Sample average approximation under heavier tails and stochastic constraints

Philip Thompson
Fond. Math. Jacques Hadamard Post-doc. Fellow at
Center for Research in Economics and Statistics (CREST) - ENSAE

PGMO Days 2017 @ EDF

14th, November, 2017

Joint work with: R.I. Oliveira (IMPA).

Sample average approximation under heavier tails and stochastic constraints

Philip Thompson
Fond. Math. Jacques Hadamard Post-doc. Fellow at
Center for Research in Economics and Statistics (CREST) - ENSAE

PGMO Days 2017 @ EDF

14th, November, 2017

Joint work with: R.I. Oliveira (IMPA).
Part of this work: Center for Mathematical Modeling (CMM) - Uni. de Chile.

Sample average approximation (SAA)

Consider:

$$
\begin{aligned}
f^* := \min_{x \in Y} \quad & f_0(x) \\
\text{s.t.} \quad & f_i(x) \le 0, \qquad \forall i \in [m].
\end{aligned}
$$

Sample average approximation (SAA)

Consider:

$$f^* := \min_{x \in Y} \quad f_0(x)$$
$$\text{s.t.} \quad f_i(x) \leq 0, \qquad \forall i \in [m].$$

with

- **Hard constraint**: compact $Y \in \mathbb{R}^d$.

Sample average approximation (SAA)

Consider:

$$f^* := \min_{x \in Y} \quad f_0(x)$$
$$\text{s.t.} \quad f_i(x) \le 0, \qquad \forall i \in [m].$$

with

- **Hard constraint**: compact $Y \in \mathbb{R}^d$.
- **Feasible set**: $X := \{x \in Y : f_i(x) \le 0, \forall i \in [m]\}$.

Sample average approximation (SAA)

Consider:

$$
\begin{aligned}
f^* := \min_{x \in Y} \quad & f_0(x) \\
\text{s.t.} \quad & f_i(x) \leq 0, \qquad \forall i \in [m].
\end{aligned}
$$

with

- **Hard constraint**: compact $Y \in \mathbb{R}^d$.
- **Feasible set**: $X := \{x \in Y : f_i(x) \leq 0, \forall i \in [m]\}$.
- ϵ-**Near optimal solution set**: $X^*_\epsilon = \{x \in X : f(x) \leq f^* + \epsilon\}$.

Sample average approximation (SAA)

Consider:

$$f^* := \min_{x \in Y} \quad f_0(x)$$
$$\text{s.t.} \quad f_i(x) \le 0, \qquad \forall i \in [m].$$

with

- **Hard constraint**: compact $Y \in \mathbb{R}^d$.
- **Feasible set**: $X := \{x \in Y : f_i(x) \le 0, \forall i \in [m]\}$.
- ϵ**-Near optimal solution set**: $X_\epsilon^* = \{x \in X : f(x) \le f^* + \epsilon\}$.

and

$$f_i(x) := \mathbf{P} F_i(x, \cdot) := \int_\Xi F_i(x, \xi) \, \mathrm{d} \, \mathbf{P}(\xi), \quad (x \in Y).$$

Sample average approximation (SAA)

Assumption

No access to distribution $\mathbf{P}$*, only to a size-*N *i.i.d. sample of* $\mathbf{P}$.

Sample average approximation (SAA)

Assumption

No access to distribution $\mathbf{P}$*, only to a size-*N *i.i.d. sample of* $\mathbf{P}$*.*

Basic framework of

- **Stochastic & Simulation Optimization**: knowledge of data is limited or computation is hard, but one can resort to samples using Monte Carlo simulation.

Sample average approximation (SAA)

Assumption

No access to distribution $\mathbf{P}$*, only to a size-*N *i.i.d. sample of* $\mathbf{P}$*.*

Basic framework of

- **Stochastic & Simulation Optimization**: knowledge of data is limited or computation is hard, but one can resort to samples using Monte Carlo simulation.
- M-**Estimation & Empirical Risk Minimization**: limited number of samples is acquired from measurements $\Rightarrow$ empirical contrast estimator is built to fit the data with respect to a chosen loss function over a hypothesis class (...)

Sample average approximation (SAA)

The SAA estimator solves the optimization problem with the "natural" plug-in rule:

$$f_i(x) := \mathbf{P} F_i(x, \cdot) \qquad \Longrightarrow \qquad \widehat{F}_i(x) := \widehat{\mathbf{P}} F_i(\cdot, x),$$

where $\widehat{\mathbf{P}} := \frac{1}{N} \sum_{j=1}^{N} \delta_{\xi_j}$ is the empirical distribution.

Sample average approximation (SAA)

The SAA estimator solves the optimization problem with the "natural" plug-in rule:

$$f_i(x) := \mathbf{P} F_i(x, \cdot) \qquad \Longrightarrow \qquad \widehat{F}_i(x) := \widehat{\mathbf{P}} F_i(\cdot, x),$$

where $\widehat{\mathbf{P}} := \frac{1}{N} \sum_{j=1}^{N} \delta_{\xi_j}$ is the empirical distribution.

For instance:

$$\begin{aligned} \widehat{F}^* := \min_{x \in Y} \quad & \widehat{F}_0(x) \\ \text{s.t.} \quad & \widehat{F}_i(x) \le 0, \qquad \forall i \in [m]. \qquad (*) \end{aligned}$$

Sample average approximation (SAA)

The SAA estimator solves the optimization problem with the "natural" plug-in rule:

$$f_i(x) := \mathbf{P} F_i(x, \cdot) \qquad \Longrightarrow \qquad \widehat{F}_i(x) := \widehat{\mathbf{P}} F_i(\cdot, x),$$

where $\widehat{\mathbf{P}} := \frac{1}{N} \sum_{j=1}^{N} \delta_{\xi_j}$ is the empirical distribution.

For instance:

$$\begin{aligned} \widehat{F}^* := \min_{x \in Y} \quad & \widehat{F}_0(x) \\ \text{s.t.} \quad & \widehat{F}_i(x) \leq 0, \qquad \forall i \in [m]. \qquad (*) \end{aligned}$$

Fundamental question: in what sense and conditions, the solutions of the data dependent problem solve the distributed dependent problem...

Sample average approximation (SAA)

The SAA estimator solves the optimization problem with the "natural" plug-in rule:

$$f_i(x) := \mathbf{P} F_i(x, \cdot) \qquad \Longrightarrow \qquad \widehat{F}_i(x) := \widehat{\mathbf{P}} F_i(\cdot, x),$$

where $\widehat{\mathbf{P}} := \frac{1}{N} \sum_{j=1}^{N} \delta_{\xi_j}$ is the empirical distribution.

For instance:

$$\begin{aligned} \widehat{F}^* := \min_{x \in Y} \quad & \widehat{F}_0(x) \\ \text{s.t.} \quad & \widehat{F}_i(x) \leq 0, \qquad \forall i \in [m]. \qquad (*) \end{aligned}$$

Fundamental question: in what sense and conditions, the solutions of the data dependent problem solve the distributed dependent problem...*uniformily* over the classes of algorithms used to solve (*).

Sample average approximation (SAA)

That's an old topic:

- BASIC FACT: *uniform* SLLN, CLT, etc...

Sample average approximation (SAA)

That's an old topic:

- BASIC FACT: *uniform* SLLN, CLT, etc...
- Asymptotic analysis ($N \to \infty$) $\Rightarrow$ Nonasymptotic analysis (Finite N uniformly on the confidence level).

Sample average approximation (SAA)

That's an old topic:

- BASIC FACT: *uniform* SLLN, CLT, etc...
- Asymptotic analysis ($N \to \infty$) $\Rightarrow$ Nonasymptotic analysis (Finite N uniformly on the confidence level).
- Stochastic optimization community: Wets et al, Shapiro, Pflug, Vogel, Dupacova, Römisch, Homem-de-Mello, and many others ...

Sample average approximation (SAA)

That's an old topic:

- BASIC FACT: *uniform* SLLN, CLT, etc...
- Asymptotic analysis ($N \to \infty$) $\Rightarrow$ Nonasymptotic analysis (Finite N uniformly on the confidence level).
- Stochastic optimization community: Wets et al, Shapiro, Pflug, Vogel, Dupacova, Römisch, Homem-de-Mello, and many others ...
- Math.-Statistics and Statistical Machine Learning community: Vapnik, Talagrand...

Sample average approximation (SAA)

That's an old topic:

- BASIC FACT: *uniform* SLLN, CLT, etc...
- Asymptotic analysis ($N \to \infty$) $\Rightarrow$ Nonasymptotic analysis (Finite N uniformly on the confidence level).
- Stochastic optimization community: Wets et al, Shapiro, Pflug, Vogel, Dupacova, Römisch, Homem-de-Mello, and many others ...
- Math.-Statistics and Statistical Machine Learning community: Vapnik, Talagrand...$\Rightarrow$ **Empirical Process Theory**: uniform concentration inequalities.

Sample average approximation (SAA)

That's an old topic:

- BASIC FACT: *uniform* SLLN, CLT, etc...
- Asymptotic analysis ($N \to \infty$) $\Rightarrow$ Nonasymptotic analysis (Finite N uniformly on the confidence level).
- Stochastic optimization community: Wets et al, Shapiro, Pflug, Vogel, Dupacova, Römisch, Homem-de-Mello, and many others ...
- Math.-Statistics and Statistical Machine Learning community: Vapnik, Talagrand...$\Rightarrow$ **Empirical Process Theory**: uniform concentration inequalities.
- **However**...

Nonasymptotic analysis of SAA: heavier tails?

In Stochastic Optimization:

- All previous exponential nonasymptotic analysis still require a uniform **sub-Gaussian** assumption: $\forall i$, $\exists \sigma_i > 0$,

$$\sup_{x \in Y} \mathbf{P}\left\{ e^{t[F_i(x,\cdot) - f_i(x)]} \right\} \leq e^{\frac{\sigma_i^2 t^2}{2}}, \qquad \forall t \in \mathbb{R}.$$

Nonasymptotic analysis of SAA: heavier tails?

In Stochastic Optimization:

- All previous exponential nonasymptotic analysis still require a uniform **sub-Gaussian** assumption: $\forall i$, $\exists \sigma_i > 0$,

$$\sup_{x \in Y} \mathbf{P}\left\{ e^{t[F_i(x,\cdot) - f_i(x)]} \right\} \leq e^{\frac{\sigma_i^2 t^2}{2}}, \qquad \forall t \in \mathbb{R}.$$

- One typical setting where this is expected to fail: **stochastic portfolio optimization**

Nonasymptotic analysis of SAA: heavier tails?

In Stochastic Optimization:

- All previous exponential nonasymptotic analysis still require a uniform **sub-Gaussian** assumption: $\forall i$, $\exists \sigma_i > 0$,

$$\sup_{x \in Y} \mathbf{P}\left\{ e^{t[F_i(x,\cdot) - f_i(x)]} \right\} \leq e^{\frac{\sigma_i^2 t^2}{2}}, \qquad \forall t \in \mathbb{R}.$$

- One typical setting where this is expected to fail: **stochastic portfolio optimization** $\Rightarrow$ risk typically aims at hedging against tails.

Nonasymptotic analysis of SAA: heavier tails?

UPSHOT: We give exponential nonasymptotic $\mathcal{O}(N^{-1/2})$ rate assuming **just** the classical assumption:

Assumption (Hölder continuous heavy-tailed functions (HT))

Let $\|\cdot\|$ *be a norm on* $\mathbb{R}^d$*. For all* $i \in [m] \cup \{0\}$, $\exists \mathsf{L}_i : \Xi \to \mathbb{R}_+$ *with*

$$\mathbf{P}\mathsf{L}_i(\cdot)^2 < \infty,$$

and $\alpha_i \in (0,1]$ *s.t.*

$$\text{a.s.} \qquad |F_i(x,\cdot) - F_i(y,\cdot)| \le \mathsf{L}_i(\cdot)\|x-y\|^{\alpha_i}, \quad \forall x, y \in Y.$$

Nonasymptotic analysis of SAA: heavier tails?

UPSHOT: We give exponential nonasymptotic $\mathcal{O}(N^{-1/2})$ rate assuming **just** the classical assumption:

Assumption (Hölder continuous heavy-tailed functions (HT))

Let $\|\cdot\|$ *be a norm on* $\mathbb{R}^d$*. For all* $i \in [m] \cup \{0\}$, $\exists \mathsf{L}_i : \Xi \to \mathbb{R}_+$ *with*

$$\mathbf{P}\mathsf{L}_i(\cdot)^2 < \infty,$$

and $\alpha_i \in (0,1]$ *s.t.*

$$a.s. \qquad |F_i(x,\cdot) - F_i(y,\cdot)| \le \mathsf{L}_i(\cdot)\|x-y\|^{\alpha_i}, \quad \forall x, y \in Y.$$

The multiplicative noise $\mathsf{L}_i(\cdot)$ can have much heavier fluctuations than the tail of a Gaussian random variable!

Nonasymptotic analysis of SAA: heavier tails?

Theorem (SAA with fixed set)

Let $x^* \in X^*$ and $z \in Y$. Then, under **HT**,

(i) *(Optimality) For any $p \in (0,1]$ and any $\epsilon > 0$, with probability $\geq 1 - p$:*

$$N \geq \mathcal{O}(1)\frac{\widehat{\sigma}_0(X)^2}{\epsilon^2}\ln(1/p) \quad \Longrightarrow \quad \widehat{X}_\epsilon^* \subset X_{2\epsilon}^*.$$

(ii) *(Optimal value) For any $p \in (0,1]$ and any $\epsilon > 0$, with probability $\geq 1 - p$:*

$$N \geq \mathcal{O}(1)\frac{\widehat{\sigma}_0(X)^2 \vee \breve{\sigma}_0(z)^2}{\epsilon^2}\ln(1/p) \quad \Longrightarrow \quad f^* - 2\epsilon \leq \widehat{F}^*,$$

$$N \geq \mathcal{O}(1)\frac{\breve{\sigma}_0(x^*)^2}{\epsilon^2}\ln(1/p) \quad \Longrightarrow \quad \widehat{F}^* \leq f^* + \epsilon.$$

Nonasymptotic analysis of SAA: heavier tails?

Where (in the Lipschitz case):

$$\breve{\sigma}_0(x)^2 := (\mathbf{P}+\widehat{\mathbf{P}})(F_0(x,\cdot)-f_0(x))^2,$$
$$\widehat{\sigma}_0(X)^2 := \mathsf{Q}(X)^2\mathcal{D}(X)^2(\mathbf{P}+\widehat{\mathbf{P}})\mathsf{L}_0(\cdot)^2,$$

Nonasymptotic analysis of SAA: heavier tails?

Where (in the Lipschitz case):

$$\breve{\sigma}_0(x)^2 \quad := \quad (\mathbf{P} + \widehat{\mathbf{P}})(F_0(x,\cdot) - f_0(x))^2,$$
$$\widehat{\sigma}_0(X)^2 \quad := \quad \mathsf{Q}(X)^2 \mathcal{D}(X)^2 (\mathbf{P} + \widehat{\mathbf{P}}) \mathsf{L}_0(\cdot)^2,$$

and $\mathsf{Q}(X)$ encodes the "integrated metric entropy" of X:

$$\mathsf{Q}(X) \quad := \quad \sum_{i=1}^{\infty} \frac{1}{2^i} \sqrt{\mathsf{H}\left(\frac{\mathcal{D}(X)}{2^i}, X\right) + \mathsf{H}\left(\frac{\mathcal{D}(X)}{2^{i-1}}, X\right) + \ln[i(i+1)]}.$$

E.g.: for any ball $X := B_p[0,R] \subset \mathbb{R}^d$, $\mathsf{Q}(X) \sim \sqrt{d}$.

Nonasymptotic analysis of SAA: heavier tails?

Main derived tool:

Theorem (sub-Gaussian uniform concentration inequality for heavy-tailed Hölder continuous functions)

Let $(\mathcal{M}, \mathrm{d})$ *is a t.b. metric space with diameter* $\mathcal{D}(\mathcal{M})$*, i.i.d. sample from* $\mathbf{P}$*, the Hölder continuous measurable function* $F : \mathcal{M} \times \Xi \to \mathbb{R}$ *satisfies*

$$\mathbf{P}\mathsf{L}(\cdot)^2 < \infty.$$

Set

$$\widehat{\delta}_F(x, y) := (\widehat{\mathbf{P}} - \mathbf{P}) \left[F(x, \cdot) - F(y, \cdot) \right].$$

Then there exists universal constant $C > 0$*, such that for any* $y \in \mathcal{M}$ *and* $t > 0$*,*

$$\mathbb{P} \left\{ \sup_{x \in \mathcal{M}} \left| \widehat{\delta}_F(x, y) \right| \geq C Q_\alpha(\mathcal{M}) \mathcal{D}(\mathcal{M})^\alpha \sqrt{\frac{(1+t)}{N} \left[\widehat{\mathsf{L}}^2 + \mathbf{P}\mathsf{L}^2(\cdot) \right]} \right\} \leq e^{-t}.$$

Nonasymptotic analysis of SAA: heavier tails?

- Chaining method,
- Self-normalization theory of empirical processes (Panchenko's inequality)
- Avoids the previous approach used in Stochastic Optimization of invoking the ready to use Talagrand's inequality for bounded empirical processes.

Nonasymptotic analysis of SAA: stochastic constraints?

FACT: Perturbed constraints is challenging: **geometry** of the set plays a role.

Nonasymptotic analysis of SAA: stochastic constraints?

FACT: Perturbed constraints is challenging: **geometry** of the set plays a role.

In Stochastic Optimization:

- Even in the context of **expected-valued constraints**, there seems to exist a gap of results concerning the nonasymptotic analysis of **SAA with random constraints**:

Nonasymptotic analysis of SAA: stochastic constraints?

FACT: Perturbed constraints is challenging: **geometry** of the set plays a role.

In Stochastic Optimization:

- Even in the context of **expected-valued constraints**, there seems to exist a gap of results concerning the nonasymptotic analysis of **SAA with random constraints**:
- Homem-de-Mello and Bayraksan, *Stochastic constraints and variance reduction techniques* (2015), in Handbook of Simulation Optimization.

Nonasymptotic analysis of SAA: stochastic constraints?

Precisely: exp. nonasymptotic deviation ineq. which

- ensure feasibility and optimality **simultaneously**.

Nonasymptotic analysis of SAA: stochastic constraints?

Precisely: exp. nonasymptotic deviation ineq. which

- ensure feasibility and optimality **simultaneously**.
- **do not** consider (nonequivalent) **problem reformulations**: penalized and Lagrangian reformulations or first-order necessary conditions

Nonasymptotic analysis of SAA: stochastic constraints?

Precisely: exp. nonasymptotic deviation ineq. which

- ensure feasibility and optimality **simultaneously**.
- **do not** consider (nonequivalent) **problem reformulations**: penalized and Lagrangian reformulations or first-order necessary conditions $\Rightarrow$ the penalizers and multipliers are data dependent whose value is a hard problem itself.

Nonasymptotic analysis of SAA: stochastic constraints?

Precisely: exp. nonasymptotic deviation ineq. which

- ensure feasibility and optimality **simultaneously**.
- **do not** consider (nonequivalent) **problem reformulations**: penalized and Lagrangian reformulations or first-order necessary conditions $\Rightarrow$ the penalizers and multipliers are data dependent whose value is a hard problem itself.
- **do not** require **metric regularity of the solution set**: which tipically enforces a strong growth assumption on the objective.

Nonasymptotic analysis of SAA: stochastic constraints?

Precisely: exp. nonasymptotic deviation ineq. which

- ensure feasibility and optimality **simultaneously**.
- **do not** consider (nonequivalent) **problem reformulations**: penalized and Lagrangian reformulations or first-order necessary conditions $\Rightarrow$ the penalizers and multipliers are data dependent whose value is a hard problem itself.
- **do not** require **metric regularity of the solution set**: which tipically enforces a strong growth assumption on the objective.
- still allow **heavier tails** over a **very large** number of constraints.

Nonasymptotic analysis of SAA: stochastic constraints?

UPSHOT: We give exponential nonasymptotic deviation inequalities assuming **just** metric regularity of the **feasible set**:

Assumption (Metric regular feasible set (MRF))

There exists $\mathfrak{c} > 0$ *s.t.*

$$\mathrm{d}(x, X) \leq \mathfrak{c} \max_{i \in \mathcal{I}} [f_i(x)]_+ , \qquad \forall x \in Y.$$

Nonasymptotic analysis of SAA: stochastic constraints?

UPSHOT: We give exponential nonasymptotic deviation inequalities assuming **just** metric regularity of the **feasible set**:

Assumption (Metric regular feasible set (MRF))

There exists $\mathfrak{c} > 0$ *s.t.*

$$\mathrm{d}(x, X) \leq \mathfrak{c} \max_{i \in \mathcal{I}} [f_i(x)]_+ , \qquad \forall x \in Y.$$

which

- ensure feasibility and optimality **simultaneously**.

Nonasymptotic analysis of SAA: stochastic constraints?

UPSHOT: We give exponential nonasymptotic deviation inequalities assuming **just** metric regularity of the **feasible set**:

Assumption (Metric regular feasible set (MRF))

There exists $\mathfrak{c} > 0$ *s.t.*

$$\mathrm{d}(x, X) \leq \mathfrak{c} \max_{i \in \mathcal{I}} [f_i(x)]_+ , \qquad \forall x \in Y.$$

which

- ensure feasibility and optimality **simultaneously**.
- in terms of the **original problem**.

Nonasymptotic analysis of SAA: stochastic constraints?

UPSHOT: We give exponential nonasymptotic deviation inequalities assuming **just** metric regularity of the **feasible set**:

Assumption (Metric regular feasible set (MRF))

There exists $\mathfrak{c} > 0$ *s.t.*

$$\mathrm{d}(x, X) \leq \mathfrak{c} \max_{i \in \mathcal{I}} [f_i(x)]_+ , \qquad \forall x \in Y.$$

which

- ensure feasibility and optimality **simultaneously**.
- in terms of the **original problem**.
- **without** any assumption on metric regularity of the **solution set** and additional **condition numbers** associated to it.

Nonasymptotic analysis of SAA: stochastic constraints?

UPSHOT: We give exponential nonasymptotic deviation inequalities assuming **just** metric regularity of the **feasible set**:

Assumption (Metric regular feasible set (MRF))

There exists $\mathfrak{c} > 0$ *s.t.*

$$\mathrm{d}(x, X) \leq \mathfrak{c} \max_{i \in \mathcal{I}} [f_i(x)]_+ , \qquad \forall x \in Y.$$

which

- ensure feasibility and optimality **simultaneously**.
- in terms of the **original problem**.
- **without** any assumption on metric regularity of the **solution set** and additional **condition numbers** associated to it.
- That is: N needed to ensure feasibility (besides optimality) is only a function of $\mathfrak{c}$.

Nonasymptotic analysis of SAA: stochastic constraints?

MRF includes classical constraint qualifications:

Assumption (Slater constraint qualification (SCQ))

$\{\widehat{F}_i\}_{i\in[m]} \cup \{f_i\}_{i\in[m]}$ *are continuous and convex on the closed and convex set* Y *and*

$$\exists \bar{x} \in Y, \qquad \mathring{\epsilon}(\bar{x}) := \min_{i\in[m]}\{-f_i(\bar{x})\} > 0.$$

Nonasymptotic analysis of SAA: stochastic constraints?

MRF includes classical constraint qualifications:

Assumption (Slater constraint qualification (SCQ))

$\{\widehat{F}_i\}_{i\in[m]} \cup \{f_i\}_{i\in[m]}$ *are continuous and convex on the closed and convex set* Y *and*

$$\exists \bar{x} \in Y, \qquad \mathring{\epsilon}(\bar{x}) := \min_{i\in[m]} \{-f_i(\bar{x})\} > 0.$$

Theorem (Robinson (1975))

SCQ + compactness $\Rightarrow$ *with*

$$\mathfrak{c} := \frac{\mathcal{D}(X)}{\mathring{\epsilon}(\bar{x})}.$$

Nonasymptotic analysis of SAA: stochastic constraints?

MRF is very general:

- used in Perturbation Theory in Optimization and Variational Analysis.

Nonasymptotic analysis of SAA: stochastic constraints?

MRF is very general:

- used in Perturbation Theory in Optimization and Variational Analysis.
- Sufficient condition on analysis of algorithms relying on constraint relaxation.

Nonasymptotic analysis of SAA: stochastic constraints?

MRF is very general:

- used in Perturbation Theory in Optimization and Variational Analysis.
- Sufficient condition on analysis of algorithms relying on constraint relaxation.
- It does **not** require **strict feasibility** in convex sets as in SCQ: e.g. any polyhedron (Hoffman's Lemma).

Nonasymptotic analysis of SAA: stochastic constraints?

MRF is very general:

- used in Perturbation Theory in Optimization and Variational Analysis.
- Sufficient condition on analysis of algorithms relying on constraint relaxation.
- It does **not** require **strict feasibility** in convex sets as in SCQ: e.g. any polyhedron (Hoffman's Lemma).
- Our analysis is easily extended to the Hölderian case:

$$\mathrm{d}(x, X) \leq \mathfrak{c} \max_{i \in \mathcal{I}} [f_i(x)]_+^{\beta}, \qquad \forall x \in Y.$$

Nonasymptotic analysis of SAA: stochastic constraints?

MRF is very general:

- used in Perturbation Theory in Optimization and Variational Analysis.
- Sufficient condition on analysis of algorithms relying on constraint relaxation.
- It does **not** require **strict feasibility** in convex sets as in SCQ: e.g. any polyhedron (Hoffman's Lemma).
- Our analysis is easily extended to the Hölderian case:

$$\mathsf{d}(x, X) \leq \mathfrak{c} \max_{i \in \mathcal{I}} [f_i(x)]_+^{\beta}, \qquad \forall x \in Y.$$

  In that case, it allows **nonconvex feasible sets**: e.g. polynomial or real-analytic constraints (Lojasiewicz's inequality).

Nonasymptotic analysis of SAA: stochastic constraints?

SAA Problem:

$$\begin{aligned} \widehat{F}^* := \min_{x \in Y} \quad & \widehat{F}_0(x) \\ \text{s.t.} \quad & \widehat{F}_i(x) \leq \widehat{\epsilon}_i, \qquad \forall i \in [m], \end{aligned}$$

where $\{\widehat{\epsilon}_i\}$ is a "tunning sequence". In fact, it is somewhat essential to be **nonzero**.

Nonasymptotic analysis of SAA: stochastic constraints?

Theorem (SAA with exterior approximation - MRF)

Suppose **HT-MRF**. *Let* $z \in Y$ *and* $x^* \in X^*$ *and*

$$\check{X}_\gamma := (X + \gamma \mathbb{B}) \cap Y.$$

For any $p \in (0,1]$ *and any* $\epsilon > 0$, *with probability* $\geq 1 - p$:

$$\widehat{\epsilon}_i \equiv \epsilon \text{ and } N \geq \frac{\mathcal{O}(\hat{\sigma}^2)}{\epsilon^2}[\ln m + \ln(1/p)] \Rightarrow \begin{cases} \mathbb{D}(\widehat{X}, X) \leq 2\mathfrak{c}\epsilon, \\ \forall \widehat{x} \in \widehat{X}^*_\epsilon, f(\widehat{x}) \leq f^* + 2\epsilon, \end{cases}$$

with $\hat{\sigma} := [\widehat{\sigma}_{\mathcal{I}}(Y) \vee \breve{\sigma}_{\mathcal{I}}(z)] \vee \breve{\sigma}_{\mathcal{I}}(x^*) \vee \widehat{\sigma}_0(\check{X}_{2\mathfrak{c}\epsilon})$.

Nonasymptotic analysis of SAA: stochastic constraints?

Theorem (SAA with **exterior** approximation - **MRF**)

Localized feasibility *under* **convex constraints***:*

$$\widehat{\sigma}_{\mathcal{I}}(Y) \vee \breve{\sigma}_{\mathcal{I}}(z)$$

replaced by

$$\widehat{\sigma}_{\mathcal{I}}(2\epsilon) \vee \breve{\sigma}_{\mathcal{I}}(y) \vee \breve{\sigma}_{\mathcal{I}}(z)$$

where the variances associated to SAA depend on diameter and metric entropy of

$$X_{i,\gamma} := \{x \in X_\gamma : f_i(x) = \gamma\} \qquad i \in [m]$$

is the γ-active level set of the ith constraint.

Nonasymptotic analysis of SAA: stochastic constraints?

Theorem (SAA with interior approximation - SCQ)

Suppose **HT-SCQ** *and* $z \in Y$. *Then, for any* $p \in (0, 1]$,

$$\epsilon \in (0, \mathring{\epsilon}(\bar{x})/2],$$

$y \in \operatorname{int}(X_{-2\epsilon})$ *and* $y^* \in \operatorname{int}(X_{-2\epsilon})^*$, *with probability* $\geq 1 - p$:

$$\widehat{\epsilon}_i \equiv -\epsilon \quad \text{and} \quad N \geq \mathcal{O}(1)\frac{\hat{\sigma}^2}{\epsilon^2}[\ln m + \ln(1/p)] \quad \Longrightarrow \quad \widehat{X}_\epsilon^* \subset X^*_{2\epsilon+\text{gap}(2\epsilon)},$$

where $\hat{\sigma} := [\widehat{\sigma}_{\mathcal{I}}(0) \vee \breve{\sigma}_{\mathcal{I}}(y) \vee \breve{\sigma}_{\mathcal{I}}(z)] \vee \widehat{\sigma}_{\mathcal{I}}(y^*) \vee \widehat{\sigma}_0(X)$ *and*

$$\text{gap}(\gamma) := \min_{X_{-\gamma}} f - f^* \geq 0.$$

$$X_{-\gamma} := \{x \in Y : f_i(x) \leq -\gamma, \forall i \in [m]\}.$$

Nonasymptotic analysis of SAA: stochastic constraints?

REMARKS: Under SCQ and tolerance level $\epsilon \sim \mathcal{O}(\mathring{\epsilon}(\bar{x}))$:

- Same **localized feasibility** in terms of variances of the exact active constraint level sets:

$$X_{i,0} := \{x \in X_\gamma : f_i(x) = 0\} \qquad i \in [m]$$

Nonasymptotic analysis of SAA: stochastic constraints?

REMARKS: Under SCQ and tolerance level $\epsilon \sim \mathcal{O}(\mathring{\epsilon}(\bar{x}))$:

- Same **localized feasibility** in terms of variances of the exact active constraint level sets:

$$X_{i,0} := \{x \in X_\gamma : f_i(x) = 0\} \qquad i \in [m]$$

- Finite-sample **exact** feasibility guarantee!

Nonasymptotic analysis of SAA: stochastic constraints?

REMARKS: Under SCQ and tolerance level $\epsilon \sim \mathcal{O}(\mathring{\epsilon}(\bar{x}))$:

- Same **localized feasibility** in terms of variances of the exact active constraint level sets:

$$X_{i,0} := \{x \in X_\gamma : f_i(x) = 0\} \qquad i \in [m]$$

- Finite-sample **exact** feasibility guarantee!
- The optimality error gap$(\cdot)$ only depends on **local Hölder continuity** around "solutions at the border":

$$\text{gap}(\gamma) \le L_\gamma \gamma^{\alpha_0}, \qquad L_\gamma \ll L_0.$$

It is zero if interior solutions exists.

Nonasymptotic analysis of SAA: **convex problems**

Questions:

(i) How can we ensure that nearly optimal solutions $\widehat{X}_{\epsilon}^{*}$ of the accessible SAA problem are nearly optimal solutions $X_{\mathcal{O}(\epsilon)}^{*}$ of the original inaccessible problem?

(ii) Related to the above questions, one of our main concerns in this paper will be of **localization in optimality & feasibility**: under **convexity**, "where in space" it is enough for the perturbations to be controlled?

Nonasymptotic analysis of SAA: **convex problems**

Questions:

(i) How can we ensure that nearly optimal solutions $\widehat{X}_{\epsilon}^{*}$ of the accessible SAA problem are nearly optimal solutions $X_{\mathcal{O}(\epsilon)}^{*}$ of the original inaccessible problem?

(ii) Related to the above questions, one of our main concerns in this paper will be of **localization in optimality & feasibility**: under **convexity**, "where in space" it is enough for the perturbations to be controlled?

UPSHOT: under **convexity** and MRF, exponential nonasymptotic joint feasibility and optimality guarantees are given for

$$N \geq \hat{\sigma}^2 \mathcal{O}(\epsilon^{-2}) \ln(m/p)$$

where $\hat{\sigma}^2$ only depends on the **diameter and metric entropy of the** $\mathcal{O}(\epsilon)$**-near optimal solution set** (or approximations of it):

$$X_{\epsilon}^{*} := \{x \in Y : f(x) \leq f^{*} + \epsilon\}.$$

Nonasymptotic analysis of SAA: **convex problems**

Questions:

(i) How can we ensure that nearly optimal solutions $\widehat{X}_\epsilon^*$ of the accessible SAA problem are nearly optimal solutions $X^*_{\mathcal{O}(\epsilon)}$ of the original inaccessible problem?

(ii) Related to the above questions, one of our main concerns in this paper will be of **localization in optimality & feasibility**: under **convexity**, "where in space" it is enough for the perturbations to be controlled?

UPSHOT: under **convexity** and MRF, exponential nonasymptotic joint feasibility and optimality guarantees are given for

$$N \geq \hat{\sigma}^2 \mathcal{O}(\epsilon^{-2}) \ln(m/p)$$

where $\hat{\sigma}^2$ only depends on the **diameter and metric entropy of the** $\mathcal{O}(\epsilon)$**-near optimal solution set** (or approximations of it):

$$X_\epsilon^* := \{x \in Y : f(x) \leq f^* + \epsilon\}.$$

In fact, even tighter subsets with **active** optimality gap and constraints are enough.

SIMPLE EXAMPLE: **Sharper sample complexity**.

$$F(x,\xi) := \mathsf{L}(\xi)\|x\|_2$$

$$X := \{x \in \mathbb{R}^d : \|x\|_2 \leq 1\}$$

Then our bounds show that $\hat{\sigma} = \mathcal{O}(\epsilon)\sqrt{\hat{L}d}$ and a sample size of

$$N \geq \hat{L}d \ln(m/p)$$

is enough to obtain $\mathcal{O}(\epsilon)$-near optimality guarantees, **without** the typical $\mathcal{O}(\epsilon^{-2})$ dependence.

Nonasymptotic analysis of SAA: **convex problems**

REMARKS:

- As before, only metric regularity of the **feasible set** is assumed allowing for convex sets **without strict feasibility**.

:

Nonasymptotic analysis of SAA: **convex problems**

REMARKS:

- As before, only metric regularity of the **feasible set** is assumed allowing for convex sets **without strict feasibility**.
- Nevertheless, **finite-sample** "transition regime" for **SCQ**:

Nonasymptotic analysis of SAA: **convex problems**

REMARKS:

- As before, only metric regularity of the **feasible set** is assumed allowing for convex sets **without strict feasibility**.
- Nevertheless, **finite-sample** "transition regime" for **SCQ**:
- For any

$$N \geq \hat{\sigma}_1^2 \mathcal{O}(\epsilon^{-2}) \ln(m/p),$$

it is guaranteed a $\mathcal{O}(\mathfrak{c}\epsilon)$-**near feasibility** and $\mathcal{O}(\epsilon)$-optimality.

Nonasymptotic analysis of SAA: **convex problems**

REMARKS:

- As before, only metric regularity of the **feasible set** is assumed allowing for convex sets **without strict feasibility**.
- Nevertheless, **finite-sample** "transition regime" for **SCQ**:
- For any

$$N \geq \hat{\sigma}_1^2 \mathcal{O}(\epsilon^{-2}) \ln(m/p),$$

  it is guaranteed a $\mathcal{O}(\mathfrak{c}\epsilon)$-**near feasibility** and $\mathcal{O}(\epsilon)$-optimality.
- For *larger*

$$N \geq \hat{\sigma}_2^2 \mathcal{O}(\epsilon^{-2}) \vee \mathcal{O}(\mathring{\epsilon}(\bar{x})^{-2}) \ln(m/p),$$

  exact feasibility and $\mathcal{O}(\epsilon) + \text{gap}(\mathcal{O}(\epsilon))$-optimality is guaranteed with high-probability.

Nonasymptotic analysis of SAA: **convex problems**

REMARKS:

- As before, only metric regularity of the **feasible set** is assumed allowing for convex sets **without strict feasibility**.
- Nevertheless, **finite-sample** "transition regime" for **SCQ**:
- For any

$$N \geq \hat{\sigma}_1^2 \mathcal{O}(\epsilon^{-2}) \ln(m/p),$$

it is guaranteed a $\mathcal{O}(\mathfrak{c}\epsilon)$-**near feasibility** and $\mathcal{O}(\epsilon)$-optimality.

- For *larger*

$$N \geq \hat{\sigma}_2^2 \mathcal{O}(\epsilon^{-2}) \vee \mathcal{O}(\mathring{\epsilon}(\bar{x})^{-2}) \ln(m/p),$$

exact feasibility and $\mathcal{O}(\epsilon) + \mathrm{gap}(\mathcal{O}(\epsilon))$-optimality is guaranteed with high-probability.

- If the problem admits an **interior solution**, then $\mathrm{gap}(\mathcal{O}(\epsilon))$ can be removed for positive small tolerances.

An application to the LASSO

Linear least-squares: $f(x) := \mathbf{P}[y(\cdot) - \langle \mathbf{x}(\xi), x \rangle]^2$.

Theorem (A persistent result for LASSO-type constraints with heavier tails)

Assume $(\mathbf{x}(\xi), y(\xi)) \in \mathbb{R}^d \times \mathbb{R}$ *is a random vector with finite* $7th$ *moments. Assume that there exist numbers* $C, u > 0$ *and* $p \in (0,1]$ *such that*

$$\forall v \in \mathbb{R}^d, \qquad \mathbf{P}\left\{\xi \in \Xi : |\langle v, \mathbf{x}(\xi)\rangle| > u\sqrt{\langle v, \mathbf{\Sigma} v\rangle}\right\} \geq p,$$

$$\forall 1 \leq \ell \leq d, \qquad \sqrt[7]{\mathbf{P}|\mathbf{x}(\cdot)[\ell]|^7} \leq C\sqrt[3]{\mathbf{P}|\mathbf{x}(\cdot)[\ell]|^3}.$$

Choose $R > 0$ *and confidence level* $\delta \in (0,1)$ *and define:*

$$\widehat{x}_{lasso} := argmin_{x \in \mathbb{R}^d}\left\{\widehat{F}(x) : \|\widehat{\mathbf{D}}_3 x\|_1 \leq R\right\}.$$

An application for the LASSO

Theorem (A persistent result for LASSO-type constraints with heavier tails)

Then $\exists C_0 = C_0(C, u, p) > 0$ *that the following holds. Let* N *so that*

$$\alpha := C_0 \left\{ \widehat{C} + 1 \right\} \sqrt{\frac{\ln(d/\delta)}{N}} \leq \frac{1}{2},$$

where $\widehat{C} := \max_{\ell \in [d]} \frac{\widehat{\mathbf{P}}\left(|\mathbf{x}(\cdot)[\ell]|^3 - \mathbf{P}|\mathbf{x}(\cdot)[\ell]|^3\right)^2}{(\mathbf{P}|\mathbf{x}(\cdot)[\ell]|^3)^2}$. *We define the "true solution"*

$$x_* := argmin_{x \in \mathbb{R}^d} \left\{ f(x) : \|\mathbf{D}_3 x\|_1 \leq (1+\alpha)R \right\},$$

and the "noise" $\epsilon(\xi) := y(\xi) - \langle x_*, \mathbf{x}(\xi) \rangle$. *Then with probability* $\geq 1 - \delta$,

$$\begin{aligned} \|\mathbf{D}_3 \widehat{x}_{lasso}\|_1 &\leq (1+\alpha)R, \\ f(\widehat{x}_{lasso}) - f(x_*) &\leq C_1 \left\{ \left[(\mathbf{P} + \widehat{\mathbf{P}}) \epsilon(\cdot)^6 \right]^{\frac{1}{6}} R \sqrt{\frac{\ln(d/\delta)}{N}} + R^2 \frac{\ln(d/\delta)}{N} \right\} \end{aligned}$$

References:

[1] R.I. Oliveira and P. Thompson (2017), *Sample average approximation with heavier tails I: non-asymptotic bounds with weak assumptions and stochastic constraints*, preprint at arxiv

[2] R.I. Oliveira and P. Thompson (2017), *Sample average approximation with heavier tails II: localization in stochastic convex optimization and persistence results for the Lasso*, preprint at arxiv

THANK YOU!