International Conference
OPTIMIZATION AND CONTROL
Dedicated to 70th Birthday of Prof. Boris Polyak
Moscow May 19 – 21, 2005

CONVEX APPROXIMATIONS OF CHANCE CONSTRAINTS

Arkadi Nemirovski, Technion, Haifa, Israel
nemirovs@ie.technion.ac.il
Joint Research with Alexander Shapiro, Georgia Institute of Technology, Atlanta, USA
ashapiro@isye.gatech.edu

Overview:

- **Chance constraints – motivation and difficulties**
- **Safe approximations of Chance constraints**
- **Bernstein Approximation**
- **Ambiguous Chance constraints**

♣ **Typically, optimization problems are of the form**

$$\min_x \{f_0(x,\xi) : f_i(x,\xi) \le 0,\ i = 1,...,I\} \qquad (P_\xi)$$

• $x \in \mathbf{R}^n$ – **decision vector** • $\xi \in \mathbf{R}^N$ – **data** • $f_i(x,\xi)$ **- given functions.**

Example: Linear Programming: $f_i(x,\xi)$ **are affine in** x**, and** ξ **is the collection of coefficients of these affine functions.**

♣ **More often than not the data in real-life optimization problems are** *uncertain* – **not known exactly when the problem should be solved. There are two major ways to treat this data uncertainty:**

♠ *Uncertain-but-bounded data: ξ runs through a given bounded "uncertainty set"* $\mathcal{U} \subset \mathbf{R}^N$.

Here it is natural to require from a candidate solution to remain feasible, whatever be a realization $\xi \in \mathcal{U}$ **(Robust Optimization).**

♠ *Stochastic data: ξ is random with a given distribution P*.

Here it is natural to replace the original constraints with their *chance* **versions:**

$$\mathrm{Prob}_{\xi\sim P}\{\xi : f_i(x,\xi) > 0\} \le \epsilon,$$

where $\epsilon << 1$ **is a given reliability.**

$$\min_x \{f_0(x,\xi) : f_i(x,\xi) \le 0,\ i=1,...,I\}, \qquad \xi \sim P$$

$$\Downarrow$$

$$\min_{x,\tau}\left\{\tau : \begin{array}{l}\mathrm{Prob}\{f_0(x,\xi) > \tau\} \le \epsilon \\ \mathrm{Prob}\{f_i(x,\xi) > 0\} \le \epsilon,\ i=1,...,I\end{array}\right\} \qquad (*)$$

♣ **<u>Difficulties with chance constraints</u>:**

♠ **Aside of few very specific cases, it is difficult to check whether a chance constraint, even a "linear" in ξ one:**

$$p(x) \equiv \mathrm{Prob}\left\{f^0(x) + \sum_{j=1}^N \xi_j f^j(x) > 0\right\} \le \epsilon$$

is satisfied at a given point:

- **when $N >> 1$, a "closed form" analytical expression for $p(x)$ is usually out of question;**
- **Monte-Carlo-based estimation of $p(x)$ is prohibitively time-consuming when ϵ is really small (like 10^{-6} or less)**

♠ **The feasible set of a chance constraint usually is nonconvex, which makes optimization over this set highly problematic.**

$$p(x) \equiv \mathrm{Prob}\{f(x,\xi) > 0\} \leq \epsilon\} \qquad (*)$$

♣ **A natural way to overcome, to some extent, severe computational difficulties with chance constraint (*) is to replace it with its** *computationally tractable safe approximation* – **an efficiently computable** *convex* **constraint**

$$F(x) \leq 0 \qquad (**)$$

which is a *sufficient condition* **for** $(*)$**:**

$$x \text{ satisfies } (**) \Rightarrow x \text{ satisfies } (*).$$

$$p(x) \equiv \text{Prob}\{f(x,\xi) > 0\} \leq \epsilon \qquad (*)$$

♣ **The simplest way to build a safe approximation of chance constraint** $(*)$ *with convex in x function $f(x,\xi)$* **is as follows:**

♠ **Let $\phi(s) : \mathbf{R} \to \mathbf{R}$ be convex, nondecreasing and such that**

$$\phi(s) \geq 0 \ \forall s \ \& \ \phi(0) \geq 1.$$

Then

$$\alpha > 0 \Rightarrow p_\alpha(x) \equiv \mathbf{E}_\xi\left\{\phi\left(\alpha^{-1} f(x,\xi)\right)\right\} \geq p(x).$$

$\Rightarrow$ **For every $\alpha > 0$ the constraint $p_\alpha(x) \leq \epsilon$ is a safe approximation of $(*)$, and therefore so is the constraint**

$$\exists \alpha > 0 : \underbrace{\alpha p_\alpha(x) - \alpha\epsilon}_{\Psi(x,\alpha)} \leq 0. \qquad (1)$$

♠ **Due to its origin, $\Psi(x,\alpha)$ is convex in $(x, \alpha > 0)$, so that the function**

$$\Phi(x) \equiv \inf_{\alpha>0} \Psi(x,\alpha) \equiv \inf_{\alpha>0}\left[\mathbf{E}_\xi\left\{\alpha\phi\left(\alpha^{-1} f(x,\xi)\right)\right\} - \alpha\epsilon\right]$$

also is convex in x. It is easily seen that the (slightly weakened, as compared to (1)) convex constraint $\Phi(x) \leq 0$ is a safe approximation of $(*)$.

$$\Phi(x) \equiv \inf_{\alpha>0} \Psi(x,\alpha) \equiv \inf_{\alpha>0} \left[\mathbf{E}_{\xi}\left\{\alpha\phi\left(\alpha^{-1}f(x,\xi)\right)\right\} - \alpha\epsilon\right] \leq 0 \quad (**)$$

$$\Downarrow$$

$$\mathrm{Prob}\{f(x,\xi) > 0\} \leq \epsilon \qquad (*)$$

[$\phi : \mathbf{R} \to \mathbf{R}$: **convex,nonnegative,nonincreasing and** $\phi(0) \geq 1$]

♠ **Function** $\Phi(x)$ **is convex. therefore,** *if* $\Phi(\cdot)$ *is efficiently computable* **(or, which is essentially the same, the expectation** $\mathbf{E}_{\xi}\{\alpha^{-1}f(x,\xi)\}$ **is so),** *convex constraint* $(**)$ *is a computationally tractable safe approximation of chance constraint* $(*)$.

♣ **Question: How to choose** $\phi(\cdot)$**?**
♣ **Answer:** *As far as tightness of approximation is concerned, the optimal choice of* $\phi(\cdot)$ *is*

$$\phi(s) = \max[0, 1+s]. \qquad (!)$$

This choice results exactly in the Rockafellar's "conditional value at risk"-based upper bound on $\mathrm{Prob}\{f(x,\xi) > 0\}$**.**
However: It usually is difficult to compute the function $\Phi(\cdot)$ **associated with (!)...**

$$p(x) \equiv \mathrm{Prob}\{f(x,\xi) > 0\} \le \epsilon \qquad (*)$$

♣ **<u>Bernstein bound:</u>** $\phi(s) = \exp\{s\}$**. Let**

- $f(x,\xi)$ **in** $(*)$ **be affine in** ξ**:**

$$f(x,\xi) = f^0(x) + \sum_{j=1}^{N} \xi_j f^j(x)$$

- $\xi_1, \xi_2, ...\xi_N$ **be independent of each other random scalars with given distributions** P_j**, and**
- P_j **possess finite explicitly computable (logarithmic) moment-generating functions:**

$$\pi_j(z) = \log\left(\mathbf{E}_{s\sim P_j}\{\exp\{zs\}\}\right) : \mathbf{R} \to \mathbf{R}.$$

For every $\alpha > 0$, we have

$$\log p(x) \leq \log\left(\mathbf{E}\left\{\exp\{\alpha^{-1} f(x,\xi)\}\right\}\right) = \alpha^{-1} f^0(x) + \sum_{j=1}^{N} \pi_j(\alpha^{-1} f^j(x))$$

$$\Downarrow$$

$$\boxed{\begin{array}{c}\exists \alpha > 0 : \alpha^{-1} f^0(x) + \sum_{j=1}^{N} \pi_j(\alpha^{-1} f^j(x)) \leq \log(1/\epsilon) \\ \Downarrow \\ \mathrm{Prob}\{f(x,\xi) > 0\} \leq \epsilon\end{array}}$$

$$\Updownarrow$$

$$\boxed{\begin{array}{c}\exists \alpha > 0 : \underbrace{f^0(x) + \sum_{j=1}^{N} \alpha \pi_j(\alpha^{-1} f^j(x)) - \alpha \log(1/\epsilon)}_{\Psi(x,\alpha)} \leq 0 \\ \Downarrow \\ \mathrm{Prob}\{f(x,\xi) > 0\} \leq \epsilon\end{array}}$$

$$\Rightarrow \left\{\Phi(x) \equiv \inf_{\alpha > 0} \Psi(x,\alpha) \leq 0\right\} \;\Rightarrow\; \mathrm{Prob}\{f(x,\xi) > 0\} \leq \epsilon$$

♠ **Assuming that • all $f^j(\cdot)$ are convex, and • $\xi_j \geq 0$ a.s. for all $j \geq 1$ with non-affine $f^j(x)$, function $\Psi(x,\alpha)$ is convex and efficiently computable $\Rightarrow$ The constraint $\Phi(x) \leq 0$ is a safe computationally tractable approximation of the chance constraint $\mathrm{Prob}\{f(x,\xi) > 0\} \leq \epsilon$.**

$$\text{Prob}\left\{f^0(x) + \sum_{j=1}^{N} \xi_j f^j(x) > 0\right\} \le \epsilon \qquad (*)$$
$$[\xi \sim P_1 \times ... \times P_N]$$

♣ **What to do when the distributions P_j are *not* known exactly, and all we know is that the tuple $P^N = (P_1, ..., P_N)$ of distributions of (independent of each other, and, say, bounded a.s. by 1) random variables $\xi_1, ..., \xi_N$ belongs to a given *-compact convex set $\mathcal{P}$?**

♠ **Let us associate with $(*)$ the** *ambiguous chance constraint*

$$\sup_{(P_1,...,P_N)\in\mathcal{P}} \text{Prob}_{\xi\sim P_1\times...\times P_N}\left\{f^0(x) + \sum_{j=1}^{N} \xi_j f^j(x) > 0\right\} \le \epsilon \qquad (!)$$

♠ **It can be shown that the convex constraint**

$$f^0(x) + \inf_{\alpha>0} \max_{(P_1,...,P_N)} \underbrace{\left[\alpha \sum_{j=1}^{N} \log\left(\mathbf{E}_{\xi_j\sim P_j}\left\{\exp\{\alpha^{-1} f^j(x)\xi_j\}\right\}\right) - \alpha\log(1/\epsilon)\right]}_{\text{convex in } (x,\alpha>0),\text{ concave in } P_j} \le 0$$

is a safe approximation of (!). This approximation is computationally tractable, provided that $\mathcal{P}$ is so.

$$\sup_{(P_1,\dots,P_N)\in\mathcal{P}} \mathrm{Prob}_{\xi\sim P_1\times\dots\times P_N}\left\{f^0(x)+\sum_{j=1}^N \xi_j f^j(x)>0\right\}\leq\epsilon \qquad (!)$$

♣ **Example: Let $\mathcal{P}=\mathcal{P}_1\times\dots\times\mathcal{P}_N$ with convex *-compact sets $\mathcal{P}_j$ of Borel probability distributions on $[-1,1]$. Then a safe convex approximation of (!) is given by**

$$f^0(x)+\inf_{\alpha>0}\left[\sum_{j=1}^N \alpha\widehat{\pi}_j\left(\alpha^{-1}f^j(x)\right)-\alpha\log(1/\epsilon)\right]\leq 0$$

$$\widehat{\pi}_j(z)=\max_{P_j\in\mathcal{P}_j}\log\left(\mathbf{E}_{s\sim P_j}\left\{\exp\{sz\}\right\}\right)$$

♠ **In many interesting cases, the function**

$$\Pi^{\mathcal{P}}(z)=\max_{P\in\mathcal{P}}\mathbf{E}_{s\sim P}\left\{\exp\{sz\}\right\}$$

is easy to compute:

$\mathcal{P}$	$\Pi^{\mathcal{P}}(z)$
$\{P:\mathrm{supp}\,(P)\subset[-1,1]\}$	$\exp\{\|z\|\}$
$\left\{P: \begin{array}{l}\mathrm{supp}\,(P)\subset[-1,1]\\ P \textbf{ is symmetric w.r.t. } \mathbf{0}\end{array}\right\}$	$\cosh(t)$
$\left\{P: \begin{array}{l}\mathrm{supp}\,(P)\subset[-1,1], P \textbf{ is}\\ \textbf{unimodal w.r.t. } \mathbf{0}\end{array}\right\}$	$\dfrac{\exp\{\|t\|\}-1}{\|t\|}$

$\mathcal{P}$	$\Pi^{\mathcal{P}}(z)$
$\left\{P: \begin{array}{l}\operatorname{supp}(P)\subset[-1,1], P \textbf{ is unimodal}\\ \textbf{w.r.t. 0 and symmetric}\end{array}\right\}$	$\dfrac{\sinh(t)}{t}$
$\left\{P: \begin{array}{l}\operatorname{supp}(P)\subset[-1,1]\\ \mathrm{Mean}[P]=\gamma\end{array}\right\}$	$\cosh(t)+\gamma\sinh(t)$
$\left\{P: \begin{array}{l}\operatorname{supp}(P)\subset[-1,1]\\ \gamma_-\leq\mathrm{Mean}[P]\leq\gamma_+\end{array}\right\}$	$\cosh(t)+\max\left[\gamma_-\sinh(t),\gamma_+\sinh(t)\right]$
$\left\{P: \begin{array}{l}\operatorname{supp}(P)\subset[-1,1]\\ \mathrm{Mean}[P]=0\\ \mathrm{Var}[P]\leq\sigma^2\end{array}\right\}$	$\dfrac{\exp\{-\lvert t\rvert\sigma^2\}+\sigma^2\exp\{\lvert t\rvert\}}{1+\sigma^2}$
$\left\{P: \begin{array}{l}\operatorname{supp}(P)\subset[-1,1], P \textbf{ is}\\ \textbf{symmetric,}\ \mathrm{Var}[P]\leq\sigma^2\end{array}\right\}$	$\sigma^2\cosh(t)+(1-\sigma^2)$
$\left\{P: \begin{array}{l}\operatorname{supp}(P)\subset[-1,1]\\ \mathrm{Mean}[P]=\gamma\\ \mathrm{Var}[P]\leq\sigma^2\end{array}\right\}$	$\begin{cases}\dfrac{(1-\gamma)^2\exp\{t\frac{\gamma-\sigma^2}{1-2\gamma+\sigma^2}\}+(\sigma^2-\gamma^2)\exp\{t\}}{1-2\gamma+\sigma^2}, & t\geq 0\\ \dfrac{(1+\gamma)^2\exp\{t\frac{\gamma+\sigma^2}{1+2\gamma+\sigma^2}\}+(\sigma^2-\gamma^2)\exp\{-t\}}{1+2\gamma+\sigma^2}, & t\leq 0\end{cases}$

♣ **Essentially, the Bernstein approximation deals with affine in ξ chance constraint**

$$p(z) \equiv \mathrm{Prob}_{\xi\sim P_1\times\ldots\times P_N}\{z_0+\xi_1 z_1+\ldots+\xi_N z_N>0\}\leq\epsilon \qquad (*)$$

in variables z and states that the feasible set $\mathcal{Z}_{\mathrm{Bern}}$ of the convex constraint

$$z_0+\inf_{\alpha>0}\left[\sum_j \alpha\log\left(\mathbf{E}_{\xi_j\sim P_j}\left\{\exp\{\xi_j\alpha^{-1}z_j\}\right\}\right)-\alpha\ln(1/\epsilon)\right]\leq 0 \qquad (**)$$

is inside the feasible set $\mathcal{Z}$ of the chance constraint $(*)$.

♣ **<u>Question:</u> How conservative is the approximation?**

♠ **When interpreting the question straightforwardly:**

How large should be the ratio $\epsilon/p(z)$ in order for Bernstein approximation to be able to certify the inequality $p(z)\leq\epsilon$?

the answer is "fully pessimistic": already in the simplest case where $N=1$ and ξ_1 takes just two values $0,1$, for every $\epsilon\in(0,1)$ and every $T>1$ it may happen that $p(z)<\epsilon/T$ and the Bernstein approximation is unable to certify that $p(z)\leq\epsilon$.

<u>However:</u> There are alternative meaningful interpretations of "level of conservatism".

♣ **Assume ξ_j are symmetrically distributed, and let us embed the original chance constraint in a natural parametric family:**

$$p_\rho(z) \equiv \mathrm{Prob}_{\xi\sim P_1\times\ldots\times P_N}\left\{z_0 + \rho\sum_j \xi_j z_j > 0\right\} \leq \epsilon$$
$$\Downarrow$$
$$\mathcal{Z}[\rho] = \{z : p_\rho(z) \leq \epsilon\}$$

• $\rho \geq 0$**: "noise intensity". Assuming** $\epsilon < 1/2$**, the sets** $\mathcal{Z}[\rho]$ **shrink as noise uncertainty** ρ **grows.**

♠ **Let**

$$\mathcal{Z}_{\mathrm{Bern}}[\rho] = \left\{z : z_0 + \inf_{\alpha>0}\left[\sum_j \alpha \log\left(\mathbf{E}_{\xi_j\sim P_j}\left\{\exp\{\rho\xi_j\alpha^{-1}z_j\}\right\}\right) - \alpha\ln(1/\epsilon)\right] \leq 0\right\}$$

be the feasible sets of the Bernstein approximation. Then

$$\mathcal{Z}_{\mathrm{Bern}}[\rho] \subset \mathcal{Z}[\rho]$$

and perhaps

$$\mathcal{Z}[\rho^+] \subset \mathcal{Z}_{\mathrm{Bern}}[\rho] \qquad (!)$$

for appropriately chosen $\rho^+ \geq \rho$**. It is natural to measure the conservatism of approximation by the quantity**

$$\frac{\inf\{\rho^+ : \mathcal{Z}[\rho^+] \subset \mathcal{Z}_{\mathrm{Bern}}[\rho]\}}{\rho}$$

$$p_\rho(z) \equiv \mathrm{Prob}_{\xi\sim P_1\times\ldots\times P_N}\left\{z_0 + \rho\sum_j \xi_j z_j > 0\right\} \leq \epsilon$$

$$\Downarrow$$

$$\mathcal{Z}[\rho] = \{z : p_\rho(z) \leq \epsilon\}$$

Theorem. *Let ξ_j be uniformly distributed in segments $[-\sigma_j, \sigma_j]$, and let $\epsilon \leq 0.05$. Then*

$$\mathcal{Z}\left[18\sqrt{\ln(1/\epsilon)}\rho\right] \subset \mathcal{Z}_{\mathrm{Bern}}[\rho] \subset \mathcal{Z}[\rho].$$