Quantifying Stochastic Model Errors via Robust Optimization

IPAM Workshop on Uncertainty Quantification for Multiscale Stochastic Systems and Applications

Jan 19, 2016

Henry Lam

Industrial & Operations Engineering

University of Michigan

Acknowledgement: National Science Foundation under grants CMMI-1436247 and CMMI-1400391

Overview

Focus:

Discrete-event stochastic systems in operations research

Motivation:

Model risk / model error/ input uncertainty arises when performance analysis is based on misspecified / uncertain input model assumptions

Goal: Performance analysis that is robust against model error

Scope:

- Input model: stochastic
- Uncertainty: nonparametric

Basic Question

Single-server queue:

- i.i.d. $Exp(0.8)$ interarrival times
- i.i.d. $Exp(1)$ service times

Target performance measure: probability of long waiting time incurred by the 50-th arriving customer, i.e.

$$P(W_{50} > 2)$$

Basic Question

Single-server queue:

- i.i.d. $Exp(0.8)$ interarrival times
- i.i.d. $Exp(1)$ service times

Target performance measure: probability of long waiting time incurred by the 50-th arriving customer, i.e.

$$P(W_{50} > 2)$$

With insufficient or unreliable historical data,

- What if service time distribution is not $Exp(1)$?

Basic Question

Single-server queue:

- i.i.d. $Exp(0.8)$ interarrival times
- i.i.d. $Exp(1)$ service times

Target performance measure: probability of long waiting time incurred by the 50-th arriving customer, i.e.

$$P(W_{50} > 2)$$

With insufficient or unreliable historical data,

- What if service time distribution is not $Exp(1)$?

- What is a reliable estimate of the performance measure, without committing to a (non-)parametric model for the service time distribution?

Question 2: Uncertain Dependency

Single-server queue:

- i.i.d. $Exp(0.8)$ interarrival times
- i.i.d. $Exp(1)$ service times

Target performance measure: probability of long waiting time incurred by the 50-th arriving customer, i.e.

$$P(W_{50} > 2)$$

Question 2: Uncertain Dependency

Single-server queue:

- i.i.d. $Exp(0.8)$ interarrival times
- i.i.d. $Exp(1)$ service times

Target performance measure: probability of long waiting time incurred by the 50-th arriving customer, i.e.

$$P(W_{50} > 2)$$

- What if the interarrival time sequence is not i.i.d.?

Question 2: Uncertain Dependency

Single-server queue:

- i.i.d. $Exp(0.8)$ interarrival times
- i.i.d. $Exp(1)$ service times

Target performance measure: probability of long waiting time incurred by the 50-th arriving customer, i.e.

$$P(W_{50} > 2)$$

- What if the interarrival time sequence is not i.i.d.?
- An estimate without restricting to any parametric class of time series models for the interarrival times?

Question 3: Input Reconstruction

Single-server queue:

- i.i.d. $Exp(0.8)$ interarrival times
- i.i.d. unknown service times

- What information can we extract on the service time distribution if we only have data on the waiting times?

Robust Stochastic Modeling

Model uncertainty arises due to modeling limitations, presence of risk, lack of direct data etc.

Robust perspective:

Robust Stochastic Modeling

Model uncertainty arises due to modeling limitations, presence of risk, lack of direct data etc.

Robust perspective:

- **Element 1:** Represent partial, nonparametric information on the input model via constraints $\Rightarrow$ uncertainty set (e.g. Ben-Tal et al. ’09, Bertsimas et al. ‘11)

Robust Stochastic Modeling

Model uncertainty arises due to modeling limitations, presence of risk, lack of direct data etc.

Robust perspective:

- **Element 1:** Represent partial, nonparametric information on the input model via constraints $\Rightarrow$ uncertainty set (e.g. Ben-Tal et al. ’09, Bertsimas et al. ‘11)
- **Element 2:** Solve optimization problem that computes the worst-case performance measure subject to the uncertainty

Basic Example

- Baseline input model: $X_t \sim P_0$
- $\boldsymbol{X} = (X_1, X_2, \dots, X_T)$
- Performance measure: $E[h(\boldsymbol{X})]$

E.g. Single-server queue: $\boldsymbol{X}$ = sequence of service times,
$P_0 = Exp(1),\ h(\boldsymbol{X}) = I(waiting\ time\ of\ 20th\ customer > 2)$

compute $E_0[h(X_1, \dots, X_T)]$
under the condition

$$X_t \overset{\text{i.i.d.}}{\sim} P_0$$

Basic Example

- Baseline input model: $X_t \sim P_0$
- $\boldsymbol{X} = (X_1, X_2, \dots, X_T)$
- Performance measure: $E[h(\boldsymbol{X})]$

E.g. Single-server queue: $\boldsymbol{X}$ = sequence of service times,

$P_0 = Exp(1),\ h(\boldsymbol{X}) = I(waiting\ time\ of\ 20th\ customer > 2)$

What if the distribution for X_t is potentially misspecified?

compute $E_0[h(X_1, \dots, X_T)]$
under the condition

$$X_t \overset{\text{i.i.d.}}{\sim} P_0$$

Basic Example

- Baseline input model: $X_t \sim P_0$
- $\boldsymbol{X} = (X_1, X_2, \dots, X_T)$
- Performance measure: $E[h(\boldsymbol{X})]$

E.g. Single-server queue: $\boldsymbol{X}$ = sequence of service times,
$P_0 = Exp(1),\ h(\boldsymbol{X}) = I(waiting\ time\ of\ 20th\ customer > 2)$

What if the distribution for X_t is potentially misspecified?

maximize $E_f[h(X_1, \dots, X_T)]$
subject to

$$X_t \overset{\text{i.i.d.}}{\sim} P_f$$
$$D\left(P_f | P_0\right) \le \eta$$

Basic Example

- Baseline input model: $X_t \sim P_0$
- $\boldsymbol{X} = (X_1, X_2, \ldots, X_T)$
- Performance measure: $E[h(\boldsymbol{X})]$

The true service time distribution P_f is in a neighborhood of the baseline model P_0, as measured by a nonparametric distance D of η units

E.g. Single-server queue: $\boldsymbol{X}$ = sequence of service times,

$P_0 = Exp(1),\ h(\boldsymbol{X}) = I(waiting\ time\ of\ 20th\ customer > 2)$

What if the distribution for X_t is potentially misspecified?

maximize $E_f[h(X_1, \ldots, X_T)]$

subject to

$$X_t \overset{\text{i.i.d.}}{\sim} P_f$$

$$D(P_f | P_0) \le \eta$$

Basic Example

- Baseline input model: $X_t \sim P_0$
- $\boldsymbol{X} = (X_1, X_2, \dots, X_T)$
- Performance measure: $E[h(\boldsymbol{X})]$

The true service time distribution P_f is in a neighborhood of the baseline model P_0, as measured by a nonparametric distance D of η units

E.g. Single-server queue: $\boldsymbol{X}$ = sequence of service times,
$P_0 = Exp(1),\ h(\boldsymbol{X}) = I(waiting\ time\ of\ 20th\ customer > 2)$

What if the distribution for X_t is potentially misspecified?

maximize $E_f[h(X_1, \dots, X_T)]$
subject to

$$X_t \overset{\text{i.i.d.}}{\sim} P_f$$
$$D\left(P_f | P_0\right) \leq \eta$$

Decision variable: P_f

Basic Example

- Baseline input model: $X_t \sim P_0$
- $\boldsymbol{X} = (X_1, X_2, \ldots, X_T)$
- Performance measure: $E[h(\boldsymbol{X})]$

The true service time distribution P_f is in a neighborhood of the baseline model P_0, as measured by a nonparametric distance D of η units

E.g. Single-server queue: $\boldsymbol{X}$ = sequence of service times,
$P_0 = Exp(1),\ h(\boldsymbol{X}) = I(waiting\ time\ of\ 20th\ customer > 2)$

What if the distribution for X_t is potentially misspecified?

maximize $E_f[h(X_1, \ldots, X_T)]$
subject to

$$X_t \overset{\text{i.i.d.}}{\sim} P_f$$
$$D\left(P_f | P_0\right) \leq \eta$$

Decision variable: P_f
$D\left(P_f | P_0\right)$ can be:

- Kullback-Leibler (KL) divergence = $\int \log \frac{dP_f}{dP_0} dP_f$
- ϕ-divergence = $\int \phi\left(\frac{dP_f}{dP_0}\right) dP_0$
- replaced/enriched by moment constraints, e.g. $\underline{\mu} \leq E_f[X_t] \leq \bar{\mu}$

Example 2: Uncertain Dependency

- Baseline input model: $X_t \sim P_0$
- $\boldsymbol{X} = (X_1, X_2, \dots, X_T)$
- Performance measure: $E[h(\boldsymbol{X})]$

E.g. Single-server queue: $\boldsymbol{X}$ = sequence of interarrival times,

$P_0 = Exp(0.8),\ h(\boldsymbol{X}) = I(waiting\ time\ of\ 20th\ customer > 2)$

What if the i.i.d. assumption for $(X_1, \dots, X_T)$ is potentially misspecified?

compute $E_0[h(X_1, \dots, X_T)]$
under the condition

$$X_t \overset{\text{i.i.d.}}{\sim} P_0$$

Example 2: Uncertain Dependency

- Baseline input model: $X_t \sim P_0$
- $\boldsymbol{X} = (X_1, X_2, \dots, X_T)$
- Performance measure: $E[h(\boldsymbol{X})]$

E.g. Single-server queue: $\boldsymbol{X}$ = sequence of interarrival times,

$P_0 = Exp(0.8),\ h(\boldsymbol{X}) = I(waiting\ time\ of\ 20th\ customer > 2)$

What if the i.i.d. assumption for $(X_1, \dots, X_T)$ is potentially misspecified?

maximize $E_f[h(X_1, \dots, X_T)]$

subject to

$(X_1, \dots, X_T) \sim \boldsymbol{P}_f$

$\boldsymbol{P}_f$ is within η-neighborhood of $\boldsymbol{P}_0 = P_0 \times \cdots \times P_0$

$\boldsymbol{P}_f$ has the same marginal as $\boldsymbol{P}_0$

$\boldsymbol{P}_f \in$ a class of stationary serially dependent processes

Example 2: Uncertain Dependency

- Baseline input model: $X_t \sim P_0$
- $\boldsymbol{X} = (X_1, X_2, \dots, X_T)$
- Performance measure: $E[h(\boldsymbol{X})]$

E.g. Single-server queue: $\boldsymbol{X}$ = sequence of interarrival times,

$P_0 = Exp(0.8),\ h(\boldsymbol{X}) = I(waiting\ time\ of\ 20th\ customer > 2)$

What if the i.i.d. assumption for $(X_1, \dots, X_T)$ is potentially misspecified?

maximize $E_f[h(X_1, \dots, X_T)]$
subject to
$(X_1, \dots, X_T) \sim \boldsymbol{P}_f$
$\boldsymbol{P}_f$ is within η-neighborhood of $\boldsymbol{P}_0 = P_0 \times \cdots \times P_0$
$\boldsymbol{P}_f$ has the same marginal as $\boldsymbol{P}_0$
$\boldsymbol{P}_f \in$ a class of stationary serially dependent processes

Example 2: Uncertain Dependency

- Baseline input model: $X_t \sim P_0$
- $\boldsymbol{X} = (X_1, X_2, \dots, X_T)$
- Performance measure: $E[h(\boldsymbol{X})]$

E.g. Single-server queue: $\boldsymbol{X}$ = sequence of interarrival times,
$P_0 = Exp(0.8),\ h(\boldsymbol{X}) = I(waiting\ time\ of\ 20th\ customer > 2)$

What if the i.i.d. assumption for $(X_1, \dots, X_T)$ is potentially misspecified?

maximize $E_f[h(X_1, \dots, X_T)]$
subject to
$(X_1, \dots, X_T) \sim \boldsymbol{P}_f$
$\boldsymbol{P}_f$ is within η-neighborhood of $\boldsymbol{P}_0 = P_0 \times \cdots \times P_0$
$\boldsymbol{P}_f$ has the same marginal as $\boldsymbol{P}_0$
$\boldsymbol{P}_f \in$ a class of stationary serially dependent processes

Example 2: Uncertain Dependency

- Baseline input model: $X_t \sim P_0$
- $\boldsymbol{X} = (X_1, X_2, \dots, X_T)$
- Performance measure: $E[h(\boldsymbol{X})]$

E.g. Single-server queue: $\boldsymbol{X}$ = sequence of interarrival times,

$P_0 = Exp(0.8),\ h(\boldsymbol{X}) = I(waiting\ time\ of\ 20th\ customer > 2)$

What if the i.i.d. assumption for $(X_1, \dots, X_T)$ is potentially misspecified?

maximize $E_f[h(X_1, \dots, X_T)]$

subject to

$(X_1, \dots, X_T) \sim \boldsymbol{P}_f$

$\boldsymbol{P}_f$ is within η-neighborhood of $\boldsymbol{P}_0 = P_0 \times \cdots \times P_0$

$\boldsymbol{P}_f$ has the same marginal as $\boldsymbol{P}_0$

$\boldsymbol{P}_f \in$ a class of stationary serially dependent processes

Example 2: Uncertain Dependency

- Baseline input model: $X_t \sim P_0$
- $\boldsymbol{X} = (X_1, X_2, \dots, X_T)$
- Performance measure: $E[h(\boldsymbol{X})]$

E.g. Single-server queue: $\boldsymbol{X}$ = sequence of interarrival times,
$P_0 = Exp(0.8),\ h(\boldsymbol{X}) = I(waiting\ time\ of\ 20th\ customer > 2)$

What if the i.i.d. assumption for $(X_1, \dots, X_T)$ is potentially misspecified?

maximize $E_f[h(X_1, \dots, X_T)]$
subject to
$(X_1, \dots, X_T) \sim \boldsymbol{P}_f$
$\boldsymbol{P}_f$ is within η-neighborhood of $\boldsymbol{P}_0 = P_0 \times \cdots \times P_0$
$\boldsymbol{P}_f$ has the same marginal as $\boldsymbol{P}_0$
$\boldsymbol{P}_f \in$ a class of stationary serially dependent processes

Decision variable: $\boldsymbol{P}_f$

Example 2: Uncertain Dependency

- Baseline input model: $X_t \sim P_0$
- $\boldsymbol{X} = (X_1, X_2, \dots, X_T)$
- Performance measure: $E[h(\boldsymbol{X})]$

E.g. Single-server queue: $\boldsymbol{X}$ = sequence of interarrival times,

$P_0 = Exp(0.8),\ h(\boldsymbol{X}) = I(waiting\ time\ of\ 20th\ customer > 2)$

What if the i.i.d. assumption for $(X_1, \dots, X_T)$ is potentially misspecified?

maximize $E_f[h(X_1, \dots, X_T)]$

subject to

$(X_1, \dots, X_T) \sim \boldsymbol{P}_f$

$\boldsymbol{P}_f$ is within η-neighborhood of $\boldsymbol{P}_0 = P_0 \times \cdots \times P_0$

$\boldsymbol{P}_f$ has the same marginal as $\boldsymbol{P}_0$

$\boldsymbol{P}_f \in$ a class of stationary 1-lag dependent processes

Decision variable: $\boldsymbol{P}_f$

Example 2: Uncertain Dependency

- Baseline input model: $X_t \sim P_0$
- $\boldsymbol{X} = (X_1, X_2, \dots, X_T)$
- Performance measure: $E[h(\boldsymbol{X})]$

E.g. Single-server queue: $\boldsymbol{X}$ = sequence of interarrival times,

$P_0 = Exp(0.8),\ h(\boldsymbol{X}) = I(waiting\ time\ of\ 20th\ customer > 2)$

What if the i.i.d. assumption for $(X_1, \dots, X_T)$ is potentially misspecified?

maximize $E_f[h(X_1, \dots, X_T)]$
subject to
$(X_1, \dots, X_T) \sim \boldsymbol{P}_f$
$D(\boldsymbol{P}_f | \boldsymbol{P}_0) \leq \eta$
$\boldsymbol{P}_f$ has the same marginal as $\boldsymbol{P}_0$
$\boldsymbol{P}_f \in$ a class of stationary 1-lag dependent processes

Decision variable: $\boldsymbol{P}_f$
$D(\boldsymbol{P}_f | \boldsymbol{P}_0)$ can be:

- mutual information for the bivariate distribution of any two consecutive states under $\boldsymbol{P}_f$

Example 2: Uncertain Dependency

- Baseline input model: $X_t \sim P_0$
- $\boldsymbol{X} = (X_1, X_2, \dots, X_T)$
- Performance measure: $E[h(\boldsymbol{X})]$

E.g. Single-server queue: $\boldsymbol{X}$ = sequence of interarrival times,

$P_0 = Exp(0.8),\ h(\boldsymbol{X}) = I(waiting\ time\ of\ 20th\ customer > 2)$

What if the i.i.d. assumption for $(X_1, \dots, X_T)$ is potentially misspecified?

maximize $E_f[h(X_1, \dots, X_T)]$

subject to

$(X_1, \dots, X_T) \sim \boldsymbol{P}_f$

$D(\boldsymbol{P}_f | \boldsymbol{P}_0) \leq \eta$

$\boldsymbol{P}_f$ has the same marginal as $\boldsymbol{P}_0$

$\boldsymbol{P}_f \in$ a class of stationary 1-lag dependent processes

Decision variable: $\boldsymbol{P}_f$

$D(\boldsymbol{P}_f | \boldsymbol{P}_0)$ can be:

- mutual information for the bivariate distribution of any two consecutive states under $\boldsymbol{P}_f$
- Pearson's ϕ^2-coefficient for the bivariate distribution of any two consecutive states under $\boldsymbol{P}_f$

Example 2: Uncertain Dependency

- Baseline input model: $X_t \sim P_0$
- $\boldsymbol{X} = (X_1, X_2, \ldots, X_T)$
- Performance measure: $E[h(\boldsymbol{X})]$

Output the worst-case performance measure among all 1-lag processes within η units of nonparametric distance from i.i.d. process, keeping marginals fixed

E.g. Single-server queue: $\boldsymbol{X}$ = sequence of interarrival times,

$P_0 = Exp(0.8),\ h(\boldsymbol{X}) = I(waiting\ time\ of\ 20th\ customer > 2)$

What if the i.i.d. assumption for $(X_1, \ldots, X_T)$ is potentially misspecified?

maximize $E_f[h(X_1, \ldots, X_T)]$

subject to

$$(X_1, \ldots, X_T) \sim \boldsymbol{P}_f$$

$$D\left(\boldsymbol{P}_f \middle| \boldsymbol{P}_0\right) \leq \eta$$

$\boldsymbol{P}_f$ has the same marginal as $\boldsymbol{P}_0$

$\boldsymbol{P}_f \in$ a class of stationary 1-lag dependent processes

Decision variable: $\boldsymbol{P}_f$

$D\left(\boldsymbol{P}_f | \boldsymbol{P}_0\right)$ can be:

- mutual information for the bivariate distribution of any two consecutive states under $\boldsymbol{P}_f$
- Pearson's ϕ^2-coefficient for the bivariate distribution of any two consecutive states under $\boldsymbol{P}_f$

Past Literature

Area	Examples of Works
Economics	Hansen & Sargent ‘01, ’08
Finance	Glasserman & Xu ’13, Lim et al. ‘12
Control theory/MDP/Markov perturbation analysis	Petersen et. al. ‘00, Nilim & El Ghaoui ‘05, Iyengar ’05, Jain et. al. ’10, Heidergott et. al. ‘10
(Distributionally) robust optimization	Delage & Ye ‘10, Goh & Sim ‘10, Ben-Tal et. al. ’13, Wiesemann et. al. ‘14, Xin et. al. ’15
Robust Monte Carlo	Glasserman & Xu ’14, Glasserman & Yang ‘15
Decision analysis/Moment method	Smith ‘93, ‘95, Birge & Wets ’87, Bertsimas & Popescu ’05, Popescu ‘05

Past Literature

Area	Examples of Works
Economics	Hansen & Sargent ‘01, ’08
Finance	Glasserman & Xu ’13, Lim et al. ‘12
Control theory/MDP/Markov perturbation analysis	Petersen et. al. ‘00, Nilim & El Ghaoui ‘05, Iyengar ’05, Jain et. al. ’10, Heidergott et. al. ‘10
(Distributionally) robust optimization	Delage & Ye ‘10, Goh & Sim ‘10, Ben-Tal et. al. ’13, Wiesemann et. al. ‘14, Xin et. al. ‘15
Robust Monte Carlo	Glasserman & Xu ’14, Glasserman & Yang ‘15
Decision analysis/Moment method	Smith ‘93, ‘95, Birge & Wets ’87, Bertsimas & Popescu ’05, Popescu ‘05

Typical focus in the literature:

- Tractable convex optimization formulation

Past Literature

Area	Examples of Works
Economics	Hansen & Sargent ‘01, ’08
Finance	Glasserman & Xu ’13, Lim et al. ‘12
Control theory/MDP/Markov perturbation analysis	Petersen et. al. ‘00, Nilim & El Ghaoui ‘05, Iyengar ’05, Jain et. al. ’10, Heidergott et. al. ‘10
(Distributionally) robust optimization	Delage & Ye ‘10, Goh & Sim ‘10, Ben-Tal et. al. ’13, Wiesemann et. al. ‘14, Xin et. al. ‘15
Robust Monte Carlo	Glasserman & Xu ’14, Glasserman & Yang ‘15
Decision analysis/Moment method	Smith ‘93, ‘95, Birge & Wets ’87, Bertsimas & Popescu ’05, Popescu ‘05

Typical focus in the literature:

- Tractable convex optimization formulation

Challenges in the stochastic modeling contexts:

- Non-standard constraints/objectives
- High or infinite dimensionality
- Stochastic/simulation-based

A Sensitivity Approach

Worst-case formulation for the Basic Example:

$$\begin{aligned} &\text{maximize} \;\; E_f[h(\boldsymbol{X})] \\ &\text{subject to} \;\; D\left(P_f | P_0\right) \leq \eta \\ &\qquad\qquad\quad X_t \overset{\text{i.i.d.}}{\sim} P_f \end{aligned}$$

where $D(\cdot \mid \cdot)$ = KL divergence

A Sensitivity Approach

Relaxed worst-case formulation for the Basic Example:

maximize $E_f[h(\boldsymbol{X})]$
subject to $D\left(\boldsymbol{P}_f|\boldsymbol{P}_0\right) \leq T\eta$

where $D(\cdot \mid \cdot)$ = KL divergence

A Sensitivity Approach

Relaxed worst-case formulation for the Basic Example:

maximize $E_f[h(\boldsymbol{X})]$

subj Too conservative… $T\eta$

where $D(\cdot \mid \cdot)$ = KL divergence

A Sensitivity Approach

Worst-case formulation:

$$\begin{aligned} &\text{maximize} \quad E_f[h(\boldsymbol{X})] \\ &\text{subject to} \quad D\left(P_f | P_0\right) \leq \eta \\ &\qquad\qquad\quad X_t \overset{\text{i.i.d.}}{\sim} P_f \end{aligned}$$

A Sensitivity Approach

Worst-case formulation:

maximize $E_f[h(\boldsymbol{X})]$
subject to $D\left(P_f|P_0\right) \leq \eta$
$X_t \overset{\text{i.i.d.}}{\sim} P_f$

Let $\eta \to 0$. Under mild regularity conditions,

$$\max E_f[h(\boldsymbol{X})] = E_0[h(\boldsymbol{X})] + \xi_1(P_0, h)\sqrt{\eta} + \xi_2(P_0, h)\eta + \cdots$$

where $\xi_1(P_0, h)$, … are computationally tractable objects that depend only on the baseline model P_0

A Sensitivity Approach

Worst-case formulation:

maximize $E_f[h(\boldsymbol{X})]$
subject to $D\left(P_f | P_0\right) \leq \eta$
$X_t \overset{\text{i.i.d.}}{\sim} P_f$

Let $\eta \to 0$. Under mild regularity conditions,

$$\max E_f[h(\boldsymbol{X})] = E_0[h(\boldsymbol{X})] + \xi_1(P_0, h)\sqrt{\eta} + \xi_2(P_0, h)\eta + \cdots$$

where $\xi_1(P_0, h)$, … are computationally tractable objects that depend only on the baseline model P_0

A Sensitivity Approach

Worst-case formulation:

$$\begin{aligned} &\text{maximize} \quad E_f[h(\boldsymbol{X})] \\ &\text{subject to} \quad D\left(P_f | P_0\right) \le \eta \\ &\qquad\qquad\quad X_t \overset{\text{i.i.d.}}{\sim} P_f \end{aligned}$$

Let $\eta \to 0$. Under mild regularity conditions,

$$\max\ E_f[h(\boldsymbol{X})] = E_0[h(\boldsymbol{X})] + \xi_1(P_0, h)\sqrt{\eta} + \xi_2(P_0, h)\eta + \cdots$$

where $\xi_1(P_0, h)$, … are computationally tractable objects that depend only on the baseline model P_0

Numerical Illustration

Performance measure: $E[h(\boldsymbol{X})] = P(W_{20} > 2)$

Baseline model P_0 for service time: $Exp(1)$

Worst-case formulation:

$$\begin{aligned} &\text{maximize} \;\; E_f[h(\boldsymbol{X})] \\ &\text{subject to} \;\; D(P_f|P_0) \leq \eta \\ &\qquad\qquad\quad X_t \overset{\text{i.i.d.}}{\sim} P_f \end{aligned}$$

and

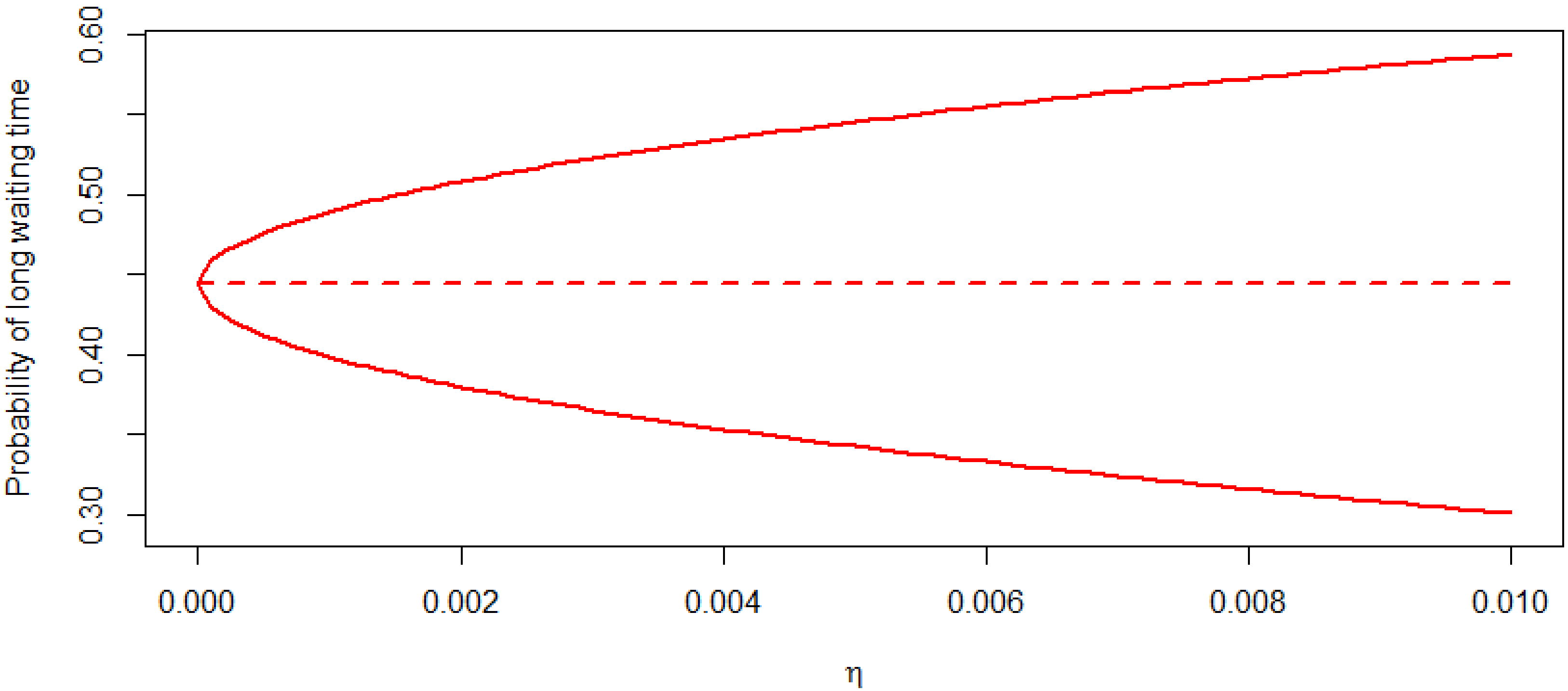
$$\begin{aligned} &\text{minimize} \;\; E_f[h(\boldsymbol{X})] \\ &\text{subject to} \;\; D(P_f|P_0) \leq \eta \\ &\qquad\qquad\quad X_t \overset{\text{i.i.d.}}{\sim} P_f \end{aligned}$$

Asymptotic approximation

$$\max E_f[h(\boldsymbol{X})] \approx E_0[h(\boldsymbol{X})] + \xi_1(P_0, h)\sqrt{\eta}$$

$$\min E_f[h(\boldsymbol{X})] \approx E_0[h(\boldsymbol{X})] - \xi_1(P_0, h)\sqrt{\eta}$$

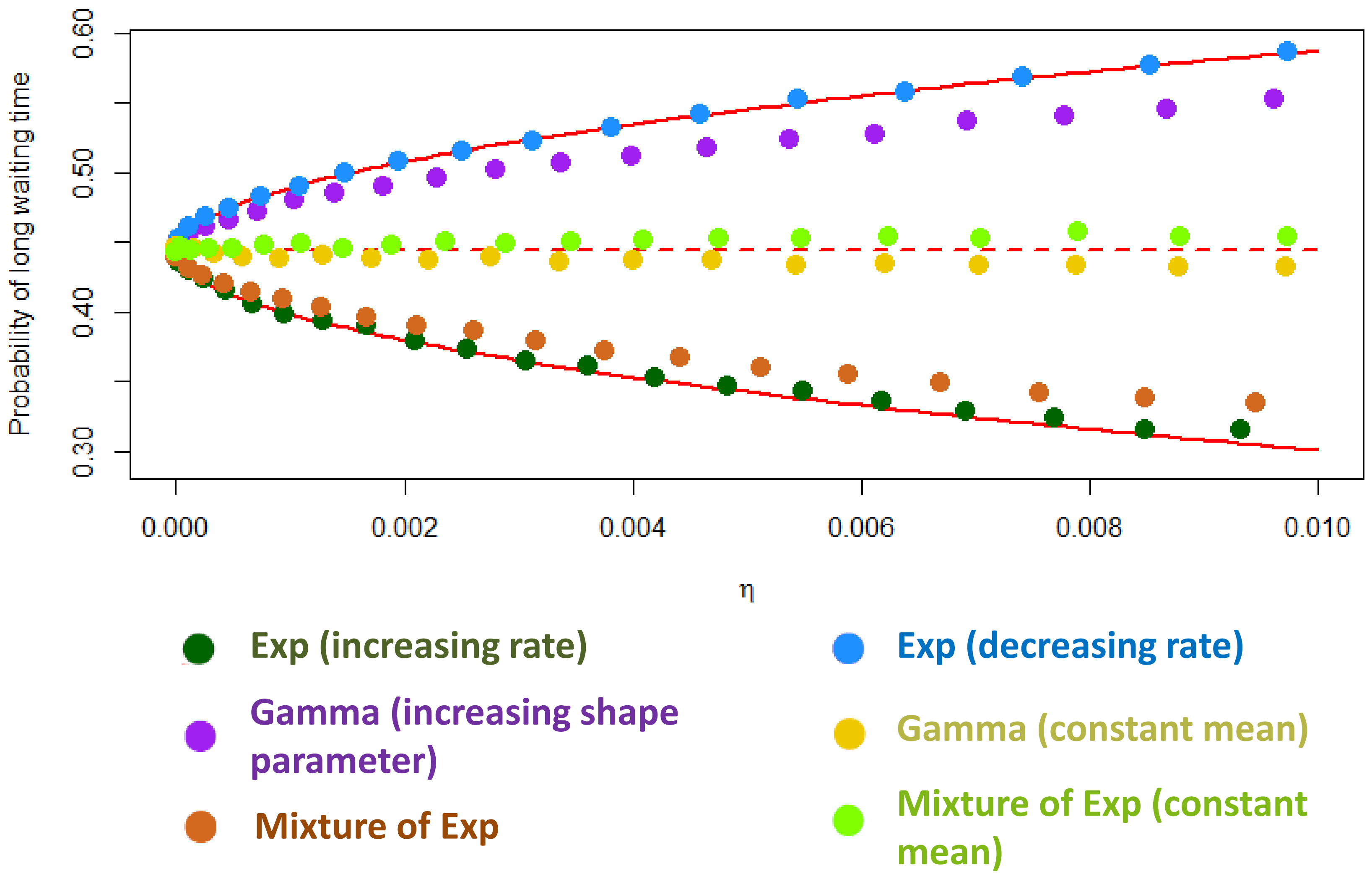
Numerical Illustration



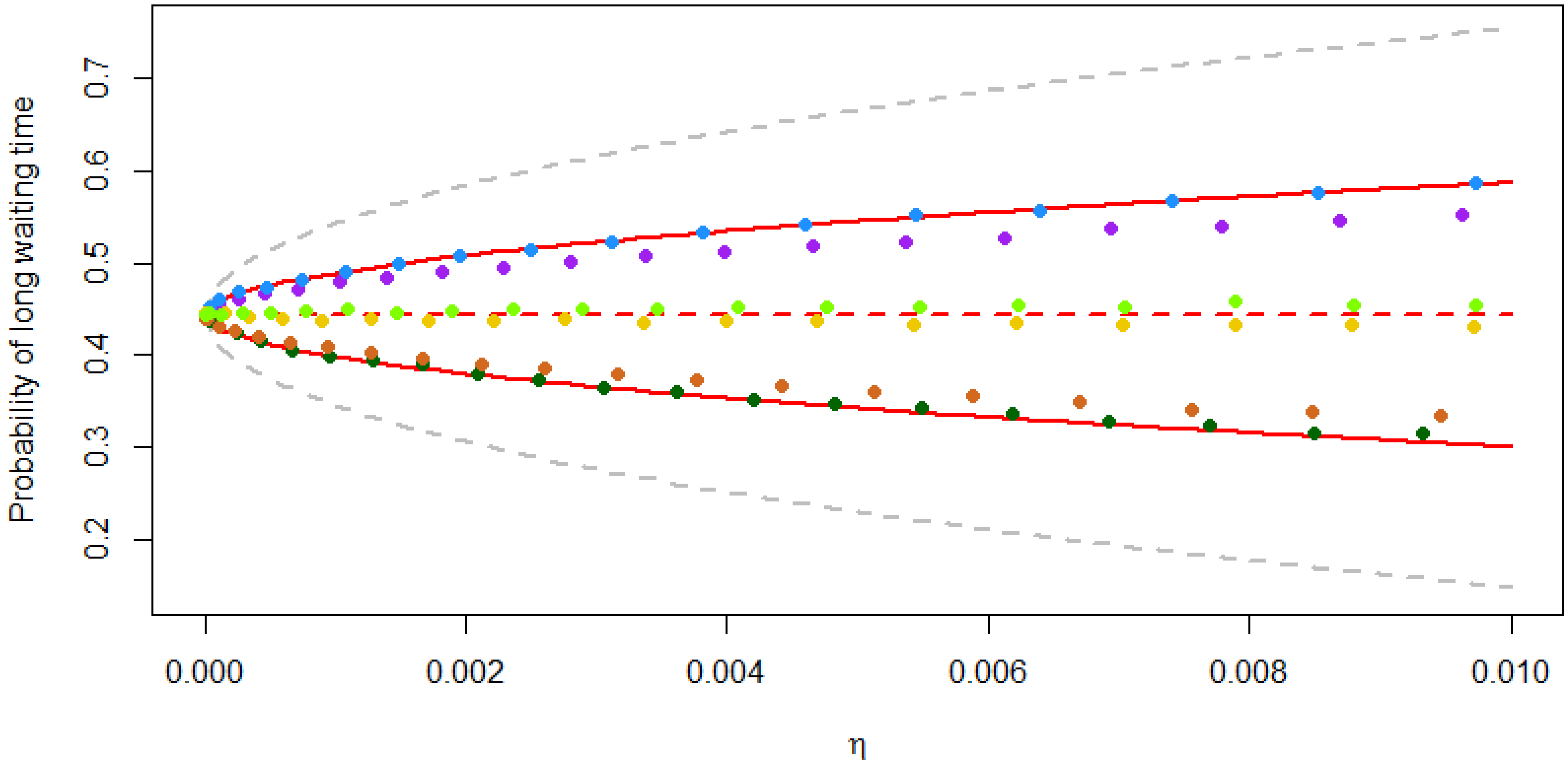
Worst-case lower and upper bounds

Baseline performance measure

Numerical Illustration



Numerical Illustration



Relaxed bounds from conventional approach

Form of Expansion Coefficients

Influence function (Hampel ’74): the first order effect on a statistical functional due to perturbation of the input probability distribution

For a functional $Z(\cdot)$ acting on P, the influence function $G(x)$ is defined such that:

$$Z\left(P_f\right) = Z(P_0) + E_f[G(X)] + \cdots$$

Form of Expansion Coefficients

Influence function (Hampel '74): the first order effect on a statistical functional due to perturbation of the input probability distribution

For a functional $Z(\cdot)$ acting on P, the influence function $G(x)$ is defined such that:

$$Z(P_f) = Z(P_0) + E_f[G(X)] + \cdots$$

For example, if $Z(P_f) = E_f[h(X_1, \dots, X_T)]$ where $X_t \overset{\text{i.i.d.}}{\sim} P_f$,

$$G(x) = \sum_{t=1}^{T} E_0[h(X_1, \dots, X_T) | X_t = x] - T E_0[h(X_1, \dots, X_T)]$$

Form of Expansion Coefficients

maximize $E_f[h(\boldsymbol{X})]$
subject to $D(P_f | P_0) \leq \eta$
$X_t \overset{\text{i.i.d.}}{\sim} P_f$

Theorem (L. '13): For KL divergence $D(\cdot \mid \cdot)$, assuming $|h(\boldsymbol{X})| \leq \sum_{i=1}^{T} \Lambda_i(X_i)$ where each $\Lambda_i(X_i)$ satisfies $E_0\left[e^{\theta \Lambda_i(X_i)}\right] < \infty$ for θ in a neighborhood of zero, and $Var_0\big(G(X)\big) > 0$, then

$$\xi_1(P_0, h) = \sqrt{2Var_0\big(G(X)\big)}$$

Form of Expansion Coefficients

maximize $E_f[h(\boldsymbol{X})]$
subject to $D\left(P_f|P_0\right) \leq \eta$
$X_t \overset{\text{i.i.d.}}{\sim} P_f$

Theorem (L. '13): For KL divergence $D(\cdot\mid\cdot)$, assuming $|h(\boldsymbol{X})| \leq \sum_{i=1}^T \Lambda_i(X_i)$ where each $\Lambda_i(X_i)$ satisfies $E_0\left[e^{\theta\Lambda_i(X_i)}\right] < \infty$ for θ in a neighborhood of zero, and $Var_0\big(G(X)\big) > 0$, then

$$\xi_1(P_0,h) = \sqrt{2Var_0\big(G(X)\big)}$$

$$\xi_2(P_0,h) = \frac{1}{Var_0\big(G(X)\big)}\left(\frac{1}{3}\kappa_3\big(G(X)\big) + \nu\right)$$

where κ_3 is the third-order cumulant and

$$\nu = E_0[(G_2(X,Y) - E_0[G_2(X,Y)])(G(X) - E_0[G(X)])(G(Y) - E_0[G(Y)])]$$

and $G_2(X,Y)$ is a "second order" influence function

Form of Expansion Coefficients

maximize $E_f[h(\boldsymbol{X})]$
subject to $D\left(P_f | P_0\right) \leq \eta$
$X_t \overset{\text{i.i.d.}}{\sim} P_f$

Theorem (L. '13): For KL divergence $D(\cdot\mid\cdot)$, assuming $|h(\boldsymbol{X})| \leq \sum_{i=1}^{T} \Lambda_i(X_i)$ where each $\Lambda_i(X_i)$ satisfies $E_0\left[e^{\theta\Lambda_i(X_i)}\right] < \infty$ for θ in a neighborhood of zero, and $Var_0\big(G(X)\big) > 0$, then

$$\xi_1(P_0, h) = \sqrt{2Var_0\big(G(X)\big)}$$

$$\xi_2(P_0, h) = \frac{1}{Var_0\big(G(X)\big)}\left(\frac{1}{3}\kappa_3\big(G(X)\big) + \nu\right)$$

where κ_3 is the third-order cumulant and

$$\nu = E_0[(G_2(X,Y) - E_0[G_2(X,Y)])(G(X) - E_0[G(X)])(G(Y) - E_0[G(Y)])]$$

and $G_2(X,Y)$ is a "second order" influence function

Simulation strategy: nested simulation, or nonparametric bootstrap

Uncertain Dependency

maximize $E_f[h(X_1, \ldots, X_T)]$

subject to

$(X_1, \ldots, X_T) \sim \boldsymbol{P}_f$

$D(\boldsymbol{P}_f | \boldsymbol{P}_0) \leq \eta$

$\boldsymbol{P}_f$ has the same marginal as $\boldsymbol{P}_0$

$\boldsymbol{P}_f \in$ a class of stationary 1-lag dependent processes

Uncertain Dependency

maximize $E_f[h(X_1, \dots, X_T)]$
subject to
$(X_1, \dots, X_T) \sim \boldsymbol{P}_f$
$D\left(\boldsymbol{P}_f \middle| \boldsymbol{P}_0\right) \le \eta$
$\boldsymbol{P}_f$ has the same marginal as $\boldsymbol{P}_0$
$\boldsymbol{P}_f \in$ a class of stationary 1-lag dependent processes

$$\max\ E_f[h(\boldsymbol{X})] = E_0[h(\boldsymbol{X})] + \xi_1(P_0, h)\sqrt{\eta} + \xi_2(P_0, h)\eta + \cdots$$

Uncertain Dependency

maximize $E_f[h(X_1, \dots, X_T)]$
subject to
$(X_1, \dots, X_T) \sim \boldsymbol{P}_f$
$D\left(\boldsymbol{P}_f \middle| \boldsymbol{P}_0\right) \leq \eta$
$\boldsymbol{P}_f$ has the same marginal as $\boldsymbol{P}_0$
$\boldsymbol{P}_f \in$ a class of stationary 1-lag dependent processes

$$\max\ E_f[h(\boldsymbol{X})] = E_0[h(\boldsymbol{X})] + \xi_1(P_0, h)\sqrt{\eta} + \xi_2(P_0, h)\eta + \cdots$$

Theorem (L. '13): Define a "bivariate" influence function

$$\tilde{G}(x, y) = \sum_{t=2}^{T} E_0[h(X_1, \dots, X_T) | X_{t-1} = x, X_t = y]$$

Uncertain Dependency

maximize $E_f[h(X_1, \ldots, X_T)]$
subject to
$(X_1, \ldots, X_T) \sim \boldsymbol{P}_f$
$D(\boldsymbol{P}_f | \boldsymbol{P}_0) \leq \eta$
$\boldsymbol{P}_f$ has the same marginal as $\boldsymbol{P}_0$
$\boldsymbol{P}_f \in$ a class of stationary 1-lag dependent processes

$$\max E_f[h(\boldsymbol{X})] = E_0[h(\boldsymbol{X})] + \xi_1(P_0, h)\sqrt{\eta} + \xi_2(P_0, h)\eta + \cdots$$

Theorem (L. '13): Define a "bivariate" influence function

$$\tilde{G}(x, y) = \sum_{t=2}^{T} E_0[h(X_1, \ldots, X_T) | X_{t-1} = x, X_t = y]$$

and its "interaction effect"

$$R(x, y) = \tilde{G}(x, y) - E_0\left[\tilde{G}(X, Y) \middle| X = x\right] - E_0\left[\tilde{G}(X, Y) \middle| Y = y\right] + E_0[\tilde{G}(X, Y)]$$

Uncertain Dependency

maximize $E_f[h(X_1, \dots, X_T)]$
subject to
$(X_1, \dots, X_T) \sim \boldsymbol{P}_f$
$D(\boldsymbol{P}_f | \boldsymbol{P}_0) \le \eta$
$\boldsymbol{P}_f$ has the same marginal as $\boldsymbol{P}_0$
$\boldsymbol{P}_f \in$ a class of stationary 1-lag dependent processes

$$\max\ E_f[h(\boldsymbol{X})] = E_0[h(\boldsymbol{X})] + \xi_1(P_0, h)\sqrt{\eta} + \xi_2(P_0, h)\eta + \cdots$$

Theorem (L. '13): Define a "bivariate" influence function

$$\tilde{G}(x, y) = \sum_{t=2}^{T} E_0[h(X_1, \dots, X_T) | X_{t-1} = x, X_t = y]$$

and its "interaction effect"

$$R(x, y) = \tilde{G}(x, y) - E_0[\tilde{G}(X, Y) | X = x] - E_0[\tilde{G}(X, Y) | Y = y] + E_0[\tilde{G}(X, Y)]$$

For mutual information $D(\cdot)$, assuming $h(\boldsymbol{X})$ is bounded, and $Var_0(R(X, Y)) > 0$, then

$$\xi_1(P_0, h) = \sqrt{2Var_0(R(X, Y))}$$

where $X, Y \overset{\text{i.i.d.}}{\sim} P_0$

Uncertain Dependency

maximize $E_f[h(X_1, \dots, X_T)]$
subject to
$(X_1, \dots, X_T) \sim \boldsymbol{P}_f$
$D(\boldsymbol{P}_f | \boldsymbol{P}_0) \leq \eta$
$\boldsymbol{P}_f$ has the same marginal as $\boldsymbol{P}_0$
$\boldsymbol{P}_f \in$ a class of stationary 1-lag dependent processes

$$\max\ E_f[h(\boldsymbol{X})] = E_0[h(\boldsymbol{X})] + \xi_1(P_0, h)\sqrt{\eta} + \xi_2(P_0, h)\eta + \cdots$$

Theorem (L. '13): Define a "bivariate" influence function

$$\tilde{G}(x, y) = \sum_{t=2}^{T} E_0[h(X_1, \dots, X_T) | X_{t-1} = x, X_t = y]$$

and its "interaction effect"

$$R(x, y) = \tilde{G}(x, y) - E_0\left[\tilde{G}(X, Y) \middle| X = x\right] - E_0\left[\tilde{G}(X, Y) \middle| Y = y\right] + E_0[\tilde{G}(X, Y)]$$

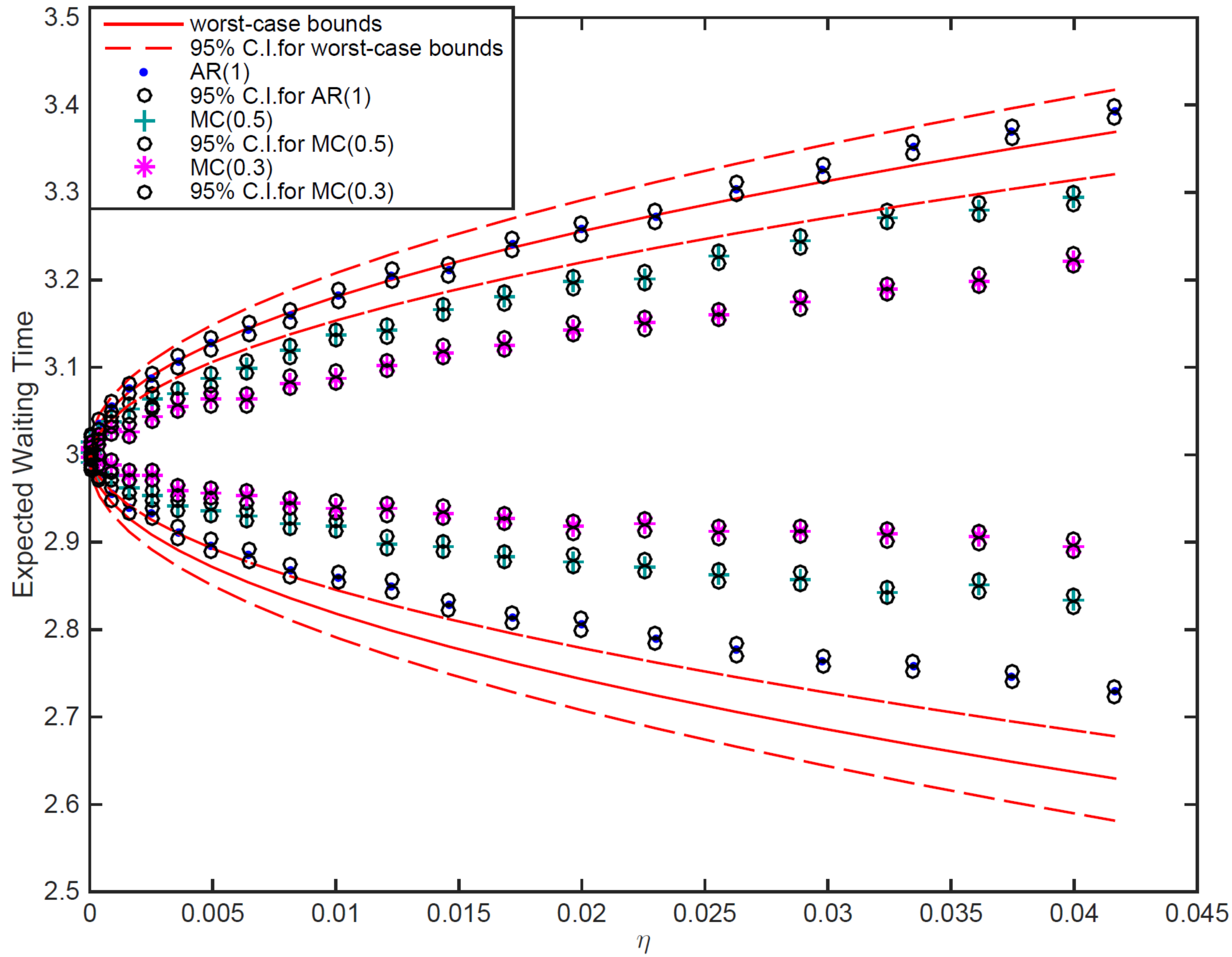
For mutual information $D(\cdot)$, assuming $h(\boldsymbol{X})$ is bounded, and $Var_0\big(R(X, Y)\big) > 0$, then

$$\xi_1(P_0, h) = \sqrt{2Var_0\big(R(X, Y)\big)}$$

where $X, Y \overset{\text{i.i.d.}}{\sim} P_0$

Simulation strategy: nested simulation + two-way analysis of variance (ANOVA)

Numerical Illustration



General Formulations and Simulation Optimization

- Beyond asymptotic approximation?
- Other types of constraints/objectives?

General Formulations and Simulation Optimization

- Beyond asymptotic approximation?
- Other types of constraints/objectives?

Example 3: Input Reconstruction

Can we infer about the service time distribution from waiting time data?

General Formulations and Simulation Optimization

- Beyond asymptotic approximation?
- Other types of constraints/objectives?

Example 3: Input Reconstruction

Can we infer about the service time distribution from waiting time data?

Find P_f

such that

$$E_f\left[\phi_j\big(h(X_1,\dots,X_T)\big)\right]=\hat{\mu}_j \text{ for } j=1,\dots,m$$
$$X_t \overset{\text{i.i.d.}}{\sim} P_f$$

- $\phi_j(\cdot), j=1,\dots,m$: "moment" functions, e.g. $I(\cdot\le q_j)$, $(\cdot)^j$
- $h(X_1,\dots,X_T)$: waiting time
- $\hat{\mu}_j$: empirical moments

General Formulations and Simulation Optimization

- Beyond asymptotic approximation?
- Other types of constraints/objectives?

Example 3: Input Reconstruction

Can we infer about the service time distribution from waiting time data?

Moment matching between the simulation output and the real output data

Find P_f

such that

$$E_f\left[\phi_j\left(h(X_1,\ldots,X_T)\right)\right]=\hat{\mu}_j \text{ for } j=1,\ldots,m$$

$$X_t \overset{\text{i.i.d.}}{\sim} P_f$$

- $\phi_j(\cdot), j=1,\ldots,m$: "moment" functions, e.g. $I(\cdot \leq q_j)$, $(\cdot)^j$
- $h(X_1,\ldots,X_T)$: waiting time
- $\hat{\mu}_j$: empirical moments

General Formulations and Simulation Optimization

- Beyond asymptotic approximation?
- Other types of constraints/objectives?

Example 3: Input Reconstruction

Can we infer about the service time distribution from waiting time data?

e.g., $R(P_f)$ = entropy of P_f

maximize $R(P_f)$

subject to

$$E_f\left[\phi_j\big(h(X_1,\ldots,X_T)\big)\right]=\hat{\mu}_j \text{ for } j=1,\ldots,m$$

$$X_t \overset{\text{i.i.d.}}{\sim} P_f$$

- $\phi_j(\cdot), j=1,\ldots,m$: "moment" functions, e.g. $I(\cdot\le q_j)$, $(\cdot)^j$
- $h(X_1,\ldots,X_T)$: waiting time
- $\hat{\mu}_j$: empirical moments

Sequential Least-Square Minimization

An optimization problem with non-convex stochastic constraints

maximize

$$R(P_f)$$

subject to

$$E_f\left[\phi_j\big(h(\boldsymbol{X})\big)\right] = \hat{\mu}_j \text{ for } j = 1, \dots, m$$

$$X_t \overset{\text{i.i.d.}}{\sim} P_f$$

Sequential Least-Square Minimization

An optimization problem with non-convex stochastic constraints

maximize

$$R(P_f)$$

subject to

$$E_f\left[\phi_j\left(h(\boldsymbol{X})\right)\right] = \hat{\mu}_j \text{ for } j = 1, \ldots, m$$

$$X_t \overset{\text{i.i.d.}}{\sim} P_f$$

$\Leftrightarrow$

A sequence of optimizations with stochastic non-convex objectives but deterministic convex constraints

Find the maximum η such that the optimal value of

minimize $Z\left(P_f\right) = \sum_{j=1}^{m}\left(E_f\left[\phi_j\left(h(\boldsymbol{X})\right)\right] - \hat{\mu}_j\right)^2$

subject to

$$R\left(P_f\right) \geq \eta$$

is 0. The P^* that is attained by η^* gives the optimal solution

Sequential Least-Square Minimization

An optimization problem with non-convex stochastic constraints

maximize

$$R(P_f)$$

subject to

$$E_f[\phi_j(h(\boldsymbol{X}))] = \hat{\mu}_j \text{ for } j = 1, \ldots, m$$

$$X_t \overset{\text{i.i.d.}}{\sim} P_f$$

$\Leftrightarrow$

A sequence of optimizations with stochastic non-convex objectives but deterministic convex constraints

Find the maximum η such that the optimal value of

minimize $Z(P_f) = \sum_{j=1}^{m}\left(E_f[\phi_j(h(\boldsymbol{X}))] - \hat{\mu}_j\right)^2$

subject to

$$R(P_f) \geq \eta$$

is 0. The P^* that is attained by η^* gives the optimal solution

Sequential Least-Square Minimization

An optimization problem with non-convex stochastic constraints

maximize

$$R(P_f)$$

subject to

$$E_f\left[\phi_j\left(h(\boldsymbol{X})\right)\right] = \hat{\mu}_j \text{ for } j = 1, \ldots, m$$

$$X_t \overset{\text{i.i.d.}}{\sim} P_f$$

$\Leftrightarrow$

A sequence of optimizations with stochastic non-convex objectives but deterministic convex constraints

Find the maximum η such that the optimal value of

minimize $Z\left(P_f\right) = \sum_{j=1}^{m}\left(E_f\left[\phi_j\left(h(\boldsymbol{X})\right)\right] - \hat{\mu}_j\right)^2$

subject to

$$R\left(P_f\right) \geq \eta$$

is 0. The P^* that is attained by η^* gives the optimal solution

Sequential Least-Square Minimization

An optimization problem with non-convex stochastic constraints

maximize

$$R(P_f)$$

subject to

$$E_f\left[\phi_j\left(h(\boldsymbol{X})\right)\right] = \hat{\mu}_j \ \text{ for } \ j = 1, \dots, m$$

$$X_t \overset{\text{i.i.d.}}{\sim} P_f$$

$\Leftrightarrow$

A sequence of optimizations with stochastic non-convex objectives but deterministic convex constraints

Find the maximum η such that the optimal value of

minimize $Z\left(P_f\right) = \sum_{j=1}^{m}\left(E_f\left[\phi_j\left(h(\boldsymbol{X})\right)\right] - \hat{\mu}_j\right)^2$

subject to

$$R\left(P_f\right) \geq \eta$$

is 0. The P^* that is attained by η^* gives the optimal solution

- Similar in spirit to quadratic penalty method (Bertsekas '95)

Sequential Least-Square Minimization

An optimization problem with non-convex stochastic constraints

maximize

$$R(P_f)$$

subject to

$$E_f\left[\phi_j\left(h(\boldsymbol{X})\right)\right] = \hat{\mu}_j \text{ for } j = 1, \ldots, m$$

$$X_t \overset{\text{i.i.d.}}{\sim} P_f$$

$\Leftrightarrow$

A sequence of optimizations with stochastic non-convex objectives but deterministic convex constraints

Find the maximum η such that the optimal value of

minimize $Z\left(P_f\right) = \sum_{j=1}^{m}\left(E_f\left[\phi_j\left(h(\boldsymbol{X})\right)\right] - \hat{\mu}_j\right)^2$

subject to

$$R\left(P_f\right) \geq \eta$$

is 0. The P^* that is attained by η^* gives the optimal solution

- Similar in spirit to quadratic penalty method (Bertsekas '95)

Simulation Optimization over Probability Distributions

Operate on a discretized space P_f as $(p_1, \dots, p_k)$ on fixed support points $(z_1, \dots, z_k)$

Constrained stochastic approximation (SA) for solving local optimum

Simulation Optimization over Probability Distributions

Operate on a discretized space P_f as $(p_1, \ldots, p_k)$ on fixed support points $(z_1, \ldots, z_k)$

Constrained stochastic approximation (SA) for solving local optimum

Gradient estimation:

- Perturb P_f to a mixture with point mass $(1-\epsilon)P_f + \epsilon e_i$ to obtain a finite-dimensional influence function
- A "nonparametric" version of the likelihood ratio / score function method (Glynn '90, Reiman & Weiss '89, L'Ecuyer '90)

Iteration on Mirror Descent Stochastic Approximation

Each iteration solves a linearization of the objective value, penalized by the KL divergence (the entropic descent algorithm; e.g., Nemirovski et. al. ’09, Beck & Teboulle ’03, Nemirovski & Yudin ‘83)

Given a current solution P_n, solve

$$P_{n+1} = \operatorname{argmin}_{P_f} \widehat{\nabla Z(P_n)}'(P_f - P_n) + \frac{1}{\gamma_n} D(P_f | P_n)$$
$$\text{subject to} \quad R(P_f) \geq \eta$$

$$\propto P_0^{\frac{\beta_n}{1+\beta_n}} P_n^{\frac{1}{1+\beta_n}} e^{\frac{\gamma_n \widehat{\nabla Z(P_n)}}{1+\beta_n}}$$

where

- $\widehat{\nabla Z(P_n)}$ is the estimated gradient
- γ_n is the step size
- β_n solves a one-dimensional root-finding problem

Convergence rate $O(1/n)$ in terms of the expected KL divergence when $\gamma_n = \theta/n$ for large enough θ and standard conditions hold (Goeva, L. & Zhang ’14, L. & Qian ‘16)

Numerical Illustration

System: M/G/1 queue with known arrival rate but unknown service time distribution

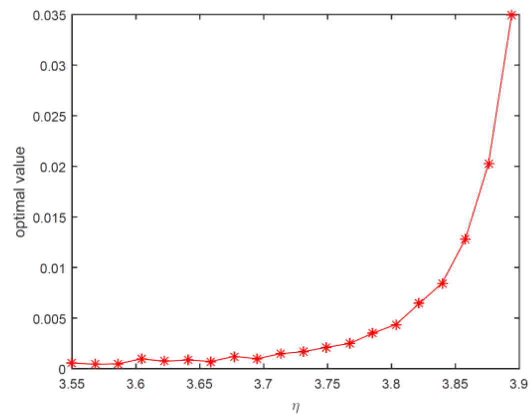
Output observations: queue length over time

Specifications:

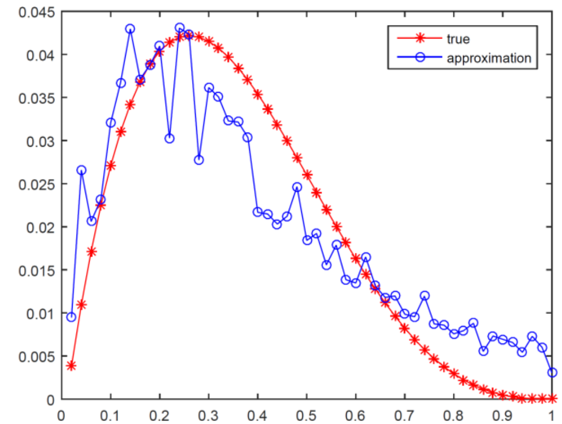
- Match 9 quantiles of the output variable “time-average queue length in a busy period”
- Discretization grid of 50 support points
- # data = $100{,}000$

Numerical Illustration

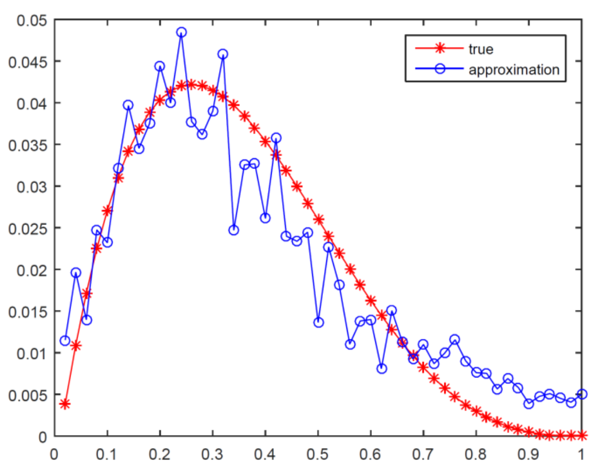
Plot of the optimal $Z(P_f)$ against η



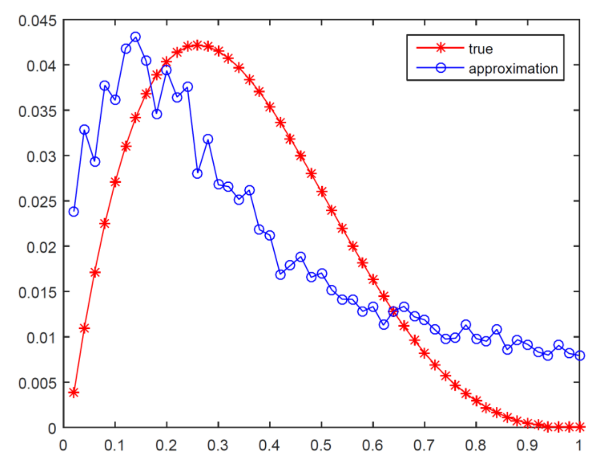
True vs reconstructed distribution



matching quantiles increased from 9 to 14

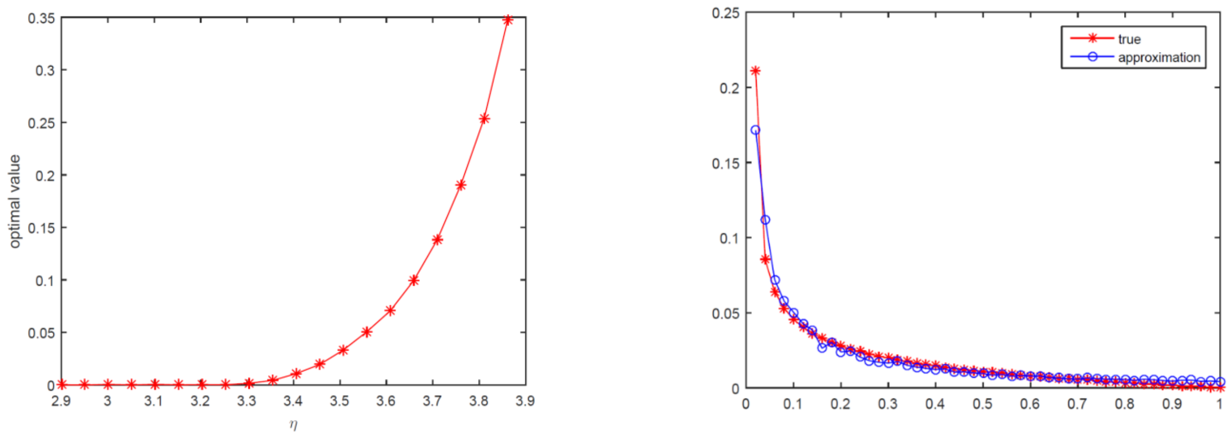


data reduced from 10^5 to 200

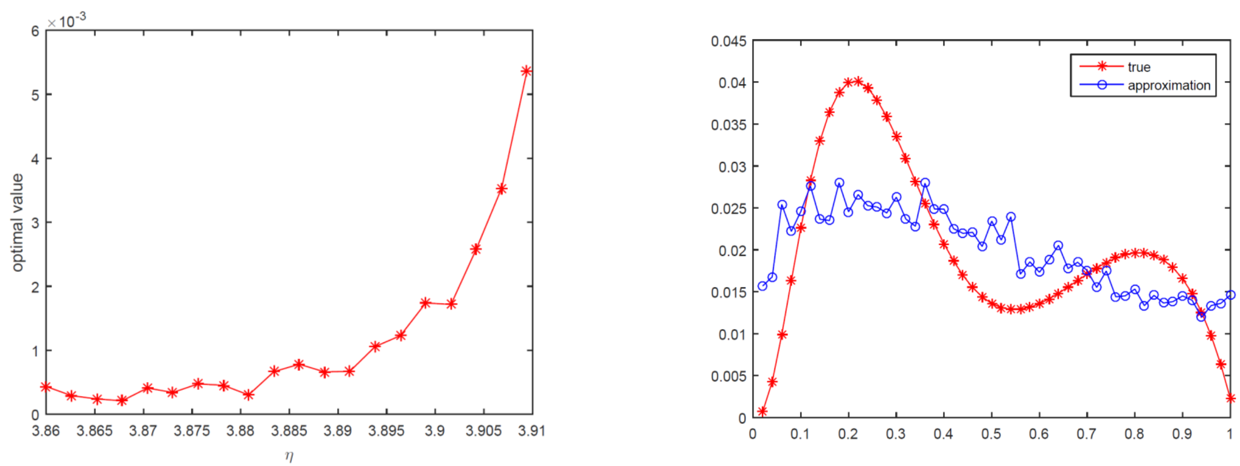


Numerical Illustration

Monotone distribution



Bimodal distribution



Discussion

Motivation: Model misspecification/uncertainty

Main Approach:

- Sensitivity/asymptotic analysis of robust optimization programs over input stochastic models
- Simulation optimization

Future Work:

- Other forms of stochastic/system uncertainty
- Relation to Bayesian methodology