1 Statistical Properties of Sample Average Approximation Estimators

We consider

$$(P) \quad \min_{x \in X} \{ f(x) = \mathbb{E}[F(x,\xi)] \}, \tag{1}$$

a stochastic optimization problem.

- Assume $\emptyset \neq X \subset \mathbb{R}^n$, closed.
- ξ is a random vector with probability distribution P, support $\Xi \subset \mathbb{R}^d$.
- $F : X \times \Xi \to \mathbb{R}$ is the optimal value of a corresponding second-stage problem.
- Assume $f(x)$ is well defined for all $x \in X$ and finite, this implies that $F(x,\xi)$ is finite a.e. for $\xi \in \Xi$.
- We call ϑ the optimal value of (1), and S the set of optimal solutions of (2).

Consider $\{\xi^i\}_{i=1}^N$ and iid sample of ξ. We can re-write (1) using the empirical distribution generated by our sample; i.e., by assuming that $P_N(\xi = \xi^i) = \frac{1}{N}$; this leads to

$$(\hat{P}_N) \quad \min_{x \in X} \left\{ \hat{f}_N(x) = \mathbb{E}_{P_N}[F(x,\xi)] = \frac{1}{N} \sum_{i=1}^{N} F(x,\xi^i) \right\}, \tag{2}$$

which is called the sample average approximation (SAA) of the "true" problem (1).

- If we denote $z_{\hat{P}_N} = \hat{\vartheta}_N$ the optimal value of (2), we see that $\hat{\vartheta}_N$ is a random variable.
- If we call $\hat{S}_N$ the set of optimal solutions of (2), we see that $\hat{S}_N$ is a random set.

The question is how $\hat{\vartheta}_N$ and $\hat{S}_N$ relate to ϑ and S respectivelly.

Theorem 1 (Convergence)**.** *When we test for convergence we have the following basic results:*

- *By the Law of Large Numbers (LLN),* $\lim_{N\to\infty} \hat{f}_N(x) = f(x)$.
- *If X is compact; then the convergence is* uniform.
- *$\hat{f}_N(x)$ is unbiased; i.e.* $\mathbb{E}[\hat{f}_N(x)] = f(x)$.
- *A common assumption is that $F(x,\xi)$ is a* Carathéodory function*, i.e. continuous in x and* measurable *in ξ;*
- *If the previous holds; then $\hat{f}_N(x) = \hat{f}_N(x,\omega)$ is also a* Carathéodory function*; and $\hat{\vartheta}_N = \hat{\vartheta}_N(\omega)$, $\hat{S}_N = \hat{S}_N(\omega)$ are measurable.*

Theorem 2 (Bounds on convergence)**.** *Note that, by definition, $\hat{\vartheta}_N \leq \hat{f}_N(x)$ for any $x \in X$. If the LLN hlds; then*

$$\limsup_{N\to\infty} \hat{\vartheta}_N \leq \lim_{N\to\infty} \hat{f}_N(x) = f(x),$$

the inequality can be strict *without aditional hypothsis.*

Theorem 3 (Consistency of $\hat{\vartheta}_N$)**.** *If $\hat{f}_N(x)$ converges to $f(x)$ w.p. 1 as $N \to \infty$ uniformly on X. Then $\hat{\vartheta}_N \to \vartheta$ w.p. 1 as $N \to \infty$.*

Theorem 4 (Consistency of $\hat{S}_N$)**.** *If exists a compact set $C \subset \mathbb{R}^n$ such that:*

1. *$S \neq \emptyset$ and $S \subset C$.*
2. *$f(x) \in \mathscr{C}^o(C)$ and $f(x) \in \mathbb{R}, \forall x \in C$.*
3. *$\hat{f}_N(x)$ converges to $f(x)$ w.p. 1 as $N \to \infty$, uniformly in C.*
4. *w.p. 1, for N large enough, $\hat{S}_N \neq \emptyset$ and $\hat{S}_N \subseteq C$.*

Then, $\hat{\vartheta}_N \to \vartheta$ and $\mathbb{D}(\hat{S}_N, S) \to 0$ w.p. 1 as $N \to \infty$.

Under convexity, we can relax some of the previous conditions.

Theorem 5 (Consistency II)**.** *Suppose that:*

1. *F is random lower semi-continuous.*
2. *For a.e. $\xi \in \Xi$, the function $F(\cdot, \xi)$ is convex.*
3. *The set X is closed and convex.*
4. *f is lower semi-continuous, and $\exists x^o \in X$, $\varepsilon > 0$ such that $\forall x \in B(x^o, \varepsilon)$, $f(x) < \infty$.*
5. *$S \neq \emptyset$ and bounded.*
6. *LLN holds pointwise.*

Then, $\hat{\vartheta}_N \to \vartheta$ and $\mathbb{D}(S, \hat{S}_N) \to 0$ w.p. 1 as $N \to \infty$.

What happen if we have a problem where X must be estimated? i.e. we are working on a problem of the form

$$(\hat{P}_N) \quad \min_{x \in X_N} \left\{ \hat{f}_N(x) = \mathbb{E}_{P_N}[F(x,\xi)] = \frac{1}{N} \sum_{i=1}^{N} F(x, \xi^i) \right\}, \tag{3}$$

where $X_N = X_N(\omega)$ i.e. depends on the sample.

Theorem 6 (Consistency III)**.** *If exists a compact set $C \subset \mathbb{R}^n$ such that:*

1. *$S \neq \emptyset$ and $S \subset C$.*
2. *$f(x) \in \mathscr{C}^o(C)$ and $f(x) \in \mathbb{R}, \forall x \in C$.*
3. *$\hat{f}_N(x)$ converges to $f(x)$ w.p. 1 as $N \to \infty$, uniformly in C.*
4. *w.p. 1, for N large enough, $\hat{S}_N \neq \emptyset$ and $\hat{S}_N \subseteq C$.*

and also,

1. *if $x_N \in X_N$ and $x_N \to x$ w.p. 1, then $x \in X$.*
2. *For some $\bar{x} \in S$, there is a sequence of $x_N \in X_N$ such that $x_N \to x$ w.p. 1*

Then, $\hat{\vartheta}_N \to \vartheta$ and $\mathbb{D}(\hat{S}_N, S) \to 0$ w.p. 1 as $N \to \infty$.

Theorem 7 (First Order Asymptotics)**.** *If we have that:*

1. *$\exists \bar{x} \in X$ such that $\mathbb{E}[F(\bar{x}, \xi)] < \infty$.*

2. $\exists C : \Xi \to \mathbb{R}_+$ *measurable such that* $\mathbb{E}[C(\xi)^2] < \infty$ *and* $|F(x,\xi) - F(x',\xi)| \leq C(\xi)\|x - x'\|$ $\forall x, x' \in X$, $\xi \in \Xi$.

Note that the previous conditions imply that $|f(x) - f(x')| \leq \kappa\|x - x'\|$ **and that** $\mathbb{V}[F(x,\xi)] < \infty$**.**

Let $Y(x) \approx \mathcal{N}(0, \sigma^2(x))$*, then by CLT is easy to see that*

$$\frac{\hat{f}_N(x) - f(x)}{\sqrt{N}} \xrightarrow[N\to\infty]{\mathscr{D}} Y(x),$$

where $\sigma^2(x) = \mathbb{V}[F(x,\xi)]$.

Then,

$$\frac{\hat{\vartheta}_N - \vartheta}{\sqrt{N}} \xrightarrow[N\to\infty]{\mathscr{D}} \inf_{x\in S} Y(x)$$

Note that the previous theorem shows that the larger the set S, the larger the bias between $\hat{\vartheta}_N$ and ϑ. It also sugest that when approximating problems with domains that depend on the sample (for example constraints with probability or in expected value), it is better to use independent iid samples for each stochastic constraint and objective function.

1.1 Convergence to ε-optimal solutions

We start by the finite-domain case, where we will assume that:

- $|X| \in \mathbb{N}$. Note that this implies that $S \neq \emptyset$ and $\hat{S}_N \neq \emptyset$, and that X, X_N are compact sets.
- We define $S^\varepsilon = \{x \in X : f(x) \leq \theta + \varepsilon\}$ and $\hat{S}_N^\delta = \{x \in X_N : \hat{f}_N(x) \leq \hat{\vartheta}_N + \delta\}$.

We want to estimate

$$\mathbb{P}[\hat{S}_N^\delta \subset S^\varepsilon],$$

for that consider the following hypothesis:

M1: Assume that $f(x) = \mathbb{E}[F(x,\xi)] \leq \infty \forall x \in X$.

M2: Ξ is a bounded sub-set of $\mathbb{R}^d$.

Theorem 8 (Exponential convergence to ε-optimal solutions)**.** *If both* **M1,M2** *holds, then*

$$1 - \mathbb{P}[\hat{S}_N^\delta \subseteq S^\varepsilon] \leq |X| e^{-N\mu(\delta,\varepsilon)}.$$

There exists a $\varepsilon^* > 0$ *such that if* $\delta < \varepsilon < \varepsilon^*$*, then* $\mu(\delta\varepsilon) > 0$*. Also, under slightly different hypothesis, for some* $\delta > 0$ *and for all* $1 > \alpha > 0$ *is possible to show that if*

$$N \geq \frac{2\sigma}{(\varepsilon - \delta)^2} \log\left(\frac{|X|}{\alpha}\right),$$

then

$$\mathbb{P}[\hat{S}_N^\delta \subseteq S] \geq 1-\alpha$$

.

We look now into the more general continuous case, we will assume that:

- $X \subset \mathbb{R}^n$ is closed and bounded.
- $f(x) = \mathbb{E}[F(x,\xi)] < \infty \forall x \in X$.

And we consider the following assumptions:

M4 For any $x, x' \in X$, $\exists \sigma_{x',x} > 0$ such that $M_{x',x}(t) := \mathbb{E}[e^{tY(x',x)}]$, where $Y(x',x) := [F(x',\xi) - f(x')] - [F(x,\xi) - f(x)]$, satisfies

$$M_{x',x} \leq e^{\sigma_{x',x}^2 t^2/2}, \forall t \in \mathbb{R}.$$

M5 There exists a measurable function $\kappa : \Xi \to \mathbb{R}_+$ such that $M_\kappa(t) := \mathbb{E}[e^{t\kappa(\xi)}]$ is finite for some neighborhood of t equal zero and

$$|F(x',\xi) - F(x,\xi)| \leq \kappa(\xi)\|x' - x\|,$$

for a.e. $\xi \in \Xi$ and $x', x \in X$.

Theorem 9 (Exponential convergence to ε-optimal solutions)**.** *Suppose that all* **M1,M4,M5** *holds with* $\sigma^2 = \sup_{x',x \in X} \sigma_{x',x}^2 < \infty$*; that the diameter* D *of* X *is finite, i.e.* $D := \sup_{x',x \in X} \|x' - x\| < \infty$*. Then, there exists constants* $, \mu_o > 0, L > 0, \beta > 0, \varepsilon^* > 0$ *such that whenever* $\delta < \varepsilon < \varepsilon^*$ *and that*

$$N \geq \frac{8\sigma^2}{(\varepsilon-\delta)^2} \max\left[n \log\left(\frac{\mu_o L D}{\varepsilon - \delta}\right) + \log\left(\frac{2}{\alpha}\right), \frac{1}{\beta}\log(2\alpha)\right],$$

it follows that

$$\mathbb{P}[\hat{S}_N^\delta \subseteq S^\varepsilon] \geq 1-\alpha.$$

2 Stochastic Approximation Method

We will make the following assumptions:

- $f(x) := \mathbb{E}[F(x,\xi)] < \infty, \forall x \in X$, and $f \in \mathscr{C}^o(X)$.
- $\emptyset \neq X \subset \mathbb{R}^n$ is closed and bounded, this implies that $\exists \overline{x} \in S \subseteq X$ and that $f(\overline{x}) = \vartheta < \infty$.
- X is a convex set and $f(x)$ is a convex function.
- We assume the existance of the following *stochastic oracle*:

Definition 1 (Stochastic Oracle $\mathcal{O}(x,\xi)$)**.** *Given $x \in X$ and $\xi \in \Xi$ the stochastic oracle returns $F(x,\xi)$ and a stochastic sub-gradient vector $G(x,\xi)$ such that $g(x) := \mathbb{E}[G(x,\xi)]$ is well defined and is a subgradient of f at x, i.e. $g(x) \in \delta f(x)$.*

Note that if $F(x,\xi)$ is convex for every $\xi \in \Xi$ and x is an interior point of X; then $\delta f(x) = \mathbb{E}[\delta_x F(x,\xi)]$. We will assume that we can generate iid samples $\{\xi^i\}_{i=1}^N$ for arbitrarily large $N \in \mathbb{N}$.

We will denote $\Pi_X(x) = \arg\min_{z\in X} \|x-z\|_2$, i.e. $\Pi_X(x)$ is the l^2 (or metric) projection of x into X. Note that since X is convex and closed $\|\Pi_X(x) - \Pi_X(x')\|_2 \leq \|x-x'\|_2$.

The basic idea is the following:

Require: $x^o \in \mathbb{R}^n$, $k = 0$, $\{\gamma_k\}_{k\in\mathbb{N}}$, $\{\xi^k\}_{k\in\mathbb{N}}$.

1: **repeat**
2: $\quad x^{k+1} \leftarrow \Pi_X\left(x^k + \gamma_k G(x^k,\xi^k)\right)$, $k \leftarrow k+1$.
3: **until** some stopping rule
4: **return** x^k

We will assume that $\mathbb{E}[\|G(x,\xi)\|_2] \leq M^2$ for some $M > 0$, this implies that $\mathbb{E}[\|G(x,\xi)\|_2] \leq M$.

Theorem 10 (Convergence of Stochastic Approximation Method)**.** *If $f(x)$ is strongly convex with parameter c, and lipschitz continuous with parameter L,(this implies that $S = \{\bar{x}\}$), and $\gamma_j = \frac{1}{cj}$, and that $\bar{x}$ is interior to X; then*

$$\mathbb{E}[f(x^j) - f(\bar{x})] \leq \frac{L\max\{M^2/c^2, \|x^o - \bar{x}\|_2\}}{2j}.$$

One disadvantage is that the previous result require a great deal of knowledge of f, and moreover, the hypothesis are very restrictive; we consider the following robust stochastic approximation method:

Require: $x^o \in \mathbb{R}^n$, $k = 0$, $N \in \mathbb{N}$, $\gamma \in \mathbb{R}_+$, $\{\xi^k\}_{k=0}^N$.

1: **repeat**
2: $\quad x^{k+1} \leftarrow \Pi_X\left(x^k + \gamma_k G(x^k,\xi^k)\right)$, $k \leftarrow k+1$.
3: **until** $k = N$
4: **return** $x^{1,N} := \frac{1}{N}\sum_{k=1}^{N} x^k$

If we define $D_X = \max_{x\in X} \|x - x^o\|_2$, is possible to prove that:

$$\mathbb{E}[f(\hat{x}^{1,N}) - f(\bar{x})] \leq \frac{D_X^2 + M^2 N\gamma^2}{2N\gamma}.$$

Optimizing over $\gamma > 0$ we obtain that by setting $\gamma = \frac{D_X}{M\sqrt{N}}$ we obtain that

$$\mathbb{E}[f(\hat{x}^{1,N}) - f(\bar{x})] \leq \frac{D_X M}{\sqrt{N}}.$$

It is also possible to prove that

$$\mathbb{E}[f(\hat{x}^{K,N}) - f(\bar{x})] \leq \frac{C_{N,K} D_X M}{\sqrt{N}},$$

where $C_{N,K} = \frac{2N}{N-K+1} + \frac{1}{2}$. If we use a different step-size, scaled by a factor $\theta > 0$; then we can prove that

$$\mathbb{E}[f(\hat{x}^{K,N}) - f(\bar{x})] \leq \max\{\theta, \frac{1}{\theta}\} \frac{C_{N,K} D_X M}{\sqrt{N}}.$$

By choosing $K \approx \frac{1}{2}N$, then we have that

$$\mathbb{E}[f(\hat{x}_{K,N}) - f(\bar{x})] \leq \max\{\theta, \frac{1}{\theta}\} \frac{5 D_X M}{\sqrt{N}}$$

and then prove that the procedure is *robust* independent of the step-length chosen.

Also, it is possible to derive the following iterate estimate:

$$\mathbb{P}[f(\hat{x}^{1,N} - f(\bar{x}) \geq \varepsilon] \leq \frac{1}{\varepsilon}\mathbb{E}[f(\hat{x}_{1,N} - f(\bar{x}) \geq \varepsilon] \leq \frac{D_X M}{\varepsilon\sqrt{N}}$$

.

From where we can derive that for $\alpha \in (0,1)$ and $N \geq \frac{D_X^2 M^2}{\varepsilon^2 \alpha^2}$ we have that

$$\mathbb{P}[\hat{x}_{1,N} \in S^{\varepsilon}] \geq 1 - \alpha$$

Note that this estimate is weaker than that from the SAA algorithm; however, under some extra mild assumptions (for exampe Ξ compact), we can prove that there exists $K_o \geq 0$ such that for $N \geq \frac{K_o M^2 D_X^2}{\varepsilon^2} \log^2(\frac{1}{\alpha})$ we can guarantee that

$$\mathbb{P}[\hat{x}_{1,N} \in S^{\varepsilon}] \geq 1 - \alpha$$

Which is a close form to the sample-size estimate for SAA.

3 Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems *Jitka Dupačová, Roger Wets*, 1988

3.1 Definitions:

1. Let $(\Omega, \mathscr{A}, P)$ be a probability space, with Ω the support of P, with $\mathscr{A}$ the borel σ-field (σ-algebra) with respect to Ω.
2. We consider the problem
$$\min_{x\in\mathbb{R}^n} \mathbb{E}(f(x))$$
where
$$\mathbb{E}(f(x)) = \int_\Omega f(x,\xi)P(d\xi) = \mathbb{E}_\xi(f(x,\xi)).$$
and where
$$f : \mathbb{R}^n \times \Omega \to \mathbb{R} \cup \{\infty\}$$
   - This allow for constraints in the domain
3. We define $\mathrm{dom}\mathbb{E}(f) = \{x : \mathbb{E}(f(x)) < \infty\}$.
Note that $\mathrm{dom}\mathbb{E}(f) \subset \{x : f(x,\xi) < \infty a.s.\}$.
4. We say that $g : \mathbb{R}^n \to \mathbb{R}$ is lower semi-continuous (lsc) iff
$$x_k \xrightarrow[k\to\infty]{} x \Rightarrow \liminf_{k\to\infty} g(x^k) \geq g(x)$$
5. $f : \mathbb{R}^n \times \Omega \to \mathbb{R} \cup \{\infty\}$ is *random lower semicontinuous function* iff for all $\xi \in \Omega$ $f(\cdot,\xi)$ is lsc. and f is mesurable.

3.2 Assumptions:

1. Continuity:
   (a) $\mathrm{dom} f := \{(x,\xi) : f(x,\xi) < \infty\} = S \times \Omega$, $S \subset \mathbb{R}^n$ closed and non-empty.
   (b) For all $x \in S, \xi \to f(x,\xi)$ is continuous on Ω.
   (c) For all $\xi \in \Omega, x \to f(x,\xi)$ is lsc on $\mathbb{R}^n$ and locally lower lipschitz on S.
2. Convergence of probability measures:
   (a) Uff....

3.3 Results

1. Under the previous assumptions, there exists $Z_o \in \mathscr{A}$ such that $\mathbb{P}(Z_o) = 1$ and such that for all $\zeta \in Z_o$, $\mathbb{E}(f)$ and $\{\mathbb{E}^\nu(f)\}_{\nu\in\mathbb{N}}$ are *proper* lsc and such that $S := \mathrm{dom}\mathbb{E}(f) = \mathrm{dom}\mathbb{E}^\nu(f), \quad \forall \nu \in \mathbb{N}$.

2. Under the previous assumptions,

$$\forall x \in S \quad \mathbb{E}(f(x)) = \lim_{\nu\to\infty} \mathbb{E}^\nu(f(x))$$

3. Under the previous assumptions, $\mathbb{E}^n(f) : \mathbb{R}^n \times Z \to \mathbb{R} \cup \{\infty\}$ is μ-surelly random lower semicontinuous.
4. Under the previous assumptions

$$\limsup_{\nu\to\infty}(\inf \mathbb{E}^\nu(f)) \leq \inf \mathbb{E}(f).$$

Moreovoer, there exist Z_o with $\mu(Z \setminus Z_o) = 0$ such that:

(a) $\forall \zeta \in Z_o$ any cluster point of the sequence $\{x^\nu\}$ with $x^\nu \in \operatorname{argmin}\mathbb{E}^\nu f^\nu(\cdot, \zeta)$ belongs to $\operatorname{argmin}\mathbb{E}f$.

(b) for $\nu \in \mathbb{N}$ the application $\zeta \to \operatorname{argmin}\mathbb{E}^\nu f(\cdot, \zeta) : Z_o \rightrightarrows \mathbb{R}^n$ is closed and $\mathscr{F}^\nu$ mesurable.

(c) If there exists $D \subseteq \mathbb{R}^n$ compact such that

$$\operatorname{argmin}\mathbb{E}^\nu f \cap D \neq \emptyset \quad \mu - a.s.$$

and $\{x^*\} = \operatorname{argmin}\mathbb{E}f \cap D$, then there exist $x^\nu \in \operatorname{argmin}\mathbb{E}^\nu f$ such that $x^* = \lim_{\nu\to\infty} x^\nu$.

4 On the rate of convergence of optimal solutions of monte carlo approximations of stochastic programs, *Alexander Shapiro, Tito Homem-de-Mello*, 2000

4.1 Problem Description

1. We have the problem

$$(P) \qquad \min_{x\in\Theta}\{f(x) := \mathbb{E}_{\mathbb{P}} h(x, \omega)\},$$

where $\mathbb{P}$ is a probability measure on a space $(\Omega, \mathscr{F})$, $\Theta \subseteq \mathbb{R}^n$.
This is called the *true* optimization problem.

2. Assuming that we generate an identically distributed random sample $\{w^i\}_{i=1}^N$ in $(\Omega, \mathscr{F})$ according to $\mathbb{P}$, we can define the following *approximated* problem

$$(P_N) : \qquad \min_{x\in\Theta}\left\{\hat{f}_N(x) := \frac{1}{N}\sum_{i=1}^{n} h(x, \omega^i)\right\}.$$

We call this the *sampled* problem.

4.2 Known results

1. Let A be the set of optimal solutions for (P), and let $\hat{x}_N$ an optimal solution of (P_N).
2. Under similar assumptions as before:

$$\text{dist}(\hat{x}_N, A) \xrightarrow[N\to\infty]{} 0, \quad a.s.$$

3. Under mild assumptions, the rate of convergence of $\text{dist}(\hat{x}_N, A)$ to zero is $\mathscr{O}\left(N^{\frac{-1}{2}}\right)$.
4. Given $\varepsilon > 0$, we have that

$$\mathbb{P}(\text{dist}(\hat{x}_N, A) > \varepsilon) \xrightarrow[N\to\infty]{} 0,$$

and its convergence is exponentially fast.

4.3 New results:

1. Suppose that:
   (a) Ω is finite.
   (b) $\forall \omega \in \Omega$ $h(\cdot, \omega)$ is pice-wise linear and convex.
   (c) Θ is closed, convex and polihedral.
   (d) (P) has a non-empty, bounded optimal set A.

   Then:
   (a) A is compact, convex and polyhedral, and w.p.1 for N large enough, the set of optimal solutions A_N for (P_N) are non-empty and are a face of A.
   (b) There exist $\beta > 0$ such that

$$\limsup_{N\to\infty} \frac{1}{N} \log(\mathbb{P}(\{A_N \neq \emptyset, A_N \cap A = A_N\}^c) \leq -\beta$$

5 The Empirical Behavior of Sampling Methods for Stochastic Programming, *Jeff Linderoth, Alexander Shapiro, Stephen Wright,* 2002

5.1 Definiciones

1. We consider

$$(P) \min_{x\in X} \{f(x) := \mathbb{E}_{\mathbb{P}}(F(x, \xi(\omega)))\}$$

where $X \subseteq \mathbb{R}^n$, $\xi \in \mathbb{R}^d$, $F : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$.

2. Two-stage stochastic linear program with recourse:

$$\min_x c^t x + \mathscr{Q}(x) : Ax = b, x \geq 0,$$

where $\mathscr{Q}(x) := \mathbb{E}(Q(x, \xi(\omega)))$ and where $Q(x, \xi)$ is the optimal value of the second-stage problem

$$\min_y q^t y : Tx + Wy = h, \quad y \geq 0.$$

Note that (q, h, T, W) may depend on ω, (i.e. random).
3. If W is not random, the problem is said to have *fixed recourse*.
4. Note that if $F(x, \xi) = c^t x + \mathscr{Q}(x, \xi)$ and $X = \{x : Ax = b, x \geq 0\}$, and $\mathbb{P}(Q(x, \xi) < \infty | x \in X) = 1$, then the two-stage stochastic problem is a particular case of our general problem.

5.2 Supuestos

1. We will assume that $\xi(\omega)$ has finite support $\{\xi_i\}_{i=1}^K$.
2. We can write $\mathscr{Q}(x) = \sum_{k=1}^K p_k Q(x, \xi_k)$.
3. Under these assumptions, we can write

$$(P) \quad \begin{array}{lrcll} \min & c^t x + \sum_{k=1}^K p_q q_k^t y_k & & & \\ s.t. & Ax & = & b & \\ & T_k x + W y_k & = & h_k, & k = 1, \ldots, K \\ & x, y & \geq & 0 & \end{array}$$

4. If $\xi \in \{0,1\}^{100}$, $k = 2^{100} \approx 10^{30}$.

5.3 Lower Bounds

1. $\mathbb{E}(\hat{v}_N) \leq v^*$.
2. We could generate and estimator of $\mathbb{E}(\hat{v}_N)$ through M sampled problems, with $\xi_{k,j}$ iid, as $L_{N,M} := \frac{1}{M} \sum_{j=1}^M \hat{v}_N^j$.
3. By *central limit theorem* we have that

$$\sqrt{M}\,(L_{N,M} - \mathbb{E}(\hat{v}_N)) \xrightarrow[M \to \infty]{} N(0, \sigma_L^2)$$

where $\sigma_L^2 = \mathbb{V}(\hat{v}_N)$.

4. We can use $s_L^2 := \frac{1}{M-1} \sum_{j-1}^{M} \left(\hat{v}_N^j - L_{N,M} \right)^2$ as an unbiased variance estimator.
5. We could assume (for small M) that we have a t-student.
6. We can use confidence intervals.

5.4 Upper Bounds

1. By definition, given $\hat{x} \in X$, $f(\hat{x}) \geq v^*$.
2. Assume we have $\{\xi^{i,j}\}_{j=1..T, i=1..N}$ iid, then we have that

$$\mathbb{E}\left(\hat{f}_N^j := \frac{1}{N} \sum_{i=1}^{N} F(x, \xi^{ij}) \right)$$

and also we can define

$$U_{N,T} := \frac{1}{T} \sum_{j=1}^{T} \hat{f}_N^j$$

3. use same trick as before.
4. We can estimate the optimality GAP now....

5.5 Problems

1. 20term: vehicle assignment with 1.2×10^{12} scenarios, 3×64 first stage, 124×756 second stage.
2. gdb: Aircraft allocation with 6.5×10^5 scenarios, 4×17 first stage, 5×10 second stage.
3. LandS: Electricity Planning with 1.0×10^6 scenarios, 2×4 first stage, 7×12 second stage.
4. ssn: Telecom Network Design with 1.0×10^{70} scenarios, 1×89 first stage, 175×706 second stage.
5. storm: Cargo Flight Scheduling with 6.0×10^{81} scenarios, 185×121 first stage, 528×1259 second stage.

5.6 Results

1. Use $T = 50, N = 2000$
2. Use $M = 7...10, N = 50, 100, 500, ...$

Prob	N	$\mathbb{E}(\hat{v}_N)$ 95%	best $\hat{f}(x_N^j)$ 95%
20term	50	$253.361 \pm .944$	254.317 ± 0.019
20term	500	$254.324 \pm .194$	254.320 ± 0.027
20term	5000	$254.340 \pm .085$	254.341 ± 0.020
gdb	50	167.862 ± 6.673	165.585 ± 0.134
gdb	500	164.966 ± 1.360	165.490 ± 0.146
gdb	5000	165.313 ± 0.437	165.640 ± 0.131
ssn	50	4.11 ± 1.23	12.68 ± 0.05
ssn	500	8.54 ± 0.34	10.28 ± 0.04
ssn	5000	9.98 ± 0.21	9.86 ± 0.05
storm	50	155.062 ± 0.220	154.990 ± 0.008
storm	500	154.981 ± 0.041	154.984 ± 0.006
storm	5000	154.981 ± 0.018	154.986 ± 0.006

6 The Sample Average Approximation Method for Stochastic Discrete Optimization, *Anton Kleywegt, Alexander Shapiro*, SIAM Journal of Optimization, 2001

6.1 The problem

$$(P) \quad \min_{x \in S} g(x) := \mathbb{E}_{\mathbb{P}}(G(x,w))$$

Con w vector aleatoreo con distribución de probabilidad $\mathbb{P}$, $G(x,w) : S \times \Omega \to \mathbb{R}$ y $|S| \in \mathbb{N}$. Asumimos que $g(x)$ esta bien definido, i.e. $G(x,\cdot)$ es $\mathbb{P}$-medible, y que $\mathbb{E}(|G(x,w)|) < \infty$. En particular estaremos interesados en problemas donde $g(x)$ es difícil de computar, pero $G(x,w)$ es fácil de evaluar.

6.2 Resultados

- Definiendo $\hat{g}_N(x) := \frac{1}{N}\sum_i = 1^N G(x,w^i)$ y el problema

  $$(\hat{P}_N) \quad \min_{x \in S} hat g_N(x)$$

  como el problema aproximado, y llamando S^* el conjunto de soluciones óptimas de (P) y $\hat{S}_N^*$ el conjunto de soluciones óptimas de $(\hat{P}_N)$, S_ε el conjunto de soluciones ε-óptimas de (P), respectivamente $\hat{S}_N^\varepsilon$ y $z_P = v^*, \hat{v}_N = z_{\hat{P}_N}$, entonces:
  - $\mathbb{E}(\hat{g}_N(x)) = g(x)$.
  - $\hat{v}_N \xrightarrow[N\to\infty]{} v^*$ w.p.1.

– $\mathbb{P}(\hat{S}_N^{\varepsilon} \subset S_{\varepsilon}) \xrightarrow[N\to\infty]{} 1$.
– Dada una variable aleatorea X, y un sample iid $\{x^i\}_{i=1}^n$, y definiendo $Z_N = \frac{1}{N}\sum_{i=1}^{N}$, entonces

$$\frac{1}{N}\log(\mathbb{P}(Z_N \geq a)) \leq -I(a)$$

donde $I(a) := \sup_{t\geq 0}\{tz - \log(\mathbb{E}(e^{tX}))\}$. Note que si $\mathbb{E}(X) = \mu$, entonces $I(\mu) = 0$. Además, si la función de momentos esta bien definida, $I'(0) = \mu$. De donde uno puede demostrar que para $a > \mu$ se tiene que $I(a) > 0$.
– Asumiendo que en torno de $t = 0$ $M(t) := \mathbb{E}(e^{tX})$ es diferenciable, entonces existen constantes $\gamma_x > 0$ y $\gamma_x' > 0$ tal que

$$\mathbb{P}(|\hat{g}_N(x) - g(x)| \geq \varepsilon/2) \leq e^{-N\gamma_x} + e^{-N\gamma_x'}$$