rangle - \Lambda_{(H_i, \boldsymbol{G}_i)}(\boldsymbol{\theta})\},$$

respectively (see, e.g., Dembo and Zeitouni, 1998, p.44). Assumption 5.1(3) is stronger than what is needed for the Gärtner-Ellis theorem to hold. However, we require $\Lambda_{H_i}(\theta)$, $\Lambda_{G_{ij}}(\theta)$, and $\Lambda_{(H_i,\boldsymbol{G}_i)}(\boldsymbol{\theta})$ to be strictly convex and $C^\infty$ on the interiors of their respective domains so that $I_i(x)$, $J_{ij}(y)$, and $I_i(x, \boldsymbol{y})$ are strictly convex and $C^\infty$ for $x \in \mathcal{F}^\circ_{H_i} = \text{int}\{\Lambda'_{H_i}(\theta) : \theta \in \mathcal{D}^\circ_{\Lambda_{H_i}}\}$, $y \in \mathcal{F}^\circ_{G_{ij}} = \text{int}\{\Lambda'_{G_{ij}}(\theta) : \theta \in \mathcal{D}^\circ_{\Lambda_{G_{ij}}}\}$, and $(x, \boldsymbol{y}) \in \mathcal{F}^\circ_{(H_i,\boldsymbol{G_i})} = \text{int}\{\nabla\Lambda_{(H_i,\boldsymbol{G}_i)}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \mathcal{D}^\circ_{\Lambda_{(H_i,\boldsymbol{G_i})}}\}$.

Let $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_s)$ and let $\mathcal{F}^c_d$ denote the closure of the convex hull of $\{(h_i, \boldsymbol{\gamma}) : (h_i, \boldsymbol{\gamma}) \in \mathbb{R}^{s+1}\}$. We assume the following.

**Assumption 5.2.** *The closure of the convex hull of all points $(h_i, \boldsymbol{\gamma}) \in \mathbb{R}^{s+1}$ is a subset of the intersection of the effective domains of the rate functions $I_i(x, \boldsymbol{y})$ for all $i \leq r$, that is, $\mathcal{F}^c_d \subset \cap_{i=1}^r \mathcal{F}^\circ_{(H_i,\boldsymbol{G_i})}$.*

As in previous chapters, assumption 5.2 ensures that there exists a nonzero probability that any system $i \leq r$ may be falsely selected as the best feasible system.

## 5.3 Rate Function of the Probability of False Selection

Let us formulate the probability of false selection as the probability that system 1 is declared infeasible or that another system $i$ is declared feasible and beats system 1 in estimated objective value. That is,

$$P\{FS\} = P\{[\underbrace{\cup_{j=1}^s \hat{G}_{1j} > \gamma_j}_{\substack{\text{system 1} \\ \text{estimated} \\ \text{infeasible}}}] \cup_{i=2}^r [(\underbrace{\cap_{j=1}^s \hat{G}_{ij} \leq \gamma_j}_{\substack{\text{system i} \\ \text{estimated} \\ \text{feasible}}}) \cap (\underbrace{\hat{H}_i \leq \hat{H}_1}_{\substack{\text{system i} \\ \text{“beats”} \\ \text{system 1}}})]\} \tag{5.1}$$

The $P\{FS\}$ in equation (5.1) has lower bound

$$\max\left(P\{\cup_{j=1}^s \hat{G}_{1j} > \gamma_j\}, \max_{2 \leq i \leq r}\left(P\{(\cap_{j=1}^s \hat{G}_{ij} \leq \gamma_j) \cap (\hat{H}_i \leq \hat{H}_1)\}\right)\right)$$

and upper bound

$$r \times \max\left(P\{\cup_{j=1}^s \hat{G}_{1j} > \gamma_j\}, \max_{2 \leq i \leq r}\left(P\{(\cap_{j=1}^s \hat{G}_{ij} \leq \gamma_j) \cap (\hat{H}_i \leq \hat{H}_1)\}\right)\right)$$

such that

$$\lim_{n\to\infty}\frac{1}{n}\log P\{FS\} = \lim_{n\to\infty}\frac{1}{n}\log\max\Big(P\{\cup_{j=1}^{s}\hat{G}_{1j} > \gamma_j\}, \max_{2\le i\le r} P\{(\cap_{j=1}^{s}\hat{G}_{ij}\le\gamma_j)\cap(\hat{H}_i\le\hat{H}_1)\}\Big). \tag{5.2}$$

As in previous chapters, there are two main terms: one for the feasibility of system 1, and another term representing the event that system 1 is "beaten" in estimated objective value by an estimated-feasible system. The following proposition states that the overall rate function is the minimum rate function of the probability of each of these two events.

**Theorem 5.1.** *The rate function for $P\{FS\}$ is,*

$$-\lim_{n\to\infty}\frac{1}{n}\log P\{FS\} = \min\Big(\min_{j\in\mathcal{C}_F^1}\alpha_1 J_{1j}(\gamma_j), \min_{2\le i\le r}\Big(\inf_{x_i\le x_1,\, \boldsymbol{y_i}\le\boldsymbol{\gamma}}(\alpha_1 I_1(x_1)+\alpha_i I_i(x_i,\boldsymbol{y_i}))\Big)\Big). \tag{5.3}$$

*Proof.* From equation (5.2) and proposition A.2, it follows that

$$\begin{aligned}-\lim_{n\to\infty}\frac{1}{n}\log P\{FS\} = \min\Big(&-\lim_{n\to\infty}\frac{1}{n}\log P\{\cup_{j=1}^{s}\hat{G}_{1j}>\gamma_j\},\\ &\min_{2\le i\le r}\big(-\lim_{n\to\infty}\frac{1}{n}\log P\{(\cap_{j=1}^{s}\hat{G}_{ij}\le\gamma_j)\cap(\hat{H}_i\le\hat{H}_1)\}\big)\Big).\end{aligned}$$

Under assumptions 5.1 and 5.2, it follows from theorem 3.4 that

$$-\lim_{n\to\infty}\frac{1}{n}\log P\{\cup_{j=1}^{s}\hat{G}_{1j}>\gamma_j\} = \min_{j\in\mathcal{C}_F^1}\alpha_1 J_{1j}(\gamma_j).$$

Let us consider $-\lim_{n\to\infty}\frac{1}{n}\log P\{(\cap_{j=1}^{s}\hat{G}_{ij}\le\gamma_j)\cap(\hat{H}_i\le\hat{H}_1)\}$ for a system $i\neq 1$. For $(\theta_1,\boldsymbol{\theta_i})\in\mathbb{R}^{s+2}$, let the cumulant generating function of $(\hat{H}_1,\hat{H}_i,\hat{\boldsymbol{G}}_i)$ be denoted $\Lambda^{(\alpha_1 n,\alpha_i n)}_{(H_1,H_i,\boldsymbol{G}_i)}(\theta_1,\boldsymbol{\theta_i})$. By the independence of system 1 and system $i$ (assumption 1),

$$\Lambda^{(\alpha_1 n,\alpha_i n)}_{(H_1,H_i,\boldsymbol{G}_i)}(\theta_1,\boldsymbol{\theta_i}) = \Lambda^{(\alpha_1 n)}_{H_1}(\theta_1)+\Lambda^{(\alpha_i n)}_{(H_i,\boldsymbol{G}_i)}(\boldsymbol{\theta_i}).$$

Under assumption 5.1(1),

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\Lambda^{(\alpha_1 n,\alpha_i n)}_{(H_1,H_i,\boldsymbol{G}_i)}(n\theta_1,n\boldsymbol{\theta_i}) &= \lim_{n\to\infty}\frac{1}{n}\Lambda^{(\alpha_1 n)}_{H_1}(n\theta_1)+\lim_{n\to\infty}\frac{1}{n}\Lambda^{(\alpha_i n)}_{(H_i,\boldsymbol{G}_i)}(n\boldsymbol{\theta_i})\\ &= \alpha_1\Lambda_{H_1}(\theta_1/\alpha_1)+\alpha_i\Lambda_{(H_i,\boldsymbol{G}_i)}(\boldsymbol{\theta_i}/\alpha_i).\end{aligned}$$

Then by the Gärtner-Ellis theorem, $(\hat{H}_1, \hat{H}_i, \hat{\boldsymbol{G}}_i)$ satisfies the LDP with good rate function

$$\begin{aligned} I_i(x_1, x_i, \boldsymbol{y_i}) &= \alpha_1 \sup_{\theta_1/\alpha_1} \left\{ \frac{\theta_1}{\alpha_1}\, x_1 - \Lambda_{H_1}\left(\frac{\theta_1}{\alpha_1}\right) \right\} + \alpha_i \sup_{\boldsymbol{\theta_i}/\alpha_i} \left\{ \left\langle \frac{\boldsymbol{\theta_i}}{\alpha_i}, (x_i, \boldsymbol{y_i}) \right\rangle - \Lambda_{(H_i, \boldsymbol{G_i})}\left(\frac{\boldsymbol{\theta_i}}{\alpha_i}\right) \right\} \\ &= \alpha_1 I_1(x_1) + \alpha_i I_i(x_i, \boldsymbol{y_i}), \end{aligned}$$

and hence

$$-\lim_{n\to\infty} \frac{1}{n} \log P\{(\cap_{j=1}^s \hat{G}_{ij} \le \gamma_j) \cap (\hat{H}_i \le \hat{H}_1)\} = \inf_{x_i \le x_1,\, \boldsymbol{y_i} \le \boldsymbol{\gamma}} \left(\alpha_1 I_1(x_1) + \alpha_i I_i(x_i, \boldsymbol{y_i})\right). \qquad \square$$

While it was possible to separate systems into sets in chapter 3 and in the one-constraint bivariate normal case in chapter 4, in this case, there are $2^{s+1}$ potential sets and rate functions that result from the Karush-Kuhn Tucker (KKT) conditions on the infimum in the expression $\inf_{x_i \le x_1,\, \boldsymbol{y_i} \le \boldsymbol{\gamma}} (\alpha_1 I_1(x_1) + \alpha_i I_i(x_i, \boldsymbol{y_i}))$. This large number of potential sets and rate functions makes it difficult to categorize the rate functions of systems in this manner.

## 5.4 Optimal Allocation Strategy

For notational simplicity, let the rate function $\inf_{x_i \le x_1,\, \boldsymbol{y_i} \le \boldsymbol{\gamma}} (\alpha_1 I_1(x_1) + \alpha_i I_i(x_i, \boldsymbol{y_i}))$ be denoted by $K_i(\alpha_1, \alpha_i)$. Let us allocate the $\alpha_i$'s to solve the following optimization problem:

$$\begin{aligned} \max \quad & \min\left( \min_{j \in \mathcal{C}_F^1} \alpha_1 J_{1j}(\gamma_j),\ \min_{2 \le i \le r} K_i(\alpha_1, \alpha_i) \right) \quad s.t. \\ & \sum_{i=1}^r \alpha_i = 1, \quad \alpha_i \ge 0. \end{aligned} \tag{5.4}$$

The problem in (5.4) can be re-expressed as:

$$\begin{aligned} \text{Problem } Q: \qquad \max \quad & z \quad s.t. \\ \alpha_1 J_{1j}(\gamma_j) &\ge z, \quad j \in \mathcal{C}_F^1 \\ K_i(\alpha_1, \alpha_i) &\ge z, \quad 2 \le i \le r, \\ \sum_{i=1}^r \alpha_i = 1,\ & \alpha_i \ge 0. \end{aligned}$$

where the values of $K_i(\alpha_1, \alpha_i)$ are obtained by solving

$$\text{Problem } K_i: \qquad \min \quad \alpha_1 I_1(x_1) + \alpha_i I_i(x_i, \boldsymbol{y_i}) \quad s.t.$$
$$x_i \leq x_1,$$
$$\boldsymbol{y_i} \leq \boldsymbol{\gamma}.$$

for each $i = 2, \ldots, r$.

Consider the structure of problem $Q$. Specifically, the rate function $K_i(\alpha_1, \alpha_i)$ is a concave function of $(\alpha_1, \alpha_i)$, which leads to the result that problem $Q$ is a concave maximization problem. Lemma 5.2 and proposition 5.3 state these results formally.

**Lemma 5.2.** *The rate function $K_i(\alpha_1, \alpha_i)$ is a concave function of $(\alpha_1, \alpha_i)$.*

*Proof.* Let $\boldsymbol{\alpha} = (\alpha_1, \alpha_i)$ and $\bar{\boldsymbol{\alpha}} = (\bar{\alpha}_1, \bar{\alpha}_i)$, such that $\boldsymbol{\alpha} \neq \bar{\boldsymbol{\alpha}}$. Further, let $\lambda \in (0, 1)$. Then

$$\begin{aligned}
&K_i(\lambda\alpha_1 + (1-\lambda)\bar{\alpha}_1, \lambda\alpha_i + (1-\lambda)\bar{\alpha}_i) \\
&\quad = \inf_{x_i \leq x_1,\, \boldsymbol{y_i} \leq \boldsymbol{\gamma}} [(\lambda\alpha_1 + (1-\lambda)\bar{\alpha}_1) I_1(x_1) + (\lambda\alpha_i + (1-\lambda)\bar{\alpha}_i) I_i(x_i, \boldsymbol{y_i})] \\
&\quad = \inf_{x_i \leq x_1,\, \boldsymbol{y_i} \leq \boldsymbol{\gamma}} [\lambda(\alpha_1 I_1(x_1) + \alpha_i I_i(x_i, \boldsymbol{y_i})) + (1-\lambda)(\bar{\alpha}_1 I_1(x_1) + \bar{\alpha}_i I_i(x_i, \boldsymbol{y_i}))] \\
&\quad \geq \inf_{x_i \leq x_1,\, \boldsymbol{y_i} \leq \boldsymbol{\gamma}} \lambda[\alpha_1 I_1(x_1) + \alpha_i I_i(x_i, \boldsymbol{y_i})] + \inf_{x_i \leq x_1,\, \boldsymbol{y_i} \leq \boldsymbol{\gamma}} (1-\lambda)[\bar{\alpha}_1 I_1(x_1) + \bar{\alpha}_i I_i(x_i, \boldsymbol{y_i})] \\
&\quad = \lambda K_i(\alpha_1, \alpha_i) + (1-\lambda) K_i(\bar{\alpha}_1, \bar{\alpha}_i).
\end{aligned}$$
□

**Proposition 5.3.** *Problem $Q$ is a concave maximization problem.*

*Proof.* The result proceeds from the following: (i) lemma 5.2 holds for all $i = 2, \ldots, r$; (ii) $\alpha_1 J_{1j}(\gamma_j)$ is a strictly concave function of $\alpha_1$ for all $j \in \mathcal{C}_F^1$, (iii) the minimum of concave functions is also concave, and (iv) the feasible set $\sum_{i=1}^r \alpha_i = 1, \boldsymbol{\alpha} > 0$ is convex. □

Under assumption 5.1, it seems that the rate function $K_i(\alpha_1, \alpha_i)$ is strictly concave in $\boldsymbol{\alpha}$ and hence problem $Q$ is a strictly concave maximization problem. However the proof of this property is not obvious, and it is presented as the following conjecture 5.4.

**Conjecture 5.4.** *The rate function $K_i(\alpha_1, \alpha_i)$ is a strictly concave function of $(\alpha_1, \alpha_i)$, and problem $Q$ is a strictly concave maximization problem.*

Let the rate function in equation (5.4) be called $z(\boldsymbol{\alpha})$, that is,

$$z(\boldsymbol{\alpha}) = \min\left( \min_{j \in \mathcal{C}_F^1} \alpha_1 J_{1j}(\gamma_j), \min_{2 \leq i \leq r} K_i(\alpha_1, \alpha_i) \right).$$

Since $\alpha_1 J_{1j}(\gamma_j)$ and $K_i(\alpha_1, \alpha_i)$ are continuous functions of $\boldsymbol{\alpha}$ for all $j \in \mathcal{C}_F^1$ and $i = 2, \ldots, r$, and the minimum of a finite number of continuous functions is continuous, then $z(\boldsymbol{\alpha})$ is a continuous function of $\boldsymbol{\alpha}$. Since problem $Q$ is the maximization of a continuous function on a compact set, a solution exists. The following proposition states this result without a formal proof.

**Proposition 5.5.** *There exists at least one optimal solution $\boldsymbol{\alpha^*} = (\alpha_1^*, \alpha_2^*, \ldots, \alpha_r^*)$ to problem $Q$.*

To obtain a solution to problem $Q$, we must also solve problem $K_i$ for each $i = 2, \ldots, r$. Under assumption 5.1, problem $K_i$ is a strictly convex minimization problem in $(x_1, x_i, \boldsymbol{y_i})$ for all $i = 2, \ldots, r$. Due to the structure of problem $K_i$ and its feasible region, the optimal solution $(x_1^*, x_i^*, \boldsymbol{y_i^*})$ exists. Proposition 5.6 formally states this result.

**Proposition 5.6.** *Under assumptions 5.1 and 5.2, problem $K_i$ is a strictly convex minimization problem in $(x_1, x_i, \boldsymbol{y_i})$, and its unique solution, $\boldsymbol{\mathfrak{z}^*}(\alpha_1, \alpha_i) = (x_1^*(\alpha_1, \alpha_i), x_i^*(\alpha_1, \alpha_i), \boldsymbol{y_i^*}(\alpha_1, \alpha_i))$, exists for all $i = 2, \ldots, r$.*

*Proof.* Under assumption 5.1, the rate function $\alpha_1 I_1(x_1) + \alpha_i I_i(x_i, \boldsymbol{y_i})$ is strictly convex and $C^\infty$ on its nonempty effective domain, and the feasible region $\mathcal{X} = \{(x_1, x_i, \boldsymbol{y_i}) \in \mathbb{R}^{s+2} : x_1 \leq x_i,\ \boldsymbol{y_i} \leq \boldsymbol{\gamma}\}$ is convex. Therefore problem $K_i$ is a strictly convex minimization problem.

Further, the objective function of problem $K_i$ is coercive. To see this, suppose there exists a sequence $(x_1(n), x_i(n), \boldsymbol{y_i}(n)) \in \mathcal{X}$ such that the euclidean norm $||(x_1(n), x_i(n), \boldsymbol{y_i}(n))|| \to \infty$ and the objective function of problem $K_i$ remains bounded. Then it must be the case that the objective function has an unbounded sub-level set. However under assumption 5.1, the rate functions in the objective function $\alpha_1 I_1(x_1) + \alpha_i I_i(x_i, \boldsymbol{y_i})$ are good, which implies that all sub-level sets are bounded. Therefore if $||(x_1(n), x_i(n), \boldsymbol{y_i}(n))|| \to \infty$, then $\alpha_1 I_1(x_1(n)) + \alpha_i I_i(x_i(n), \boldsymbol{y_i}(n)) \to \infty$ and we conclude that the objective function is coercive.

Under assumption 5.2 the intersection of $\mathcal{X}$ and the effective domain of $\alpha_1 I_1(x_1) + \alpha_i I_i(x_i, \boldsymbol{y_i})$ is nonempty. Further, since $\mathcal{X}$ is closed and the objective function is coercive, then the minimum is attained and an optimal solution $\boldsymbol{\mathfrak{z}^*}(\alpha_1, \alpha_i)$ exists. By Bazaraa et al. (2006, theorem 3.4.2, p. 125), since the optimal solution exists and problem $K_i$ is strictly convex, $\boldsymbol{\mathfrak{z}^*}(\alpha_1, \alpha_i)$ is unique. $\square$

Each $\boldsymbol{\alpha}$-step in the concave "outer" optimization problem $Q$ requires solving $r - 1$ versions of problem $K_i$. Due to the nested-optimization structure of problem $Q$, we propose that the solution be obtained using a solver. We note that our allocation framework will work best with a "small" number of systems, where the quantifier "small" depends on the resources available to solve the optimal allocation problem.

## 5.5 Multivariate Normal Assumption

While the theory presented in sections 5.3 and 5.4 holds generally, we make a multivariate normal assumption to allow the development of a useful consistent estimator and a sequential allocation algorithm. Let the exponent $T$ denote the matrix transpose. Henceforth, let us redefine $\boldsymbol{G}_i$, $\boldsymbol{g}_i$, and $\boldsymbol{\gamma}$ as column vectors, and assume the following.

**Assumption 5.3.** *We assume we may obtain iid replicates of the multivariate normal random vector $(H_i, \boldsymbol{G}_i^T)$. That is,*

$$\begin{bmatrix} H_i \\ \boldsymbol{G}_i \end{bmatrix} \sim MVN\left(\begin{bmatrix} h_i \\ \boldsymbol{g}_i \end{bmatrix}, \Sigma_i\right),$$

*where $\Sigma_i$ is the variance-covariance matrix*

$$\Sigma_i = \begin{pmatrix} \sigma_{h_i}^2 & \rho_{h_i,g_{i1}}\sigma_{h_i}\sigma_{g_{i1}} & \cdots & \rho_{h_i,g_{is}}\sigma_{h_i}\sigma_{g_{is}} \\ \rho_{h_i,g_{i1}}\sigma_{h_i}\sigma_{g_{i1}} & \sigma_{g_{i1}}^2 & \cdots & \rho_{g_{i1},g_{is}}\sigma_{g_{i1}}\sigma_{g_{is}} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{h_i,g_{is}}\sigma_{h_i}\sigma_{g_{is}} & \rho_{g_{i1},g_{is}}\sigma_{g_{i1}}\sigma_{g_{is}} & \cdots & \sigma_{g_{is}}^2 \end{pmatrix}$$

*with correlation $\rho_{h_i,g_{ij}}, |\rho_{h_i,g_{ij}}| < 1$ and variances $0 < \sigma_{h_i}^2 < \infty$ and $0 < \sigma_{g_{ij}}^2 < \infty$ for all $i \leq r$ and $j \leq s$.*

Assumption 5.3 satisfies assumptions 5.1 and 5.2 in section 5.2. Thus recalling that $(\hat{H}_i, \hat{\boldsymbol{G}}_i^T) = (\bar{H}_i(\alpha_i n), \bar{\boldsymbol{G}}_i(\alpha_i, n)^T)$, it follows that $(\hat{H}_i, \hat{\boldsymbol{G}}_i^T)$ satisfies the LDP with good rate function

$$\alpha_i I_i(x, \boldsymbol{y}) = \frac{\alpha_i}{2} \begin{bmatrix} x - h_i \\ \boldsymbol{y} - \boldsymbol{g}_i \end{bmatrix}^T \Sigma_i^{-1} \begin{bmatrix} x - h_i \\ \boldsymbol{y} - \boldsymbol{g}_i \end{bmatrix}.$$

where $\boldsymbol{y} = (y_1, y_2, \ldots, y_s)^T$ and the exponent $-1$ denotes the matrix inverse. Since the marginal distributions for each replicate of $H_i$ and $G_{ij}$ are also normal, $\hat{H}_i$ and $\hat{G}_{ij}$ likewise satisfy the LDP with good rate functions

$$\alpha_i I_i(x) = \frac{\alpha_i (x - h_i)^2}{2\sigma_{h_i}^2} \quad \text{and} \quad \alpha_i J_{ij}(y) = \frac{\alpha_i (y - g_{ij})^2}{2\sigma_{g_{ij}}^2}$$

for all $i \leq r, j \leq s$.

Under assumption 5.3, we can substitute the multivariate normal rate functions into the expression for $z(\boldsymbol{\alpha})$. The resulting rate function for the $P\{FS\}$ is presented in the following corollary to theorem 5.1.

**Corollary 5.7.** *Under assumption 5.3, the rate function of the probability of false selection is*

$$-\lim_{n\to\infty}\frac{1}{n}\log P\{FS\} = \min\Bigg(\min_{j\in\mathcal{C}_F^1}\frac{\alpha_1(\gamma_j - g_{1j})^2}{2\sigma^2_{g_{1j}}},$$
$$\min_{2\le i\le r}\inf_{\substack{x_i\le x_1\\ \boldsymbol{y_i}\le\boldsymbol{\gamma}}}\left(\frac{\alpha_1(x_1-h_1)^2}{2\sigma^2_{h_1}} + \frac{\alpha_i}{2}\begin{bmatrix} x_i - h_i \\ \boldsymbol{y_i} - \boldsymbol{g}_i\end{bmatrix}^T \Sigma_i^{-1}\begin{bmatrix} x_i - h_i \\ \boldsymbol{y_i} - \boldsymbol{g}_i\end{bmatrix}\right)\Bigg).$$

Likewise, we may express problem $Q$ as,

$$\begin{aligned}\text{Problem } Q: \qquad & \max \quad z \quad s.t. \\ & \frac{\alpha_1(\gamma_j - g_{1j})^2}{2\sigma^2_{g_{1j}}} \ge z, \quad \text{for all } j\in\mathcal{C}_F^1 \\ & K_i(\alpha_1,\alpha_i)\ge z, \quad 2\le i\le r, \\ & \sum_{i=1}^r \alpha_i = 1,\ \alpha_i\ge 0.\end{aligned}$$

where the values of $K_i(\alpha_1,\alpha_i)$ are obtained by solving

$$\begin{aligned}\text{Problem } K_i: \qquad \min \quad & \frac{\alpha_1(x_1-h_1)^2}{2\sigma^2_{h_1}} + \frac{\alpha_i}{2}\begin{bmatrix} x_i - h_i \\ \boldsymbol{y_i} - \boldsymbol{g}_i\end{bmatrix}^T \Sigma_i^{-1}\begin{bmatrix} x_i - h_i \\ \boldsymbol{y_i} - \boldsymbol{g}_i\end{bmatrix} \quad s.t. \\ & x_i \le x_1, \\ & \boldsymbol{y_i}\le\boldsymbol{\gamma}.\end{aligned}$$

for each $i = 2,\ldots,r$. Due to the quadratic structure of problem $K_i$, much can be said about the optimal solution to this problem. However this quadratic form also makes finding a solution to problem $K_i$ quite easy in a solver. Therefore we refrain from further exploring the KKT conditions for this problem.

## 5.6 Consistency and Implementation

This section contains an estimator for the optimal allocation and results regarding its consistency, as well as a fully sequential algorithm fit for implementation under the multivariate normal assumption (assumption 5.3).

### 5.6.1 Consistency of Optimal Allocation Estimator

To solve problem $Q$, one must have full knowledge of the rate functions. However in practice, these rate functions must be estimated. Let us suppose that each system receives $m$ samples, where the dependence of the estimators on $m$ is written explicitly in this section. Let

$$(\hat{H}_i(m),\ \hat{\boldsymbol{G}}_i(m)) = \left(\frac{1}{m}\sum_{k=1}^{m} H_{ik},\ \frac{1}{m}\sum_{k=1}^{m} G_{i1k},\ \ldots,\ \frac{1}{m}\sum_{k=1}^{m} G_{isk}\right)$$

denote the estimators of $(h_i, g_{i1}, \ldots, g_{is})$.

To estimate the rate functions, the best feasible system must be estimated. Let the unique best feasible system be denoted

$$1 = \arg\min_i \{h_i : g_{ij} \leq \gamma_j \text{ for all } j \leq s\},$$

and the estimate of the best feasible system be denoted

$$\hat{1}(m) := \arg\min_i \{\hat{H}_i(m) : \hat{G}_{ij}(m) \leq \gamma_j \text{ for all } j \leq s\},$$

where any ties are broken randomly. Under assumptions 2 and 5.1, by proposition 3.10, we have system $\hat{1}(m) \to$ system 1 with probability one (wp1) as $m \to \infty$. Then under the multivariate normal assumption (assumption 5.3), consistent estimators of the relevant rate functions can be constructed from simulation observations $(H_{ik}, G_{i1k}, \ldots, G_{isk})$, $k = 1, \ldots, m$ for the $i$th system as follows.

$$\alpha_1 \hat{J}_{1j,m}(\gamma_j) = \frac{\alpha_1(\gamma_j - \hat{G}_{1j}(m))^2}{2\hat{\sigma}^2_{G_{1j}}}, \tag{5.5}$$

$$\alpha_1 \hat{I}_1(x_1) + \alpha_i \hat{I}_i(x_i, \boldsymbol{y_i}) = \frac{\alpha_1(x_1 - \hat{H}_1(m))^2}{2\hat{\sigma}^2_{H_1}} + \frac{\alpha_i}{2}\begin{bmatrix} x_i - \hat{H}_i(m) \\ \boldsymbol{y_i} - \hat{\boldsymbol{G}}_i(m)\end{bmatrix}^T \hat{\Sigma}_i^{-1}(m) \begin{bmatrix} x_i - \hat{H}_i(m) \\ \boldsymbol{y_i} - \hat{\boldsymbol{G}}_i(m)\end{bmatrix} \tag{5.6}$$

$$\hat{K}_{i,m}(\alpha_1, \alpha_i) = \inf_{\substack{x_i \leq x_1 \\ \boldsymbol{y_i} \leq \boldsymbol{\gamma}}} \left(\alpha_1 \hat{I}_1(x_1) + \alpha_i \hat{I}_i(x_i, \boldsymbol{y_i})\right), \tag{5.7}$$

where $\hat{\sigma}^2_{H_1}$ denotes the sample variance of replicates $H_{1k}$ and $\hat{\Sigma}_i(m)$ denotes the sample covariance matrix for replicates $(H_{ik}, G_{i1k}, \ldots, G_{isk})$.

Recall that problem $Q$ is to minimize $z(\boldsymbol{\alpha})$ over compact support $\sum_{i=1}^{r} \alpha_i = 1, \boldsymbol{\alpha} \geq 0$. Let $\mathcal{S}_{\boldsymbol{\alpha}^*}$ be the set of optimal solutions to problem $Q$, which is nonempty by proposition 5.5. Let $\hat{z}_m(\boldsymbol{\alpha})$ be as $z(\boldsymbol{\alpha})$, except with all random quantities replaced by their estimators as specified in equations (5.5)–(5.7). Then the sample average approximation (SAA) problem, called problem $\hat{Q}$, is to minimize $\hat{z}_m(\boldsymbol{\alpha})$ over the compact set $\sum_{i=1}^{r} \alpha_i = 1, \boldsymbol{\alpha} \geq 0$. Within problem $\hat{Q}$, we have the nested SAA problem $\hat{K}_i$, which is identical to problem $K_i$, except with random quantities replaced by their estimators as specified in equation (5.7).

Since the search space in problem $Q$ is convex and compact, and the estimated (consistent) rate functions are finite-valued, continuous, and concave, it broadly follows that for large enough $m$, the set $\hat{\mathcal{S}}_{\hat{\boldsymbol{\alpha}}^*_m}$ of optimal solutions to problem $\hat{Q}$ exists, and as $m \to \infty$, $\hat{\mathcal{S}}_{\hat{\boldsymbol{\alpha}}^*_m} \to \mathcal{S}_{\boldsymbol{\alpha}^*}$ wp1 (for a definition of set convergence, see section 1.5). Theorem 5.10 states this result formally, which follows as a direct application of results in the SAA literature (see, e.g., Shapiro et al., 2009, theorem 5.4, p. 159). Before the formal presentation of theorem 5.10, we present lemmas 5.8 and 5.9 regarding the convergence of the optimal value and solution in problem $\hat{K}_i$. The results presented in these lemmas are required for the postulates of theorem 5.10.

**Lemma 5.8.** *Let assumption 5.3 hold. Then for any fixed $\boldsymbol{\alpha}$ and for all $i = 2, \ldots, r$,*

*(i)* $\alpha_1 \hat{I}_1(x_1) + \alpha_i \hat{I}_i(x_i, \boldsymbol{y_i}) \to \alpha_1 I_1(x_1) + \alpha_i I_i(x_i, \boldsymbol{y_i})$ *pointwise wp1 as $m \to \infty$, and*

*(ii) problem $\hat{K}_i$ has a unique solution* $\hat{\boldsymbol{\mathfrak{z}}}^*_{\boldsymbol{m}}(\alpha_1, \alpha_i) = (\hat{x}^*_{1,m}(\alpha_1, \alpha_i), \hat{x}^*_{i,m}(\alpha_1, \alpha_i), \hat{\boldsymbol{y}}^*_{\boldsymbol{i,m}}(\alpha_1, \alpha_i))$ *wp1 for large enough $m$.*

*Proof.* Under assumption 5.3, the sample averages and variances in the expression of $\alpha_1 \hat{I}_1(x_1) + \alpha_i \hat{I}_i(x_i, \boldsymbol{y_i})$ converge to their true values wp1 as $m \to \infty$. Further, the covariance estimates in $\hat{\Sigma}_i(m)$ converge to their true values wp1 as $m \to \infty$. Since $\Sigma_i$ is positive definite, so that $\Sigma_i^{-1}$ is well-defined, and since $\hat{\Sigma}_i(m)$ is strongly consistent and the matrix inverse is a continuous function, it follows that $\hat{\Sigma}_i^{-1}(m)$ is well-defined for large enough $m$ and hence $\hat{\Sigma}_i^{-1}(m) \to \Sigma_i^{-1}$ wp1 as $m \to \infty$. Then for any fixed $x$ and $\boldsymbol{\alpha}$, $\alpha_1 \hat{I}_1(x_1) + \alpha_i \hat{I}_i(x_i, \boldsymbol{y_i}) \to \alpha_1 I_1(x_1) + \alpha_i I_i(x_i, \boldsymbol{y_i})$ wp1 as $m \to \infty$ (see, e.g., Serfling, 1980, p. 24).

Due to its quadratic form, for $m$ large enough that $\hat{\Sigma}_i^{-1}(m)$ is well-defined, the estimated-objective function $\alpha_1 \hat{I}_1(x_1) + \alpha_i \hat{I}_i(x_i, \boldsymbol{y_i})$ (expanded in equation (5.6)) is a strictly convex function of $(x_1, x_i, \boldsymbol{y_i})$ for each $m$. Further, the feasible set is convex for each $m$. By a proof similar to that of proposition 5.6, it holds that problem $\hat{K}_i$ has a unique solution $\hat{\boldsymbol{\mathfrak{z}}}^*_{\boldsymbol{m}}(\alpha_1, \alpha_i) = (\hat{x}^*_{1,m}(\alpha_1, \alpha_i), \hat{x}^*_{i,m}(\alpha_1, \alpha_i), \hat{\boldsymbol{y}}^*_{\boldsymbol{i,m}}(\alpha_1, \alpha_i))$ wp1 for large enough $m$. □

**Lemma 5.9.** *Let assumption 5.3 hold. Then for any fixed $\boldsymbol{\alpha}$, and for all $i = 2, \ldots, r$,*

*(i)* $\hat{K}_{i,m}(\alpha_1, \alpha_i) \to K_{i,m}(\alpha_1, \alpha_i)$ *wp1 as $m \to \infty$, and*

*(ii)* $\hat{\boldsymbol{\mathfrak{z}}}^*_{\boldsymbol{m}}(\alpha_1, \alpha_i) \to \boldsymbol{\mathfrak{z}}^*(\alpha_1, \alpha_i)$ *wp1 as $m \to \infty$.*

*Proof.* Under assumption 5.3, the objective function in problem $K_i$, $\alpha_1 I_1(x_1) + \alpha_i I_i(x_i, \boldsymbol{y_i})$, is a continuous function of $(x_1, x_i, \boldsymbol{y_i})$ and is finite-valued for all $(x_1, x_i, \boldsymbol{y_i}) \in \mathbb{R}^{(s+2)}$. By lemma 5.8, the law of large numbers holds pointwise for the objective function in problem $\hat{K}_i$, $\alpha_1 \hat{I}_1(x_1) + \alpha_i \hat{I}_i(x_i, \boldsymbol{y_i})$. Further, the feasible set $\{(x_1, x_i, \boldsymbol{y_i}) : x_i \leq x_1, \boldsymbol{y_i} \leq \boldsymbol{\gamma}\}$ is closed and convex, and by proposition 5.6, the unique optimal solution $\boldsymbol{\mathfrak{z}}^*(\alpha_1, \alpha_i)$ to the true problem $K_i$ exists. Due to its quadratic form, the estimated-objective function $\alpha_1 \hat{I}_1(x_1) + \alpha_i \hat{I}_i(x_i, \boldsymbol{y_i})$ (expanded in equation (5.6)) is a continuous, convex function of $(x_1, x_i, \boldsymbol{y_i})$ for each large-enough $m$ that $\hat{\Sigma}_i^{-1}(m)$ is well-defined. Thus we have satisfied the postulates of Shapiro et al. (2009, theorem 5.4, p. 159), and the conclusions follow. □

**Theorem 5.10.** *The estimated optimal value and set of estimated optimal allocations of problem $\hat{Q}$ are strongly consistent, that is, $\hat{z}_m(\hat{\boldsymbol{\alpha}}^*_m) \to z(\boldsymbol{\alpha}^*)$ and $\hat{\mathcal{S}}_{\hat{\boldsymbol{\alpha}}^*_m} \to \mathcal{S}_{\boldsymbol{\alpha}}$ wp1 as $m \to \infty$.*

*Proof.* As argued previously, $z(\boldsymbol{\alpha})$ is a continuous, concave, finite-valued, function of $\boldsymbol{\alpha}$, the feasible region in problem $Q$ is compact and convex, and the optimal solution exists. The estimated function $\hat{z}_m(\boldsymbol{\alpha})$ is also a continuous, concave function of $\boldsymbol{\alpha}$ for each large-enough $m$. It follows from the strong consistency of the estimator $\alpha_1 \hat{J}_{1j,m}(\gamma_j)$ and from lemma 5.9 that the law of large numbers holds pointwise for $\hat{z}_m(\boldsymbol{\alpha})$. Thus we have satisfied the postulates of Shapiro et al. (2009, theorem 5.4, p. 159), and the conclusion follows. □

### 5.6.2 A Sequential Algorithm for Implementation

In the following algorithm 5.4, we present a fully sequential algorithm based on the consistent estimators outlined in section 5.6.1, where we let $n$ be the total simulation budget and $n_i$ be the total sample expended at system $i$.

The essential idea in algorithm 5.4 is straightforward. After an initialization step in which we calculate initial estimates $(\hat{H}_i(n_i), \hat{\boldsymbol{G}}_i(n_i))$ and $\hat{\Sigma}_i(n_i)$ for all systems $i \leq r$, and $\hat{1}(n)$, the estimate of system 1, the eigenvalues of the estimated covariance matrix $\hat{\Sigma}_i(n_i)$ are checked for all $i \neq \hat{1}(n)$ to determine if the estimated matrix can be reliably inverted. If the eigenvalues are large enough, as determined by a pre-specified tolerance, then $\hat{\Sigma}_i(n_i)$ is inverted and problem $\hat{Q}$ is solved by solving problem $\hat{K}_i$ for all $i \neq \hat{1}(n)$ at each $\boldsymbol{\alpha}$-step. The subsequent solution $\hat{\boldsymbol{\alpha}}^*_n$ is used as a sampling distribution to select the next system from which to sample. Sampling continues in this manner until $\delta$ new samples are observed, at which time the algorithm updates the estimators.

After updating the estimates $(\hat{H}_i(n_i), \hat{\boldsymbol{G}}_i(n_i))$ and $\hat{\Sigma}_i(n_i)$ for all $i \leq r$, if no systems are estimated feasible or if any eigenvalue is too small and hence an estimated covariance matrix cannot be inverted, we check if problem $\hat{Q}$ has been successfully solved before. If not, we set $\hat{\boldsymbol{\alpha}}^*_n$ to equal allocation. If problem $\hat{Q}$ has been successfully solved before, we simply keep sampling using the previous estimator.

Since the true optimal allocation $\boldsymbol{\alpha}^* > 0$, it follows that the sequential algorithm should sample from each system infinitely often. To ensure systems with small allocations continue to be sampled,

**Algorithm 5.4** Sequential algorithm for the case in which underlying random variables have multivariate normal distributions

**Require:** Number of pilot samples $\delta_0 > 0$; number of samples between allocation vector updates $\delta > 0$; an eigenvalue tolerance vector $\boldsymbol{\epsilon} > 0$; and a minimum-sample vector $\boldsymbol{\varepsilon} > 0$.

1: Initialize: collect $\delta_0$ samples from each system $i \leq r$.
2: Initialize: total simulation effort $n = r\delta_0$, effort for each system $n_i = \delta_0$, and an indicator that problem $\hat{Q}$ has not been solved yet, $\mathbb{I} = 0$.
3: Update the estimators $(\hat{H}_i(n_i), \hat{\boldsymbol{G}}_i(n_i))$ and $\hat{\Sigma}_i(n_i)$ for all $i \leq r$.
4: **if** no systems are estimated feasible **then**
5:     go to step 12.
6: **else**
7:     update $\hat{1}(n)$, the estimated system 1.
8:     For each system $i \neq \hat{1}(n)$, calculate the vector of eigenvalues of $\hat{\Sigma}_i(n_i)$, denoted $\boldsymbol{e}_i$.
9:     **if** $\boldsymbol{e}_i \geq \boldsymbol{\epsilon}$ for all $i \leq r$ **then**
10:         Solve problem $\hat{Q}$ to obtain $\hat{\boldsymbol{\alpha}}_n^*$. {At each step in problem $\hat{Q}$, solve problem $\hat{K}_i$ for all $i \neq \hat{1}(n)$.} Set $\mathbb{I} = 1$ to indicate that problem $\hat{Q}$ has been solved.
11:     **else**
12:         **if** $\mathbb{I} = 0$ **then**
13:             Set $\hat{\boldsymbol{\alpha}}_n^* = (1/r, 1/r, \ldots, 1/r)$.
14:         **else**
15:             Set $\hat{\boldsymbol{\alpha}}_n^* = \hat{\boldsymbol{\alpha}}_{n-(\delta+\delta^+)}^*$. {If we cannot solve problem $\hat{Q}$, use the previous solution.}
16:         **end if**
17:     **end if**
18: **end if**
19: Collect one sample at each system $X_k, k = 1, 2, \ldots, \delta$, where the $X_k$'s are iid random variates with probability mass function $\hat{\boldsymbol{\alpha}}_n^*$ and support $\{1, 2, \ldots, r\}$. Update $n_{X_k} = n_{X_k} + 1$.
20: Set $n = n + \delta$ and update $\bar{\boldsymbol{\alpha}}_{\boldsymbol{n}} = \{n_1/n, n_2/n, \ldots, n_r/n\}$.
21: **if** $\bar{\boldsymbol{\alpha}}_{\boldsymbol{n}} > \boldsymbol{\varepsilon}$ **then**
22:     Set $\delta^+ = 0$.
23: **else**
24:     Collect one sample from each system in the set of systems receiving insufficient sample $\mathcal{I}_n$.
25:     Update $n_i = n_i + 1$ for all $i \in \mathcal{I}_n$. Let $\delta^+ = |I_n|$.
26: **end if**
27: Set $n = n + \delta^+$ and go to step 3.

we assume knowledge of an "indifference zone" vector $\boldsymbol{\varepsilon} > 0$ such that if the actual proportion of sample expended at each system in algorithm 5.4, defined as $\bar{\boldsymbol{\alpha}}_{\boldsymbol{n}} = \{n_1/n, n_2/n, \ldots, n_r/n\}$, falls below $\boldsymbol{\varepsilon}$, we sample once from each system receiving insufficient sample. All elements of $\boldsymbol{\varepsilon}$ should be "small" relative to $1/r$. Therefore the algorithm essentially guarantees that a minimal amount of sample is spent at each system in the limit.

Algorithm 5.4 provides flexibility in how often the optimal allocation vector is re-estimated through the algorithm parameter $\delta$. The choice of parameter $\delta$ will depend on the problem, particularly how many systems and constraints are in contention, as well as how expensive simulation execution is. As algorithm 5.4 requires fully sequential and simultaneous observation of the objective and constraint functions, deviations from these assumptions renders it inapplicable.

## 5.7 Numerical Results

Given a problem $P$ with a pre-specified number of systems and constraints (see the problem statement in section 1.2), and a simulation oracle capable of sequentially generating iid replicates of multivariate normal random variables according to assumptions 1 and 5.3, we can employ algorithm 5.4 to estimate the optimal allocation. Therefore we wish to show the numerical performance of algorithm 5.4 for a variety of problems with varying numbers of systems and constraints.

Algorithm B.5 (see appendix B.3) was developed to generate random versions of problem $P$ for a specified number of systems and constraints in the multivariate normal context. Specifically, given $r$ systems and $s$ constraints, algorithm B.5 generates systems "located" at $(h_i, \boldsymbol{g_i})$ with $(s+1)$-by-$(s+1)$ covariance matrices $\Sigma_i$ for all $i \le r$. The system "locations" and covariance matrices are used in the simulation oracle to generate iid multivariate normal replicates in steps 1 and 3 of algorithm 5.4. For the numerical experiments, algorithm B.5 generated one hundred problems $P_k$, $k = 1, \ldots, 100$ of each type (i) 2 systems with 2 constraints, (ii) 5 systems with 2 constraints, (iii) 2 systems with 5 constraints, and (iv) 5 systems with 5 constraints. The shorthand notation $2 \times 2$, $5 \times 2$, $2 \times 5$, and $5 \times 5$, respectively, refers to these generated problem scenarios.

Given a sampling algorithm and problem $P_k$, let us define the actual proportion of the total sampling budget $n_k$ expended at each system $i = 1, 2, \ldots, r$ in problem $P_k$ as

$$\bar{\boldsymbol{\alpha}}_{\boldsymbol{n,k}} = \left\{ \frac{n_{1,k}}{n_k}, \frac{n_{2,k}}{n_k}, \ldots, \frac{n_{r,k}}{n_k} \right\},$$

and the optimality gap in the rate of decay of the $P\{FS\}$ of the sampling algorithm for problem $P_k$ as $z(\boldsymbol{\alpha}^*_{\boldsymbol{k}}) - z(\bar{\boldsymbol{\alpha}}_{\boldsymbol{n,k}})$. Since $z(\boldsymbol{\alpha}^*_{\boldsymbol{k}})$ is the fixed optimal rate of decay of the $P\{FS\}$ for each problem $P_k$, the optimality gap is necessarily positive.

The proposed algorithm 5.4 was used to sample from the one hundred randomly generated problems of each type $2 \times 2$, $5 \times 2$, $2 \times 5$, and $5 \times 5$, with pilot samples $\delta_0 = 2$, number of samples between allocation vector updates $\delta = 20$, and eigenvalue tolerance vector $\boldsymbol{\epsilon} = \{\epsilon_1, \epsilon_2, \ldots, \epsilon_{s+1}\} = \{0.01, 0.01, \ldots, 0.01\}$. At each sample size $n_k = \{r\delta_0, r\delta_0 + 1, \ldots, 1000\}$, $k = 1, 2, \ldots, 100$, the optimality gap $z(\boldsymbol{\alpha}^*_{\boldsymbol{k}}) - z(\bar{\boldsymbol{\alpha}}_{\boldsymbol{n,k}})$ was calculated for the proposed algorithm 5.4. Due to the nature of algorithm 5.4, $z(\boldsymbol{\alpha}^*_{\boldsymbol{k}}) - z(\bar{\boldsymbol{\alpha}}_{\boldsymbol{n,k}})$ is a random variable for each $k = 1, 2, \ldots, 100$. The optimality gap $z(\boldsymbol{\alpha}^*_{\boldsymbol{k}}) - z(\bar{\boldsymbol{\alpha}}_{\boldsymbol{n,k}})$ was also calculated for sampling with equal allocation. For each $k = 1, 2, \ldots, 100$, the equal allocation optimality gap is a fixed quantity. Then at each sample size $n_k$, the 90th, 75th, 50th, and 25th percentiles of the distribution of the optimality gaps across problems $P_k$ for the proposed allocation and equal allocation were calculated. The results of these numerical experiments are presented in figure 5.1.

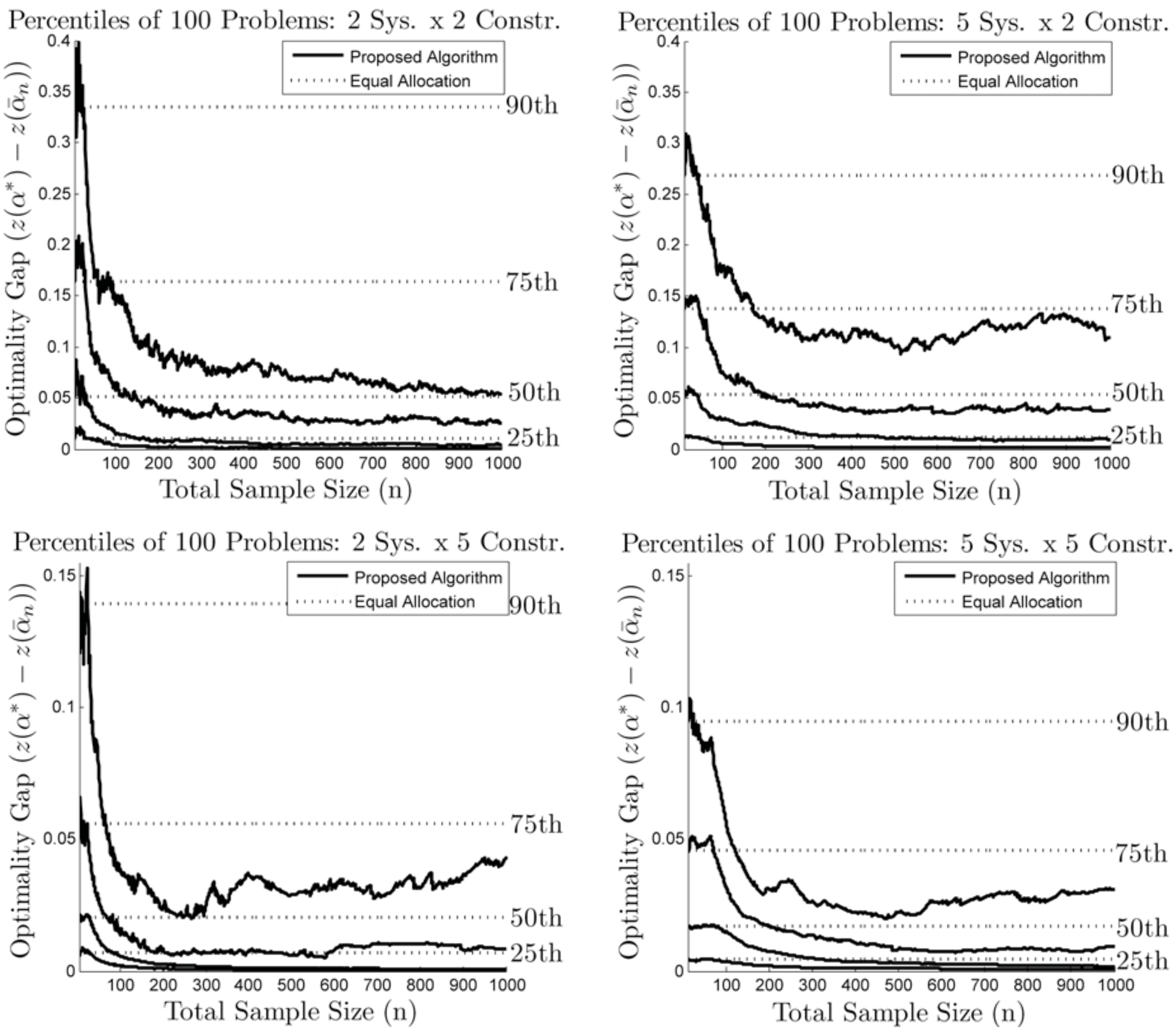


Figure 5.1: The graphs show a comparison between the 90th, 75th, 50th, and 25th percentiles of the distribution of the optimality gap of the rate of decay of the $P\{FS\}$, $(z(\boldsymbol{\alpha}^*) - z(\bar{\boldsymbol{\alpha}}_{\boldsymbol{n}}))$, resulting from (i) the proposed allocation in algorithm 5.4; and (ii) equal allocation; each calculated across 100 randomly generated 2-system and 2-constraint, 5-system and 2-constraint, 2-system and 5-constraint, and 5-system and 5-constraint problems, respectively, at sample size values $n = r\delta_0, r\delta_0 + 1, \ldots, 1000$.

Figure 5.1 shows that, as predicted by theory, the optimality gap of the actual allocation used in algorithm 5.4 converges to zero as the sampling budget $n$ increases. While the optimality gap of the proposed algorithm converges to zero, it does not do so since the optimality gaps $(z(\boldsymbol{\alpha}^*) - z(\bar{\boldsymbol{\alpha}}_{\boldsymbol{n}}))$

calculated at each value of $n$ are highly autocorrelated. Also as expected, the optimality gap of equal allocation remains fixed as the sampling budget $n$ increases.

For problems of each type, $2 \times 2$, $5 \times 2$, $2 \times 5$, and $5 \times 5$, respectively, the gains in the rate of decay of the $P\{FS\}$ of the proposed algorithm over equal allocation across a total sample size of only $n = 1000$ are significant. In each problem type, the 90th percentile of problems achieve a proposed-algorithm optimality gap that is lower than the 75th percentile equal-allocation optimality gap within a sampling budget of approximately $n = 200$. However the number of samples required for the proposed algorithm to reliably perform better than the competing equal-allocation scheme varies by the number of systems: the more systems, the larger the required number of samples. This result is intuitive since the more systems in consideration, the larger the required overall sampling budget to gather information from each system (e.g., two systems require $n = 4$ for each system to receive two samples, while ten systems require $n = 20$ for each system to receive two samples).

This phenomenon is further illustrated in figure 5.2, which displays exactly the same problem setup as figure 5.1, except for problems containing 10 systems and 10 constraints and with updating parameter $\delta = 100$. Figure 5.2 shows that it takes substantially more total sampling budget for the proposed allocation to reliably outperform equal allocation than for 2-system or 5-system problems, although the overall optimality gap of the proposed allocation does tend to zero. While figure 5.2 was relatively computationally intense to create, it does demonstrate that we are able to solve problems of higher dimensions reliably using the proposed algorithm.

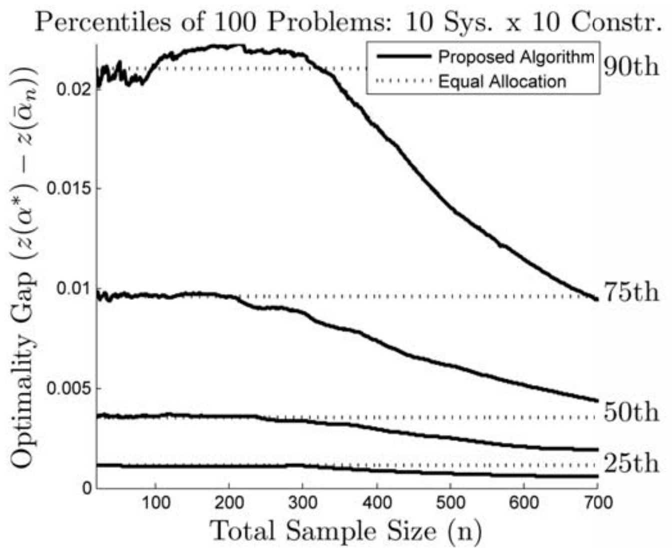


Figure 5.2: The graphs show a comparison between the 90th, 75th, 50th, and 25th percentiles of the distribution of the optimality gap of the rate of decay of the $P\{FS\}$, $(z(\boldsymbol{\alpha}^*) - z(\bar{\boldsymbol{\alpha}}_{\boldsymbol{n}}))$, resulting from (i) the proposed allocation in algorithm 5.4; and (ii) equal allocation; each calculated across 100 randomly generated 10-system and 10-constraint problems at sample size values $n = r\delta_0, r\delta_0 + 1, \ldots, 700$.

# Chapter 6

# Concluding Remarks

The constrained SO problem on finite sets is an important SO variation about which little is currently known. Questions surrounding the relationship between sampling and error-probability decay, sampling rates to ensure optimal convergence to the correct solution, and minimum sample size rules that probabilistically guarantee attainment of the correct solution remain largely unexplored. Following recent work by Glynn and Juneja (2004) for the unconstrained SO context and Szechtman and Yücesan (2008) for the context of detecting feasibility, we take key steps toward answering these questions.

In particular, in chapter 3, we fully characterize the rate of decay of the probability of false selection for the case of general distributions. We use this characterization to derive an asymptotically optimal sampling framework, which is the solution to a concave maximization problem. We then develop a sequential algorithm designed to sample from amongst the systems in the optimal proportions in the limit. No competing algorithm exists for this general distribution case; our results are the first on optimal sampling allocation in this context.

An important assumption made in chapter 3 is that of limiting independence between the objective function and constraint estimators for each system. While this assumption is true in certain contexts, it may be violated in a number of other "real-world" contexts. In contexts in which the assumption does not hold, the framework presented in this chapter should be seen as an approximate guide to simulation allocation obtained through the analysis of an imperfect but tractable model.

To account for the effect of dependence on the optimal allocation, we investigate the optimal allocation under the assumption that the underlying random variables follow a multivariate normal distribution (chapters 4 and 5). This work is aimed at guiding sampling in contexts where dependence between objective and constraint estimators cannot be ignored, and where the underlying distributions can be assumed as normal. In chapter 4, for one constraint, we characterize the rate of decay of the probability of false selection as an explicit function of correlation. We then express the optimal allocation as the solution to a concave maximization problem that is also an explicit function of correlation. In chapter 5, for multiple constraints, we characterize the rate of decay of the probability of false selection and express the asymptotically exact optimal allocation as the solution to a concave maximization problem in which the constraints are themselves $r-1$ convex minimization problems. We present a fully sequential algorithm based on this optimal allocation scheme, and demonstrate that it reliably performs better than equal allocation in finite time for a "small" number of systems and constraints, where the quantifier "small" depends on available computing resources. Thus we demonstrate that not only are we able to reliably solve for the optimal allocation, and hence the asymptotically optimal allocation yields "implementable" results, but the proposed sequential algorithm provides finite-time gains in the rate of decay of the probability of false selection over naïve allocation.

The two models offer an interesting dichotomy in the required assumptions. In the first model (the general model), no distributional assumptions are required, but the potentially-impractical assumption of limiting independence between the objective function and constraint estimators holds. In the second model (the multivariate normal model), the random variables may be correlated, but a distributional assumption of multivariate normality is required. Thus the selection of an appropriate model in practice will require a trade-off between correlation and non-normality. The general model is perhaps most useful when correlation is mild, but the random variates exhibit non-normality. This model provides the only guidance in the literature for sampling in the context SO on finite sets with stochastic constraints and general distributions. The multivariate normal model may be useful when the multivariate normality assumption is "reasonable" or when the correlation is strong. When the multivariate normal assumption is satisfied, this model provides the only asymptotically exact allocation that directly accounts for correlation. (While Glynn and Juneja (2004) describe batching as scaling the rate function and hence not altering the optimal allocation, batching can alter the optimal allocation if we consider the limit as the batch size tends to infinity, not as the overall sample size tends to infinity. Thus batching to achieve normality is possible.)

The usefulness of the general model may also rely on knowledge of the distributional family, since papers such as Broadie et al. (2007) find that naïve estimators of the rate functions, typically computed as the Legendre-Fenchel transform (see, e.g., Dembo and Zeitouni, 1998, p. 26) of the estimated cumulant generating function, are computationally impractical. Further, Glynn and Juneja (2011) make the point that for unbounded random variables, the naïve estimator of the rate function is heavy-tailed. This observation implies that the error in estimating the rate function swamps the gains made by optimizing with respect to sample size. However, given knowledge of the distributional family, is the resulting rate function estimator light-tailed? Also, if we assume no knowledge of the distributional family, can computational complexity be reduced by estimating an approximation to the rate function, e.g., estimating several terms in the Taylor series expansion (Rudin, 1976, p. 110) and bounding the resulting error? Preliminary work on deriving a general expression for the Taylor series expansion of an analytic rate function is provided in appendix D. Interesting questions to explore in this context include, how many terms of the Taylor series expansion should be used to estimate the rate function and ensure a reasonable approximation to the actual optimal allocation? Further, how does this method compare in computational time and in estimation accuracy with traditional (yet computationally intense) methods of estimating rate functions — specifically, is the estimator light-tailed?

Numerous other ongoing research topics naturally stem from this work. For example, in the unconstrained setting, we may derive an indifference zone formulation to answer the question of how to sample when we are interested in selecting a system within some tolerance value of the best feasible system. Further, we may move the constrained optimal allocation formulation into an unconstrained realm by including the constraints in the objective function with weights or penalties for violation of the constraints. We may also attempt to extend work on optimal allocation in the unconstrained, heavy-tailed context by Blanchet et al. (2008) to the constrained, heavy-tailed context.

# Bibliography

# Bibliography


Andradóttir, S. 2006. An overview of simulation optimization via random search. S. G. Henderson, B. L. Nelson, eds., *Simulation*. Handbooks in Operations Research and Management Science, Elsevier, 617–631.

Andradóttir, S., S.-H. Kim. 2010. Fully sequential procedures for comparing constrained systems via simulation. *Naval Research Logistics* **57**(5) 403–421.

Barton, R. R., M. Meckesheimer. 2006. Metamodel-based simulation optimization. S. G. Henderson, B. L. Nelson, eds., *Simulation*. Handbooks in Operations Research and Management Science, Elsevier, 535–574.

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**(3) 13:1–26.

Bazaraa, M. S., H. D. Sherali, C. M. Shetty. 2006. *Nonlinear Programming: Theory and Algorithms*. 3rd ed. Wiley-Interscience.

Blanchet, J., J. Liu, B. Zwart. 2008. Large deviations perspective on ordinal optimization of heavy-tailed systems. S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler, eds., *Proceedings of the 2008 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers, Inc., Piscataway, NJ, 489–494.

Boyd, S., L. Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press, New York.

Broadie, M., M. Han, A. Zeevi. 2007. Implications of heavy tails on simulation-based ordinal optimization. S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, R. R. Barton, eds., *Proceedings of the 2007 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers, Inc., Piscataway, NJ, 439–447.

Bucklew, J. A. 2003. *Introduction to Rare Event Simulation*. Springer, New York.

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**(3) 251–270.

Dembo, A., O. Zeitouni. 1998. *Large Deviations Techniques and Applications*. 2nd ed. Springer, New York.

Fu, M. C. 2002. Optimization for simulation: Theory vs. practice. *INFORMS Journal on Computing* **14** 192–215.

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., *Proceedings of the 2004 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers, Inc., Piscataway, New Jersey, 577–585.

Glynn, P. W., S. Juneja. 2006. Ordinal optimization: A large deviations perspective Working Paper Series, Indian School of Business.

Glynn, P. W., S. Juneja. 2011. Ordinal optimization: A nonparametric framework. S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, M. Fu, eds., *Proceedings of the 2011 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers, Inc., Piscataway, NJ.

Henderson, S. G., R. Pasupathy. 2011. Simulation optimization library. URL `http://www.simopt.org`. Accessed March 26, 2011.

Hunter, S. R., R. Pasupathy. 2010. Large-deviation sampling laws for constrained simulation optimization on finite sets. B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, E. Yücesan, eds., *Proceedings of the 2010 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers, Inc., Piscataway, New Jersey, 995–1002.

Hunter, S. R., R. Pasupathy. 2011. Optimal sampling laws for stochastically constrained simulation optimization on finite sets. Under Review.

Hunter, S. R., N. A. Pujowidianto, C.-H. Chen, L. H. Lee, R. Pasupathy, C. M. Yap. 2011. Optimal sampling laws for constrained simulation optimization on finite sets: The bivariate normal case. S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, M. Fu, eds., *Proceedings of the 2011 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers, Inc., Piscataway, New Jersey. To Appear.

Johnson, W. P. 2002. The curious history of Faà di Bruno's formula. *The American Mathematical Monthly* **109**(3) 217–234.

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.

Lee, L. H., N. A. Pujowidianto, L.-W. Li, C.-H. Chen, C. M. Yap. 2011. Asymptotic simulation budget allocation for selecting the best design in the presence of stochastic constraints Submitted to *IEEE Transactions on Automatic Control*.

Ólafsson, S., J. Kim. 2002. Simulation optimization. E. Yucesan, C. H. Chen, J. L. Snowdon, J. M. Charnes, eds., *Proceedings of the 2002 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers: Piscataway, New Jersey, 79–84.

Pasupathy, R., S. Kim. 2011. The stochastic root-finding problem: overview, solutions, and open questions. *ACM Transactions on Modeling and Computer Simulation* Under revision.

Rudin, W. 1976. *Principles of Mathematical Analysis*. International series in pure and applied mathematics, McGraw-Hill.

Serfling, R. J. 1980. *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Mathematical Statistics, Wiley & Sons, New York.

Shapiro, A., D. Dentcheva, A. Ruszczyński. 2009. *Lectures on stochastic programming: modeling and theory*. MPS-SIAM Series on Optimization, SIAM and MPS.

Spall, J. C. 2003. *Introduction to Stochastic Search and Optimization*. John Wiley & Sons, Inc., Hoboken, NJ.

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., *Proceedings of the 2008 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers, Inc., Piscataway, New Jersey, 273–280.

# Appendix A

# Useful Results

## A.1 Principle of the Slowest Term

The following proposition appears as the "Principle of the Largest Term" in Ganesh et al. (2004, lemma 2.1).

**Proposition A.1.** *Let $a_n^i, i = 1, 2, \ldots, k$, be a finite number of sequences such that $0 \le a_n^i \le 1$ for all $n$ and $\lim_{n\to\infty} \frac{1}{n} \log a_n^i = -a^i$, where $a^i > 0$. Let $a_n^1$ be the slowest converging sequence, that is, $a^1 = -\min_i a^i$, and let $a_n^1 > 0$ for all $n$. Then,*

$$\lim_{n\to\infty} \frac{1}{n} \log \sum_{i=1}^{m} a_n^i = \max_i \left( \lim_{n\to\infty} \frac{1}{n} \log a_n^i \right) = -\min_i a^i.$$

*Proof.* First,

$$\lim_{n\to\infty} \frac{1}{n} \log \sum_{i=1}^{k} a_n^i = \lim_{n\to\infty} \frac{1}{n} \log \left( a_1^i \left( 1 + \sum_{i>1} \frac{a_n^i}{a_n^1} \right) \right) = \lim_{n\to\infty} \frac{1}{n} \log a_n^1 + \lim_{n\to\infty} \frac{1}{n} \log \left( 1 + \sum_{i>1} \frac{a_n^i}{a_n^1} \right).$$

Since $\lim_{n\to\infty} \frac{1}{n} \log a_n^i = -a^i$, for all $\epsilon > 0$, there exists an $N_i(\epsilon)$ such that for all $n > N_i(\epsilon)$, we have $|\frac{1}{n} \log a_n^i + a^i| < \epsilon$. Then $a_n^i \in \left( e^{-n(a^i+\epsilon)}, e^{-n(a^i-\epsilon)} \right)$.

Let $\epsilon > 0$ and $n > \max\left\{k, \max_i\{N_i(\epsilon)\}, \max_i\{(-\log \epsilon)/(a^i - a^1 - 2\epsilon)\}\right\}$. Then,

$$\begin{aligned} \left| \frac{1}{n} \log \left( 1 + \sum_{i>1} \frac{a_n^i}{a_n^1} \right) \right| &\le \left| \frac{1}{n} \log \left( 1 + \sum_{i>1} \frac{e^{-n(a^i-\epsilon)}}{e^{-n(a^1+\epsilon)}} \right) \right| = \left| \frac{1}{n} \log \left( 1 + \sum_{i>1} e^{-n(a^i-a^1-2\epsilon)} \right) \right| \\ &\le \left| \frac{1}{n} \log \left( 1 + \sum_{i>1} \epsilon \right) \right| \le \left| \frac{1}{n} \log (1 + k\epsilon) \right| \le \left| \frac{1}{n} (k\epsilon) \right| \le \epsilon. \end{aligned}$$

□

## A.2 Limit of the Maximum Equals Maximum of the Limits

**Proposition A.2.** *Let $a_n^i, i = 1, 2, \ldots, k$, be a finite number of sequences such that $0 < a_n^i \le 1$ for all $n$ and $\lim_{n\to\infty} \frac{1}{n} \log a_n^i = a^i$, where $a^i < 0$. Let $a_n^1$ be the slowest converging sequence, that is, $a^1 = \max_i\{a^i\}$. Then,*

$$\max_i \left( \lim_{n\to\infty} \frac{1}{n} \log a_n^i \right) = \lim_{n\to\infty} \frac{1}{n} \log \left( \max_i a_n^i \right)$$

*Proof.* By the principle of the largest term, the lower bound is

$$\max_i \left( \lim_{n\to\infty} \frac{1}{n} \log a_n^i \right) = \lim_{n\to\infty} \frac{1}{n} \log \sum_{i=1}^{k} a_n^i \ge \lim_{n\to\infty} \frac{1}{n} \log \max_i a_i^n.$$

Now the upper bound is given by

$$\max_i \left( \lim_{n\to\infty} \frac{1}{n} \log a_n^i \right) = \lim_{n\to\infty} \frac{1}{n} \log \sum_{i=1}^{k} a_n^i \leq \lim_{n\to\infty} \frac{1}{n} \log \left( k \max_i a_i^n \right) = \lim_{n\to\infty} \frac{1}{n} \log \max_i a_i^n. \qquad \square$$

## Appendix B

# Items Omitted in the Main Text

## B.1 Proof of Theorem 3.5

The rate function for $P\{FS_2\}$ is the rate function for the probability that system 1 is estimated feasible, but another estimated-feasible system has a better estimated objective value. Since the estimated set of feasible systems $\bar{\Gamma}$ may contain worse feasible systems ($i \in \Gamma$), better infeasible systems ($i \in \mathcal{S}_b$), and worse infeasible systems ($i \in \mathcal{S}_w$), in lemma B.1 we strategically consider the rate functions for the probability that system 1 is beaten by a system in $\bar{\Gamma} \cap \Gamma$, $\bar{\Gamma} \cap \mathcal{S}_b$, or $\bar{\Gamma} \cap \mathcal{S}_w$ separately. Lemmas B.3 – B.5 provide specific statements of these three rate functions over the sets $\Gamma, \mathcal{S}_b$, and $\mathcal{S}_w$, respectively. Lemma B.2 provides a useful bookkeeping-type result that is the starting point for lemmas B.3 – B.5.

Assuming for now that the required limits exist, lemma B.1 states that the rate function of $P\{FS_2\}$ is determined by the slowest-converging probability that system 1 will be "beaten" by an estimated-feasible system from $\Gamma, \mathcal{S}_b$, or $\mathcal{S}_w$.

**Lemma B.1.** *The rate function for $P\{FS_2\}$ is given by the minimum rate function of the probability that system 1 is beaten by an estimated-feasible system that is (i) feasible and worse, (ii) infeasible and better, or (iii) infeasible and worse. That is,*

$$- \lim_{n\to\infty} \frac{1}{n} \log P\{FS_2\} = \min\Bigg( - \lim_{n\to\infty} \frac{1}{n} \log P\{\cup_{i\in\bar{\Gamma}\cap\Gamma} \hat{H}_1 \geq \hat{H}_i\}, \\ - \lim_{n\to\infty} \frac{1}{n} \log P\{\cup_{i\in\bar{\Gamma}\cap\mathcal{S}_b} \hat{H}_1 \geq \hat{H}_i\}, - \lim_{n\to\infty} \frac{1}{n} \log P\{\cup_{i\in\bar{\Gamma}\cap\mathcal{S}_w} \hat{H}_1 \geq \hat{H}_i\}\Bigg). \tag{B.1}$$

*Proof.* From equation (3.3), the probability that system 1 is beaten by another estimated-feasible system can be written as

$$P\{\cup_{i\in\bar{\Gamma}} \hat{H}_1 \geq \hat{H}_i\} = P\{(\cup_{i\in\bar{\Gamma}\cap\Gamma} \hat{H}_1 \geq \hat{H}_i) \cup (\cup_{i\in\bar{\Gamma}\cap\mathcal{S}_b} \hat{H}_1 \geq \hat{H}_i) \cup (\cup_{i\in\bar{\Gamma}\cap\mathcal{S}_w} \hat{H}_1 \geq \hat{H}_i)\}.$$

We have

$$\begin{aligned} &\frac{1}{n} \log\Big(\max\Big(P\{\cup_{i\in\bar{\Gamma}\cap\Gamma} \hat{H}_1 \geq \hat{H}_i\}, P\{\cup_{i\in\bar{\Gamma}\cap\mathcal{S}_b} \hat{H}_1 \geq \hat{H}_i\}, P\{\cup_{i\in\bar{\Gamma}\cap\mathcal{S}_w} \hat{H}_1 \geq \hat{H}_i\}\Big)\Big) \\ &\leq \frac{1}{n} \log P\{\cup_{i\in\bar{\Gamma}} \hat{H}_1 \geq \hat{H}_i\} \\ &\leq \frac{1}{n} \log\Big(P\{\cup_{i\in\bar{\Gamma}\cap\Gamma} \hat{H}_1 \geq \hat{H}_i\} + P\{\cup_{i\in\bar{\Gamma}\cap\mathcal{S}_b} \hat{H}_1 \geq \hat{H}_i\} + P\{\cup_{i\in\bar{\Gamma}\cap\mathcal{S}_w} \hat{H}_1 \geq \hat{H}_i\}\Big). \end{aligned}$$

Assuming the relevant limits exist, the conclusion is reached by noting that the limit of the left-hand and right-hand sides are equivalent by proposition A.2 and the principle of the slowest term, respectively. $\square$

Next, we will individually consider each of the terms on the right-hand side of equation (B.1), and establish their respective limits. However before proceeding to these results, we first present the following lemma which is a preliminary step for the proofs that follow. Lemma B.2 uses the law of total probability to further separate the events involved in system 1 being "beaten" by another estimated-feasible system.

**Lemma B.2.** *For sets of systems* $\mathcal{S} \in \{\Gamma, \mathcal{S}_b, \mathcal{S}_w\}$ *and* $C \subseteq \mathcal{S}$,

$$P\{\cup_{i \in \bar{\Gamma} \cap \mathcal{S}} \hat{H}_1 \geq \hat{H}_i\}$$
$$= \sum_C P\Big\{\underbrace{(\cup_{i \in C} \hat{H}_1 \geq \hat{H}_i)}_{\substack{\text{1 beaten by} \\ \text{system(s) in } C}} \cap \underbrace{(\cap_{i \in C} \; \cap_{j \in \mathcal{C}_F^i} \hat{G}_{ij} \leq \gamma_j)}_{\substack{\text{all systems in } C \\ \text{declared feasible on} \\ \text{all feasible constraints}}} \cap \underbrace{(\cap_{i \in C} \; \cap_{j \in \mathcal{C}_I^i} \hat{G}_{ij} \leq \gamma_j)}_{\substack{\text{all systems in } C \\ \text{declared feasible on all} \\ \text{infeasible constraints}}} \cap \underbrace{(\cap_{i \in \mathcal{S} \setminus C} \cup_j \hat{G}_{ij} > \gamma_j)}_{\substack{\text{all systems in } \mathcal{S}\setminus C \\ \text{declared infeasible} \\ \text{on at least one} \\ \text{constraint}}}\Big\}.$$

*Proof.* By the law of total probability, for some set of systems $C \subseteq \mathcal{S}$,

$$P\{\cup_{i \in \bar{\Gamma} \cap \mathcal{S}} \hat{H}_1 \geq \hat{H}_i\} = \sum_C P\{(\cup_{i \in \bar{\Gamma} \cap \mathcal{S}} \hat{H}_1 \geq \hat{H}_i) \cap (\bar{\Gamma} \cap \mathcal{S} = C)\}$$
$$= \sum_C P\{(\cup_{i \in C} \hat{H}_1 \geq \hat{H}_i) \cap (\bar{\Gamma} \cap \mathcal{S} = C)\} = \sum_C P\{(\cup_{i \in C} \hat{H}_1 \geq \hat{H}_i) \cap_{i \in C} \; (i \in \bar{\Gamma}) \cap_{i \in \mathcal{S} \setminus C} (i \notin \bar{\Gamma})\}$$
$$= \sum_C P\{(\cup_{i \in C} \hat{H}_1 \geq \hat{H}_i) \cap (\cap_{i \in C} \; (\cap_{j \in \mathcal{C}_F^i} \hat{G}_{ij} \leq \gamma_j \cap_{j \in \mathcal{C}_I^i} \hat{G}_{ij} \leq \gamma_j)) \cap (\cap_{i \in \mathcal{S} \setminus C} \cup_j \hat{G}_{ij} > \gamma_j)\}. \qquad \square$$

Let us now consider the rate function of the probability that system 1 is "beaten" by a worse estimated-feasible system from $\Gamma$. Since $\bar{\Gamma}$ is equivalent to $\Gamma$ in the limit, and we are considering only the probability that system 1 is beaten by another truly feasible system, we expect that the rate function will be the same as in the unconstrained case presented by Glynn and Juneja (2004). Also, since system 1 can be beaten by any system in $\bar{\Gamma} \cap \Gamma$, we intuitively expect the rate function to be the minimum rate function across all systems in $\Gamma$, corresponding to the system that is "best" at crossing the optimality hurdle. Lemma B.3 states that this is indeed the case.

**Lemma B.3.** *The rate function for the probability that system 1 is estimated feasible and has a worse estimated objective value than an estimated-feasible system from* $\Gamma$ *(feasible and worse) is*

$$-\lim_{n \to \infty} \frac{1}{n} \log P\{\cup_{i \in \bar{\Gamma} \cap \Gamma} \hat{H}_1 \geq \hat{H}_i\} = \min_{i \in \Gamma} \left( \inf_x (\alpha_1 I_1(x) + \alpha_i I_i(x)) \right).$$

*Proof.* From lemma B.2, let $\mathcal{S} = \Gamma$ and therefore $C \subseteq \Gamma$. Then

$$P\{\cup_{i\in\bar{\Gamma}\cap\Gamma}\ \hat{H}_1 \geq \hat{H}_i\}$$
$$= \sum_{C} P\{(\cup_{i\in C}\hat{H}_1 \geq \hat{H}_i) \cap (\cap_{i\in C} \cap_{j\in\mathcal{C}_F^i} \hat{G}_{ij} \leq \gamma_j) \cap (\cap_{i\in C} \cap_{j\in\mathcal{C}_I^i} \hat{G}_{ij} \leq \gamma_j) \cap (\cap_{i\in\Gamma\setminus C} \cup_j \hat{G}_{ij} > \gamma_j)\}$$

We derive a lower bound bound by letting $C = \Gamma$ and noticing that all constraints are feasible for all $i \in \Gamma$. Then

$$\begin{aligned} P\{\cup_{i\in\bar{\Gamma}\cap\Gamma}\ \hat{H}_1 \geq \hat{H}_i\} &\geq P\{(\cup_{i\in\Gamma}\hat{H}_1 \geq \hat{H}_i) \cap (\cap_{i\in\Gamma} \cap_j \hat{G}_{ij} \leq \gamma_j)\} \\ &\geq \max_{i\in\Gamma} P\{(\hat{H}_1 \geq \hat{H}_i) \cap (\cap_{i\in\Gamma} \cap_j \hat{G}_{ij} \leq \gamma_j)\}. \end{aligned}$$

We derive an upper bound by noting that,

$$P\{\cup_{i\in\bar{\Gamma}\cap\Gamma}\ \hat{H}_1 \geq \hat{H}_i\} \leq P\{\cup_{i\in\Gamma}\ \hat{H}_1 \geq \hat{H}_i\} \leq |\Gamma| \max_{i\in\Gamma} P\{\hat{H}_1 \geq \hat{H}_i\}. \tag{B.2}$$

Then

$$\max_{i\in\Gamma} P\{(\hat{H}_1 \geq \hat{H}_i) \cap (\cap_{i\in\Gamma} \cap_j \hat{G}_{ij} \leq \gamma_j)\} \leq P\{\cup_{i\in\bar{\Gamma}\cap\Gamma}\ \hat{H}_1 \geq \hat{H}_i\} \leq |\Gamma| \max_{i\in\Gamma} P\{\hat{H}_1 \geq \hat{H}_i\}.$$

By proposition A.2 and the assumptions regarding independence (assumptions 1 and 3.1), the rate function for the lower bound is,

$$\lim_{n\to\infty} \frac{1}{n} \log \max_{i\in\Gamma} P\{(\hat{H}_1 \geq \hat{H}_i) \cap (\cap_{i\in\Gamma} \cap_j \hat{G}_{ij} \leq \gamma_j)\}$$
$$= \max_{i\in\Gamma} \lim_{n\to\infty} \frac{1}{n} \log P\{\underbrace{(\hat{H}_1 \geq \hat{H}_i)}_{\text{pr } \to 0} \cap \underbrace{(\cap_{i\in\Gamma}\ \cap_j \hat{G}_{ij} \leq \gamma_j)}_{\text{pr } \to 1}\} = \max_{i\in\Gamma} \lim_{n\to\infty} \frac{1}{n} \log P\{\hat{H}_1 \geq \hat{H}_i\}.$$

Likewise applying proposition A.2 to the upper bound in the right-hand side of equation (B.2), we find that the rate function for the upper bound is equivalent to the rate function for the lower bound. By Glynn and Juneja (2004),

$$-\lim_{n\to\infty} \frac{1}{n} \log P\{\hat{H}_1 \geq \hat{H}_i\} = \inf_x (\alpha_1 I_1(x) + \alpha_i I_i(x)),$$

and hence the conclusion follows. □

We now consider the rate function of the probability that system 1 has a worse estimated objective value than an estimated-feasible system from $\mathcal{S}_b$ (infeasible but better). Since the only hurdle to an infeasible, better system being falsely selected as the best system is feasibility, the only terms that appear in the rate function in the prior lemma are with regard to feasibility.

**Lemma B.4.** *The rate function for the probability that system 1 is estimated feasible and has a worse estimated objective value than an estimated-feasible system from $\mathcal{S}_b$ (infeasible and better) is*

$$-\lim_{n\to\infty}\frac{1}{n}\log P\{\cup_{i\in\bar{\Gamma}\cap\mathcal{S}_b}\hat{H}_1 \geq \hat{H}_i\} = \min_{i\in\mathcal{S}_b}\alpha_i \sum_{j\in\mathcal{C}_I^i} J_{ij}(\gamma_j).$$

*Proof.* From lemma B.2, let $\mathcal{S} = \mathcal{S}_b$ and therefore $C \subseteq \mathcal{S}_b$. An upper bound is given by

$$P\{\cup_{i\in\bar{\Gamma}\cap\mathcal{S}_b}\hat{H}_1 \geq \hat{H}_i\}$$
$$= \sum_C P\{(\cup_{i\in C}\hat{H}_1 \geq \hat{H}_i) \cap (\cap_{i\in C}\cap_{j\in\mathcal{C}_F^i}\hat{G}_{ij} \leq \gamma_j) \cap (\cap_{i\in C}\cap_{j\in\mathcal{C}_I^i}\hat{G}_{ij} \leq \gamma_j) \cap (\cap_{i\in\mathcal{S}_b\setminus C}\cup_j \hat{G}_{ij} > \gamma_j)\} \tag{B.3}$$

$$\leq \sum_C P\{\cap_{i\in C}\ \cap_{j\in\mathcal{C}_I^i}\hat{G}_{ij} \leq \gamma_j\} \leq \sum_C \max_{i\in C} P\{\cap_{j\in\mathcal{C}_I^i}\hat{G}_{ij} \leq \gamma_j\} \leq \sum_{d=1}^{|\mathcal{S}_b|}\binom{|\mathcal{S}_b|}{d}\max_{i\in\mathcal{S}_b}P\{\cap_{j\in\mathcal{C}_I^i}\hat{G}_{ij} \leq \gamma_j\}$$
$$\leq 2^{|\mathcal{S}_b|}\max_{i\in\mathcal{S}_b}P\{\cap_{j\in\mathcal{C}_I^i}\hat{G}_{ij} \leq \gamma_j\}.$$

Therefore the rate function for the upper bound is

$$\lim_{n\to\infty}\frac{1}{n}\log\max_{i\in\mathcal{S}_b}P\{\cap_{j\in\mathcal{C}_I^i}\hat{G}_{ij} \leq \gamma_j\}. \tag{B.4}$$

Let $k^* = \arg\max_{i\in\mathcal{S}_b} P\{\cap_{j\in\mathcal{C}_I^i}\hat{G}_{ij} \leq \gamma_j\}$. A lower bound may be found by letting $k^*$ be the only element in $C$. Continuing from equation (B.3), the lower bound is given by,

$$\sum_C P\{(\cup_{i\in C}\hat{H}_1 \geq \hat{H}_i) \cap (\cap_{i\in C}\cap_{j\in\mathcal{C}_F^i}\hat{G}_{ij} \leq \gamma_j) \cap (\cap_{i\in C}\cap_{j\in\mathcal{C}_I^i}\hat{G}_{ij} \leq \gamma_j) \cap (\cap_{i\in\mathcal{S}_b\setminus C}\cup_j \hat{G}_{ij} > \gamma_j)\}$$
$$\geq P\{(\hat{H}_1 \geq \hat{H}_{k^*}) \cap (\cap_{j\in\mathcal{C}_F^{k^*}}\hat{G}_{k^*j} \leq \gamma_j) \cap (\cap_{j\in\mathcal{C}_I^{k^*}}\hat{G}_{k^*j} \leq \gamma_j) \cap (\cap_{i\in\mathcal{S}_b\setminus\{k^*\}}\cup_j \hat{G}_{ij} > \gamma_j)\}.$$

Under the assumptions regarding independence (assumptions 1 and 3.1), for the lower bound,

$$\lim_{n\to\infty}\frac{1}{n}\log P\{\underbrace{(\hat{H}_1 \geq \hat{H}_{k^*}) \cap (\cap_{j\in\mathcal{C}_F^{k^*}}\hat{G}_{k^*j} \leq \gamma_j)}_{\text{pr }\to 1} \cap \underbrace{(\cap_{j\in\mathcal{C}_I^{k^*}}\hat{G}_{k^*j} \leq \gamma_j)}_{\text{pr }\to 0} \cap \underbrace{(\cap_{i\in\mathcal{S}_b\setminus\{k^*\}}\cup_j \hat{G}_{ij} > \gamma_j)}_{\text{pr }\to 1}\}$$
$$= \lim_{n\to\infty}\frac{1}{n}\log P\{\cap_{j\in\mathcal{C}_I^{k^*}}\hat{G}_{k^*j} \leq \gamma_j\} \equiv \lim_{n\to\infty}\frac{1}{n}\log\max_{i\in\mathcal{S}_b}P\{\cap_{j\in\mathcal{C}_I^i}\hat{G}_{ij} \leq \gamma_j\},$$

which is the rate function for the probability that system $k^*$ is falsely estimated as feasible on all constraints for which it is truly infeasible. This rate function is equivalent to the rate function for the upper bound in equation (B.4). By proposition A.2 and the assumptions regarding independence, the rate function for the upper and lower bounds is,

$$\lim_{n\to\infty}\frac{1}{n}\log\max_{i\in\mathcal{S}_b}P\{\cap_{j\in\mathcal{C}_I^i}\hat{G}_{ij}\leq\gamma_j\}=\max_{i\in\mathcal{S}_b}\sum_{j\in\mathcal{C}_I^i}\lim_{n\to\infty}\frac{1}{n}\log P\{\hat{G}_{ij}\leq\gamma_j\}=-\min_{i\in\mathcal{S}_b}\alpha_i\sum_{j\in\mathcal{C}_I^i}J_{ij}(\gamma_j),$$

and the conclusion follows. □

Finally, consider the rate function for the probability that system 1 has a worse estimated objective value than an estimated-feasible system from $\mathcal{S}_w$ (infeasible and worse). To be falsely classified as the best feasible system, systems in $\mathcal{S}_w$ must be falsely estimated as optimal and falsely estimated as feasible on all constraints for which they are truly infeasible. There are two hurdles for a system in $\mathcal{S}_w$ to be declared the best feasible system: optimality and feasibility. Thus the overall rate function for systems in $\mathcal{S}_w$ is determined by the system that is the best at "pretending" to be optimal *and* feasible. Lemma B.5 states this result formally.

**Lemma B.5.** *The rate function for the probability that system 1 is estimated feasible and has a worse estimated objective value than an estimated-feasible system from $\mathcal{S}_w$ (infeasible and worse) is*

$$-\lim_{n\to\infty}\frac{1}{n}\log P\{\cup_{i\in\bar{\Gamma}\cap\mathcal{S}_w}\hat{H}_1\geq\hat{H}_i\}=\min_{i\in\mathcal{S}_w}\left(\inf_x(\alpha_1 I_1(x)+\alpha_i I_i(x))+\alpha_i\sum_{j\in\mathcal{C}_I^i}J_{ij}(\gamma_j)\right).$$

*Proof.* From lemma B.2, let $\mathcal{S}=\mathcal{S}_w$ and therefore $C\subseteq\mathcal{S}_w$. An upper bound is given by

$$\begin{aligned}
&P\{\cup_{i\in\bar{\Gamma}\cap\mathcal{S}_w}\hat{H}_1\geq\hat{H}_i\}\\
&=\sum_C P\{(\cup_{i\in C}\hat{H}_1\geq\hat{H}_i)\cap(\cap_{i\in C}\cap_{j\in\mathcal{C}_F^i}\hat{G}_{ij}\leq\gamma_j)\cap(\cap_{i\in C}\cap_{j\in\mathcal{C}_I^i}\hat{G}_{ij}\leq\gamma_j)\cap(\cap_{i\in\mathcal{S}_w\setminus C}\cup_j\hat{G}_{ij}>\gamma_j)\} \qquad \text{(B.5)}\\
&\leq\sum_C P\{(\cup_{i\in C}\hat{H}_1\geq\hat{H}_i)\cap(\cap_{i\in C}\cap_{j\in\mathcal{C}_I^i}\hat{G}_{ij}\leq\gamma_j)\}\leq\sum_C P\{\cup_{i\in C}(\hat{H}_1\geq\hat{H}_i\cap(\cap_{j\in\mathcal{C}_I^i}\hat{G}_{ij}\leq\gamma_j))\}\\
&\leq\sum_C |C|\max_{i\in C}P\{(\hat{H}_1\geq\hat{H}_i)\cap(\cap_{j\in\mathcal{C}_I^i}\hat{G}_{ij}\leq\gamma_j)\}\\
&\leq\sum_{d=1}^{|\mathcal{S}_w|}\binom{|\mathcal{S}_w|}{d}|\mathcal{S}_w|\max_{i\in\mathcal{S}_w}P\{(\hat{H}_1\geq\hat{H}_i)\cap(\cap_{j\in\mathcal{C}_I^i}\hat{G}_{ij}\leq\gamma_j)\}\\
&\leq 2^{|\mathcal{S}_w|}|\mathcal{S}_w|\max_{i\in\mathcal{S}_w}P\{(\hat{H}_1\geq\hat{H}_i)\cap(\cap_{j\in\mathcal{C}_I^i}\hat{G}_{ij}\leq\gamma_j)\}.
\end{aligned}$$

Therefore the rate function for the upper bound is

$$\lim_{n\to\infty} \frac{1}{n} \log \max_{i\in\mathcal{S}_w} P\{(\hat{H}_1 \geq \hat{H}_i) \cap (\cap_{j\in\mathcal{C}_I^i} \hat{G}_{ij} \leq \gamma_j)\}. \tag{B.6}$$

Let $k^* = \arg\max_{i\in\mathcal{S}_w} P\{(\hat{H}_1 \geq \hat{H}_i) \cap (\cap_{j\in\mathcal{C}_I^i} \hat{G}_{ij} \leq \gamma_j)\}$. A lower bound is found by letting $k^*$ be the only element in $C$. Continuing from equation (B.5),

$$\begin{aligned}
&\sum_C P\{(\cup_{i\in C} \hat{H}_1 \geq \hat{H}_i) \cap (\cap_{i\in C} \cap_{j\in\mathcal{C}_F^i} \hat{G}_{ij} \leq \gamma_j) \cap (\cap_{i\in C} \cap_{j\in\mathcal{C}_I^i} \hat{G}_{ij} \leq \gamma_j) \cap (\cap_{i\in\mathcal{S}_w\setminus C} \cup_j \hat{G}_{ij} > \gamma_j)\} \\
&\geq P\{(\hat{H}_1 \geq \hat{H}_{k^*}) \cap (\cap_{j\in\mathcal{C}_F^{k^*}} \hat{G}_{k^*j} \leq \gamma_j) \cap (\cap_{j\in\mathcal{C}_I^{k^*}} \hat{G}_{k^*j} \leq \gamma_j) \cap (\cap_{i\in\mathcal{S}_w\setminus\{k^*\}} \cup_j \hat{G}_{ij} > \gamma_j)\}.
\end{aligned}$$

Under the assumptions regarding independence (assumptions 1 and 3.1),

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} \log P\{\underbrace{(\hat{H}_1 \geq \hat{H}_{k^*})}_{\text{pr }\to 0} \cap \underbrace{(\cap_{j\in\mathcal{C}_F^{k^*}} \hat{G}_{k^*j} \leq \gamma_j)}_{\text{pr }\to 1} \cap \underbrace{(\cap_{j\in\mathcal{C}_I^{k^*}} \hat{G}_{k^*j} \leq \gamma_j)}_{\text{pr }\to 0} \cap \underbrace{(\cap_{i\in\mathcal{S}_w\setminus\{k^*\}} \cup_j \hat{G}_{ij} > \gamma_j)}_{\text{pr }\to 1}\} \\
&= \lim_{n\to\infty} \frac{1}{n} \log P\{\underbrace{(\hat{H}_1 \geq \hat{H}_{k^*})}_{\text{pr }\to 0} \cap \underbrace{(\cap_{j\in\mathcal{C}_I^{k^*}} \hat{G}_{k^*j} \leq \gamma_j)}_{\text{pr }\to 0}\} \\
&\equiv \lim_{n\to\infty} \frac{1}{n} \log \max_{i\in\mathcal{S}_w} P\{(\hat{H}_1 \geq \hat{H}_i) \cap (\cap_{j\in\mathcal{C}_I^i} \hat{G}_{ij} \leq \gamma_j)\},
\end{aligned}$$

which is the rate function for the probability that system $k^*$ is falsely estimated as optimal and feasible on all constraints for which it is truly infeasible. This rate function is equivalent to the rate function for the upper bound in equation (B.6). By proposition A.2 and the assumptions regarding independence, the rate function for the upper and lower bounds is,

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} \log \max_{i\in\mathcal{S}_w} P\{(\hat{H}_1 \geq \hat{H}_i) \cap (\cap_{j\in\mathcal{C}_I^i} \hat{G}_{ij} \leq \gamma_j)\} \\
&= \max_{i\in\mathcal{S}_w} \lim_{n\to\infty} \frac{1}{n} \log P\{(\hat{H}_1 \geq \hat{H}_i) \cap (\cap_{j\in\mathcal{C}_I^i} \hat{G}_{ij} \leq \gamma_j)\} \\
&= \max_{i\in\mathcal{S}_w} \lim_{n\to\infty} \left( \frac{1}{n} \log P\{\hat{H}_1 \geq \hat{H}_i\} + \sum_{j\in\mathcal{C}_I^i} \frac{1}{n} \log P\{\hat{G}_{ij} \leq \gamma_j\} \right)
\end{aligned}$$

Applying previous results, the conclusion follows. □

*Proof of Theorem 3.5.* Theorem 3.5 follows by substituting the results from lemmas B.3–B.5 into the result presented in lemma B.1. □

## B.2 Alternate Proof of Proposition 4.4

We provide the following alternate, algebraic proof for proposition 4.4. Loosely speaking, proposition 4.4 states that if $\rho_1 < 0$ in the formulation of the bivariate normal case presented in section 4.3.1, then the sets $\Gamma_u(\rho,\alpha)$ and $\mathcal{S}_{w,u}(\rho,\alpha)$ are empty. This result implies that when $\rho_1 < 0$, $\rho_1$ does not affect the rate function of any system.

*Proof of proposition 4.4. Case* $\Gamma_u(\rho,\alpha)$. Recall the following definition of $\Gamma_u(\rho,\alpha)$:

$$\Gamma_u(\rho,\alpha) = \Bigg\{ i : h_i > h_1 + \rho_1\sigma_{h_1}\frac{(\gamma - g_1)}{\sigma_{g_1}},\ \gamma - \rho_1\frac{\sigma_{g_1}}{\sigma_{h_1}}\frac{(\sigma^2_{h_1}/\alpha_1)}{(\sigma^2_{h_1}/\alpha_1) + (\sigma^2_{h_i}/\alpha_i)}(h_i - h_1) < g_1 \le \gamma,$$
$$g_i \le \gamma + \rho_i\frac{\sigma_{g_i}}{\sigma_{h_i}}\frac{(\sigma^2_{h_i}/\alpha_i)}{((1-\rho_1^2)\sigma^2_{h_1}/\alpha_1) + (\sigma^2_{h_i}/\alpha_i)}\left(h_i - \big(h_1 + \rho_1\sigma_{h_1}\frac{(\gamma-g_1)}{\sigma_{g_1}}\big)\right)\Bigg\}.$$

Suppose that $\Gamma_u(\rho,\alpha)$ is nonempty and $\rho_1 < 0$. From the condition on $g_1$, it must be the case that $h_i - h_1 < 0$. Now let us consider the conditions on $h_i$ and $g_1$ together, writing them as

$$\rho_1\frac{\sigma_{h_1}}{(h_i - h_1)}\frac{(\gamma - g_1)}{\sigma_{g_1}} > 1 \ \text{ and } \ \frac{\sigma_{h_1}}{(h_i - h_1)}\frac{(\gamma - g_1)}{\sigma_{g_1}} > \rho_1\frac{\sigma_{g_1}}{\sigma_{h_1}}\frac{(\sigma^2_{h_1}/\alpha_1)}{(\sigma^2_{h_1}/\alpha_1) + (\sigma^2_{h_i}/\alpha_i)}. \tag{B.7}$$

Multiplying the right-hand inequality in equation (B.7) by $\rho_1$ on both sides and combining the statements yields

$$\rho_1^2\frac{(\sigma^2_{h_1}/\alpha_1)}{(\sigma^2_{h_1}/\alpha_1) + (\sigma^2_{h_i}/\alpha_i)} > \rho_1\frac{\sigma_{h_1}}{(h_i - h_1)}\frac{(\gamma - g_1)}{\sigma_{g_1}} > 1. \tag{B.8}$$

However now we have a contradiction, since the left-hand side of equation (B.8) must be less than one.

*Case* $\mathcal{S}_{w,u}(\rho,\alpha)$. The proof of the $\mathcal{S}_{w,u}(\rho,\alpha)$ case proceeds in an similar way to the proof of $\Gamma_u(\rho,\alpha)$. Recall the following definition of $\mathcal{S}_{w,u}(\rho,\alpha)$:

$$\mathcal{S}_{w,u}(\rho,\alpha) = \Bigg\{ i : h_i - \rho_i\sigma_{h_i}\frac{(g_i - \gamma)}{\sigma_{g_i}} > h_1 + \rho_1\sigma_{h_1}\frac{(\gamma - g_1)}{\sigma_{g_1}},$$
$$\gamma - \rho_1\frac{\sigma_{g_1}}{\sigma_{h_1}}\frac{(\sigma^2_{h_1}/\alpha_1)}{(\sigma^2_{h_1}/\alpha_1) + ((1-\rho_i^2)\sigma^2_{h_i}/\alpha_i)}\left(\big(h_i - \rho_i\sigma_{h_i}\frac{(g_i-\gamma)}{\sigma_{g_i}}\big) - h_1\right) < g_1 \le \gamma,$$
$$g_i > \gamma + \rho_i\frac{\sigma_{g_i}}{\sigma_{h_i}}\frac{\sigma^2_{h_i}/\alpha_i}{((1-\rho_1^2)\sigma^2_{h_1}/\alpha_1) + (\sigma^2_{h_i}/\alpha_i)}\left(h_i - \big(h_1 + \rho_1\sigma_{h_1}\frac{(\gamma-g_1)}{\sigma_{g_1}}\big)\right)\Bigg\}.$$

Suppose that $\mathcal{S}_{w,u}(\rho,\alpha)$ is nonempty and $\rho_1 < 0$. From the condition on $g_1$, it must be the case that $\left(\big(h_i - \rho_i\sigma_{h_i}\frac{(g_i-\gamma)}{\sigma_{g_i}}\big) - h_1\right) < 0$. Now let us consider the conditions on $h_i$ and $g_1$ together, writing them as

$$\rho_1 \frac{\sigma_{h_1}}{\left(\left(h_i - \rho_i \sigma_{h_i} \frac{(g_i - \gamma)}{\sigma_{g_i}}\right) - h_1\right)} \frac{(\gamma - g_1)}{\sigma_{g_1}} > 1 \tag{B.9}$$

and

$$\frac{\sigma_{h_1}}{\left(\left(h_i - \rho_i \sigma_{h_i} \frac{(g_i - \gamma)}{\sigma_{g_i}}\right) - h_1\right)} \frac{(\gamma - g_1)}{\sigma_{g_1}} > \rho_1 \frac{(\sigma_{h_1}^2/\alpha_1)}{(\sigma_{h_1}^2/\alpha_1) + ((1-\rho_i^2)\sigma_{h_i}^2/\alpha_i)}. \tag{B.10}$$

Multiplying both sides of equation (B.10) by $\rho_1$ and combining the resulting inequality with the inequality in equation (B.9),

$$\rho_1^2 \frac{(\sigma_{h_1}^2/\alpha_1)}{(\sigma_{h_1}^2/\alpha_1) + ((1-\rho_i^2)\sigma_{h_i}^2/\alpha_i)} > \rho_1 \frac{\sigma_{h_1}}{\left(\left(h_i - \rho_i \sigma_{h_i} \frac{(g_i - \gamma)}{\sigma_{g_i}}\right) - h_1\right)} \frac{(\gamma - g_1)}{\sigma_{g_1}} > 1. \tag{B.11}$$

Now there is a contradiction, since the left-hand side of equation (B.11) must be less than one. $\square$

## B.3 Algorithm for Creating Random Multivariate Normal Problems (Problems $\boldsymbol{P}$)

The following algorithm is used to create random instances of multivariate normal problems $P$ (see the problem statement in section 1.2). This algorithm is referenced in section 5.7.

---

**Algorithm B.5** Algorithm that generates a random Problem $P$ in the multivariate normal context

---

**Require:** Number of systems $numsys > 0$; Number of constraints $numcon > 0$.

1: Initialize: the constraint values $\gamma_j = 0$ for all $j \leq s$; set the tolerance value $tolerance = 0.05$.
2: **for** $i = 1$ to $numsys$ **do**
3:   Generate the objective value $h_i$ as a uniformly distributed random variate in $[-3, 3]$
4:   **for** $j = 1$ to $numcon$ **do**
5:     **if** $i <=$ ceiling($numsys/3$) **then**
6:       Generate the constraint value $g_{ij}$ as a uniformly distributed random variate in $[-3, 0]$. {Put approximately the first one-third of systems in the feasible set.}
7:     **else**
8:       Generate the constraint value $g_{ij}$ as a uniformly distributed random variate in $[-3, -3]$.
9:     **end if**
10:     **if** $|g_{ij} - \gamma_j| < tolerance$ **then**
11:       Go to step 5 {If the system is too close to the constraint, move it.}
12:     **end if**
13:   **end for**
14: **end for**
15: Find the location of system 1.
16: **for** $i = 1$ to $numsys$ **do**
17:   **while** system $i$ is not system 1 and $|h_i - h_1| < tolerance$ **do**
18:     Generate a new $h_i$ value as a uniformly distributed random variate in $[-3, 3]$.
19:   **end while**
20:   **if** system $i$ is feasible and $h_i < h_1$ **then**
21:     Go to step 15. {The location of system 1 has changed.}
22:   **end if**
23: **end for**
24: **for** $i = 1$ to $numsys$ **do**
25:   Create a $(numcon + 1)$-by-$(numcon + 1)$ matrix $X$ with all entries generated as uniformly distributed random variates in $[-1, 1]$.
26:   Calculate $X^T X$. {The matrix $X^T X$ is at least positive semi-definite, if not positive definite.}
27:   Calculate the eigenvalues of $X^T X$.
28:   **if** any eigenvalue of $X^T X < tolerance$ **then**
29:     Go to step 25.
30:   **else**
31:     Assign $X^T X$ as the covariance matrix for system $i$.
32:   **end if**
33: **end for**

---

## Appendix C

# Source Code

## C.1 Mathematica Code for Symbolic Solution of Bivariate Normal Rate Functions, First Formulation

```
In[1]:= Clear["`*"];(* Clear["`*"] clears all input variables *)
Sig1 = {{sh1^2,p1*sh1*sg1},{p1*sh1*sg1,sg1^2}};
Sigi = {{shi^2,pi*shi*sgi},{pi*shi*sgi,sgi^2}} ;
x1y1 = Transpose[{{x1,y1}}]; xiyi = Transpose[{{xi,yi}}];
h1g1 = Transpose[{{h1,g1}}]; higi = Transpose[{{hi,gi}}];
K[x1y1v_,xiyiv_]:=(a1/2)*Transpose[(x1y1v-h1g1)].Inverse[Sig1].(x1y1v-
    h1g1)+(ai/2)*Transpose[(xiyiv-higi)].Inverse[Sigi].(xiyiv-higi);
L[]:=K[x1y1,xiyi]+lamx*(xi-x1)+lamy1*(y1-gam)+lamyi*(yi-gam);
gradL = {{Part[Simplify[D[L[],x1]],1,1]},{Part[Simplify[D[L[],xi
    ]],1,1]},{Part[Simplify[D[L[],y1]],1,1]},{Part[Simplify[D[L[],yi
    ]],1,1]}}//MatrixForm
Simplify[Solve[{D[L[],x1]==0},{lamx}]]
lamx1e=lamx/.% ; (*get expression for solution*)
Simplify[Solve[{D[L[],xi]==0},{lamx}]]
lamxie=lamx/.%;
Simplify[Solve[{D[L[],y1]==0},{lamy1}]]
lamy1e = lamy1/.%;
Simplify[Solve[{D[L[],yi]==0},{lamyi}]]
lamyie = lamyi/.%;
Simplify[Solve[{lamx1e==lamxie},{x1}]]
x1e=x1/.%;
lamx1eFn[x1_,y1_]:=(a1 (h1 sg1-g1 p1 sh1-sg1 x1+p1 sh1 y1))/((-1+p1^2)
    sg1 sh1^2);
lamxieFn[xi_,yi_]:=(ai (-hi sgi+gi pi shi+sgi xi-pi shi yi))/((-1+pi^2)
     sgi shi^2);
lamy1eFn[x1_,y1_]:=(a1 (h1 p1 sg1-g1 sh1-p1 sg1 x1+sh1 y1))/((-1+p1^2)
    sg1^2 sh1);
lamyieFn[xi_,yi_]:= (ai (hi pi sgi-gi shi-pi sgi xi+shi yi))/((-1+pi^2)
     sgi^2 shi);

Out[8]//MatrixForm= (
```

```
(−lamx (−1+p1^2) sg1 sh1^2+a1 (h1 sg1−g1 p1 sh1−sg1 x1+p1 sh1 y1))
    /((−1+p1^2) sg1 sh1^2)
(lamx (−1+pi^2) sgi shi^2+ai (hi sgi−gi pi shi−sgi xi+pi shi yi))/((−1+
    pi^2) sgi shi^2)
(lamy1 (−1+p1^2) sg1^2 sh1+a1 (−h1 p1 sg1+g1 sh1+p1 sg1 x1−sh1 y1))
    /((−1+p1^2) sg1^2 sh1)
(lamyi (−1+pi^2) sgi^2 shi+ai (−hi pi sgi+gi shi+pi sgi xi−shi yi))
    /((−1+pi^2) sgi^2 shi)
)
Out[9]= {{lamx−>(a1 (h1 sg1−g1 p1 sh1−sg1 x1+p1 sh1 y1))/((−1+p1^2) sg1
     sh1^2)}}
Out[11]= {{lamx−>(ai (−hi sgi+gi pi shi+sgi xi−pi shi yi))/((−1+pi^2)
    sgi shi^2)}}
Out[13]= {{lamy1−>(a1 (h1 p1 sg1−g1 sh1−p1 sg1 x1+sh1 y1))/((−1+p1^2)
    sg1^2 sh1)}}
Out[15]= {{lamyi−>(ai (hi pi sgi−gi shi−pi sgi xi+shi yi))/((−1+pi^2)
    sgi^2 shi)}}
Out[17]= {{x1−>(a1 (−1+pi^2) sgi shi^2 (h1 sg1+p1 sh1 (−g1+y1))+ai (−1+
    p1^2) sg1 sh1^2 (hi sgi−gi pi shi−sgi xi+pi shi yi))/(a1 (−1+pi^2)
    sg1 sgi shi^2)}}

In[23]:= (* ——————— SOLVE FOR VARIOUS CASES ——————— *)
(*CONDITIONS: INFEASIBLE *)
lam000=Simplify[Solve[{lamx1e==0,lamxie==0,lamy1e==0,lamyie==0},{x1,xi,
    y1,yi}]]
x1y1ans=Transpose[{x1,y1}/.lam000]; xiyians=Transpose[{xi,yi}/.lam000];
Part[K[x1y1ans,xiyians],1,1 ](*GIVES RATE FUNCTION*)
x1ans=Part[x1/.lam000,1]; xians = Part[xi/.lam000,1];
y1ans = Part[y1/.lam000,1]; yians = Part[yi/.lam000,1];
GamCond = Reduce[{xians<=x1ans,y1ans<=gam,yians<=gam,h1<hi,g1<gam,gi<
    gam,0<a1<1,0<ai<1,a1+ai<1,sh1>0,shi>0,sg1>0,sgi>0,−1<p1<1,−1<pi<1},{
    h1,g1,gi} ]
SbSwCond = Reduce[{xians<=x1ans,y1ans<=gam,yians<=gam,g1<gam,gi>gam,0<
    a1<1,0<ai<1,a1+ai<1,sh1>0,shi>0,sg1>0,sgi>0,−1<p1<1,−1<pi<1},{h1,g1,
    gi} ]
```

```
Out[23]= {{x1->h1, xi->hi, y1->g1, yi->gi}}
Out[25]= 0
Out[28]= False
Out[29]= False

In[30]:= (* ——— S_b(\rho) ——— *)
Clear[x1y1ans, xiyians, x1ans, y1ans, xians, yians, x1, xi, y1, yi, h1, g1, hi, gi]
lam001=Simplify[Solve[{lamx1e==0,lamxie==0,lamy1e==0,yi==gam},{x1, xi, y1
    , yi}]]
x1y1ans=Transpose[{x1, y1}/.lam001]; xiyians=Transpose[{xi, yi}/.lam001];
Part[Simplify[K[x1y1ans, xiyians]], 1, 1]
x1ans=Part[x1/.lam001, 1]; xians = Part[xi/.lam001, 1];
y1ans = Part[y1/.lam001, 1]; yians = Part[yi/.lam001, 1];
lambdayi = Simplify[lamyieFn[xians, yians]]

Out[31]= {{x1->h1, xi->(hi sgi+(gam-gi) pi shi)/sgi, y1->g1, yi->gam}}
Out[33]= (ai (gam-gi)^2)/(2 sgi^2)
Out[36]= (ai (-gam+gi))/sgi^2

In[37]:= (*CONDITIONS: INFEASIBLE *)
Clear[x1y1ans, xiyians, x1ans, y1ans, xians, yians, x1, xi, y1, yi, h1, g1, hi, gi,
    lambdayi]
lam010=Simplify[Solve[{lamx1e==0,lamxie==0,y1==gam, lamyie==0},{x1, xi, y1
    , yi}]]
x1y1ans=Transpose[{x1, y1}/.lam010]; xiyians=Transpose[{xi, yi}/.lam010];
Part[Simplify[K[x1y1ans, xiyians]], 1, 1]
x1ans=Part[x1/.lam010, 1]; xians = Part[xi/.lam010, 1];
y1ans = Part[y1/.lam010, 1]; yians = Part[yi/.lam010, 1];
lambday1 = Simplify[lamy1eFn[x1ans, y1ans]];
GamCond = Simplify[Reduce[{xians<=x1ans, lambday1>0, yians<=gam, h1<hi, g1<
    gam, gi<gam,0<a1<1,0<ai<1,a1+ai<1,sh1>0,shi>0,sg1>0,sgi>0,-1<p1<1,-1<
    pi<1},{h1, g1, gi} ]]
SbSwCond = Simplify[Reduce[{xians<=x1ans, lambday1>0, yians<=gam, g1<gam,
    gi>gam,0<a1<1,0<ai<1,a1+ai<1,sh1>0,shi>0,sg1>0,sgi>0,-1<p1<1,-1<pi
```

```
    <1},{h1,g1,gi}  ]]

Out[38]= {{x1->(h1 sg1+(-g1+gam) p1 sh1)/sg1,xi->hi,y1->gam,yi->gi}}
Out[40]= (a1 (g1-gam)^2)/(2 sg1^2)
Out[44]= False
Out[45]= False

In[46]:= (*CONDITIONS: INFEASIBLE *)
Clear[x1y1ans,xiyians,x1ans,y1ans,xians,yians,x1,xi,y1,yi,h1,g1,hi,gi,
    lambday1]
lam011=Simplify[Solve[{lamx1e==0,lamxie==0,y1==gam,yi==gam},{x1,xi,y1,
    yi}]]
x1y1ans=Transpose[{x1,y1}/.lam011]; xiyians=Transpose[{xi,yi}/.lam011];
Part[Simplify[K[x1y1ans,xiyians]],1,1]
x1ans=Part[x1/.lam011,1]; xians = Part[xi/.lam011,1];
y1ans = Part[y1/.lam011,1]; yians = Part[yi/.lam011,1];
lambday1 = Simplify[lamy1eFn[x1ans,y1ans]]; lambdayi = Simplify[
    lamyieFn[xians,yians]];
GamCond = Simplify[Reduce[{xians<=x1ans,lambday1>0,lambdayi>0, h1<hi,g1
    <gam,gi<gam,0<a1<1,0<ai<1,a1+ai<1,sh1>0,shi>0,sg1>0,sgi>0,-1<p1
    <1,-1<pi<1},{h1,g1,gi}  ]]
SbSwCond = Simplify[Reduce[{xians<=x1ans,lambday1>0,lambdayi>0,g1<gam,
    gi>gam,0<a1<1,0<ai<1,a1+ai<1,sh1>0,shi>0,sg1>0,sgi>0,-1<p1<1,-1<pi
    <1},{h1,g1,gi}  ]]

Out[47]= {{x1->(h1 sg1+(-g1+gam) p1 sh1)/sg1,xi->(hi sgi+(gam-gi) pi
    shi)/sgi,y1->gam,yi->gam}}
Out[49]= (ai (gam-gi)^2 sg1^2+a1 (g1-gam)^2 sgi^2)/(2 sg1^2 sgi^2)
Out[53]= False
Out[54]= False

In[55]:= (* ------ Gamma_\ell ------ *)
Clear[x1y1ans,xiyians,x1ans,y1ans,xians,yians,x1,xi,y1,yi,h1,g1,hi,gi,
    lambdayi,lambday1]
```

```
lam100=Simplify[Solve[{x1e==xi,x1==xi,lamy1e==0,lamyie==0},{x1,xi,y1,yi
    }]]
x1y1ans=Transpose[{x1,y1}/.lam100]; xiyians=Transpose[{xi,yi}/.lam100];
Part[Simplify[K[x1y1ans,xiyians]],1,1]
x1ans=Part[x1/.lam100,1]; xians = Part[xi/.lam100,1];
y1ans = Part[y1/.lam100,1]; yians = Part[yi/.lam100,1] ;
lambdax = Simplify[lamx1eFn[x1ans,y1ans]]

Out[56]= {{x1->(ai hi sh1^2+a1 h1 shi^2)/(ai sh1^2+a1 shi^2),xi->(ai hi
     sh1^2+a1 h1 shi^2)/(ai sh1^2+a1 shi^2),y1->(ai sh1 (-h1 p1 sg1+hi
    p1 sg1+g1 sh1)+a1 g1 shi^2)/(ai sh1^2+a1 shi^2),yi->(ai gi sh1^2+a1
    shi (h1 pi sgi-hi pi sgi+gi shi))/(ai sh1^2+a1 shi^2)}}
Out[58]= (a1 ai (h1-hi)^2)/(2 (ai sh1^2+a1 shi^2))
Out[61]= (a1 ai (-h1+hi))/(ai sh1^2+a1 shi^2)

In[62]:= (* ------ Gamma_u ------ *)
Clear[x1y1ans,xiyians,x1ans,y1ans,xians,yians,x1,xi,y1,yi,h1,g1,hi,gi,
    lambdax,lambdayi]
lam110=Simplify[Solve[{x1e==xi,x1==xi,y1==gam,lamyie==0},{x1,xi,y1,yi
    }]]
x1y1ans=Transpose[{x1,y1}/.lam110]; xiyians=Transpose[{xi,yi}/.lam110];
Collect[Part[Simplify[K[x1y1ans,xiyians]],1,1],g1-gam]
x1ans=Part[x1/.lam110,1]; xians = Part[xi/.lam110,1];
y1ans = Part[y1/.lam110,1]; yians = Part[yi/.lam110,1];
lambdax = Simplify[lamx1eFn[x1ans,y1ans]]
lambday1 = Simplify[lamy1eFn[x1ans,y1ans]]

Out[63]= {{x1->(ai hi (-1+p1^2) sg1 sh1^2-a1 (h1 sg1+(-g1+gam) p1 sh1)
    shi^2)/(sg1 (ai (-1+p1^2) sh1^2-a1 shi^2)),xi->(ai hi (-1+p1^2) sg1
    sh1^2-a1 (h1 sg1+(-g1+gam) p1 sh1) shi^2)/(sg1 (ai (-1+p1^2) sh1^2-
    a1 shi^2)),y1->gam,yi->(ai gi (-1+p1^2) sg1 sh1^2-a1 shi (h1 pi sg1
    sgi-hi pi sg1 sgi-g1 p1 pi sgi sh1+gam p1 pi sgi sh1+gi sg1 shi))/(
    sg1 (ai (-1+p1^2) sh1^2-a1 shi^2))}}
Out[65]= -((a1 (ai h1^2 sg1^2-2 ai h1 hi sg1^2+ai hi^2 sg1^2))/(2 sg1^2
     (ai (-1+p1^2) sh1^2-a1 shi^2)))-(a1 (g1-gam) (-2 ai h1 p1 sg1 sh1+2
```

```
    ai hi p1 sg1 sh1))/(2 sg1^2 (ai (-1+p1^2) sh1^2-a1 shi^2))-(a1 (g1-
   gam)^2 (ai sh1^2+a1 shi^2))/(2 sg1^2 (ai (-1+p1^2) sh1^2-a1 shi^2))
Out[68]= (a1 ai (h1 sg1-hi sg1+(-g1+gam) p1 sh1))/(sg1 (ai (-1+p1^2)
   sh1^2-a1 shi^2))
Out[69]= -((a1 (ai sh1 (-h1 p1 sg1+hi p1 sg1+g1 sh1-gam sh1)+a1 (g1-gam
   ) shi^2))/(sg1^2 (ai (-1+p1^2) sh1^2-a1 shi^2)))

In[70]:= (* ------ S_w,\ell ------*)
Clear[x1y1ans,xiyians,x1ans,y1ans,xians,yians,x1,xi,y1,yi,h1,g1,hi,gi,
   lambdax]
lam101=Simplify[Solve[{x1e==xi,x1==xi,lamy1e==0,yi==gam},{x1,xi,y1,yi
   }]]
x1y1ans=Transpose[{x1,y1}/.lam101]; xiyians=Transpose[{xi,yi}/.lam101];
Collect[Part[Simplify[K[x1y1ans,xiyians]],1,1],gam-gi]
x1ans=Part[x1/.lam101,1]; xians = Part[xi/.lam101,1];
y1ans = Part[y1/.lam101,1] ;yians = Part[yi/.lam101,1];
lambdax = Simplify[lamx1eFn[x1ans,y1ans]]
lambdayi = Simplify[lamyieFn[xians,yians]]

Out[71]= {{x1->(-a1 h1 (-1+pi^2) sgi shi^2+ai sh1^2 (hi sgi+(gam-gi) pi
    shi))/(sgi (ai sh1^2-a1 (-1+pi^2) shi^2)),xi->(-a1 h1 (-1+pi^2) sgi
    shi^2+ai sh1^2 (hi sgi+(gam-gi) pi shi))/(sgi (ai sh1^2-a1 (-1+pi
   ^2) shi^2)),y1->(-a1 g1 (-1+pi^2) sgi shi^2+ai sh1 (-h1 p1 sg1 sgi+
   hi p1 sg1 sgi+g1 sgi sh1+gam p1 pi sg1 shi-gi p1 pi sg1 shi))/(sgi (
   ai sh1^2-a1 (-1+pi^2) shi^2)),yi->gam}}
Out[73]= (ai (a1 h1^2 sgi^2-2 a1 h1 hi sgi^2+a1 hi^2 sgi^2))/(2 sgi^2 (
   ai sh1^2-a1 (-1+pi^2) shi^2))+(ai (gam-gi) (-2 a1 h1 pi sgi shi+2 a1
    hi pi sgi shi))/(2 sgi^2 (ai sh1^2-a1 (-1+pi^2) shi^2))+(ai (gam-gi
   )^2 (ai sh1^2+a1 shi^2))/(2 sgi^2 (ai sh1^2-a1 (-1+pi^2) shi^2))
Out[76]= -((a1 ai (h1 sgi-hi sgi+(-gam+gi) pi shi))/(sgi (ai sh1^2-a1
   (-1+pi^2) shi^2)))
Out[77]= -((ai (ai (gam-gi) sh1^2+a1 shi (-h1 pi sgi+hi pi sgi+gam shi-
   gi shi)))/(sgi^2 (ai sh1^2-a1 (-1+pi^2) shi^2)))

In[78]:= (* ------ S_wu ------ *)
```

```
Clear[x1y1ans, xiyians, x1ans, y1ans, xians, yians, x1, xi, y1, yi, h1, g1, hi, gi,
    lambdax, lambday1]
lam111=Simplify[Solve[{x1e==xi, x1==xi, y1==gam, yi==gam}, {x1, xi, y1, yi}]]
x1y1ans=Transpose[{x1, y1}/.lam111]; xiyians=Transpose[{xi, yi}/.lam111];
Collect[Part[Simplify[K[x1y1ans, xiyians]], 1, 1], gam-gi];
KSim = Collect[%, g1-gam];
Kmod = ((hi-h1)^2 -(2*p1*sh1*(hi-h1)*(gam-g1)/sg1) -(2*pi*shi*(hi-h1)*(
    gi - gam)/sgi)+a1*((gam-g1)^2/sg1^2)*(sh1^2/a1 + (1-pi^2)*shi^2/ai)
    + 2*p1*pi*sh1*shi*((gam-g1)/sg1)*((gi-gam)/sgi)+ai*((gi-gam)^2/sgi
    ^2)*((1-p1^2)*sh1^2/a1+shi^2/ai) )/(2*((1-p1^2)*sh1^2/a1+(1-pi^2)*
    shi^2/ai))
If[Simplify[KSim -Kmod]==0, True, False]
x1ans=Part[x1/.lam111, 1]; xians = Part[xi/.lam111, 1];
y1ans = Part[y1/.lam111, 1]; yians = Part[yi/.lam111, 1];
lambdax = Simplify[lamx1eFn[x1ans, y1ans]]
lambday1 = Simplify[lamy1eFn[x1ans, y1ans]]
lambdayi = Simplify[lamyieFn[xians, yians]]

Out[79]= {{x1->(a1 (-1+pi^2) sgi (h1 sg1+(-g1+gam) p1 sh1) shi^2+ai
    (-1+p1^2) sg1 sh1^2 (hi sgi+(gam-gi) pi shi))/(sg1 sgi (ai (-1+p1^2)
     sh1^2+a1 (-1+pi^2) shi^2)), xi->(a1 (-1+pi^2) sgi (h1 sg1+(-g1+gam)
    p1 sh1) shi^2+ai (-1+p1^2) sg1 sh1^2 (hi sgi+(gam-gi) pi shi))/(sg1
    sgi (ai (-1+p1^2) sh1^2+a1 (-1+pi^2) shi^2)), y1->gam, yi->gam}}
Out[83]= ((-h1+hi)^2-(2 (-g1+gam) (-h1+hi) p1 sh1)/sg1-(2 (-gam+gi) (-
    h1+hi) pi shi)/sgi+(2 (-g1+gam) (-gam+gi) p1 pi sh1 shi)/(sg1 sgi)+(
    ai (-gam+gi)^2 (((1-p1^2) sh1^2)/a1+shi^2/ai))/sgi^2+(a1 (-g1+gam)^2
     (sh1^2/a1+((1-pi^2) shi^2)/ai))/sg1^2)/(2 (((1-p1^2) sh1^2)/a1+((1-
    pi^2) shi^2)/ai))
Out[84]= True
Out[87]= (a1 ai (h1 sg1 sgi-hi sg1 sgi-g1 p1 sgi sh1+gam p1 sgi sh1-gam
     pi sg1 shi+gi pi sg1 shi))/(sg1 sgi (ai (-1+p1^2) sh1^2+a1 (-1+pi
    ^2) shi^2))
Out[88]= (a1 (a1 (g1-gam) (-1+pi^2) sgi shi^2+ai sh1 (h1 p1 sg1 sgi-hi
    p1 sg1 sgi-g1 sgi sh1+gam sgi sh1-gam p1 pi sg1 shi+gi p1 pi sg1 shi
    )))/(sg1^2 sgi (ai (-1+p1^2) sh1^2+a1 (-1+pi^2) shi^2))
```

```
Out[89]= -((ai (ai (gam-gi) (-1+p1^2) sg1 sh1^2+a1 shi (h1 pi sg1 sgi-
    hi pi sg1 sgi-g1 p1 pi sgi sh1+gam p1 pi sgi sh1-gam sg1 shi+gi sg1
    shi)))/(sg1 sgi^2 (ai (-1+p1^2) sh1^2+a1 (-1+pi^2) shi^2)))
```

# Appendix D

# Preliminary Work on the Implementability of Optimal Sampling Laws

## D.1 Taylor Series Expansion of the Rate Function About its Mean

**Proposition D.1.** *Let $X$ be a random variable with mean $\mu$, variance $\sigma^2$, and analytic rate function $I(x)$ having the form $I(x) = \sup_\theta\{\theta x - \log E[e^{\theta X}]\}$. Then the Taylor series expansion of $I(x)$ about $\mu$ is*

$$I(x) = \sum_{n=1}^{\infty} \frac{(x-\mu)^n}{\sigma^2} \left[ \sum_{\Theta_n \setminus \{b_n=1\}} \frac{\left(\mu E\left[X^{\sum_{i=1}^{n} b_i}\right] - E\left[X^{1+\sum_{i=1}^{n} b_i}\right]\right)}{b_1! b_2! \ldots b_{n-1}!} \prod_{i=1}^{n-1} \left(\frac{\theta^{(i)}(\mu)}{i!}\right)^{b_i} \right.$$
$$\left. + \sum_{\Theta_{n-1}} \frac{E\left[X^{\sum_{i=1}^{n-1} b_i}\right]}{b_1! b_2! \ldots b_{n-1}!} \prod_{i=1}^{n-1} \left(\frac{\theta^{(i)}(\mu)}{i!}\right)^{b_i} \right]$$

*where for any $m$, $\Theta_m$ denotes all different solutions in nonnegative integers $b_1, b_2, \ldots, b_m$ of the equation $b_1 + 2b_2 + \ldots + mb_m = m$, and $\Theta_m \setminus \{b_m = 1\}$ denotes the set of solutions without the solution in which $b_m = 1$.*

*Proof.* The general form of the Taylor series expansion of $I(x)$ about $\mu$ is

$$I(x) = \sum_{n=0}^{\infty} I^{(n)}(\mu) \frac{(x-\mu)^n}{n!}. \tag{D.1}$$

Since $I(\mu) = I'(\mu) = 0$, let us consider terms beginning at $n = 2$ in the equation (D.1).

Let $M(\theta) = E[e^{\theta X}]$ denote the moment generating function of $X$. Then

$$I(x) = \theta(x)x - \log M(\theta(x)),$$

where $\theta(x)$ satisfies $x = M'(\theta)/M(\theta)$. Taking the derivative of $I(x)$ yields

$$I'(x) = \theta(x) + x\theta'(x) - \frac{M'(\theta(x))}{M(\theta(x))}\theta'(x) = \theta(x).$$

Then it follows that $I^{(n)}(x) = \theta^{(n-1)}(x)$. Noting that $I(\mu) = I'(\mu) = \theta(\mu) = 0$, equation (D.1) may be rewritten as,

$$I(x) = \sum_{n=1}^{\infty} \theta^{(n-1)}(\mu) \frac{(x-\mu)^n}{n!} \tag{D.2}$$

To obtain a general expression for $\theta^{(n)}(x)$, let us take the $n$th derivative on both sides of the equation $xM(\theta(x)) = M'(\theta(x))$. That is, we wish to solve for $\theta^{(n)}(x)$ using the equation

$$(xM(\theta(x)))^{(n)} = (M'(\theta(x)))^{(n)} \tag{D.3}$$

By the general Leibniz rule, and since $x^{(0)} = x, x^{(1)} = 1, x^{(2)} = 0, \ldots, x^{(n)} = 0$, we have

$$(xM(\theta(x)))^{(n)} = \sum_{k=0}^{n} \binom{n}{k} x^{(k)} M(\theta(x))^{(n-k)} = xM(\theta(x))^{(n)} + nM(\theta(x))^{(n-1)}.$$

Now Faà di Bruno's Formula (see, e.g., Johnson, 2002) may be used to obtain expressions for $M(\theta(x))^{(n)}$, $M(\theta(x))^{(n-1)}$, and $(M'(\theta(x)))^{(n)}$. Substituting these expressions into equation (D.3) yields

$$\begin{aligned} x&\left[\sum_{\Theta_n} \frac{n!}{b_1!b_2!\ldots b_n!} M^{(\sum_{i=1}^n b_i)}(\theta(x)) \prod_{i=1}^{n} \left(\frac{\theta^{(i)}(x)}{i!}\right)^{b_i}\right] \\ &\quad + n\left[\sum_{\Theta_{n-1}} \frac{(n-1)!}{b_1!b_2!\ldots b_{n-1}!} M^{(\sum_{i=1}^{n-1} b_i)}(\theta(x)) \prod_{i=1}^{n-1} \left(\frac{\theta^{(i)}(x)}{i!}\right)^{b_i}\right] \\ &= \sum_{\Theta_n} \frac{n!}{b_1!b_2!\ldots b_n!} M^{(1+\sum_{i=1}^n b_i)}(\theta(x)) \prod_{i=1}^{n} \left(\frac{\theta^{(i)}(x)}{i!}\right)^{b_i}, \end{aligned}$$

where for any $m$, $\Theta_m$ denotes all different solutions in nonnegative integers $b_1, b_2, \ldots, b_m$ of the equation $b_1 + 2b_2 + \ldots + mb_m = m$. Noting that the only solutions to the equations in $\Theta_n$ which yield the terms $\theta^{(n)}(x)$ are $b_1 = 0, b_2 = 0, \ldots, b_{n-1} = 0, b_n = 1$, and that except in this solution $b_n = 0$, we may simplify the above equation to

$$\begin{aligned} x&\left[M'(\theta(x))\theta^{(n)}(x)\right] + x\left[\sum_{\Theta_n \setminus \{b_n=1\}} \frac{n!}{b_1!b_2!\ldots b_n!} M^{(\sum_{i=1}^n b_i)}(\theta(x)) \prod_{i=1}^{n-1} \left(\frac{\theta^{(i)}(x)}{i!}\right)^{b_i}\right] \\ &\quad + n\left[\sum_{\Theta_{n-1}} \frac{(n-1)!}{b_1!b_2!\ldots b_{n-1}!} M^{(\sum_{i=1}^{n-1} b_i)}(\theta(x)) \prod_{i=1}^{n-1} \left(\frac{\theta^{(i)}(x)}{i!}\right)^{b_i}\right] \\ &= M''(\theta(x))\theta^{(n)}(x) + \sum_{\Theta_n \setminus \{b_n=1\}} \frac{n!}{b_1!b_2!\ldots b_n!} M^{(1+\sum_{i=1}^n b_i)}(\theta(x)) \prod_{i=1}^{n-1} \left(\frac{\theta^{(i)}(x)}{i!}\right)^{b_i} \end{aligned}$$

Combining terms yields the following general expression.

$$\theta^{(n)}(x) = \Bigg[\sum_{\Theta_n \setminus \{b_n=1\}} \frac{n!}{b_1!b_2!\ldots b_{n-1}!}\left(xM^{(\sum_{i=1}^{n} b_i)}(\theta(x)) - M^{(1+\sum_{i=1}^{n} b_i)}(\theta(x))\right)\prod_{i=1}^{n-1}\left(\frac{\theta^{(i)}(x)}{i!}\right)^{b_i}$$
$$+ \sum_{\Theta_{n-1}} \frac{n!}{b_1!b_2!\ldots b_{n-1}!} M^{(\sum_{i=1}^{n-1} b_i)}(\theta(x)) \prod_{i=1}^{n-1}\left(\frac{\theta^{(i)}(x)}{i!}\right)^{b_i}\Bigg] \times \left[M''(\theta(x)) - xM'(\theta(x))\right]^{-1}.$$

Substituting in $x = \mu$ and noting that $\theta(\mu) = 0$, we find

$$\theta^{(n)}(\mu) = \frac{n!}{\sigma^2}\Bigg[\sum_{\Theta_n \setminus \{b_n=1\}} \frac{\left(\mu E\left[X^{\sum_{i=1}^{n} b_i}\right] - E\left[X^{1+\sum_{i=1}^{n} b_i}\right]\right)}{b_1!b_2!\ldots b_{n-1}!}\prod_{i=1}^{n-1}\left(\frac{\theta^{(i)}(\mu)}{i!}\right)^{b_i}$$
$$+ \sum_{\Theta_{n-1}} \frac{E\left[X^{\sum_{i=1}^{n-1} b_i}\right]}{b_1!b_2!\ldots b_{n-1}!}\prod_{i=1}^{n-1}\left(\frac{\theta^{(i)}(\mu)}{i!}\right)^{b_i}\Bigg],$$

and the result follows. □

**Corollary D.2.** *The first three nonzero terms of the Taylor series expansion of $I(x)$ given in proposition D.1 are*

$$I(x) = \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 - \frac{1}{3!}\, E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]\left(\frac{x-\mu}{\sigma}\right)^3$$
$$+ \frac{1}{4!}\left[3E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]^2 - E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] + 3\right]\left(\frac{x-\mu}{\sigma}\right)^4.$$

*Proof.* Taking derivatives on both sides of the equation $xM(\theta(x)) = M'(\theta(x))$, we find

$$\theta'(x) = \frac{M(\theta(x))}{M''(\theta(x)) - xM'(\theta(x))}$$
$$\theta''(x) = \frac{2M'(\theta(x))\theta'(x) + \theta'(x)^2[xM''(\theta(x)) - M'''(\theta(x))]}{M''(\theta(x)) - xM'(\theta(x))}$$
$$\theta'''(x) = \Big[3M'(\theta(x))\theta''(x) + 3M''(\theta(x))\theta'(x)^2 + 3\theta'(x)\theta''(x)[xM''(\theta(x)) - M'''(\theta(x))] +$$
$$\theta'(x)^3[xM'''(\theta(x)) - M^{(4)}(\theta(x))]\Big] \times [M''(\theta(x)) - xM'(\theta(x))]^{-1}$$

which, after substitution and some algebra in the first two equations above, implies

$$I''(\mu) = \theta'(\mu) = \frac{1}{\sigma^2} \quad \text{and} \quad I'''(\mu) = \theta''(\mu) = -\left(\frac{1}{\sigma^2}\right)^3 \mathrm{E}[(X-\mu)^3].$$

However, the algebra is cumbersome for finding $\theta'''(\mu)$ and any larger derivatives since we are using an unsimplified version of proposition D.1. Thus let us use proposition D.1. We find,

$$\begin{aligned}
\theta''(\mu) &= \frac{2!}{\sigma^2}\left[\frac{\mu\mathrm{E}[X^2]-\mathrm{E}[X^3]}{2!}\left(\frac{1}{\sigma^2}\right)^2 + \mathrm{E}[X]\left(\frac{1}{\sigma^2}\right)\right] = \left(\frac{1}{\sigma^2}\right)^3\left[\mu\mathrm{E}[X^2]-\mathrm{E}[X^3]+2\mu\sigma^2\right] \\
&= -\left(\frac{1}{\sigma^2}\right)^3\left[\mathrm{E}[X^3]-3\mu\mathrm{E}[X^2]+2\mu^3\right] = -\left(\frac{1}{\sigma^2}\right)^3\mathrm{E}[(X-\mu)^3].
\end{aligned}$$

Likewise, the result follows from noting that,

$$\begin{aligned}
\theta'''(\mu) &= \frac{3!}{\sigma^2}\Bigg[\frac{\mu\mathrm{E}[X^3]-\mathrm{E}[X^4]}{3!}\left(\frac{1}{\sigma^2}\right)^3 + \left(\mu\mathrm{E}[X^2]-\mathrm{E}[X^3]\right)\frac{1}{\sigma^2}\left(-\left(\frac{1}{\sigma^2}\right)^3\frac{\mathrm{E}[(X-\mu)^3]}{2!}\right) \\
&\qquad -\mu\left(\frac{1}{\sigma^2}\right)^3\frac{\mathrm{E}[(X-\mu)^3]}{2!} + \frac{E[X^2]}{2!}\left(\frac{1}{\sigma^2}\right)^2\Bigg] \\
&= \left(\frac{1}{\sigma^2}\right)^2\Bigg[\left(\mu\mathrm{E}[X^3]-\mathrm{E}[X^4]\right)\left(\frac{1}{\sigma^2}\right)^2 + 3\left(\mathrm{E}[X^3]-\mu\mathrm{E}[X^2]\right)\left(\frac{1}{\sigma^2}\right)^3\mathrm{E}[(X-\mu)^3] \\
&\qquad -3\mu\left(\frac{1}{\sigma^2}\right)^2\mathrm{E}[(X-\mu)^3] + 3\mathrm{E}[X^2]\left(\frac{1}{\sigma^2}\right)\Bigg] \\
&= \left(\frac{1}{\sigma^2}\right)^5\Big[\sigma^2(\mu\mathrm{E}[X^3]-\mathrm{E}[X^4]) + 3\left(\mathrm{E}[X^3]-\mu\mathrm{E}[X^2]\right)\mathrm{E}[(X-\mu)^3] \\
&\qquad -3\mu\sigma^2\mathrm{E}[(X-\mu)^3] + 3(\sigma^2)^2\mathrm{E}[X^2]\Big] \\
&= \left(\frac{1}{\sigma^2}\right)^5\left[\sigma^2(\mu\mathrm{E}[X^3]-\mathrm{E}[X^4]) + 3\left(\mathrm{E}[X^3]-\mu\mathrm{E}[X^2]-\mu\sigma^2\right)\mathrm{E}[(X-\mu)^3] + 3(\sigma^2)^2\mathrm{E}[X^2]\right] \\
&= \left(\frac{1}{\sigma^2}\right)^5\left[\sigma^2(\mu\mathrm{E}[X^3]-\mathrm{E}[X^4]) + 3\left(\mathrm{E}[(X-\mu)^3]+\mu\sigma^2\right)\mathrm{E}[(X-\mu)^3] + 3(\sigma^2)^2\mathrm{E}[X^2]\right] \\
&= \left(\frac{1}{\sigma^2}\right)^5\left[\sigma^2\left(\mu\mathrm{E}[X^3]-\mathrm{E}[X^4]+3\mu\mathrm{E}[(X-\mu)^3]+3\sigma^2\mathrm{E}[X^2]\right) + 3\mathrm{E}[(X-\mu)^3]^2\right] \\
&= \left(\frac{1}{\sigma^2}\right)^5\left[\sigma^2(3(\sigma^2)^2-\mathrm{E}[(X-\mu)^4]) + 3\mathrm{E}[(X-\mu)^3]^2\right] \\
&= \left(\frac{1}{\sigma^2}\right)^2\left[3\mathrm{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]^2 - \mathrm{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] + 3\right].
\end{aligned}$$

□