Stochastic programming

- stochastic programming
- chance constraints and percentile optimization
- violation/shortfall constraints and penalties
- convex approximation of chance constraints
- Monte Carlo sampling methods
- validation

sources: Rockafellar & Uryasev, Nemirovsky & Shapiro

Stochastic programming

- objective and constraint functions $f_i(x, \omega)$ depend on optimization variable x *and* a random variable ω
- ω models
  - parameter variation and uncertainty
  - random variation in implementation, manufacture, operation
- value of ω is not known, but its distribution is
- goal: choose x so that
  - constraints are satisfied on average, or with high probability
  - objective is small on average, or with high probability

Stochastic programming

- basic stochastic programming problem:

$$\begin{array}{ll} \text{minimize} & F_0(x) = \mathbf{E}\, f_0(x,\omega) \\ \text{subject to} & F_i(x) = \mathbf{E}\, f_i(x,\omega) \leq 0, \quad i = 1,\ldots,m \end{array}$$

  - variable is x
  - problem data are f_i, distribution of ω

- if $f_i(x,\omega)$ are convex in x for each ω
  - F_i are convex
  - hence stochastic programming problem is convex

- F_i have analytical expressions in only a few cases; in other cases we will solve the problem approximately

Example with analytic form for F_i

- $f(x) = \|Ax - b\|_2^2$, with A, b random

- $F(x) = \mathbf{E}\, f(x) = x^T P x - 2q^T x + r$, where

$$P = \mathbf{E}(A^T A), \quad r = \mathbf{E}(A^T b), \quad r = \|b\|_2^2$$

- only need second moments of (A, b)

- stochastic constraint $\mathbf{E}\, f(x) \leq 0$ can be expressed as standard quadratic inequality

‘Certainty-equivalent’ problem

- ‘certainty-equivalent’ (a.k.a. ‘mean field’) problem:

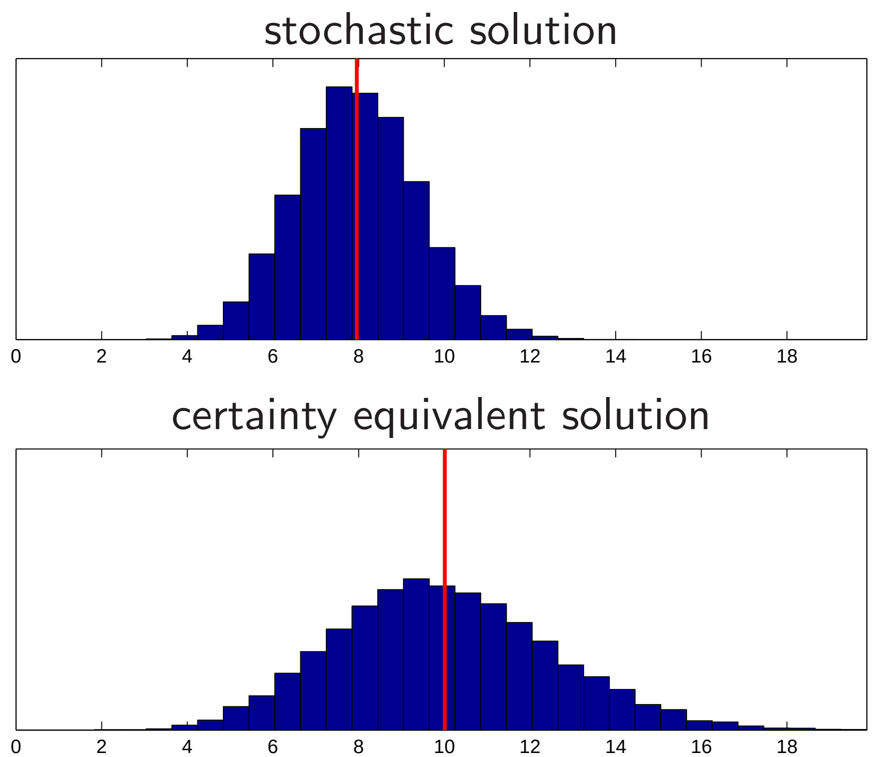
$$\begin{array}{ll} \text{minimize} & f_0(x, \mathbf{E}\,\omega) \\ \text{subject to} & f_i(x, \mathbf{E}\,\omega) \leq 0, \quad i = 1, \ldots, m \end{array}$$

- roughly speaking: ignore parameter variation

- if f_i convex in ω for each x, then
  - $f_i(x, \mathbf{E}\,\omega) \leq \mathbf{E}\, f_i(x, \omega)$
  - so optimal value of certainty-equivalent problem is lower bound on optimal value of stochastic problem

Stochastic programming example

- minimize $\mathbf{E}\,\|Ax - b\|_1$; A_{ij} uniform on $\bar{A}_{ij} \pm \gamma_{ij}$; b_i uniform on $\bar{b}_i \pm \delta_i$
- objective PDFs for stochastic optimal and certainty-equivalent solutions



Expected violation/shortfall constraints/penalties

- replace $\mathbf{E}\, f_i(x,\omega) \leq 0$ with
  - $\mathbf{E}\, f_i(x,\omega)_+ \leq \epsilon$ (LHS is expected violation)
  - $\mathbf{E}\,(\max_i f_i(x,\omega)_+) \leq \epsilon$ (LHS is expected worst violation)

- variation: add violation/shortfall penalty to objective

$$\text{minimize} \quad \mathbf{E}\left(f_0(x,\omega) + \textstyle\sum_{i=1}^m c_i f_i(x,\omega)_+\right)$$

  where $c_i > 0$ are penalty rates for violating constraints

- these are convex problems if f_i are convex in x

Chance constraints and percentile optimization

- 'chance constraints' (η is 'confidence level'):

$$\mathbf{Prob}(f_i(x,\omega) \le 0) \ge \eta$$

  - convex in some cases (later)
  - generally interested in $\eta = 0.9,\ 0.95,\ 0.99$
  - $\eta = 0.999$ is meaningless (unless you're sure about the distribution tails)

- percentile optimization (γ is 'η-percentile'):

$$\begin{array}{ll} \text{minimize} & \gamma \\ \text{subject to} & \mathbf{Prob}(f_0(x,\omega) \le \gamma) \ge \eta \end{array}$$

  - convex or quasi-convex in some cases (later)

Value-at-risk and conditional value-at-risk

- value-at-risk of random variable z, at level η:

$$\mathbf{VaR}(z;\eta) = \inf\{\gamma \mid \mathbf{Prob}(z \leq \gamma) \geq \eta\}$$

  - chance constraint $\mathbf{Prob}(f_i(x,\omega) \leq 0) \geq \eta$ same as $\mathbf{VaR}(f_i(x,\omega);\eta) \leq 0$

- conditional value-at-risk:

$$\mathbf{CVaR}(z) = \inf_{\beta} \left(\beta + 1/(1-\eta)\,\mathbf{E}(z-\beta)_+\right)$$

  - $\mathbf{CVaR}(z;\eta) \geq \mathbf{VaR}(z;\eta)$ (more on this later)

Chance constraints for log-concave distributions

- suppose
  - ω has log-concave density $p(\omega)$
  - $C = \{(x, \omega) \mid f(x, \omega) \leq 0\}$ is convex in (x, ω)

- then

$$\mathbf{Prob}(f(x, \omega) \leq 0) = \int 1((x, \omega) \in C) p(\omega)\, d\omega$$

  is log-concave, since integrand is

- so chance constraint $\mathbf{Prob}(f(x, \omega) \leq 0) \geq \eta$ can be expressed as convex constraint

$$\log \mathbf{Prob}(f(x, \omega) \leq 0) \geq \log \eta$$

Linear inequality with normally distributed parameter

- consider $a^T x \leq b$, with $a \sim \mathcal{N}(\bar{a}, \Sigma)$
- then $a^T x - b \sim \mathcal{N}(\bar{a}^T x - b, x^T \Sigma x)$
- hence

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

- and so

$$\mathbf{Prob}(a^T x \leq b) \geq \eta \iff b - \bar{a}^T x \geq \Phi^{-1}(\eta) \|\Sigma^{1/2} x\|_2$$

a second-order cone constraint for $\eta \geq 0.5$ ($i.e.$, $\Phi^{-1}(\eta) \geq 0$)

Portfolio optimization example

- $x \in \mathbf{R}^n$ gives portfolio allocation; x_i is (fractional) position in asset i
- x must satisfy $\mathbf{1}^T x = 1$, $x \in \mathcal{C}$ (convex portfolio constraint set)
- portfolio return (say, in percent) is $p^T x$, where $p \sim \mathcal{N}(\bar{p}, \Sigma)$
  (a more realistic model is p log-normal)
- maximize expected return subject to limit on probability of loss

- problem is

$$\begin{array}{ll} \text{maximize} & \mathbf{E}\, p^T x \\ \text{subject to} & \mathbf{Prob}(p^T x \leq 0) \leq \beta \\ & \mathbf{1}^T x = 1, \quad x \in \mathcal{C} \end{array}$$

- can be expressed as convex problem (provided $\beta \leq 1/2$)

$$\begin{array}{ll} \text{maximize} & \bar{p}^T x \\ \text{subject to} & \bar{p}^T x \geq \Phi^{-1}(1-\beta)\|\Sigma^{1/2} x\|_2 \\ & \mathbf{1}^T x = 1, \quad x \in \mathcal{C} \end{array}$$

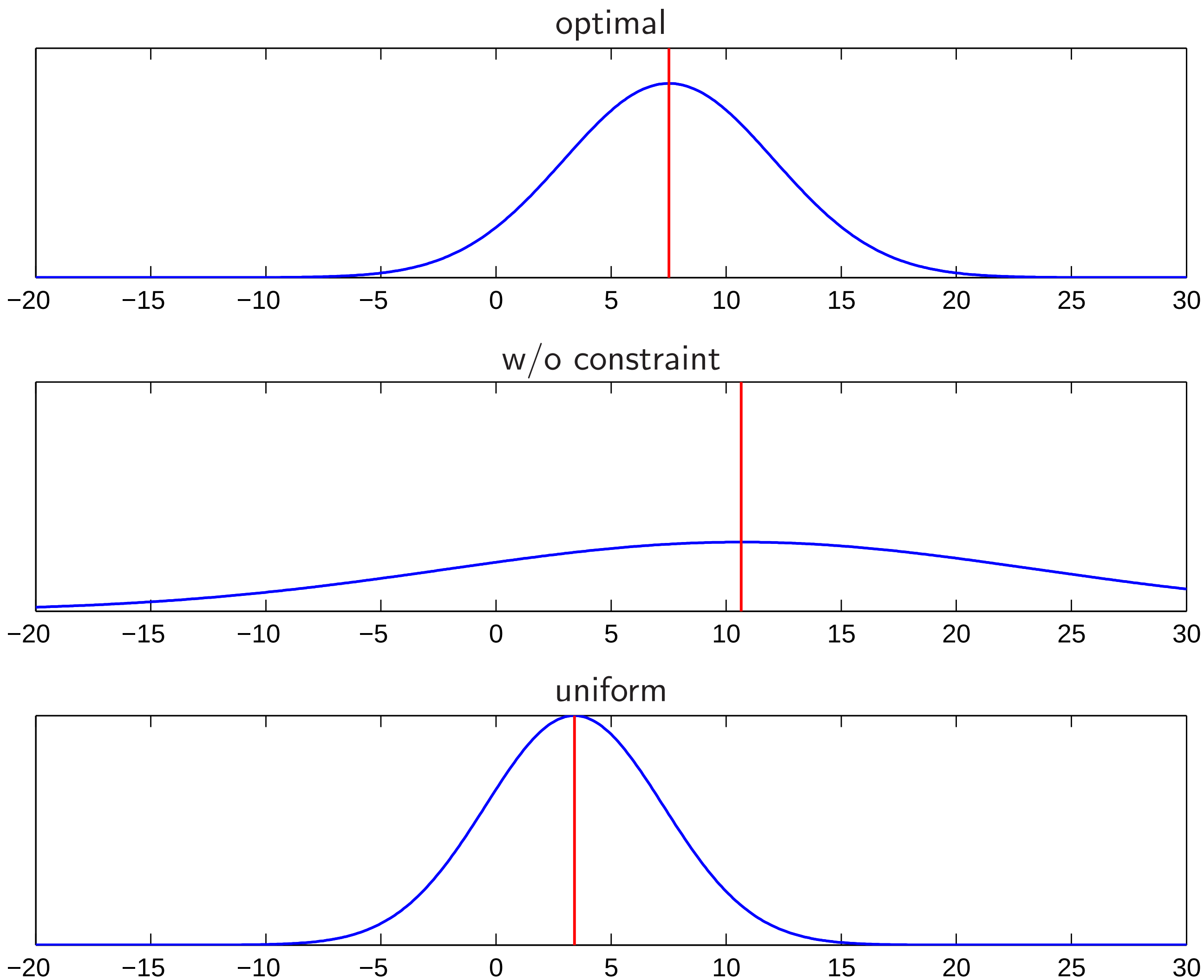
(an SOCP when $\mathcal{C}$ is polyhedron)

Example

- $n = 10$ assets, $\beta = 0.05$, $\mathcal{C} = \{x \mid x \succeq -0.1\}$
- compare
  - optimal portfolio
  - optimal portfolio w/o loss risk constraint
  - uniform portfolio $(1/n)\mathbf{1}$

portfolio	$\mathbf{E}\, p^T x$	$\mathbf{Prob}(p^T x \leq 0)$
optimal	7.51	5.0%
w/o constraint	10.66	20.3%
uniform	3.41	18.9%

return distributions:



Convex approximation of chance constraint bound

- assume $f_i(x, \omega)$ is convex in x
- suppose ϕ : **R** $\rightarrow$ **R** is nonnegative convex nondecreasing, with $\phi(0) = 1$
- for any $\alpha_i > 0$, $\phi(z/\alpha_i) \geq 1(z > 0)$ for all z, so

$$\mathbf{E}\,\phi(f_i(x, \omega)/\alpha_i) \geq \mathbf{Prob}(f_i(x, \omega) > 0)$$

- hence (convex) constraint

$$\mathbf{E}\,\phi(f_i(x, \omega)/\alpha_i) \leq 1 - \eta$$

  ensures chance constraint $\mathbf{Prob}(f_i(x, \omega) \leq 0) \geq \eta$ holds

- this holds for any $\alpha_i > 0$; we now show how to optimize over α_i

- write constraint as

$$\mathbf{E}\,\alpha_i\phi(f_i(x,\omega)/\alpha_i) \le \alpha_i(1-\eta)$$

  - (perspective function) $v\phi(u/v)$ is convex in (u,v) for $v>0$, nondecreasing in u
  - so composition $\alpha_i\phi(f_i(x,\omega)/\alpha_i)$ is convex in (x,α_i) for $\alpha_i>0$
  - hence constraint above is convex in x *and* α_i
  - so we can optimize over x *and* $\alpha_i>0$ via convex optimization

- yields a convex stochastic optimization problem that is a conservative approximation of the chance-constrained problem

- we'll look at some special cases

Markov chance constraint bound

- taking $\phi(u) = (u+1)_+$ gives Markov bound: for any $\alpha_i > 0$,

$$\mathbf{Prob}(f_i(x,\omega) > 0) \leq \mathbf{E}(f_i(x,\omega)/\alpha_i + 1)_+$$

- convex approximation constraint

$$\mathbf{E}\,\alpha_i(f_i(x,\omega)/\alpha_i + 1)_+ \leq \alpha_i(1-\eta)$$

can be written as

$$\mathbf{E}(f_i(x,\omega) + \alpha_i)_+ \leq \alpha_i(1-\eta)$$

- we can optimize over x and $\alpha_i \geq 0$

Interpretation via conditional value-at-risk

- write conservative approximation as

$$\frac{\mathbf{E}(f_i(x,\omega)+\alpha_i)_+}{1-\eta} - \alpha_i \leq 0$$

- LHS is convex in (x, α_i), so minimum over α_i,

$$\min_{\alpha_i > 0} \left(\frac{\mathbf{E}(f_i(x,\omega)+\alpha_i)_+}{1-\eta} - \alpha_i \right)$$

  is convex in x

- this is $\mathbf{CVaR}(f_i(x,\omega);\eta)$ (can show $\alpha_i > 0$ can be dropped)
- so convex approximation replaces $\mathbf{VaR}(f_i(x,\omega);\eta) \leq 0$ with $\mathbf{CVaR}(f_i(x,\omega);\eta) \leq 0$ which is convex in x

Chebyshev chance constraint bound

- taking $\phi(u) = (u+1)_+^2$ yields Chebyshev bound: for any $\alpha_i > 0$,

$$\mathbf{Prob}(f_i(x,\omega) > 0) \leq \mathbf{E}(f_i(x,\omega)/\alpha_i + 1)_+^2$$

- convex approximation constraint

$$\mathbf{E}\,\alpha_i(f_i(x,\omega)/\alpha_i + 1)_+^2 \leq \alpha_i(1-\eta)$$

can be written as

$$\mathbf{E}(f_i(x,\omega) + \alpha_i)_+^2/\alpha_i \leq \alpha_i(1-\eta)$$

Traditional Chebyshev bound

- dropping subscript + we get more conservative constraint

$$\mathbf{E}\,\alpha_i(f_i(x,\omega)/\alpha_i + 1)^2 \le \alpha_i(1-\eta)$$

  which we can write as

$$2\,\mathbf{E}\,f_i(x,\omega) + (1/\alpha_i)\,\mathbf{E}\,f_i(x,\omega)^2 + \alpha_i\eta \le 0$$

- minimizing over α_i gives $\alpha_i = \left(\mathbf{E}\,f_i(x,\omega)^2/\eta\right)^{1/2}$; yields constraint

$$\mathbf{E}\,f_i(x,\omega) + \left(\eta\,\mathbf{E}\,f_i(x,\omega)^2\right)^{1/2} \le 0$$

  which depends only on first and second moments of f_i

Example

- $f_i(x) = a^T x - b$, where a is random with $\mathbf{E}\, a = \bar{a}$, $\mathbf{E}\, aa^T = \Sigma$
- traditional Chebyshev approximation of chance constraint is

$$\bar{a}^T x - b + \left(x^T \Sigma x - 2b\bar{a}^T x + b^2\right)^{1/2} \leq 0$$

- can write as second-order cone constraint

$$\bar{a}^T x - b + \|(z, y)\|_2 \leq 0$$

  with $z = \Sigma^{1/2} x - b\Sigma^{-1/2}\bar{a}$, $\quad y = b\left(1 - \bar{a}^T \Sigma^{-1} \bar{a}\right)^{1/2}$

- can interpret as certainty-equivalent constraint, with norm term as 'extra margin'

Chernoff chance constraint bound

- taking $\phi(u) = \exp u$ yields Chernoff bound: for any $\alpha_i > 0$,

$$\mathbf{Prob}(f_i(x,\omega) > 0) \leq \mathbf{E} \exp(f_i(x,\omega)/\alpha_i)$$

- convex approximation constraint

$$\mathbf{E}\, \alpha_i \exp(f_i(x,\omega)/\alpha_i) \leq \alpha_i(1-\eta)$$

can be written as

$$\log \mathbf{E} \exp(f_i(x,\omega)/\alpha_i) \leq \log(1-\eta)$$

(LHS is cumulant generating function of $f_i(x,\omega)$, evaluated at $1/\alpha_i$)

Solving stochastic programming problems

- analytical solution in special cases, $e.g.$, when expectations can be found analytically
  - ω enters quadratically in f_i
  - ω takes on finitely many values
- general case: approximate solution via (Monte Carlo) sampling

Finite event set

- suppose $\omega \in \{\omega_1, \ldots, \omega_N\}$, with $\pi_j = \mathbf{Prob}(\omega = \omega_j)$
- sometime called 'scenarios'; often we have $\pi_j = 1/N$
- stochastic programming problem becomes

$$\begin{array}{ll} \text{minimize} & F_0(x) = \sum_{j=1}^N \pi_j f_0(x, \omega_j) \\ \text{subject to} & F_i(x) = \sum_{j=1}^N \pi_j f_i(x, \omega_j) \leq 0, \quad i = 1, \ldots, m \end{array}$$

- a (standard) convex problem if f_i convex in x
- computational complexity grows *linearly* in the number of scenarios N

Monte Carlo sampling method

- a general method for (approximately) solving stochastic programming problem

- generate N samples (realizations) $\omega_1, \ldots, \omega_N$, with associated probabilities $\pi_1, \ldots, \pi_N$ (usually $\pi_j = 1/N$)

- form sample average approximations

$$\hat{F}_i(x) = \sum_{j=1}^{N} \pi_j f_i(x, \omega_j), \quad i = 0, \ldots, m$$

- these are RVs with mean $\mathbf{E}\, f_i(x)$

- now solve finite event problem

$$\begin{array}{ll} \text{minimize} & \hat{F}_0(x) \\ \text{subject to} & \hat{F}_i(x) \leq 0, \quad i = 1, \ldots, m \end{array}$$

- solution $x^\star_{\rm mcs}$ is a random variable, hopefully close to $x^\star$

- theory says
  - (with some technical conditions) as $N \to \infty$, $x^\star_{\rm mcs} \to x^\star$
  - $\mathbf{E}\, f_0(x^\star_{\rm mcs}) \leq \mathbf{E}\, f_0(x^\star)$

Out-of-sample validation

- a practical method to check if N is 'large enough'
- use a second set of samples ('validation set') $\omega_1^{\rm val}, \ldots, \omega_M^{\rm val}$, with probabilities $\pi_1^{\rm val}, \ldots, \pi_M^{\rm val}$ (usually $M \gg N$)

  (original set of samples called 'training set')
- evaluate

$$\hat{F}_i^{\rm val}(x^\star_{\rm mcs}) = \sum_{j=1}^{M} \pi_j^{\rm val} f_i(x^\star_{\rm mcs}, \omega_j^{\rm val}), \quad i = 0, \ldots, m$$

- if $\hat{F}_i(x^\star_{\rm mcs}) \approx \hat{F}_i^{\rm val}(x^\star_{\rm mcs})$, our confidence that $x^\star_{\rm mcs} \approx x^\star$ is enhanced
- if not, increase N and re-compute $x^\star_{\rm mcs}$

Example

- we consider problem

$$\begin{array}{ll} \text{minimize} & F_0(x) = \mathbf{E} \max_i (Ax+b)_i \\ \text{subject to} & F_1(x) = \mathbf{E} \max_i (Cx+d)_i \leq 0 \end{array}$$

  with optimization variable $x \in \mathbf{R}^n$

  $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $C \in \mathbf{R}^{k \times n}$, $d \in \mathbf{R}^k$ are random

- we consider instance with $n = 10$, $m = 20$, $k = 5$
- certainty-equivalent optimal value yields lower bound 19.1
- we use Monte Carlo sampling with $N = 10,\ 100,\ 1000$
- validation set uses $M = 10000$

	$N = 10$	$N = 100$	$N = 1000$	CE
F_0 (training)	51.8	54.0	55.4	19.1
F_0 (validation)	56.0	54.8	55.2	162.0
F_1 (training)	0	0	0	0
F_1 (validation)	1.3	0.7	-0.03	85.0

we conclude:

- $N = 10$ is too few samples
- $N = 100$ is better, but not enough
- $N = 1000$ is probably fine

Production planning with uncertain demand

- manufacture quantities $q = (q_1, \dots, q_m)$ of m finished products
- purchase raw materials in quantities $r = (r_1, \dots, r_n)$ with costs $c = (c_1, \dots, c_n)$, so total cost is $c^T r$
- manufacturing process requires $r \succeq Aq$

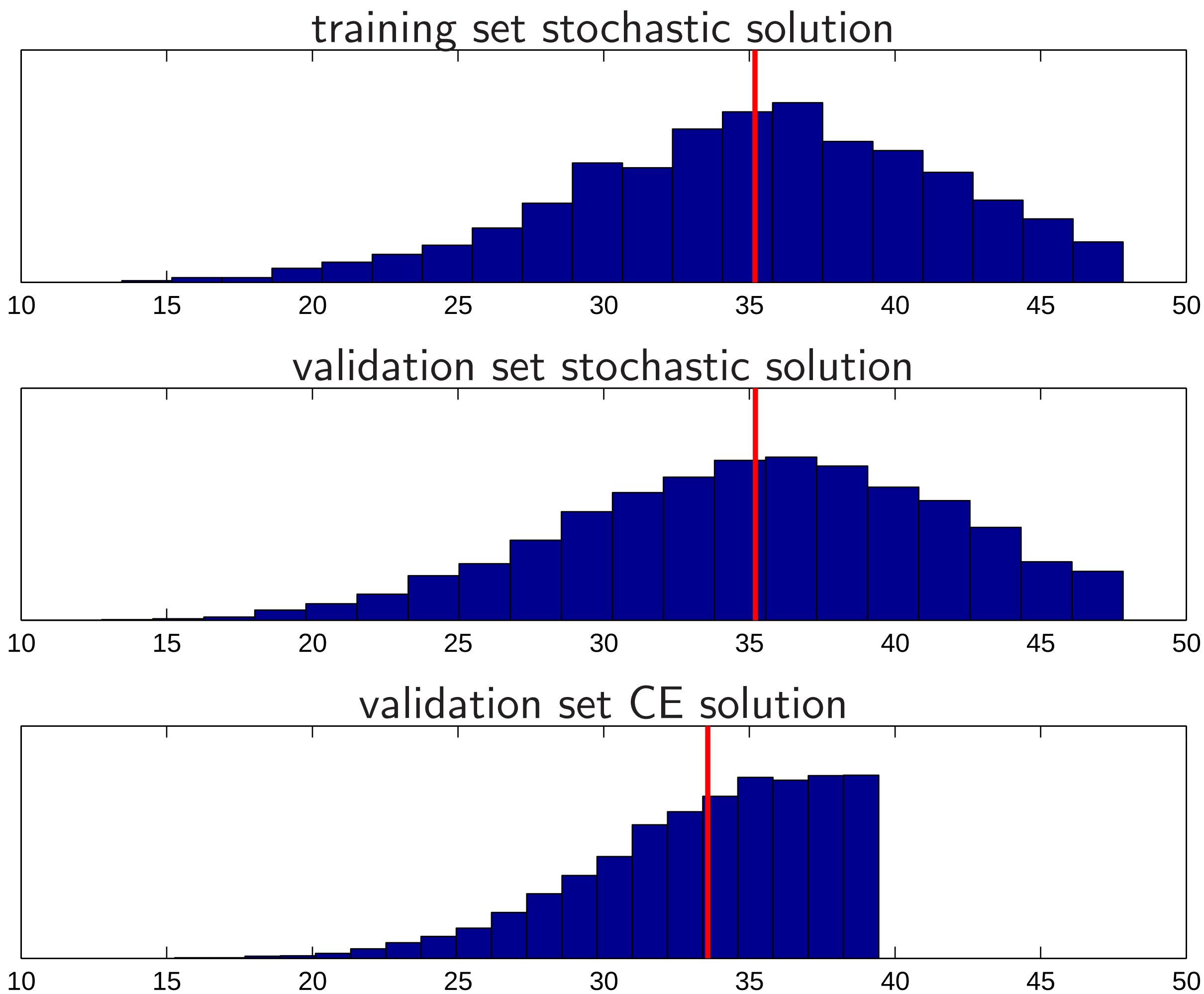
  A_{ij} is amount of raw material i needed per unit of finished product j
- product demand $d = (d_1, \dots, d_m)$ is random, with known distribution
- product prices are $p = (p_1, \dots, p_m)$, so total revenue is $p^T \min(d, q)$
- maximize (expected) net revenue (over optimization variables q, r):

$$\begin{array}{ll} \text{maximize} & \mathbf{E}\, p^T \min(d, q) - c^T r \\ \text{subject to} & r \succeq Aq, \quad q \succeq 0, \quad r \succeq 0 \end{array}$$

Problem instance

- problem instance has $n = 10$, $m = 5$, d log-normal
- certainty-equivalent problem yields upper bound 39.4
- we use Monte Carlo sampling with $N = 1000$ training samples
- validated with $M = 10000$ validation samples

	F_0
training	35.2
validation	35.2
CE training	39.4
CE validation	33.6


training set stochastic solution
10 15 20 25 30 35 40 45 50
validation set stochastic solution
10 15 20 25 30 35 40 45 50
validation set CE solution
10 15 20 25 30 35 40 45 50