Chance constraints and distributionally robust optimization

- chance constraints
- approximations to chance constraints
- distributional robustness

Chance constraints

non-convex problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & \mathbf{Prob}(f_i(x,U) > 0) \le \epsilon, \quad i = 1, \ldots, m \end{array}$$

safe approximation: find convex $g_i : \mathbf{R}^n \to \mathbf{R}$ such that

$$g_i(x) \le 0 \ \text{ implies } \ \mathbf{Prob}(f_i(x,U) > 0) \le \epsilon$$

- sufficient condition: find set $\mathcal{U}$ such that

$$\mathbf{Prob}(U \in \mathcal{U}) \ge 1 - \epsilon \ \text{ set } \ g_i(x) = \sup_{u \in \mathcal{U}} f_i(x,u)$$

Bounds on probability of error

- sometimes useful to directly bound $\mathbf{Prob}(f_i(x,U) > 0)$ instead of $\mathbf{Prob}(U \in \mathcal{U})$

- **Value at risk** of random Z is

$$\mathbf{VaR}(Z;\epsilon) = \inf\{\gamma \mid \mathbf{Prob}(Z \leq \gamma) \geq 1-\epsilon\} = \inf\{\gamma \mid \mathbf{Prob}(Z > \gamma) \leq \epsilon\}$$

- equivalence of $\mathbf{VaR}$ and deviation:

$$\mathbf{VaR}(Z;\epsilon) \leq 0 \;\text{ if and only if }\; \mathbf{Prob}(Z > 0) \leq \epsilon$$

- Gaussians: if $U \sim \mathcal{N}(\mu, \Sigma)$ then $x^T U - \gamma \sim \mathcal{N}(\mu^T x - \gamma, x^T \Sigma x)$ and

$$\mathbf{Prob}(x^T U \leq \gamma) = \Phi\left(\frac{\gamma - x^T U}{\sqrt{x^T \Sigma x}}\right)$$

so

$$\mathbf{VaR}(U^T x - \gamma; \epsilon) \leq 0 \;\text{ iff }\; \gamma \geq \mu^T x + \Phi^{-1}(1-\epsilon)\left\|\Sigma^{1/2} x\right\|_2$$

convex iff $\epsilon \leq 1/2$

Convex bounds on probability of error

- sometimes more usable idea: convex upper bounds on probability of error
- simple observation: if $\phi : \mathbf{R} \to \mathbf{R}$ is non-negative, non-decreasing

$$1\,(z \geq 0) \leq \phi(z)$$

- consequence: for all $\alpha > 0$,

$$\mathbf{Prob}(Z \geq 0) \leq \mathbf{E}\phi(\alpha^{-1}Z),$$

- so if

$$\mathbf{E}\phi(\alpha^{-1}Z) \leq \epsilon, \quad \text{then} \quad \mathbf{Prob}(Z \geq 0) \leq \epsilon$$

Perspective transforms and convex bounds

- perspective transform of function $f : \mathbf{R}^n \to \mathbf{R}$ is

$$f_{\text{per}}(x, \lambda) = \lambda f\left(\frac{x}{\lambda}\right)$$

- jointly convex in $(x, \lambda) \in \mathbf{R}^n \times \mathbf{R}_+$ when f convex

- so for $\phi(\cdot) \geq 1\,(\cdot)$, better convex constraint (valid for all $\alpha > 0$)

$$\alpha \mathbf{E} \phi\left(\frac{f(x, U)}{\alpha}\right) \leq \alpha \epsilon$$

  is convex in x, α if f is

- optimize this bound,

$$\inf_{\alpha \geq 0} \left\{ \alpha \mathbf{E} \phi \left(\frac{f(x,U)}{\alpha} \right) - \alpha\epsilon \right\} \leq 0$$

- *convex* constraint satisfied implies that

$$\mathbf{Prob}(f(x,U) > 0) \leq \epsilon$$

Tightest convex relaxation

- set $\phi(z) = [1+z]_+$, where $[x]_+ = \max\{x, 0\}$

$$\inf_{\alpha \geq 0} \left\{ \alpha \mathbf{E} \left[\frac{f(x,U)}{\alpha} + 1 \right]_+ - \alpha\epsilon \right\} = \inf_{\alpha \geq 0} \left\{ \mathbf{E} \left[f(x,U) + \alpha \right]_+ - \alpha\epsilon \right\}$$

- **conditional value at risk** is

$$\mathbf{CVaR}(Z;\epsilon) = \inf_{\alpha} \left\{ \frac{1}{\epsilon} \mathbf{E} \left[Z - \alpha \right]_+ + \alpha \right\},$$

- key inequalities:

$$\mathbf{Prob}(Z \geq 0) - \epsilon \leq \epsilon \mathbf{CVaR}(Z;\epsilon)$$

Interpretation of conditional value at risk

- minimize out α and find

$$0 = \frac{\partial}{\partial \alpha}\left\{\alpha + \frac{1}{\epsilon}\mathbf{E}\left[Z-\alpha\right]_+\right\} = 1 - \frac{1}{\epsilon}\mathbf{E}1\left(Z \geq \alpha\right) = 1 - \frac{1}{\epsilon}\mathbf{Prob}(Z \geq \alpha).$$

- value at risk plus upward deviations: set $\alpha^\star$ s.t. $\epsilon = \mathbf{Prob}(Z \geq \alpha^\star)$,

$$\mathbf{CVaR}(Z;\epsilon) = \frac{1}{\epsilon}\mathbf{E}\left[Z-\alpha^\star\right]_+ + \alpha^\star = \frac{1}{\epsilon}\mathbf{E}\left[Z-\alpha^\star\right]_+ + \mathbf{VaR}(Z;\epsilon)$$

- conditional expectation version:

$$\begin{aligned}\mathbf{E}[Z \mid Z \geq \alpha^\star] &= \mathbf{E}[\alpha^\star + (Z-\alpha^\star) \mid Z \geq \alpha^\star] \\ &= \alpha^\star + \frac{\mathbf{E}\left[Z-\alpha^\star\right]_+}{\mathbf{Prob}(Z \geq \alpha^\star)} = \alpha^\star + \frac{\mathbf{E}\left[Z-\alpha^\star\right]_+}{\epsilon} = \mathbf{CVaR}(Z;\epsilon).\end{aligned}$$

Benefits and drawbacks of CVaR

- easy to simulate, approximate well
- typically hard to evaluate exactly; not very tractable bounds
- e.g. $f(x, U) = U^T x$ and $U \sim$ Uniform$\{-1, 1\}^n$ gives combinatorial sum
- if available, analytic approximations using moment generating function (MGF) can give better behavior (notes)

Portfolio optimization example

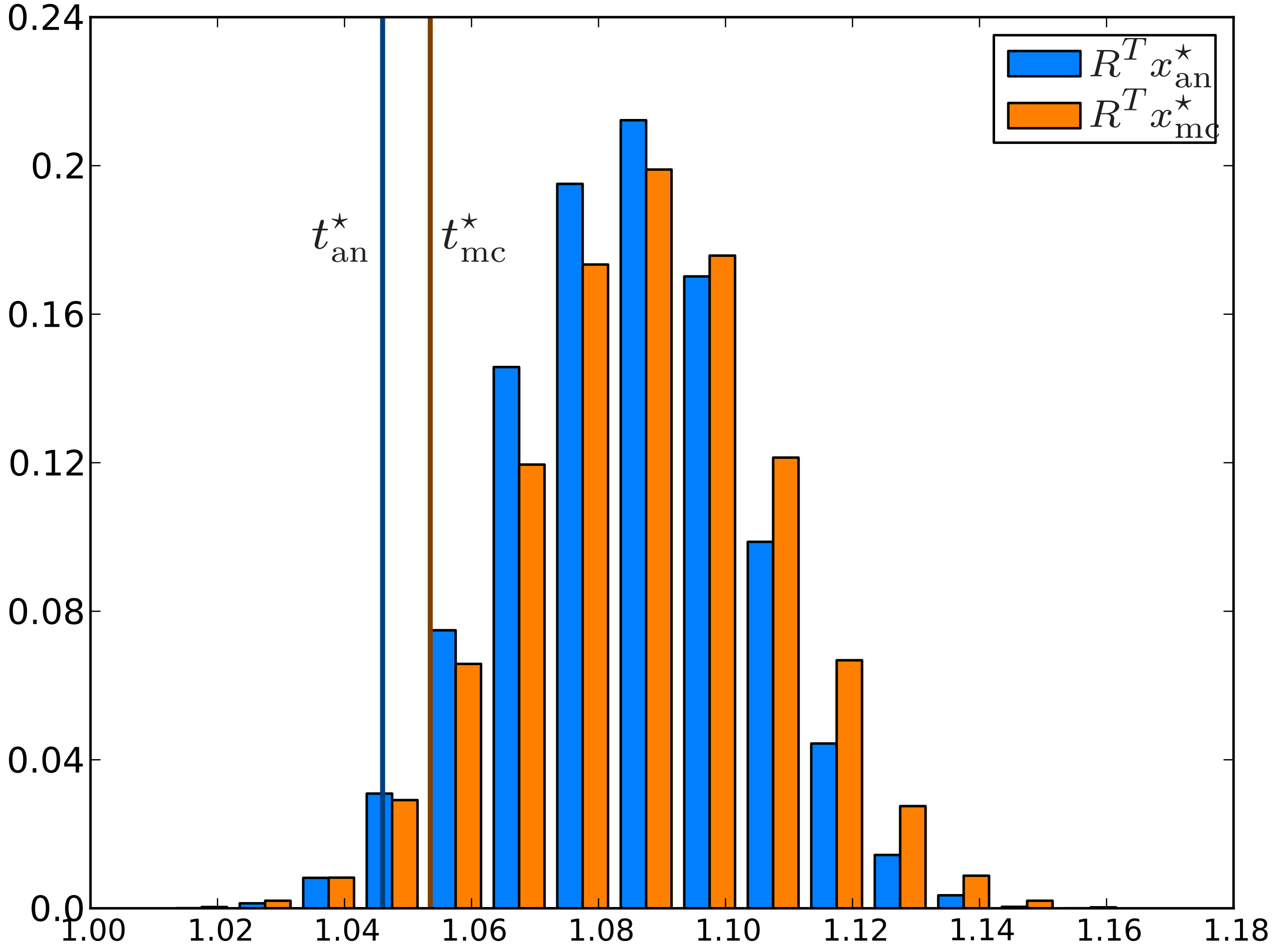
- n assets $i = 1, \dots, n$, random multiplicative return R_i with $\mathbf{E}[R_i] = \mu_i \geq 1$, $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$

- asset i return varies in range $R_i \in [\mu_i - \mathsf{u}_i, \mu_i + \mathsf{u}_i]$

- data $\mu_i = 1.05 + \frac{3(n-i)}{10n}$, uncertainty $|u_i| \leq \mathsf{u}_i = .05 + \frac{n-i}{2n}$ and $\mathsf{u}_n = 0$

- moment generating function approximation to $\mathbf{VaR}(R^T x - t; \epsilon) \leq 0$ is

$$t - \mu^T x + \sqrt{\frac{1}{2} \log \frac{1}{\epsilon}} \|\mathbf{diag}(\mathsf{u}) x\|_2 \leq 0$$



Monte-Carlo CVaR solution $x^\star_{\rm mc}$ vs. analytic (MGF) approximation $x^\star_{\rm an}$

Distributionally robust optimization

stochastic optimization problems:

$$\underset{x}{\text{minimize}} \; \mathbf{E}_P f(x, S) = \int f(x, s) dP(s)$$

distributionally robust formulation:

$$\underset{x}{\text{minimize}} \quad \sup_{P \in \mathcal{P}} \mathbf{E}_P f(x, S) = \sup_{P \in \mathcal{P}} \int f(x, s) dP(s)$$

new question: how should we choose $\mathcal{P}$?

Choices of uncertainty sets

- moment-based conditions, e.g.

$$\mathcal{P} = \{P \mid \mathbf{E}_P[h_i(S)] \preceq b_i\}$$

  for function $h_i : \mathcal{S} \to \mathbf{R}^{n_i}$

- "nonparametric" sets, using divergence-like quantities, e.g.

$$\mathcal{P} = \{P \mid \mathbf{D}_{\rm kl}(P \| P_0) \le \rho\}$$

- often estimate these based on sample $S_1, \ldots, S_m$

Moment-based uncertainty

- general moment constraints: let $K_i \subset \mathbf{R}^{n_i}$ be convex cones, $i = 1, \ldots, m$, $h_i : \mathcal{S} \to \mathbf{R}^{n_i}$, and

$$\mathcal{P} = \left\{ P \mid \int h_i(s) dP(s) \preceq_{K_i} b_i \right\}$$

- under constraint qualification (Rockafellar 1970, Isii 1963, Shapiro 2001, Delage & Ye 2010)

$$\sup_{P \in \mathcal{P}} \int f(s) dP(s) = \inf_{r,t,z} \left\{ \begin{array}{ll} r + \sum_i t_i \text{ s.t.} & r \geq f(s), \; z_i \in K_i^* \\ & t_i \geq z_i^T b_i - z_i^T h_i(s), \text{ all } s \in \mathcal{S} \end{array} \right\}$$

Uncertainty from central limit theorems

- typical idea: start with some probabilistic understanding, work to get uncertainty set
- central limit theorem: for $S_i \in \mathbf{R}$, drawn i.i.d., Φ Gaussian CDF

$$\mathbf{Prob}\left(\frac{1}{m}\sum_{i=1}^m S_i \geq \mathbf{E}S + \frac{1}{\sqrt{m}}\sqrt{\mathbf{Var}(S)}t\right) \to \Phi(-t)$$

- empirical confidence set: let $\mathcal{U}_\rho = \{u \in \mathbf{R}^m \mid \mathbf{1}^T u = 0, \|u\|_2 \leq \rho\}$

$$\left\{\frac{1}{m}\sum_{i=1}^m S_i + \frac{1}{m}\sum_{i=1}^m u_i S_i \mid u \in \mathcal{U}_\rho\right\} = \left\{\overline{S}_m \pm \frac{\rho}{\sqrt{m}}\sqrt{\frac{1}{m}\sum_{i=1}^m (S_i - \overline{S}_m)^2}\right\}$$

- slight extension of CLT implies

$$\mathbf{Prob}\left(\overline{S}_m - \frac{\rho}{\sqrt{m}}\sqrt{\mathbf{Var}_m(S)} \le \mathbf{E}S \le \overline{S}_m + \frac{\rho}{\sqrt{m}}\sqrt{\mathbf{Var}_m(S)}\right)$$
$$= \mathbf{Prob}\left(\mathbf{E}S \in \left\{\overline{S}_m + u^T S/m \mid u \in \mathcal{U}_\rho\right\}\right)$$
$$\to \mathbf{Prob}(-\rho \le \mathcal{N}(0,1) \le \rho) = \Phi(\rho) - \Phi(-\rho)$$

- natural confidence set for distributionally robust optimization:

$$\mathcal{P}_{m,\rho} = \left\{p \in \mathbf{R}^n \mid \mathbf{1}^T p = 1, \|p - \mathbf{1}/m\|_2 \le \rho/m\right\}$$

and

$$\sup_{p \in \mathcal{P}_{m,\rho}} \sum_{i=1}^{m} p_i f(x, S_i)$$

Divergence-based confidence sets

- general form of robustness set: ϕ-divergences

$$\mathbf{D}_\phi\left(P \| Q\right) = \int \phi\left(\frac{p(s)}{q(s)}\right) q(s)$$

where $\phi : \mathbf{R}_+ \to \mathbf{R}$ is convex, $\phi(1) = 0$

- uncertainty set for $P_m = \frac{1}{m}\sum_{i=1}^m \delta_{S_i}$ (empirical distribution)

$$\mathcal{P}_m = \{P : \mathbf{D}_\phi\left(P \| P_m\right) \le \rho/m\}$$

- examples: $\phi(t) = (t-1)^2$, $\phi(t) = t\log t$, $\phi(t) = -\log t$

- can show (Duchi, Glynn, Namkoong) that for ϕ with $\phi''(0) = 2$ that for $S_1, \ldots, S_m \sim P_0$,

$$\mathbf{Prob}\left(\sup_{P \in \mathcal{P}_m} \mathbf{E}_P f(x, S) \leq \mathbf{E}_{P_0} f(x, S) \right) \to \Phi(-\rho)$$

(and versions uniform in x too)

Dual representation of divergence-based confidence sets

if

$$\mathcal{P} = \{P \mid \mathbf{D}_\phi\left(P \| P_0\right) \leq \rho\}$$

then

$$\sup_{P \in \mathcal{P}} \mathbf{E}_P Z = \inf_{\alpha \geq 0, \eta} \left\{ \alpha \mathbf{E} \phi^* \left(\frac{Z - \eta}{\alpha} \right) + \rho\alpha + \eta \right\}$$

- example: $\phi(t) = \frac{1}{2}t^2 - 1$ has

$$\phi^*(u) = \frac{1}{2}(u)_+^2 + 1$$

so

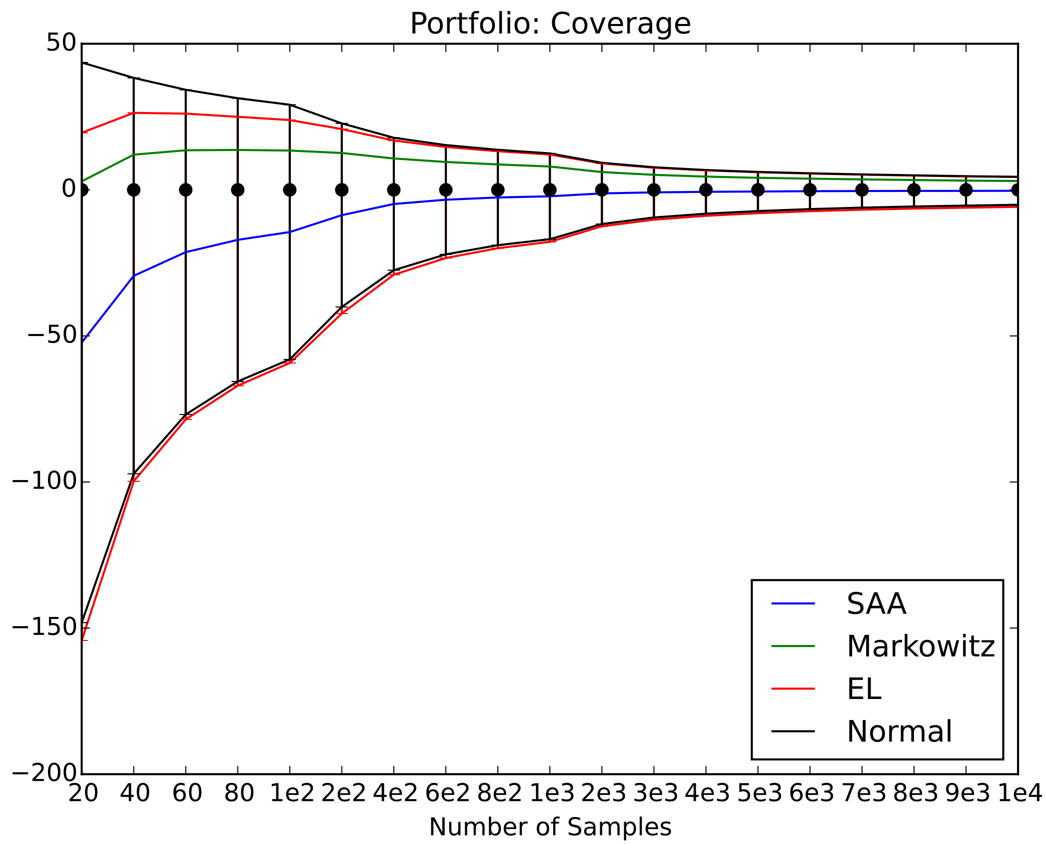
$$\sup_{P \in \mathcal{P}} \mathbf{E}_P f(x, S) = \inf_{\eta} \left\{ \sqrt{1 + \rho} (\mathbf{E}_{P_0}(f(x, S) - \eta)_+^2)^{1/2} + \eta \right\}$$

Portfolio optimization revisited

- returns $R \in \mathbf{R}^n$, $n = 20$, domain $X = \{x \in \mathbf{R}^n \mid \mathbf{1}^T x = 1, x \in [-10, 10]\}$ (leveraging allowed)
- returns $R \sim \mathcal{N}(\mu, \Sigma)$
- within simulation, μ, Σ chosen randomly
- recall Markowitz portfolio problem to

$$\text{maximize } \frac{1}{m} \sum_{i=1}^{m} R_i^T x - \sqrt{\rho/m} \sqrt{x^T \Sigma_m x}$$

where $\Sigma_m = \frac{1}{m} \sum_i (R_i - \overline{R}_m)(R_i - \overline{R}_m)^T$ is empirical covariance



compare returns of robust/empirical likelihood method, Markowitz portfolio, true gaussian results, sample average, 95% confidence