CHICAGO BOOTH

# Stochastic Programming: Basic Theory

John R. Birge
John.Birge@ChicagoBooth.edu
www.chicagobooth.edu/fac/john.birge

# Goals

- Provide background on the basics of stochastic programming theory
- Answer the following questions:
  - When is a stochastic programming model consistent?
  - When is a stochastic program solvable?
  - What should be true about an SP solution?
  - What can be inferred from solving one problem to another (sensitivity)?

# General Model

- Choose $x \in X$ to:

$$\text{minimize } E_w[f(x,w)]$$

- where:
- $X$ can be a general space (and might include dynamics probabilistic constraints)
- $f$ can include some form of risk measure

Categories: Two or multi-stage, recourse (or not), probabilistic constrained, robust..

# Two-Stage Stochastic Linear Program with (Fixed) Recourse

- Key: Decisions now ($x$), observe an uncertain outcome ($\xi(\omega)$), take a recourse action $y(\omega)$
- Formulation:

$$\min c^T x + E_\xi [ Q(x,\xi) ]$$

$$s.t. \;\; Ax=b,\; x \geq 0$$

where $Q(x,\xi) = \min\{ q^T y \mid Wy = h - Tx,\, y \geq 0 \}$

$\mathcal{Q}(x) = E_\xi [ Q(x,\xi) ]$ is the recourse function and $\xi$ consists of random components of $q$, $(W)$, $h$, $T$

For the farmer in BL book, randomness is in $T$ (yield).

# Multi-Stage and Nonlinear Models

- Find $x_1,\ldots.,x_t,\ldots. \in X$

to minimize

$$\sum_{t=1}^{\infty} E[f_t(x_t,x_{t+1})]$$

- DP/Bellman form:

$$\Psi_t(x,w_t)=\min_u f_{t+1}(x,u)+E[\Psi_t((x,u),w_{t+1})]$$

$X$, $f$ include constraints and any restrictions

(Note: could also have continuous time.)

# Relationships

- Statistical decision theory:
  - Usually emphasizes information discovery and low dimensions
- Decision analysis
  - Usually few alternatives
- Dynamic programming/Markov decision processes
  - Usually low dimension (often finite state/action)

# More Relationships

- Stochastic control
  - Usually continuous time and very low dimension
- Machine learning
  - Usually no distributions (online) and focus on regret (relative to best possible)
- Robust optimization
  - No distribution (but an uncertainty set) and measured by the worst possible outcome

# Basic Modeling Questions

- What makes a model consistent?
- What form should objective take?
- What form should the constraints take?
- What can be assumed about the distributions?

(Note: criticism of SP: not knowing distributions?)

# Model Consistency

- The model should not allow for solutions that cannot be implemented in reality

Example: A financial model should not allow arbitrage, i.e., ability to buy and sell and make an infinite profit

Conditions:

Share price=$1

Risk-free rate=10%

Share can go to 0.5 or 2 with equal probability

Call option with $1.50 exercise price available for $0.20

Problem: Maximize value of portfolio at time 1
subject to limited downside risk

# Consistency Example

- Solve with upper bounds of 1 on each asset
- Limit downside risk:

  $(E((\text{Bond} - W(1))^{+}) \leq 0.10$

Problems?

Caution: "hidden arbitrage" may lead to quite different solutions from what was intended.

# Discovering Arbitrage

- Example with 2 branches
- Start at S
- Equally likely to uS or dS
- Exercise K: dS<K<uS
- Arbitrage free if $\nexists\ x_1, x_2, x_3$ s. t.

$x_1(uS)+x_2(1+r)+x_3(uS-K)\geq 0$

$x_1(dS)+x_2(1+r)\geq 0$

$Sx_1+x_2+Cx_3=0$

$x_1(p\ u + (1-p)d)S + x_2(1+r)+x_3(p(uS-K))>0$ for any $0<p<1$

# Using Theorem of Alternative

(ToA: $\nexists$ $Ax \geq 0$, $b^T x > 0 \Leftrightarrow \exists\, \pi \geq 0$, s.t. $-\pi^T A = b^T$ (=> C must make A lower rank) )

or $\exists\, \pi_1 \geq 0$, $\pi_2 \geq 0$, $\pi_3$ s.t.

$(p+\pi_1)u+(1-p+\pi_2)d=\pi_3$, $(p+\pi_1+(1-p)+\pi_2)(1+r)=\pi_3$, $(\pi_1+p)((uS-K)/C)=\pi_3$

or for $q=(p+\pi_1)(1+r)/\pi_3$, $1-q= (1-p+\pi_2)(1+r)/\pi_3$;

$R_i(s)$=Value in state s/Value at 0=$S_1(i)/S_0(i)$

$q\, R_i(High) + (1-q)\, R_i(Lo)=(1+r)$ for any asset i

**General: $\exists$ Q (risk-neutral or equivalent martingale measure) s.t. $E_Q[S_t(i)/S_{t-1}(i)]=(1+r_t)$, for all i**

# Check Example

If C=0.2, what happens?

Try solving for q?

Finding consistent value:

$x_3=x_2(1+r)(u-d)/(d(uS-K))$, $x_1=-(1+r)x_2/dS$,

$\Rightarrow C=((1+r)-dS)(uS-K))/((u-d)(1+r))$

Here, u=2,d=0.5, S=1, K=1.5

$x_1=2.2x_2$, $x_3=6.6x_2$,C=2/11=0.181818…

q=0.4, 1-q=0.6

# General Results

- For each set of branches,

The no-arbitrage condition must hold; so,

$\exists$ consistent Q

If not, need to modify prices on each branch.

Otherwise, results may have a bias that is hard to detect (arbitrage will over-whelm any other part of solution).

Practical approach: relax constraints and check for unbounded and non-intuitive results

Klaassen (1998) also shows how to collapse branches together (aggregation) and maintain consistency.

# Other Forms of Consistency

- When are models consistent with rational preferences?
  - – Axioms (e.g., von Neumann-Morganstern)
  - ⇒Expected utility
  - - More information is better
- When are models consistent other forms of behavioral choice?
- Can we learn about the model's form from choices?

# Information Consistency: Paradoxes and Pitfalls

Assumption: More information improves decision making or EVPI$\geq$ 0

Value of Information: “Blau’s dilemma”

Suppose demand=$b=0$ w.p. 0.9 and $1$ w.p. 0.1

Problem:

$$\min x \text{ s.t. } P[x \geq b] \geq 0.9$$

Solution: $x^*=0$

With perfect information: $x^P=0$ w.p. 0.9 and $1$ w.p. 0.1

EVPI = Exp. Value without Perfect Information – Exp. Value with Perfect Information

= 0 – 0.1 = -0.1 < 0

(Same may be true with EVSampleInformation)

For RO, let $\Xi = \{b \mid P[b] \geq 0.9\} = \{0\}$

# Problems with "Paradox"

- Utility may depend on information level
  - With no information, 0.9 may be acceptable but is not the same with more information
  - Cannot make direct comparisons in information value
- Not including role of competitor (something not in model but in consideration)
  - Competitor may gain information as well
  - In this case, more information may not always be beneficial

# What is Missing with Probabilistic Constraint?

- May not correspond to "axioms of choice" or other properties
- Example: Value-at-Risk:

$VaR_{1-\alpha}(x(\xi))$=-inf$\{t|P(x(\xi)\leq t)\geq \alpha\}$

Non-convexity $x_1$={-1 w.p. 0.005, 0 w.p. 0.995}

$x_2$ ={-1 w.p. 0.005, 0 w.p. 0.995}

$VaR_{.99}(X_1)=0, VaR_{.99}(X_2)=0$

Possible: $VaR_{.99}((X_1+X_2))=1$ (with correlation)

Alternatives? Coherent risk measures (but are they consistent with actual choice?)

# Axioms and Coherent Risk Measures

- Von Neumann-Morgenstern (rational) utility (negative risk):
  Complete, Transitive, Continuous, Monotonic, Substitutable (Independent)
  Implies convexity (concavity) of preferences
  Independence is like the additive term for coherent risk
  Both are unclear in practice.
- *R* is a *coherent risk measure* if
  - *R* is convex and decreasing
  - $R(x(\xi)+a)=R(x(\xi)) - a, a\in \Re$
  - $R(\lambda x(\xi)) = \lambda R(x(\xi))$

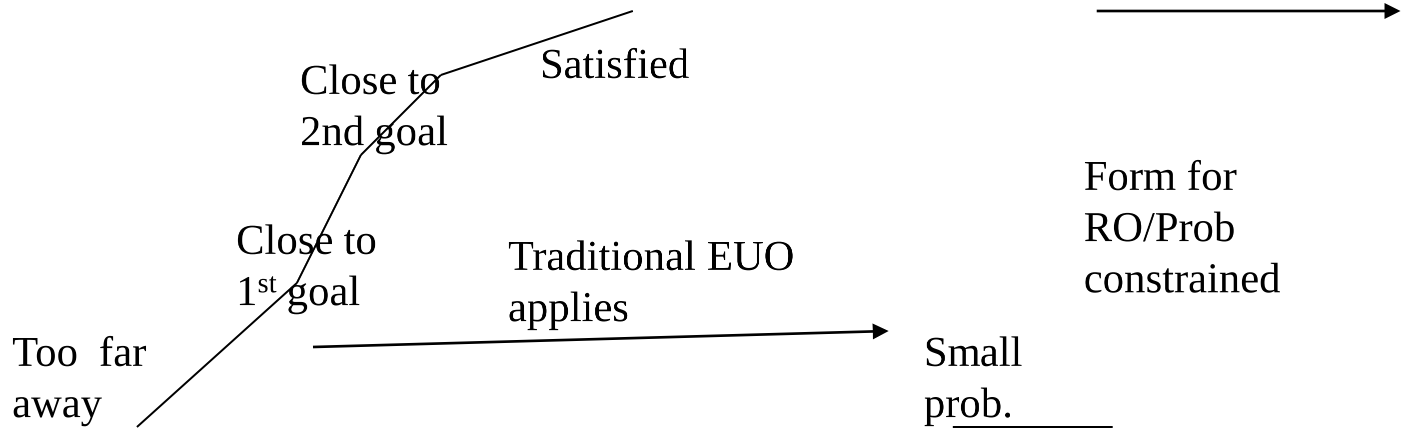
  *Note: this is risk for positive outcomes (gains); if* $\xi$ *is a random loss, use* $x(\xi)+a=-\xi-a$.

# Issues with Axioms

- They may not represent actual choices
  - Prospect theory
- They may be require information (to distinguish among choices) than what we can measure
  - Heyde and Kou (2004) show that distinguishing exponential from power tails may require an excessive number of observations

# Do Axioms Capture Real Choices?

- What is observed? (Kahnemann-Tversky prospect theory)
  - Targets define utility
  - Preference depends on closeness to targets



# Consistent Model: Formulation

- Two-stage stochastic linear program:

$$\min z = c^T x + E_\xi [\ \min q(\omega)^T y(\omega)\ ]$$

$$\text{s.t.}\ \ A x\ = b$$

$$T(\omega)\, x + W\, y(\omega) = h(\omega)\ (\text{a.s.})$$

$$x \geq 0,\ \ y(\omega) \geq 0$$

Example: News vendor

# News Vendor Formulation

- x = number of papers to buy
- c = cost of the papers minus price (-price+cost)
- Ax =b (maybe limit on how many to buy)
- Tx + Wy = h , x + y1 – y2 = h (demand)
- y1 = unmet demand (no revenue)
- y2 = papers not sold
- q = (0,price-salvage) for y1 and y2

# Recourse Function

- **Deterministic Equivalent Program**

$\min z = c^T x + \mathcal{Q}(x)$

s.t. $Ax = b, x \geq 0$

where

$\mathcal{Q}(x) = E_\xi Q(x, \xi(\omega))$ and

$Q(x, \xi(\omega))$

$= \min_y \{ q(\omega)^T y | W y = h(\omega) - T(\omega) x, y \geq 0 \}$

# When Does a Solution Exist? (Feasibility)

- Definitions

  $K_1 = \{ x \mid Ax = b, x \geq 0 \}$

  $K_2 = \{ x \mid \mathcal{Q}(x) < \infty \}$, $K_2^P = \cap_{\xi \in \Xi} \{x \mid Q(x,\xi) < \infty\}$

  $x \in K_1 \cap K_2$

- Results:

  $K_2^P = K_2$ if $\Xi$ finite or W is fixed and $\boldsymbol{\xi}$ has finite second moments, which also means:

  $K_2$ is closed and convex.

  If T is fixed, $K_2$ is polyhedral.

  Let $\Xi_T$ be the support of the distribution of **T**. If $h(\boldsymbol{\xi})$ and $T(\boldsymbol{\xi})$ are independent and $\Xi_T$ is polyhedral, then $K_2$ is polyhedral.

# Properties of the Objective Function

- For a stochastic program with fixed recourse, $Q(x,\xi)$ is

  a piecewise linear convex function in (h,T);

  a piecewise linear concave function in q;

  a piecewise linear convex function in x for all x in $K = K_1 \cap K_2$.

- Proof: Linear supports – use duality.
- With finite second moments, $\mathcal{Q}(x)$ is:

  Lipschitzian, convex, finite on $K_2$, p.l. if $\Xi$ finite, differentiable on $K_2$ if $F(\xi)$ abs. continuous

# Basic Properties

# Special Cases

- Complete recourse

  $K_2 = \Re^{n1}$

- Relatively complete recourse

  $K_2 \supset K_1$

- Simple recourse

  $W=[\ I,\ -I],\ q = (q^+, q^-)$

  Holds for the news vendor problem

# Optimality Conditions

- Existence: suppose $\xi$ has finite second moments and either K is bounded or $\mathcal{Q}$ becomes linear eventually, then optimum attained if it exists.
- Conditions:

  Suppose a finite optimal value. A solution $x^* \in K_1$, is optimal in DEP if and only if

  there exists some $\lambda^* \in \Re^{m1}$, $\mu^* \in \Re^{n1}{}_+$, $\mu^{*T}x^*=0$, such that,

  $-c+ A^T\lambda^* + \mu^* \in \partial\mathcal{Q}(x^*)$.

  $\partial\mathcal{Q}(x^*) = E[\partial Q(x^*,\xi)] + N(K_2,x^*)$

# Duality

Assume X= $\mathcal{L}_\infty(\Omega,\mathcal{B},\mu;\Re^{n1+n2})$, the SP is feasible, has a bounded optimal value, and satisfies relatively complete recourse, a solution

$(x^*(\omega),y^*(\omega))$ is optimal if and only if there exist integrable functions on $\Omega$, $(\lambda^*(\omega),\rho^*(\omega),\pi^*(\omega))$, such that $c_j - \lambda^*(\omega) A_{.\,j} - \rho^*(\omega) - \pi^{*T}(\omega)T_{.\,j}(\omega) \geq 0$, a.s., $j=1,\ldots,n_1$

$(c_j - \lambda^*(\omega) A_{.\,j} - \rho^*(\omega) - \pi^{*T}(\omega)T_{.\,j}(\omega))x^*_j(\omega)= 0$, a.s., $j=1,\ldots,n_1$,

$$q_j(\omega) - \pi^{*T}(\omega)W_{.\,j} \geq 0, \text{ a.s.}, j=1,\ldots,n_2$$

$(q_j(\omega) - \pi^{*T}(\omega)W_{.\,j})y_j^*(\omega) = 0$, a.s., $j=1,\ldots,n_2$,

and $E_\omega[\rho^*(\omega)] = 0$.

# More Duality

- General duals

$z^{*}=\sup \langle c,x\rangle$ s.t. $x\in (N + b)\cap(C+d)$

*where N is a subspace, C is a cone*

$w^{*}= \inf \langle b,\pi\rangle+ \langle d,\rho\rangle$ s.t. $c=\pi + \rho,\ \pi \perp N, \rho \in C^{*}$

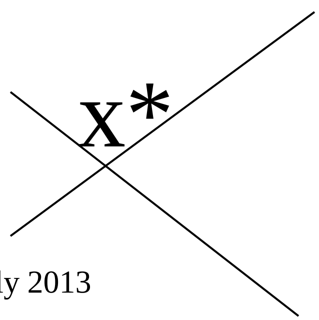
*where* $C^{*}$ *is the polar cone to C, i.e.,* $C^{*}=\{y \mid y^{T} x\leq 0\ \forall x\in C\}$

- Why $z^* \leq w^*$?

$z^*=\langle c,x^*\rangle = \langle \pi^{*}+\rho^{*},x^*\rangle$
$=\langle \pi^{*},x^{*}\rangle + \langle \rho^{*},x^{*}\rangle$
$=\langle \pi^{*},n^{*}+b\rangle+\langle \rho^{*},v^{*}+d\rangle$ ($n^*\in N, v^*\in C$)
$\leq \langle \pi^{*}, b\rangle+\langle \rho^{*}, d\rangle = w^*$

- Why $z^*=w^*$?

(If not, some form of separation for linear problems but may have issues for general nonlinear problems.)



# Dual SLP

- Find $(\lambda(\omega),\rho(\omega),\pi(\omega))$ to

$$\max E[\lambda(\omega)^T b + \pi(\omega)^T h(\omega)]$$

$$\text{s.t. } \lambda(\omega)^T A + \pi(\omega)^T T(\omega) + \rho(\omega)^T \leq c^T, \text{ a.s.}$$

$$\pi(\omega)^T W \leq q(\omega)^T, \text{ a.s.}$$

$$E(\rho(\omega))=0$$

Note: possibly multiple rho values but all only differ in the constant.

# Problems with No Complete Recourse

- min x s.t. $x \geq 0 + \mathcal{Q}(x)$

$\mathcal{Q}(x) = 0$ if $y = x - \xi(\omega) \geq 0$

$\infty$ o.w.

$\xi \sim U(k/K)$, k=0,…,K-1

Multipliers: $\rho^k=1$, $\pi^k=0$, k=0,…,K-2

$\rho^{K-1}=1-K, \pi^{K-1}=K$

As $K \rightarrow \infty$, $\sup(|\rho|) \rightarrow \infty$

# Multistage Formulation (Linear Problems

Linear Program:

$$\min z \;=\; c^1 x^1 \;+\; E_{\xi^2}[\min c^2(\omega)\, x^2(\omega^2)$$

$$+ \ldots \;+\; E_{\xi^H}[\min c^H(\omega)\, x^H(\omega^H)]]$$

$$\text{s.t.} \quad W^1 x^1 = h^1,$$

$$T^1(\omega)\, x^1 \;+\; W^2 x^2(\omega^2) = h^2(\omega),$$

$$\ldots..$$

$$T^{H-1}(\omega)\, x^{H-1}(\omega^{H-1}) \;+\; W^H x^H(\omega^H) = h^H(\omega),$$

$$x^1 \geq 0;\;\; x^t(\omega^t) \geq 0,\;\; t=2,\ldots,H$$

# General Optimality Conditions

General Form:

$\min z = \sum_{t=1}^{\infty} f_t(x_t, x_{t+1})$

s.t $x_t - E[x_t \mid \Sigma_t] = 0$ ,a.s., $\forall t \geq 0$

Need:

nonanticipative feasibility, strict feasibility, finite horizon continuation

Then $x^*$ is optimal with given initial conditions $x^0$ iff there exist $\pi^t \in L^n{}_1(\Sigma)$, $t=0,\ldots$such that $\pi^t$ is nonanticipative

$E^0(f^0(x^0,x^1)-\pi^0 x^0+\pi^1 x^1)$ is a.s. minimized by $x^{*1}$ over $x^1=E[x^1|\Sigma^1]$, and, for $t > 0$

$E[f^t(x^t,x^{t+1}) - \pi^t x^t + \pi^{t+1} x^{t+1}]$ is a.s. minimized by $(x^{*t},x^{*t+1})$over $x^t=E[x^t/\Sigma^t]$ and $x^{t+1} =E[x^{t+1}/\Sigma^{t+1}]$

and $E\pi^{t_k}(x^{t_k} - x^{*t_k}) \rightarrow 0$ as $t_k \rightarrow \infty$, for all $x \in dom\ z$.

# DP Version

- Bellman Equation

$V_t(x_t) = \min f_t(x_t,u_t) + E[V_{t+\delta}(x_{t+\delta})|x_t,u_t]$

Use: $V_{t+\delta}(x_{t+\delta}) =$

$V_t(x_t)+(\partial V_t/\partial t)(\delta t)$

$+ (\partial V_t / \partial x)(\delta x)$

$+(1/2)(\partial^2 V_t / \partial x^2)(\delta x)^2 + o(\delta_t^2)$

(Can use this to obtain continuous-time results)

# Probabilistic Constraints

- Basic (P) Model:

min $c^T x$

s.t. $P(A^i(\omega)\, x \geq h^i(\omega)) \geq \alpha^i$, i=1,…,$m_1$

Examples: Probability of loss greater than some level is at most 1-$\alpha_i$;

Probability of not meeting demand is at most 1-$\alpha_i$;

Note: sometimes these are given as separate (easier) and sometimes joint – unclear on utility implications
(may be easier to estimate than others)

# Issues with Probabilistic Constraints

- **Convexity:**

In general, the feasible region is not necessarily convex –

$A=[1; -1]$

$h(\omega_1)=[0; -1]$ $h(\omega_2)=[2; -3]$

$P(\omega_1)=P(\omega_2)=0.5$

Nice property: quasi-concavity (including log-concavity) :

$P((1-\lambda)U+\lambda V) \geq \min(P(U),P(V))$

Convex: if A fixed and h q-concave, then convex.

Nice properties with normal distribution, simple recourse

In general, find some deterministic approximation.

Results: equivalent recourse formulation (but often relies on knowing the solution).

# Additional Theory of Probabilistic Constraints

- Finding deterministic (or stochastic) approximations:

- or they hold with given confidence.

# Forms of Sensitivity

- For specific parameters:

Convexity in constraint parameters (h, T)

Concavity in linear objective (q)

- Also, note with respect to changes in distribution:

# Additional Inferences: Inverse Optimization

- Suppose given observed decisions $x^*$:

# Questions and Answers

- When is a stochastic programming model consistent?
  - When it doesn’t contradict observations of behavior (e.g., prices and quantities in markets)
- When is a stochastic program solvable?
  - When it has consistent properties (e.g., compact regions and continuous and bounded objectives)

# More Q and Q

- What should be true about an SP solution (and value)? (With some conditions:)
  - Convexity/concavity in certain parameters
  - Differentiability of objective wrt decisions
  - Properties linking prices (dual variables) and quantities (primal variables)
- What can be inferred from solving one problem?
  - Bounds based on solutions (primal, dual) and distances between distributions

# Conclusions/Further Results

- Basic properties enable:
  - Discovery about choices:
    - Preferences (including risk)
    - Constraints
  - Bases for computational methods
  - Inferences about solutions from samples
  - Implications on what information to gather
- More results on basic theory can help improve decisions and our understanding of how decisions are made

- Thanks and questions?