Aggressive Probabilistic Certification of Complex Control Systems

Teodoro Alamo

Dep.Ingeniería de Sistemas y Automática
Universidad de Sevilla

First half-day workshop on
Challenges in Robust and Complex Systems
Torino, 20th October 2022

Collaborators

The contents of this presentation are the result of the collaboration with:

- **Victor Mirasierra**. Departamento de Ingeniería de Sistemas y Automática, Universidad de Sevilla, Spain.
- **Martina Mammarella**. CNR-IEIIT, c/o Politecnico di Torino. Italia.
- **Fabrizio Dabbene**. CNR-IEIIT, c/o Politecnico di Torino. Italia.

Outline of the presentation

- Problem formulation: maximization of a performance index.
- Empirical maximum.
- Generalized maximum.
- Characterizing the critical region.
- Aggressive scheme.
- Conclusions.

Problem Formulation

Suppose that

- $w \in \mathbb{W}$ is a random variable with probabilistic distribution $\text{Pr}_{\mathbb{W}}$.
- We are able to generate i.i.d. samples from $\text{Pr}_{\mathbb{W}}$.
- $\Psi : \mathbb{W} \to \mathbb{R} \cup +\infty$ is a performance index, where the value $+\infty$ is attained when the specifications are not meet:

$$\Psi(w) = \begin{cases} J(w) & \text{if the constraints are satisfied,} \\ +\infty & \text{otherwise.} \end{cases}$$

Problem Formulation

We consider the following problem:

- **Probabilistic Maximization:** Compute ρ_ε such that, with probability no smaller than $1 - \delta$,

$$\mathrm{Pr}_{\mathbb{W}}\{\Psi(w) > \rho_\varepsilon\} \leq \varepsilon.$$

- $\delta \in (0, 1)$ is a confidence parameter.
- $\varepsilon \in (0, 1)$ determines the level of violation that we are willing to accept.

Example

- Consider the following system:

$$x_{k+1} = \begin{bmatrix} 1 & 0.9 \\ 0 & 1 \end{bmatrix} x_k + \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix} \varphi(Kx_k) + \begin{bmatrix} 0 \\ 0.2 \end{bmatrix} d_k,$$

  where $x_0 = 0$.
- $\varphi(\cdot)$ is the normalized saturation.
- $\mathbf{w} \in \mathbb{R}^M = \mathbb{W}$ corresponds to the sequence of disturbances $\{d_k\}_{k=0}^{M-1}$, where $M = 100$.
- $\Pr_{\mathbb{W}}$ is a normalized Gaussian distribution with zero mean.
- The cost function is

$$\Psi(\mathbf{w}) = \sum_{k=1}^{M} x_k^T \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} x_k + 15\varphi^2(Kx_k).$$

- $K = -\begin{bmatrix} 0.2212 & 1.4181 \end{bmatrix}$ optimizes this performance index for the nominal unsaturated system.
- The objective is, given δ and ε, to found ρ_ε such that, with probability no smaller than $1 - \delta$,

$$\Pr_{\mathbb{W}}\{\Psi(\mathbf{w}) > \rho_\epsilon\} \leq \varepsilon.$$

Example (generalized)

- Consider the **closed loop** time dynamics

$$\left\{ \begin{array}{rcl} x_{k+1} & = & f(x_k, d_k) \\ y_k & = & g(x_k, d_k), \end{array} \right.$$

- d_k serves to model disturbances and/or parameters of the model.
- The trajectory is said to be admissible if $y_k \in \mathbb{Y}$, $k = 1, \dots, M$.
- The uncertainty vector $\mathbf{w}$ is compounded by the initial condition and the sequence $\{d_k\}_{k=1}^{M-1}$.
- Since the trajectories depend on $\mathbf{w}$, we make explicit this dependence: $y_k = y_k(\mathbf{w})$.
- The cost function is

$$\Psi(\mathbf{w}) = \left\{ \begin{array}{rl} \sum\limits_{k=0}^{M} \ell_k(y_k(\mathbf{w})) & \text{if } y_k(\mathbf{w}) \in \mathbb{Y},\ k = 1, \dots N \\ +\infty & \text{otherwise.} \end{array} \right.$$

Probabilistic Certification through Empirical Maximum

Algorithm Empirical Maximum

1: Given a cost function $\Psi : \mathbb{W} \to \mathbb{R} \bigcup +\infty$, and probability levels $\varepsilon \in (0,1)$ and $\delta \in (0,1)$, choose

$$N = \left\lceil \frac{1}{\varepsilon} \ln \frac{1}{\delta} \right\rceil.$$

2: Draw N i.i.d. samples $\{\mathbf{w}_i\}_{i=1}^{N}$ according to $\mathrm{Pr}_{\mathbb{W}}$.

3: Compute $q_i = \Psi(\mathbf{w}_i)$, $i = 1, \ldots, N$.

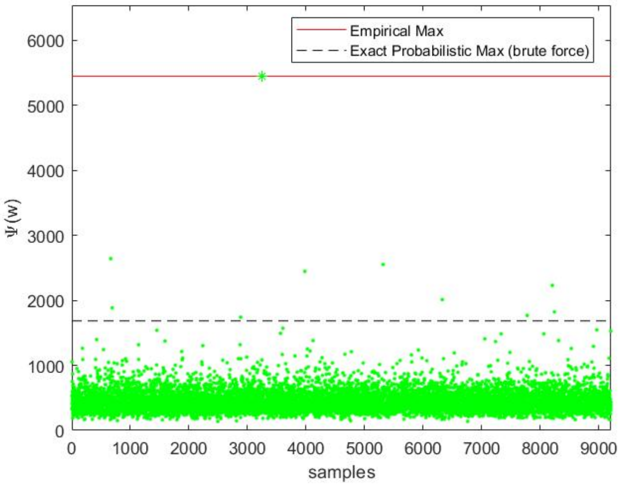
4: Return $\rho = \max\limits_{i=1,\ldots,N} q_i$ as the probabilistic maximum of $\Psi(\mathbf{w})$.

- With probability at least $1-\delta$, $\mathrm{Pr}_{\mathbb{W}} \{\Psi(\mathbf{w}) > \rho\} \leq \varepsilon$.
- The number of samples N depends only on ε and δ.
- No specific assumptions are required on $\Psi(\mathbf{w})$ or $\mathrm{Pr}_{\mathbb{W}}$.

R. Tempo, E.W. Bai, and F. Dabbene. Probabilistic robustness analysis: explicit bounds for the minimum number of samples. *Systems & Control Letters*, 1997.

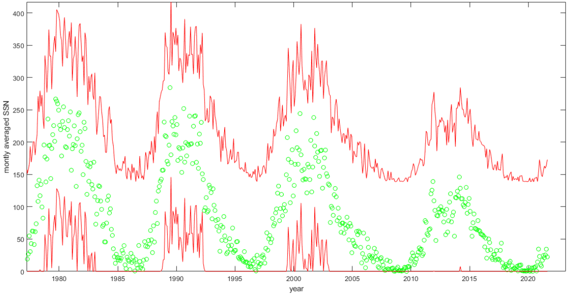
Algorithm 1: Saturated system performance

- Probabilistic specifications: $\varepsilon = 10^{-3}$, $\delta = 10^{-4}$.
- Resulting value for N: $N = 9211$.



Algorithm 1: Application to the SSN time series

- Objective: Obtain the largest error ρ of a given predictor for the Sun Spot Number time series.
- Probabilistic specifications: $\varepsilon = 0.05$, $\delta = 10^{-4}$.
- Resulting value for N and ρ: $N = 185$, $\rho = 138.5$.
- Mean interval width in validation set: 209.3.
- Violation fraction in the validation set: 0.0013.



Generalized max

The following definition is borrowed from the field of order statistics.

Definition (Generalized Max)

Given a collection of N scalars $Q = \{q_1, q_2, \ldots, q_N\} = \{q_i\}_{i=1}^N$, and an integer $r \in [N]$, we say that $q_r^+ \in Q$ is the r-largest value of Q if there is no more than $r-1$ elements of Q strictly larger than q_r^+.

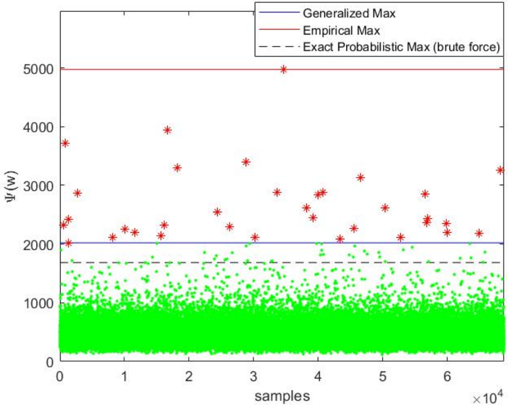
Hence,

- q_1^+ denotes the largest value in Q,
- q_2^+ the second largest one, and so on until q_N^+, which is equal to the smallest one.
- We also use the alternative notation $q_r^+ = \max^{(r)}\{q_i\}_{i=1}^N$.
- In matlab, if q is a vector and the integer r is given, q_r^+ (qr in the code) is obtained from

```
q_sorted=sort(q,'descend');
qr=q_sorted(r);
```

Example: Generalized max for the saturated system

- Imagine that we take N samples of $\Psi(\mathbf{w})$ and take the r largest value ($N = 9.211$, $r = 34$).



- We obtain a better estimation. However: which are the probabilistic guarantees ?

Generalized Max

Property (Generalized Max)

- *Suppose that* $q \in \mathbb{Q} \subseteq \mathbb{R}$ *is a random scalar variable with probability distribution* $\mathrm{Pr}_{\mathbb{Q}}$.
- *Given* $\varepsilon \in (0,1)$, $\delta \in (0,1)$ *and* $r \geq 1$, *let* $N \geq r$ *be such that*

$$\mathbf{B}(r-1; N, \varepsilon) \doteq \sum_{i=0}^{r-1} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i} \leq \delta. \qquad (1)$$

- *Draw* N *i.i.d. samples* $\{q_i\}_{i=1}^N$ *from distribution* $\mathrm{Pr}_{\mathbb{Q}}$.

Then, with a probability no smaller than $1-\delta$,

$$\mathrm{Pr}_{\mathbb{Q}}\{q > \max^{(r)}\{q_i\}_{i=1}^N\} \leq \varepsilon. \qquad (2)$$

T. Alamo, J.M Manzano, and E.F. Camacho. Robust design through probabilistic maximization.In *Uncertainty in Complex Networked Systems. In Honor of Roberto Tempo*, pages 247–274. Birkhäuser, 2018.

Generalized max

- The proof of the generalized max property can be found in
  - T. Alamo, J.M Manzano, and E.F. Camacho. Robust design through probabilistic maximization.In *Uncertainty in Complex Networked Systems. In Honor of Roberto Tempo*, pages 247–274. Birkhäuser, 2018.
- The result is derived using techniques from the field of order statistics, see e.g.
  - M. Ahsanullah, V.B. Nevzorov, and M. Shakil. *An introduction to order statistics*. Atlantis Press, 2013.
- Another possibility is to apply the scenario approach with discarded constraints:
  - M.C. Campi and S. Garatti. The exact feasibility of randomized solutions of robust convex programs. *SIAM Journal of Optimization*, 19:1211–1230, 2008.
  - G. Calafiore. Random convex programs. *SIAM Journal of Optimization*, 20:3427–3464, 2010.

Applications of the generalized max property

Adaptations of this result have been used in the context of:

- Chance constrained optimization:

  T. Alamo, V. Mirasierra, F. Dabbene, and M. Lorenzen. Safe approximations of chance constrained sets by probabilistic scaling. In *2019 18th European Control Conference (ECC)*, pages 1380–1385. IEEE, 2019.

  M. Mammarella, V. Mirasierra, M. Lorenzen, T. Alamo, and F. Dabbene. Chance constrained sets approximation: A probabilistic scaling approach. *Automatica* 137, 110108, 2022.

- Stochastic model predictive control:

  M. Mammarella, T. Alamo, F. Dabbene, and M. Lorenzen. Computationally efficient stochastic MPC: A probabilistic scaling approach. In *Proc. of 4th IEEE Conference on Control Technology and Applications (CCTA)*, pages 25–30, 2020.

  M. Martina, T. Alamo, S. Lucia, and F. Dabbene. A probabilistic validation approach for penalty function design in stochastic model predictive control. In *IFAC World Congress*, 53(2), pages 516–521, 2020.

- Uncertainty Quantification:

  V. Mirasierra, M. Mammarella, F. Dabbene, Fabrizio and T. Alamo. Prediction error quantification through probabilistic scaling. *IEEE Control Systems Letters* volume 6, pagea 1118–1123, 2021.

Algorithm based on Generalized Max

Algorithm Probabilistic Certification through Generalized Max

1: Given a cost function $\Psi : \mathbb{W} \to \mathbb{R} \bigcup +\infty$, and probability levels $\varepsilon \in (0,1)$ and $\delta \in (0,1)$, choose

$$N \geq \frac{7.47}{\varepsilon} \ln \frac{1}{\delta} \quad \text{and} \quad r = \left\lfloor \frac{\varepsilon N}{2} \right\rfloor.$$

2: Draw N i.i.d. samples $\{\mathbf{w}_i\}_{i=1}^N$ according to $\mathrm{Pr}_{\mathbb{W}}$.

3: Compute $q_i = \Psi(\mathbf{w}_i)$, $i \in [N]$.

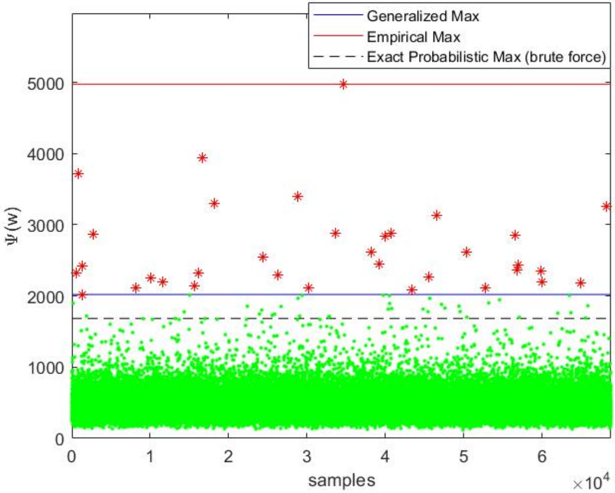
4: Return $\rho = \max^{(r)}\{q_i\}_{i=1}^N$ as the probabilistic upper bound for $\Psi(\mathbf{w})$.

- With probability at least $1-\delta$, $\mathrm{Pr}_{\mathbb{W}}\{\Psi(\mathbf{w}) > \rho\} \leq \varepsilon$.

M. Mammarella, V. Mirasierra, M. Lorenzen, T. Alamo, and F. Dabbene. Chance constrained sets approximation: A probabilistic scaling approach. *Automatica* 137, 110108, 2022.

Algorithm 2: Saturated system performance

- Probabilistic specifications: $\varepsilon = 10^{-3}$, $\delta = 10^{-4}$.
- Resulting value for N and r: $N = 9211$, $r = 34$.



Empirical Max vs Generalized Max: SSN example

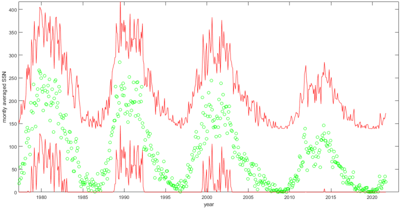


Figure: Alg, 1: Empirical Max

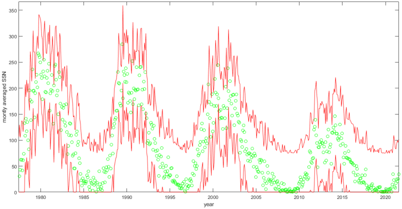


Figure: Alg. 2: Generalized Max

$\varepsilon = 0.01$, $\delta = 1e-4$	Empirical max	Generalized max
N	185	1377
r	1	34
ρ	138.5	74.60
mean interval width	209.29	125.18
violation ratio	0.0013	0.0169

Probabilistic certification when $\varepsilon \to 0$

- Given ρ, certifying $\mathrm{Pr}_{\mathbb{W}}\{\Psi(\mathbf{w}) > \rho\} \leq \varepsilon$ is *generally* not possible if ε is too small ($\varepsilon \to 0 \Rightarrow N \to \infty$).
- Exceptions:
  - When ρ is large enough and the support of $\mathrm{Pr}_{\mathbb{W}}$ is finite, robust techniques could be used to try to prove that the probability of violation is 0.
  - Similarly, one could try to obtain a set $\Omega_\epsilon \subset \mathbb{W}$ such that

$$\mathrm{Pr}_{\mathbb{W}}\{\mathbf{w} \in \Omega_\varepsilon\} \geq 1 - \varepsilon.$$

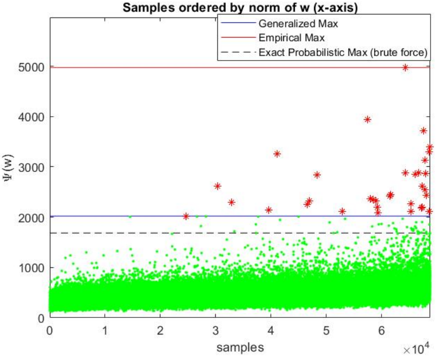
Then it suffices to show (with robustness techniques) that

$$\Psi(\mathbf{w}) \leq 0, \ \forall \mathbf{w} \in \Omega_\varepsilon.$$

K. Margellos, P. Goulart and J. Lygeros. On the road between robust optimization and the scenario approach for chance constrained optimization problems *IEEE Transactions on Automatic Control*, 59(8):2258–2263, 2014.

Indicators of extreme values

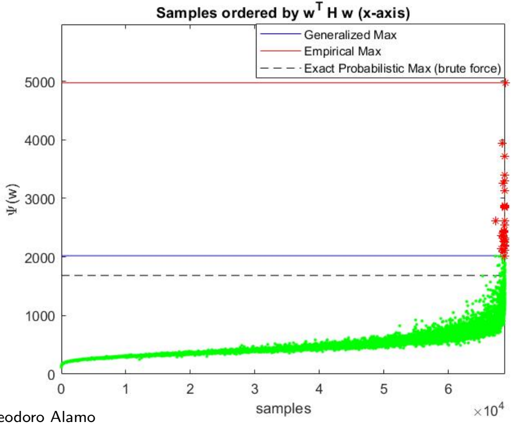
- From an intuitive point of view, the larger **w** is, the larger its potential effect on the system.
- We sort the samples by their norm and see how this correlates with $\Psi(w)$.



- We observe that the extreme red points are somewhat shifted to the right.

Indicators of extreme values

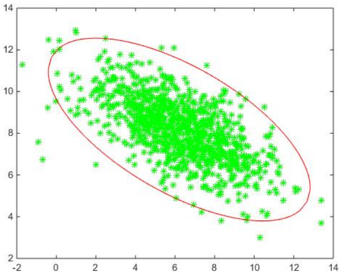
- The effect of $\mathbf{w}$ on $\Psi(\mathbf{w})$ can often be approximated, at least for small values of $\mathbf{w}$, by a quadratic function.
- For example, for a nonlinear closed loop system:
  1. Linearize the system around the operating point.
  2. For the linearized system, approximate $\Psi(\mathbf{w})$ with a quadratic function $\mathbf{w}^T H \mathbf{w}$.



- $\mathbf{w}^T H \mathbf{w}$ is much more informative than the norm of $\mathbf{w}$!

Aggressive Sampling Scheme

- A methodology, based on the computation of three probabilistic maximization problems, will be proposed in the following slides.
- The proposed scheme:
  1. Bounds the non-critical region.
  2. Obtains an estimation of the median in the critical region.
  3. Obtains an estimation of the max in the non critical region.
- A certified probabilistic maximum results of the scheme.



Aggresive sampling

- Suppose that given δ and ε:
  1. We obtain r_ε such that with probability no smaller than $1-\frac{\delta}{3}$:

$$\mathrm{Pr}_{\mathbb{W}}\{\mathbf{w}^T H\mathbf{w} > r_\varepsilon\} \leq \varepsilon.$$

  2. Denote $\Omega_\varepsilon = \{\ \mathbf{w}\ :\ \mathbf{w}^T H\mathbf{w} \leq r_\varepsilon\ \}$.
  3. Obtain ρ_a such that, with probability no smaller than $1-\frac{\delta}{3}$:

$$\mathrm{Pr}_{\mathbb{W}}\{\Psi(\mathbf{w}) > \rho_a \,|\, \mathbf{w} \notin \Omega_\varepsilon\} \leq \frac{1}{2}.$$

  4. Obtain ρ_n such that, with probability no smaller than $1-\frac{\delta}{3}$:

$$\mathrm{Pr}_{\mathbb{W}}\{\Psi(\mathbf{w}) > \rho_n \,|\, \mathbf{w} \in \Omega_\varepsilon\} \leq \frac{\varepsilon}{2}.$$

- Then, $\rho = \max\{\rho_a, \rho_b\}$ satisfies, with probability no smaller than $1-\delta$:

$$\mathrm{Pr}_{\mathbb{W}}\{\Psi(\mathbf{w}) > \rho\} \leq \varepsilon.$$

Proof

- Since each maximization problems is solved with a confidence $\delta/3$ we have that the probability that the three probabilistic inequalities are satisfied is no smaller than $(1-\delta/3)^3 \geq 1-\delta$.
- Thus, with probability no smaller than $1-\delta$:

$$\begin{aligned} \mathrm{Pr}_{\mathbb{W}}\{\Psi(\mathbf{w}) > \rho\} &= \mathrm{Pr}_{\mathbb{W}}\{\Psi(\mathbf{w}) > \rho \,|\, \mathbf{w} \notin \Omega_\varepsilon\}\mathrm{Pr}_{\mathbb{W}}\{\mathbf{w} \notin \Omega_\varepsilon\} \\ &+ \mathrm{Pr}_{\mathbb{W}}\{\Psi(\mathbf{w}) > \rho \,|\, \mathbf{w} \in \Omega_\varepsilon\}\mathrm{Pr}_{\mathbb{W}}\{\mathbf{w} \in \Omega_\varepsilon\}. \end{aligned}$$

- Since $\rho = \max\{\rho_a, \rho_n\}$, and $\mathrm{Pr}_{\mathbb{W}}\{\mathbf{w} \in \Omega_\varepsilon\} \leq 1$:

$$\begin{aligned} \mathrm{Pr}_{\mathbb{W}}\{\Psi(\mathbf{w}) > \rho\} &\leq \mathrm{Pr}_{\mathbb{W}}\{\Psi(\mathbf{w}) > \rho_a \,|\, \mathbf{w} \notin \Omega_\varepsilon\}\mathrm{Pr}_{\mathbb{W}}\{\mathbf{w} \notin \Omega_\varepsilon\} \\ &+ \mathrm{Pr}_{\mathbb{W}}\{\Psi(\mathbf{w}) > \rho_n \,|\, \mathbf{w} \in \Omega_\varepsilon\} \\ &\leq \frac{1}{2}\mathrm{Pr}_{\mathbb{W}}\{\mathbf{w} \notin \Omega_\varepsilon\} + \frac{\varepsilon}{2} \\ &= \frac{1}{2}\mathrm{Pr}_{\mathbb{W}}\{\mathbf{w}^T H \mathbf{w} > r_\varepsilon\} + \frac{\varepsilon}{2} \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Aggresive sampling: The three maximization problems

- The proposed scheme relies on three probabilistic maximization problems.
- That is, given ϵ and δ, one needs to compute r_ε, ρ_a and ρ_n such that

$$\mathrm{Pr}_{\mathbb{W}}\{\mathbf{w}^T H \mathbf{w} > r_\varepsilon\} \leq \varepsilon.$$

$$\mathrm{Pr}_{\mathbb{W}}\{\Psi(\mathbf{w}) > \rho_a \,|\, \mathbf{w} \not\in \Omega_\varepsilon\} \leq \frac{1}{2}.$$

$$\mathrm{Pr}_{\mathbb{W}}\{\Psi(\mathbf{w}) > \rho_n \,|\, \mathbf{w} \in \Omega_\varepsilon\} \leq \frac{\varepsilon}{2}.$$

- We analyze in what follows these three problems.

Aggresive sampling: First maximization problem

- The first maximization problem is to compute r_ε such that, with confidence $\delta/3$:

$$\mathrm{Pr}_{\mathbb{W}}\{\mathbf{w}^T H \mathbf{w} > r_\varepsilon\} \leq \varepsilon.$$

- If the distribution is simple enough, one could obtain r_ε analytically.
- If not, one could resort to generalized maximum.
- In any case, no simulations are required ($\Psi(\mathbf{w})$ is not required).
- The computational time of this step will be significantly smaller than the one of the third problem (in which simulations are required).

Aggresive sampling: Second maximization problem

- The second maximization problem is to compute ρ_a such that, with confidence $\delta/3$:

$$\mathrm{Pr}_{\mathbb{W}}\{\Psi(\mathbf{w}) > \rho_a \,|\, \mathbf{w} \notin \Omega_\varepsilon\} \leq \frac{1}{2}.$$

- We notice that this corresponds to over bound the median of $\Psi(\mathbf{w})$ in the critical region ($\mathbf{w} \notin \Omega_\varepsilon$).
- The number of samples required do not depend on ε !
- Accurate estimations can be obtained with relatively small number of samples.
- Conditioned sampling is required.
- The samples are easy to obtain for some distributions. In other cases one could resort to rejection approaches.

Aggresive sampling: Third maximization problem

- The third maximization problem is to compute ρ_n such that, with confidence $\delta/3$:

$$\mathrm{Pr}_{\mathbb{W}}\{\Psi(\mathbf{w}) > \rho_n \,|\, \mathbf{w} \in \Omega_\varepsilon\} \leq \frac{\varepsilon}{2}.$$

- This problem has some advantages when compared with the original maximization problem:
  1. The maximization is in a bounded set. Thus, the variance of $\Psi(\mathbf{w})$ in Ω_ε can be significantly smaller than the one corresponding to the complete support of $\mathrm{Pr}_{\mathbb{W}}$.
  2. Because of the boundness of the set, one could try robust approaches to prove that $\rho_n \leq \rho_a$ (i.e. to show that the maximum is attained at the second optimization problem).
  3. In any case, probabilistic scaling can always be used.
- The generation of samples is simple because rejection will work very well (most of the samples belong to Ω_ε).

Conclusions

- The notion of probabilistic max has been recalled.
- We characterize the critical region for the probabilistic maximization of a performance index.
- A methodology, based on the computation of three probabilistic maximization problems, is proposed.
- The proposed scheme:
  1. Bounds the non-critical region.
  2. Obtains an estimation of the median in the critical region.
  3. Obtains an estimation of the max in the non critical region.
- A certified probabilistic maximum results of the scheme.