Robust Confidence Bands for Stochastic Processes Using Simulation

Timothy C. Y. Chan, Jangwon Park*, Vahid Sarhangian

Department of Mechanical and Industrial Engineering, University of Toronto, Canada

Abstract

We propose a robust optimization approach for constructing confidence bands for stochastic processes using a finite number of simulated sample paths. Our methodology addresses optimization bias within the constraints, avoiding overly narrow confidence bands of existing methods. In our first case study, we show that our approach achieves the desired coverage probabilities with an order-of-magnitude fewer sample paths than the state-of-the-art baseline approach. In our second case study, we illustrate how our approach can validate stochastic simulation models.

Keywords: Stochastic simulation, confidence bands, validation, uncertainty quantification, robust optimization

1. Introduction

Stochastic simulation [1, 2] is a primary tool for performance evaluation of stochastic dynamical systems, in particular under counter-factual or "what-if" scenarios. In many applications, simulation outputs are sample paths of stochastic processes realized over a finite horizon. For instance, these sample paths may correspond to hospital occupancy levels [3, 4], the price of financial products [5], or the number of infected patients in an infectious disease model [6]. A natural way to validate a simulation model in this setting is to construct a *confidence band* over the sample paths at a specified level of coverage probability, and check whether the historical paths from the actual system are "covered" by the confidence band (see Section 2 for the formal definition).

In addition to validation, confidence bands can also be used to quantify the uncertainty in realizations of a stochastic process over a finite horizon. This has received much attention in the context of estimating impulse response functions in vector autoregressive (VAR) models in the Economics literature [e.g., 7, 8, 9, 10, 11, 12]. Confidence bands on impulse response functions are commonly estimated to examine the effects of shocks to a system over time. However, the majority of approaches are either heuristics or asymptotic methods, which may provide an overly wide confidence band when the number of observations (sample paths) is small. Moreover, these approaches are not universally applicable given their specific application to VAR. The best known approach that is most relevant to our work is [12], who propose a mixed-integer program (MIP) to construct minimum-width confidence bands. As we show in this work, however, this approach may produce a biased confidence band, i.e., one with a smaller coverage probability than desired. Although generating more sample paths reduces this bias, the number of sample paths required to achieve a sufficiently small bias may result in a large MIP that is not solvable to optimality in practical time, as noted in [12].

Related methods for specific stochastic processes include [13], who formulate an optimization problem based on local time arguments for constructing confidence bands on Brownian motion and perturbed Brownian paths. While the approach is relevant when considering Brownian approximations, it is not applicable to general simulation output that may not be well-approximated by Brownian motion. On the other hand, [14] and [15] propose various heuristics and an MIP approach, respectively, which are applicable for stochastic processes such as stock market data, temperature data, and medical data (e.g., heartbeat). However, these works pursue an alternative definition of coverage, which is arguably less practical and less amenable to theoretical analysis; see also the discussion in Section 2. Therefore, we do not consider it here. Furthermore, they do not directly address the potential optimization bias that may lead to an overly narrow confidence band.

Also related is the literature on conformal prediction [e.g., 16, 17, 18, 19, 20, 21] and prediction intervals for metamodels [e.g., 22]. In this context, the aim is to construct prediction intervals that capture both the intrinsic (aleatoric) and extrinsic (epistemic) uncertainty, arising from model mis-specification, limited data, and the inherent stochasticity of the model. As such, these methods typically involve training and validation phases to compute the residual errors of predictors to calibrate the width of the intervals. Our work is different because we focus only on the intrinsic uncertainty and more data can be generated at will (i.e., by simulation). Thus, we cast the task of constructing intervals as an empirical constrained optimization problem where we minimize the width of the intervals directly, subject to a desired coverage probability over a set of generated sample paths. [22] solves a similar problem for simulation metamodeling, whose solution is the prediction interval for the response surface $\mathbb{E}[Y(x)]$, where x is an input parameter

*Corresponding author at: Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario, Canada M5S 3G8.

Email address: `jangwon.park@mail.utoronto.ca` (Jangwon Park)

(e.g., time) and $Y(x)$ is a random output from simulation (e.g., number of customers). The main difference in our work is the ability to solve the empirical constrained optimization problem directly, rather than an approximation of it, due to our specific focus on (discrete-time) stochastic processes and the intrinsic uncertainty. Additionally, while [22] adds a positive constant to the coverage probability to avoid overly narrow intervals, we add a positive constant to the total width of the intervals instead, which allows more flexibility since one can increase the width without affecting the coverage probability, but not vice versa.

In this paper, we propose a novel approach for constructing confidence bands for general discrete-time stochastic processes using a finite number of simulated sample paths. Unlike existing approaches in the literature, our methodology is widely applicable and directly addresses optimization bias through a robust optimization approach. It is tractable, being only slightly more complex than the state-of-the-art baseline approach, and easy to use, as it employs standard techniques. Additionally, our approach is also applicable to continuous-time processes after appropriately discretizing time. In our first case study, we demonstrate that our confidence bands achieve the desired coverage probabilities with an order-of-magnitude fewer sample paths than the best known approach for the same problem. In the second case study, we illustrate how our method can be used to validate stochastic simulation models. The data and code for our numerical experiments are available on GitHub[1].

2. Preliminaries

Let $\{X_t; t \geq 0\}$ denote the (discrete-time) stochastic process of interest and let $X \equiv (X_1, \ldots, X_H)$ denote the finite-dimensional random variable corresponding to the value of the process at times $\{1, \ldots, H\}$, taking values in $\mathbb{R}^H$. The path X could, for example, represent the output of a simulation model we are interested in validating, or an impulse response function. Denote a confidence band by (l, u) where $l, u \in \mathbb{R}^H$. We formally define "coverage" as follows.

Definition 1 (Coverage). We say $x \equiv (x_1, \ldots, x_H) \in \mathbb{R}^H$ is *covered* by (l, u) if $l_t \leq x_t \leq u_t$ for all $t = 1, \ldots, H$.

Definition 1 is a natural definition for coverage. We do not pursue the alternative definition in [14] and [15], which states that a sample path is covered if it does not lie outside the confidence band by more than l times (where l is user-specified). Although more general, it is less practical since it is not possible to know at *which* time steps a given sample path will lie inside the confidence band. Moreover, computing its coverage probability requires considering all possible combinations of $s \in \{0, \ldots, l\}$ time steps that lie outside the band, which is less conducive to theoretical analysis, especially when coverage at different time steps is correlated.

Let $C_t = \{l_t \leq X_t \leq u_t\}$ be the event that X is covered by the given confidence band (l, u) at time t. We make the following assumption on C_t.

Assumption 1. For any t' and t such that $t' > t$, $P(C_{t'}|C_t) < 1$.

Assumption 1 is not restrictive. It states that there is uncertainty in coverage over time so that if X is covered at time t, this does not guarantee coverage at a future time. In the rest of the paper, for ease of exposition, we use the shorthand notation $P(l, u) = P(C_1, \ldots, C_H)$ to denote the coverage probability of (l, u). We omit its dependence on H, as it should be clear from context.

When $H = 1$, the confidence band is simply the interval enclosed between the $(1 - \alpha/2)$- and $(\alpha/2)$-quantiles, where $1 - \alpha \in [0, 1]$ is the specified coverage rate. When $H > 1$, one may consider the "naive confidence band" $(\bar{l}, \bar{u})$ by extending the $H = 1$ case as follows: first estimate the $(1 - \alpha/2)$- and $(\alpha/2)$-quantiles at each time step and connect the upper and lower points of the adjacent intervals. Although intuitive and often used in applied work, this solution provides a smaller coverage probability than desired, as formalized below.

Proposition 1. *Let $1 - \alpha$ be the desired coverage rate. Under Assumption 1, the naive confidence band under-covers, i.e., $P(\bar{l}, \bar{u}) < 1 - \alpha$.*

Proof. The proof is by induction on H, the length of the horizon. Let C_t be the event that X is covered by $(\bar{l}, \bar{u})$ at time t. Suppose $H = 2$. We observe that

$$P(\bar{l}, \bar{u}) = P(C_1, C_2) = P(C_2|C_1)P(C_1) = P(C_2|C_1)(1 - \alpha) < 1 - \alpha,$$

where the last inequality holds by Assumption 1. Now, suppose the same holds for $H = n$. Then for $H = n + 1$, we observe that

$$\begin{aligned} P(\bar{l}, \bar{u}) = P(C_1, \ldots, C_{n+1}) &= P(C_{n+1}|C_1, \ldots, C_n)P(C_1, \ldots, C_n) \\ &< P(C_{n+1}|C_1, \ldots, C_n)(1 - \alpha) \\ &< 1 - \alpha, \end{aligned}$$

where the first inequality is by the induction hypothesis and the second inequality is again by Assumption 1. □

Proposition 1 states that in the presence of uncertainty (Assumption 1), the naive solution $(\bar{l}, \bar{u})$ fails to provide the desired coverage probability, thus warranting a more rigorous approach.

3. Baseline: nominal MIP

We first present the model in [12], which we refer to as the *nominal problem* and use as the baseline model. Suppose we have generated n iid sample paths. Let $x^i \in \mathbb{R}^H$ be the ith sample path and denote by $1 - \alpha$ the desired coverage rate. Let q_t^u and q_t^l represent the $(1 - \alpha/2)$- and $(\alpha/2)$-quantile estimates at time t based on the n sample paths. Let $\delta(l, u, x^i)$ denote an indicator function that evaluates to 1 if and only if x^i is covered by the confidence band (l, u), where $l = (l_1, \ldots, l_H)$ and $u = (u_1, \ldots, u_H)$. The nominal problem finds a minimum-width confidence band such that it covers at least $(1 - \alpha) \times 100\%$ of

[1] https://github.com/parkjan4/RobustConfidenceBands

the sample paths:

$$
\begin{aligned}
w^\star = \min_{u,l} \quad & \sum_{t=1}^{H} (u_t - l_t) \\
\text{s.t.} \quad & u_t \geq q_t^u, \quad \forall t, \\
& l_t \leq q_t^l, \quad \forall t, \\
& \frac{1}{n}\sum_{i=1}^{n} \delta(l, u, x^i) \geq 1 - \alpha.
\end{aligned}
\tag{NP}
$$

We note that (NP) can be reformulated as a MIP using n binary variables and an appropriate large constant as shown in [12]. Hereafter, we refer to an optimal solution of (NP) as the *nominal confidence band*. For convenience, we focus on the case where $n(1-\alpha)$ is an integer. If it is not an integer, the constraint $(1/n)\sum_{i=1}^{n} \delta_i \geq 1 - \alpha$ holds as a strict inequality, and therefore, one may consider the smallest $\beta \in (\alpha, 1]$ such that $n(1-\beta)$ is an integer.

In the rest of the section, we analyze the coverage probability of the nominal confidence band. To this end, we first show that a nominal confidence band satisfies a coverage rate of exactly $1 - \alpha$ over the sample paths used to construct it.

Lemma 1. *Let* $(\hat{l}, \hat{u})$ *be an optimal solution to* (NP). *Then* $\frac{1}{n}\sum_{i=1}^{n} \delta(\hat{l}, \hat{u}, x^i) = 1 - \alpha$.

Proof. Suppose for the sake of contradiction $\frac{1}{n}\sum_{i=1}^{n} \delta(\hat{l}, \hat{u}, x^i) > 1-\alpha$, i.e., the nominal confidence band covers strictly more than $n(1-\alpha)$ sample paths. The constraints $u_t \geq q_t^u$ and $l_t \leq q_t^l$ cannot be binding for all t, since by Proposition 1, such a confidence band under-covers and is therefore infeasible in (NP). So, let t' be some time where $u_{t'} > q_{t'}^u$ (or $l_{t'} < q_{t'}^l$). Then we can lower $u_{t'}$ (or raise $l_{t'}$) until a new sample path is crossed from above (from below) while still satisfying $u_{t'} \geq q_{t'}^u$ ($l_{t'} \leq q_{t'}^l$). The resulting solution is still feasible but strictly better since the total width is reduced, contradicting the optimality of $(\hat{l}, \hat{u})$. □

The key result we establish is that although a nominal confidence band achieves a coverage rate of $1 - \alpha$ on the samples used in (NP) (as stated in Lemma 1), its true coverage probability may be less than $1-\alpha$. To show this, analyzing (NP) directly is inconvenient because the coverage probability appears as a constraint, not an objective function. Therefore, we introduce a related MIP where the objective and constraint are swapped.

Lemma 2. *Consider the problem,*

$$
\begin{aligned}
c^\star = \max_{u,l} \quad & \frac{1}{n}\sum_{i=1}^{n} \delta(l, u, x^i) \\
s.t. \quad & u_t \geq q_t^u, \quad \forall t, \\
& l_t \leq q_t^l, \quad \forall t, \\
& \sum_{t=1}^{H} (u_t - l_t) \leq w^\star.
\end{aligned}
\tag{1}
$$

Then $c^\star = 1 - \alpha$.

Proof. First, note that there exists an optimal solution to (1) whose width is exactly equal to $w^\star$. Otherwise, one can simply increase its width towards $w^\star$, and since the objective function is non-decreasing in the total width of the confidence band, the resulting solution can only be better. This solution is then optimal in (NP), and by Lemma 1, $c^\star = 1 - \alpha$. □

We now establish the following result about the nominal confidence band.

Theorem 3. *An optimal solution to* (NP) *is biased towards under-coverage, i.e.,* $P(\hat{l}, \hat{u}) \leq 1 - \alpha$.

Proof. Denote by $(\hat{l}, \hat{u})$ an optimal solution to (NP). Let $\mathcal{W} \equiv \left\{(l, u) : u \geq q^u, l \leq q^l, \sum_{t=1}^{H}(u_t - l_t) \leq w^\star\right\}$ where $w^\star$ is the width of $(\hat{l}, \hat{u})$. Then,

$$
\begin{aligned}
P(\hat{l}, \hat{u}) = \mathbb{E}[\delta(\hat{l}, \hat{u}, X)] \leq \max_{(l,u)\in\mathcal{W}} \mathbb{E}[\delta(l, u, X)] & \\
= \max_{(l,u)\in\mathcal{W}} \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \delta(l, u, x^i)\right] & \\
\leq \mathbb{E}\left[\max_{(l,u)\in\mathcal{W}} \frac{1}{n}\sum_{i=1}^{n} \delta(l, u, x^i)\right] & \\
= 1 - \alpha. &
\end{aligned}
$$

The first inequality holds since $(\hat{l}, \hat{u}) \in \mathcal{W}$. The last equality holds by Lemma 2. □

Intuitively, the nominal problem "over-optimizes" with respect to the given sample and produces an optimistic, or overly narrow, confidence band. This phenomenon is observed in other stochastic optimization contexts and referred to as *optimization bias* [23] or the *optimizer's curse* [24] in the literature. Although this bias diminishes as the number of sample paths increases, the required size of the MIP to achieve a sufficiently small bias may be too large to solve to optimality in practical time. This motivates the development of a new approach to generate confidence bands that can achieve the desired coverage probabilities even with relatively small numbers of samples.

4. Robust MIP

So far, we established that an optimal solution to the nominal problem may provide a smaller coverage probability than expected because it is too narrow. To protect against this phenomenon, we take a robust optimization perspective. Specifically, we introduce the following budget uncertainty set [25], where the parameter Γ takes values in $[0, 1]$:

$$
\mathcal{Z}(\Gamma) \equiv \left\{ z \in [0, 1]^H : \frac{1}{H}\sum_{t=1}^{H} z_t \leq \Gamma \right\} \tag{2}
$$

Intuitively, for any $z \in \mathcal{Z}(\Gamma)$, at most $\Gamma \times 100\%$ of its components can be set to 1. We then introduce the following constraints to (NP):

$$\sum_{t=1}^{H} u_t \geq \sum_{t=1}^{H} \left(q_t^u + c_t^u z_t^u\right), \quad \forall z^u \in \mathcal{Z}(\Gamma), \tag{3}$$

$$\sum_{t=1}^{H} l_t \leq \sum_{t=1}^{H} \left(q_t^l - c_t^l z_t^l\right), \quad \forall z^l \in \mathcal{Z}(\Gamma), \tag{4}$$

where c_t^u and c_t^l are positive constants such that $q_t^u + c_t^u$ and $q_t^l - c_t^l$ serve as upper and lower bounds on x_t^i, respectively. These constraints help protect against under-coverage by forcing $\sum_{t=1}^{H} u_t$ to be larger than $\sum_{t=1}^{H} q_t^u$ (and $\sum_{t=1}^{H} l_t$ to be smaller than $\sum_{t=1}^{H} q_t^l$) by the extent allowed by Γ, thereby increasing the total width of the confidence band. Thus, the *robust problem* is formulated as follows:

$$\begin{aligned}
\min_{u,l} \quad & \sum_{t=1}^{H} (u_t - l_t) \\
\text{s.t.} \quad & u_t \geq q_t^u, \quad \forall t, \\
& l_t \leq q_t^l, \quad \forall t, \\
& \sum_{t=1}^{H} u_t \geq \sum_{t=1}^{H} \left(q_t^u + c_t^u z_t^u\right), \quad \forall z^u \in \mathcal{Z}(\Gamma), \\
& \sum_{t=1}^{H} l_t \leq \sum_{t=1}^{H} \left(q_t^l - c_t^l z_t^l\right), \quad \forall z^l \in \mathcal{Z}(\Gamma), \\
& \frac{1}{n} \sum_{i=1}^{n} \delta(l, u, x^i) \geq 1 - \alpha,
\end{aligned} \tag{RP}$$

where the definition of $\delta(l, u, x^i)$ can again be enforced using n binary variables and an appropriate large constant. Denote by (l^Γ, u^Γ) a confidence band obtained by solving the robust problem with a given Γ. We refer to this as a *robust confidence band*.

Remark 1. If $\Gamma = 0$, the problems (NP) and (RP) are equivalent. This implies $P(l^0, u^0) \leq 1 - \alpha$ by Theorem 3.

Ideally, if $\Gamma = 1$, the robust confidence band (l^1, u^1) should provide an upper bound on the coverage probability so that through appropriate tuning of Γ, we can obtain an unbiased confidence band. To ensure $\Gamma = 1$ leads to such an *over-covering* confidence band, we make the following assumption on the parameters c_t^u and c_t^l.

Assumption 2. For all t, $P(q_t^l - c_t^l \leq X_t \leq q_t^u + c_t^u) \geq 1 - \alpha$.

We note that this assumption is not practically restrictive, as in most applications one can set natural upper and lower bounds for the sample paths. For example, for hospital occupancy, $q_t^u + c_t^u$ can be set to the hospital capacity while $q_t^l - c_t^l = 0$ for all t, in which case $P(q_t^l - c_t^l \leq X_t \leq q_t^u + c_t^u) = 1$. In other applications where the sample paths are technically unbounded, e.g., financial prices, the upper (lower) bounds can be set sufficiently large (small) to ensure Assumption 2 holds.

4.1. Reformulation

In this section, we present an equivalent formulation for the robust problem (RP) that accounts for the infinitely many constraints introduced by (3) and (4).

Denote by $c_{(t)}^u$ the tth *largest* parameter, i.e., the tth value from the left of the ordered parameters $c_{(1)}^u > c_{(2)}^u > \cdots > c_{(H)}^u$. By perturbing each parameter by a small amount as needed, the strict inequality holds without loss of generality. We define $c_{(t)}^l$ analogously. We will use $\lceil \cdot \rceil$ to indicate the ceiling operator, which rounds up the expression to the nearest integer. Define $t^\star \equiv \max\{\lceil \Gamma H \rceil, 1\}$ and

$$\beta_t^u(\Gamma) \equiv \left(c_t^u - c_{(t^\star)}^u\right)^+ + \Gamma c_{(t^\star)}^u, \quad \forall t, \tag{5}$$

$$\beta_t^l(\Gamma) \equiv \left(c_t^l - c_{(t^\star)}^l\right)^+ + \Gamma c_{(t^\star)}^l, \quad \forall t. \tag{6}$$

If $\Gamma = 0$, then $t^\star = 1$ and we have $\beta_t^u(\Gamma) = \beta_t^l(\Gamma) = 0$ for all t, which will make (7) and (8) redundant in the reformulation. If $\Gamma = 1$, then $t^\star = H$ and we have $\beta_t^u(\Gamma) = c_t^u$ and $\beta_t^l(\Gamma) = c_t^l$ for all t, which will widen the total width of the confidence band to the maximal extent. We present the reformulation below, where we replace constraints (3) and (4) with

$$\sum_{t=1}^{H} u_t \geq \sum_{t=1}^{H} \left(q_t^u + \beta_t^u(\Gamma)\right), \tag{7}$$

$$\sum_{t=1}^{H} l_t \geq \sum_{t=1}^{H} \left(q_t^l - \beta_t^l(\Gamma)\right), \tag{8}$$

respectively:

$$\begin{aligned}
\min_{u,l} \quad & \sum_{t=1}^{H} (u_t - l_t) \\
\text{s.t.} \quad & u_t \geq q_t^u, \quad \forall t, \\
& l_t \leq q_t^l, \quad \forall t, \\
& \sum_{t=1}^{H} u_t \geq \sum_{t=1}^{H} \left(q_t^u + \beta_t^u(\Gamma)\right), \\
& \sum_{t=1}^{H} l_t \leq \sum_{t=1}^{H} \left(q_t^l - \beta_t^l(\Gamma)\right), \\
& \frac{1}{n} \sum_{i=1}^{n} \delta(l, u, x^i) \geq 1 - \alpha.
\end{aligned} \tag{RP'}$$

Proposition 2. *Problems* (RP) *and* (RP') *are equivalent.*

Proof. We prove the equivalence by showing that constraint (3) may be replaced by (7) without loss of optimality. The same can be shown very similarly between constraints (4) and (8), and we omit its proof for brevity.

Starting from (3), by the theorem of the alternative with appropriate primal and dual linear programs, we replace it with the following system of linear inequalities:

$$\sum_{t=1}^{H} u_t \geq \sum_{t=1}^{H} \left(q_t^u + \lambda_t^u + \Gamma \rho^u\right) \tag{9}$$

$$\lambda_t^u + \rho^u \geq c_t^u, \quad \forall t \tag{10}$$

$$\lambda^u, \rho^u \geq 0. \tag{11}$$

We further simplify these constraints by deducing the optimal values of λ_t^u and ρ^u. Recall that $t^\star \equiv \max\{\lceil \Gamma H \rceil, 1\}$ and $c^u_{(t)}$ represents the tth largest parameter among c_t^u. Suppose we enforce $\rho^u = c^u_{(t^\star)}$ as a constraint. We will show that this is not restrictive, i.e., does not harm the optimal objective value.

With $\rho^u = c^u_{(t^\star)}$, we must have at optimality:

$$\lambda^u_{(t)} = \left(c^u_{(t)} - c^u_{(t^\star)}\right)^+ = \begin{cases} c^u_{(t)} - c^u_{(t^\star)}, & \text{if } t < t^\star, \\ 0, & \text{if } t \geq t^\star. \end{cases}$$

Let $\epsilon > 0$ and consider the following two cases to observe their impact on the objective value:

Case 1: Increase ρ^u by ϵ. Then for small enough ϵ, we must have $\lambda^u_{(t)} = c^u_{(t)} - c^u_{(t^\star)} - \epsilon$ at optimality for all $t < t^\star$; $\lambda^u_{(t)} = 0$ otherwise. The change in the objective value is

$$\Delta = \Gamma H \epsilon - (t^\star - 1)\epsilon > 0,$$

since $\Gamma H > t^\star - 1$. This indicates that the new objective value is worse.

Case 2: Decrease ρ^u by ϵ. Then for small enough ϵ, we must have $\lambda^u_{(t)} = c^u_{(t)} - c^u_{(t^\star)} + \epsilon$ at optimality for $t \leq t^\star$; $\lambda^u_{(t)} = 0$ otherwise. The change in the objective value is

$$\Delta = t^\star \epsilon - \Gamma H \epsilon \geq 0,$$

since $t^\star \geq \Gamma H$. This indicates that the new objective value is not better.

In either case, we cannot strictly improve the objective value. Therefore, there exists an optimal solution where $\rho^u = c^u_{(t^\star)}$ and $\lambda^u_t = (c^u_t - c^u_{(t^\star)})^+$ for all t. Letting $\beta^u_t(\Gamma) = \lambda^u_t + \Gamma \rho^u$ for all t concludes the proof. $\square$

Because (RP) and (RP') are equivalent, Proposition 2 shows that the robust problem is only slightly more complex than the nominal problem, with just two more constraints and no additional decision variables.

4.2. *Tuning* Γ

By solving (RP') for different values of Γ, we can obtain confidence bands of varying widths, starting from an under-covering solution at $\Gamma = 0$ to an over-covering one at $\Gamma = 1$ (Assumption 2). In this section, we present an algorithm to tune Γ.

We cast the task of tuning Γ as a root-finding problem. Let $f(\Gamma) = P(l^\Gamma, u^\Gamma) - (1-\alpha)$. Then Theorem 3 and Assumption 2 imply $f(0) \leq 0$ and $f(1) \geq 0$, respectively. Furthermore, we note that $P(l^\Gamma, u^\Gamma)$ is continuous and increasing in Γ, and consequently, there must exist $\Gamma^\star \in [0, 1]$ such that $f(\Gamma^\star) = 0$. Therefore, the bisection method on $[0, 1]$ is guaranteed to converge to $\Gamma^\star$.

In practice, we can only estimate $f(\Gamma)$ using finitely many sample paths since $P(l^\Gamma, u^\Gamma)$ is unknown. Therefore, an estimator $\hat{f}(\Gamma)$ may have jump discontinuities in Γ, possibly implying $\hat{f}(\Gamma^\star) \neq 0$. Below, we propose a bisection method with K-fold cross-validation to estimate $\Gamma^\star$ using an estimator of $f(\Gamma)$. In Section 5, we demonstrate that despite the said difficulties, our approach still converges to a value that produces robust confidence bands whose coverage probabilities are close to the desired values.

Algorithm 1: Bisection method with K-fold cross-validation for estimating $\Gamma^\star$

Input: n sample paths $x^1, \ldots, x^n$, coverage rate $1-\alpha$, number of iterations $\bar{N}$, number of folds K

Initialize: $\Gamma_a = 0$, $\Gamma_b = 1$, $N = 0$, K random partitions of n sample paths, $\mathcal{P}_1, \ldots, \mathcal{P}_K$, each containing $m \equiv n/K$ paths (assume n is divisible by K).
while $N < \bar{N}$ **do**
    $N = N + 1$
    $\hat{\Gamma} = (\Gamma_a + \Gamma_b)/2$
    for $k = 1, \ldots, K$ **do**
        Solve (RP') with $\hat{\Gamma}$ and all partitions except $\mathcal{P}_k$ to obtain $(l^{\hat{\Gamma}}, u^{\hat{\Gamma}})$.
        $C_k = \frac{1}{m}\sum_{j:x^j \in \mathcal{P}_k} \delta(l^{\hat{\Gamma}}, u^{\hat{\Gamma}}, x^j)$, i.e., coverage rate over the kth partition $\mathcal{P}_k$
    end for
    $\hat{f}(\hat{\Gamma}) = \frac{1}{K}\sum_{k=1}^{K} C_k - (1-\alpha)$
    if $\hat{f}(\hat{\Gamma}) < 0$ **do**
        $\Gamma_a = \hat{\Gamma}$
    else
        $\Gamma_b = \hat{\Gamma}$
    end if
end while
Output: $\hat{\Gamma}$

Algorithm 1 requires solving (RP') $\bar{N}K$ times in total to estimate $\Gamma^\star$, where $\bar{N}$ is the maximum number of iterations. Upon estimating $\Gamma^\star$, we solve the robust problem once more to obtain the confidence band. For the experiments in the next section, the total run time of this procedure is only a few seconds when $n \leq 500$ at the 1% optimality gap criterion, making it a practical approach. Nevertheless, for a potential speed-up, we observe the following.

Remark 2. The feasible set of (RP') is decreasing in Γ.

This implies that for a fixed partition of n sample paths, the confidence band obtained by solving (RP') using Γ_2 is a feasible solution for (RP') that uses Γ_1 for any $\Gamma_2 \geq \Gamma_1$. Therefore, solutions can be stored and used as feasible warm-starts throughout Algorithm 1 wherever applicable.

5. Case study: estimating a confidence band for a vector autoregressive (VAR) model

VAR models are commonly used in Economics and the natural sciences to examine impulse responses, which measure the impact of an external shock to one variable on others over time. To quantify the uncertainty in these responses, a confidence band must be estimated. In this section, we apply our methodology to the same two-dimensional VAR(1) example used in

[10] and [12] to illustrate the advantage of our approach in accurately estimating a confidence band. Consider

$$x_t = A_0 + A_1 x_{t-1} + \epsilon_t, \tag{12}$$

where

$$A_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 0.5 & 0.3 \\ -0.6 & 1.3 \end{bmatrix}, \quad \Sigma_\epsilon = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

We simulate n sample paths, for different values of n, over 12 periods for the first variable of the VAR(1) process. We use both the baseline approach and our proposed methodology to obtain two confidence bands at $\alpha = 0.1$, which we refer to as the nominal and robust confidence bands. Same as [12], we construct four random sets #1, ..., #4, each with 1,000 sample paths from the VAR(1) process with the same number of periods to estimate the coverage probability. In Table 1, we summarize the coverage probability on each of the four sets by the two confidence bands for $n \in \{100, 200, 500, 1000, 5000\}$.

Our approach clearly produces higher quality confidence bands whose estimated coverage probabilities are much closer to 90%, especially with limited samples (e.g., $n = 100$). At $n = 200$, the robust confidence band already achieves an average coverage probability of approximately 90%, whereas the nominal confidence band does not, even with 5,000 sample paths, indicating an order-of-magnitude improvement in the required sample size. Moreover, the nominal confidence band is clearly biased, as its coverage probabilities are always smaller than 90%, whereas the coverage probabilities for the robust confidence band hover around 90%, either slightly above or slightly below. The nominal confidence band can achieve a small enough bias with 10,000 sample paths, as noted in [12], but the resulting MIP can take several hours to solve to optimality; even with a more relaxed optimality gap criterion, generating so many sample paths may also be very time-consuming for complex simulation models. In contrast, with a 1% optimality gap criterion, the entire procedure of estimating $\Gamma^\star$ and constructing the robust solution takes roughly two and five seconds with $n = 200$ and $n = 500$, respectively, without using warm-starts.

The differences in the nominal and robust confidence bands are illustrated in Figure 1 for 500 and 5,000 sample paths. With a relatively small n, the nominal confidence band is considerably narrower than the robust one, as it is significantly biased. As n increases, the difference between the two solutions diminishes, and with $n = 5,000$, they are very similar.

6. Case study: validating a queueing model of patient flow during a Mass Casualty Event (MCE)

We now demonstrate how our methodology can be used to validate a queueing model of patient flow. We consider the Erlang-R queue in [26], which models a time-varying queue with reentrant customers who can return multiple times during their sojourn within the system. The Erlang-R queue has diverse applications with various extensions studied in the literature, see, e.g., [27, 28].

The example we consider is a hospital emergency ward during an MCE. We use the same data and parameters from [26], which describe a chemical MCE drill that took place in July 2010 at 11:00 and lasted until 13:15. Figure 2 shows the actual cumulative arrivals and departures of patients at the emergency department (left) and the estimated non-stationary arrival rate function (right). Their model specified four servers. The average treatment time was 5.4 minutes ($\mu = 11.06$), average time until readmission 24.6 minutes ($\delta = 2.44$), and the probability of readmission $p = 0.662$. [26] use confidence bands from a diffusion approximation of the Erlang-R model to assess the validity of their model. Here, we use a simulation model of the Erlang-R queue to construct the confidence bands.

We first generate 300 sample paths from the Erlang-R queue. We discretize the time horizon into $H = 30$ equidistant intervals and extract the value of each sample path at the start of each interval. Using this data, we construct the confidence bands by running Algorithm 1 with $K = 3$ and $\bar{N} = 10$ and solve (RP') with the obtained estimate of $\Gamma^\star$. In Figure 3a, we present confidence bands with $\alpha = 0.5$ and 0.05, both of which cover the actual sample path, thus supporting the validity of the specified queueing model in [26]. In contrast, the diffusion approximation-based confidence bands shown in [26] do not fully cover the sample path.

Lastly, we apply our methodology to a simpler model that assumes a stationary arrival process with mean equal to the average arrival rate indicated in Figure 2b. Figure 3b presents the confidence bands with $\alpha = 0.5$ and 0.05. The results clearly suggest that there is model mis-specification when assuming stationary arrivals, since neither band fully contains the actual sample path and they fail to capture the time-varying nature of the process.

7. Conclusion

We present a methodology for constructing accurate confidence bands for general discrete-time stochastic processes using a finite and small number of simulated sample paths. Importantly, our methodology can achieve better performance with significantly less data than existing approaches. Our approach improves upon existing methods by addressing optimization bias directly in the constraints, thus producing confidence bands that achieve the desired coverage probability without having to simulate many samples. We demonstrate the effectiveness of our methodology on two case studies from the literature, a vector autoregressive model and a queuing model. Note that our approach is also applicable to continuous-time processes after appropriately discretizing time.

References

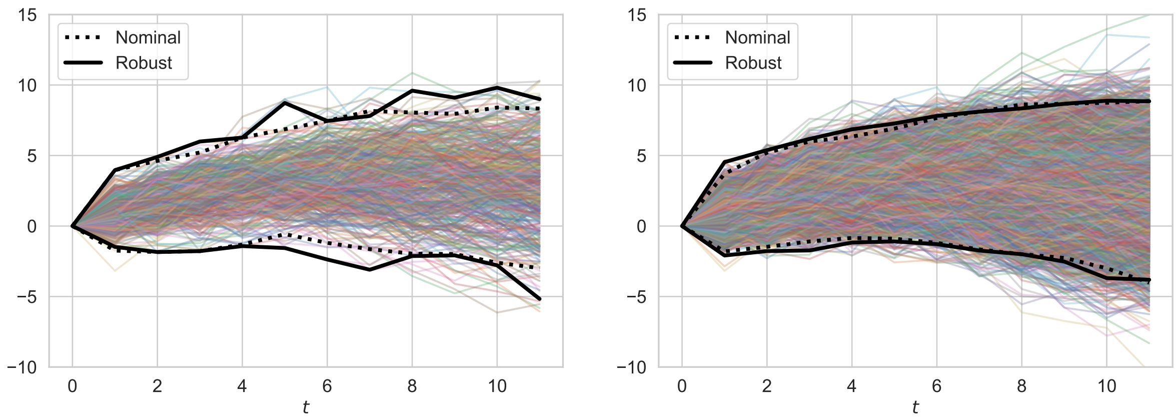
[1] S. Asmussen, P. W. Glynn, Stochastic simulation: algorithms and analysis, Vol. 57, Springer, 2007.

[2] B. Nelson, L. Pei, Foundations and methods of stochastic simulation, Springer, 2021.

[3] T. Chan, J. Park, F. Pogacar, V. Sarhangian, E. Hellsten, F. Razak, A. Verma, Optimizing inter-hospital patient transfer decisions during a

n	Nominal confidence band					Robust confidence band				
	#1	#2	#3	#4	Avg.	#1	#2	#3	#4	Avg.
100	63.8%	63.1%	65.4%	65.5%	64.4%	93.1%	92.0%	91.4%	90.3%	91.7%
200	74.2%	74.6%	76.1%	77.1%	75.5%	90.1%	90.7%	90.2%	88.5%	89.9%
500	83.5%	80.9%	84.7%	83.3%	83.1%	91.0%	91.0%	89.2%	91.2%	90.6%
1,000	85.0%	83.5%	86.5%	84.6%	84.9%	90.9%	88.7%	88.9%	90.8%	89.8%
5,000	87.2%	86.9%	86.5%	89.6%	87.8%	90.4%	89.5%	89.4%	90.3%	89.9%

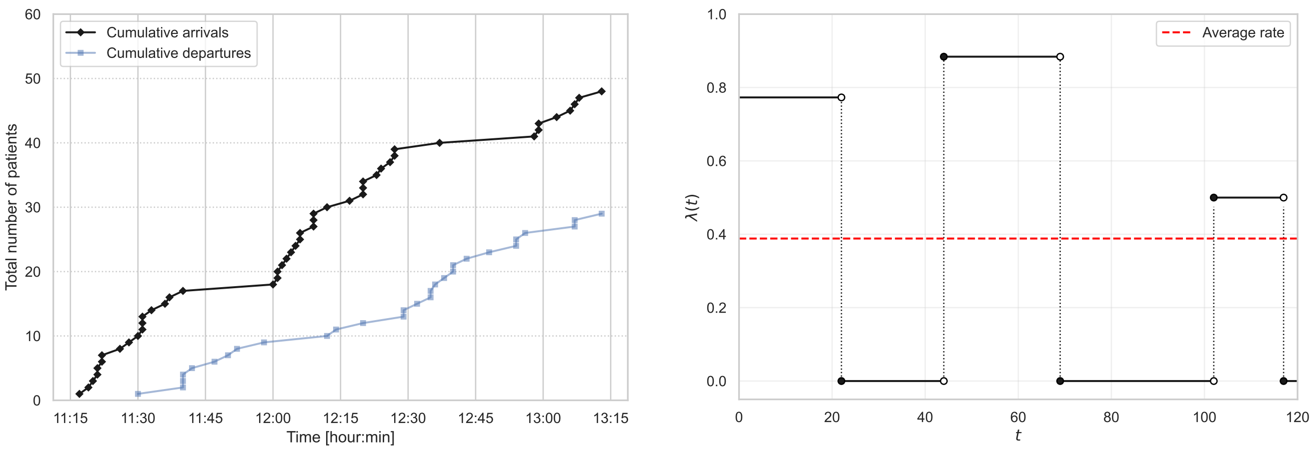
Table 1: Results of the experiment for the first variable of the VAR(1) process with $\alpha = 0.1$. Each set #1, . . . , #4 contains 1,000 sample paths. The table summarizes the coverage probability along with the average over the four sets. We use $K = 2$ in Algorithm 1 for $n = 100$ and $n = 200$ and use $K = 4$ for the rest.



(a) $n = 500$

(b) $n = 5,000$

Figure 1: The nominal and robust confidence bands for the first variable of the VAR(1) process constructed using 500 (left) and 5,000 (right) sample paths.



(a) Actual cumulative arrivals and departures in MCE drill

(b) Estimated arrival rate function (patients per minute): $\lambda(t) = 0.773 \times 1\{0 \le t < 22\} + 0.884 \times 1\{44 \le t < 69\} + 0.5 \times 1\{102 \le t < 117\}$, where 0 corresponds to 11:16 and 120 corresponds to 13:15. The average rate corresponds to 0.388 patients per minute.

Figure 2: Arrival and departure data and the estimated arrival rate function for the MCE.

pandemic: A queueing network approach, Available at SSRN 3975839 (2021).

[4] J. E. Helm, P. Shi, M. Drewes, J. Cecil, Delta coverage: The analytics journey to implement a novel nurse deployment program.

[5] D. Lamberton, B. Lapeyre, Introduction to stochastic calculus applied to finance, Chapman and Hall/CRC, 2011.

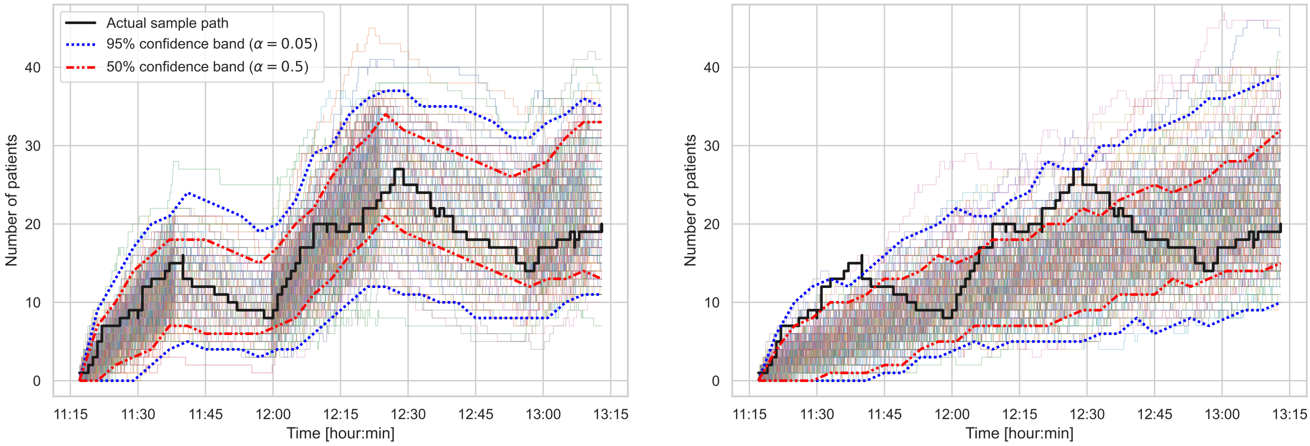
[6] E. Y. Cramer, E. L. Ray, V. K. Lopez, J. Bracher, A. Brennen, A. J. Castro Rivadeneira, A. Gerding, T. Gneiting, K. H. House, Y. Huang, et al., Evaluation of individual and ensemble probabilistic forecasts of covid-19 mortality in the united states, Proceedings of the National Academy of Sciences 119 (15) (2022) e2113561119.

[7] C. A. Sims, T. Zha, Error Bands for Impulse Responses, Econometrica 67 (5) (1999) 1113–1155. `arXiv:2999514`.

[8] A. Staszewska, Representing uncertainty about response paths: The use of heuristic optimisation methods, Computational Statistics & Data Analysis 52 (1) (2007) 121–132.

[9] Ò. Jordà, Simultaneous Confidence Regions for Impulse Responses, The Review of Economics and Statistics 91 (3) (2009) 629–647. `arXiv: 25651363`.

[10] A. Staszewska-Bystrova, P. Winker, Constructing narrowest pathwise



(a) Assuming time-varying arrival rate function $\lambda(t)$ in Figure 2b

(b) Assuming an avarege (stationary) arrival rate of 0.388 patients per minute

Figure 3: Robust confidence band with the actual sample path.

bootstrap prediction bands using threshold accepting, International Journal of Forecasting 29 (2) (2013) 221–233.

[11] H. Lütkepohl, A. Staszewska-Bystrova, P. Winker, Comparison of methods for constructing joint confidence bands for impulse response functions, International Journal of Forecasting 31 (3) (2015) 782–798.

[12] R. Schüssler, M. Trede, Constructing minimum-width confidence bands, Economics Letters 145 (2016) 182–185.

[13] W. S. Kendall, J.-M. Marin, C. P. Robert, Confidence bands for Brownian motion and applications to Monte Carlo simulation, Statistics and Computing 17 (1) (2007) 1–10.

[14] J. Korpela, E. Oikarinen, K. Puolamäki, A. Ukkonen, Multivariate Confidence Intervals, in: Proceedings of the 2017 SIAM International Conference on Data Mining (SDM), Proceedings, Society for Industrial and Applied Mathematics, 2017, pp. 696–704.

[15] J. Berg, E. Oikarinen, M. Järvisalo, K. Puolamäki, Minimum-Width Confidence Bands via Constraint Optimization, in: J. C. Beck (Ed.), Principles and Practice of Constraint Programming, Lecture Notes in Computer Science, Springer International Publishing, Cham, 2017, pp. 443–459.

[16] J. Lei, A. Rinaldo, L. Wasserman, A conformal prediction approach to explore functional data, Annals of Mathematics and Artificial Intelligence 74 (1) (2015) 29–43.

[17] Y. Romano, E. Patterson, E. Candes, Conformalized Quantile Regression, Advances in Neural Information Processing Systems 32 (2019).

[18] K. Stankeviciute, A. M. Alaa, M. van der Schaar, Conformal Time-series Forecasting, in: Advances in Neural Information Processing Systems, Vol. 34, Curran Associates, Inc., 2021, pp. 6216–6228.

[19] M. Zaffran, O. Feron, Y. Goude, J. Josse, A. Dieuleveut, Adaptive Conformal Predictions for Time Series, in: Proceedings of the 39th International Conference on Machine Learning, PMLR, 2022, pp. 25834–25866.

[20] C. Xu, Y. Xie, Conformal Prediction for Time Series, IEEE Transactions on Pattern Analysis and Machine Intelligence 45 (10) (2023) 11575–11587.

[21] A. Auer, M. Gauch, D. Klotz, S. Hochreiter, Conformal Prediction for Time Series with Modern Hopfield Networks, Advances in Neural Information Processing Systems 36 (2023) 56027–56074.

[22] H. Lam, H. Zhang, Prediction Intervals for Simulation Metamodeling (Apr. 2022). `arXiv:2204.01904`.

[23] T. Homem-de Mello, G. Bayraksan, Monte carlo sampling-based methods for stochastic optimization, Surveys in Operations Research and Management Science 19 (1) (2014) 56–85.

[24] J. E. Smith, R. L. Winkler, The optimizer's curse: Skepticism and postdecision surprise in decision analysis, Management Science 52 (3) (2006) 311–322.

[25] D. Bertsimas, M. Sim, The Price of Robustness, Operations Research 52 (1) (2004) 35–53.

[26] G. B. Yom-Tov, A. Mandelbaum, Erlang-R: A Time-Varying Queue with Reentrant Customers, in Support of Healthcare Staffing, Manufacturing & Service Operations Management 16 (2) (2014) 283–299.

[27] C. W. Chan, G. Yom-Tov, G. Escobar, When to use speedup: An examination of service systems with returns, Operations Research 62 (2) (2014) 462–482.

[28] T. C. Chan, S. Y. Huang, V. Sarhangian, Dynamic control of service systems with returns: Application to design of post-discharge hospital readmission prevention programs, Operations Research (2024).