Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

ROBUST CONFIDENCE BANDS FOR STOCHASTIC PROCESSES USING SIMULATION

Jangwon Park

Dept. of Mechanical and Industrial Eng., University of Toronto, Toronto, ON, CANADA

ABSTRACT

In many applications of stochastic simulation, outputs are sample paths of stochastic processes. 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 historical paths from the actual system fall within this band. We propose a robust optimization approach for constructing confidence bands, which, contrary to existing methods, directly addresses optimization bias within the constraints to prevent overly narrow confidence bands. 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.

1 INTRODUCTION

Constructing confidence bands that guarantee a specified coverage probability over sample paths from a general stochastic process is difficult. Existing heuristics or asymptotic methods are not easily generalizable to other problems and often provide overly wide bands, whereas mathematical programming methods suffer from optimization bias, leading to overly narrow bands. In this work, we propose a robust optimization approach that directly addresses the optimization bias within the constraints. Our approach produces confidence bands that achieve the desired coverage probabilities with far fewer sample paths than existing approaches, and is a broadly applicable method for validating stochastic simulation models.

2 METHODOLOGY

Having generated n sample paths $x^1,\ldots,x^n \in \mathbb{R}^H$ where $H < \infty$ is the length of the time horizon, the task of constructing a confidence band at a specified coverage probability can be formulated as a constrained optimization problem whose objective is to minimize the total width of the band, subject to a coverage probability constraint:

$$\begin{aligned}
\min_{l,u\in\mathbb{R}^H} \quad & \sum_{t=1}^{H}(u_t - l_t) \\
\text{subject to} \quad & \frac{1}{n}\sum_{i=1}^{n} 1\{l_t \le x_t^i \le u_t, \forall t\} \ge 1-\alpha, \\
& \sum_{t=1}^{H} u_t \ge \sum_{t=1}^{H}\left(q_t^u + c_t^u z_t^u\right), \quad \forall z^u \in \mathscr{Z}(\Gamma), \\
& \sum_{t=1}^{H} l_t \le \sum_{t=1}^{H}\left(q_t^l - c_t^l z_t^l\right), \quad \forall z^l \in \mathscr{Z}(\Gamma).
\end{aligned} \tag{1}$$

$1-\alpha$ is the specified coverage rate, q_t^u and q_t^l are the $(1-\alpha)$- and α-quantiles on the sample paths at time t respectively, c_t^u and c_t^l are constant parameters, and $\mathscr{Z}(\Gamma) := \left\{z \in [0,1]^H : \frac{1}{H}\sum_{t=1}^{H} z_t \le \Gamma\right\}$ is a budget uncertainty set whose size increases with the parameter $\Gamma \in [0,1]$. While (1) is not immediately solvable due to infinitely many constraints, we can reformulate it as a mixed-integer linear program (MILP).

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: Coverage probabilities evaluated on out-of-sample paths. Closer to 90% is better.

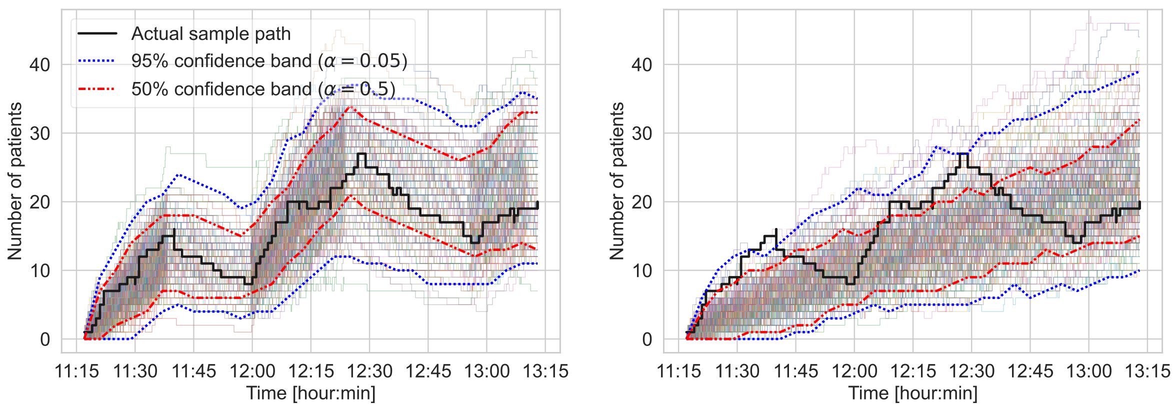

Figure 1: Robust confidence bands with the actual sample path. Left: simulated using a time-varying arrival rate function (correctly specified). Right: simulated using an average (stationary) arrival rate (misspecified).

Varying Γ varies the width of the resulting confidence band. If $\Gamma = 0$, (1) is equivalent to the non-robust MILP in Schüssler and Trede (2016), which we prove produces an overly *narrow* confidence band. On the other hand, setting $\Gamma = 1$ results in an overly *wide* confidence band, with mild assumptions on c_t^u and c_t^l. We cast the task of finding the optimal value $\Gamma^\star$ as a root-finding problem on $[0,1]$. We use a bisection method with K-fold cross-validation to estimate $\Gamma^\star$.

3 CASE STUDIES

In our first case study, we simulate the vector autoregressive (VAR) model from Schüssler and Trede (2016) to illustrate the advantage of our approach. We obtain two confidence band, nominal and robust, at $\alpha = 0.1$ for various sample sizes (n). We construct four random sets #1, ..., #4, each with 1,000 sample paths from the VAR process with $H = 12$ periods to estimate the coverage probability. Results in Table 1 clearly demonstrate that our robust approach produces higher quality confidence bands, whose estimated coverage probability is much closer to 90%, especially with limited samples (e.g., $n = 100, 200$).

In our second case study, we demonstrate that our approach can validate queueing models. We consider the Erlang-R queue in Yom-Tov and Mandelbaum (2014), which models a queue with reentrant patients in a mass casualty event setting with significant time-varying arrivals. Figure 1 shows that with a carefully estimated time-varying arrival rate function, the model is indeed valid as the confidence bands fully enclose the actual sample path (left); however, the simpler model with the average (stationary) arrival rate (right) is clearly misspecified, as it fails to capture the time-varying dynamics of the actual sample path.

REFERENCES

Schüssler, R. and M. Trede. 2016, August. “Constructing Minimum-Width Confidence Bands”. *Economics Letters* 145:182–185.

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