ORIGINS OF STOCHASTIC PROGRAMMING

Early 1950's: in applications of Linear Programming –

unknown values of coefficients:

demands, technological coefficients, yields, etc.

QUOTATION – Dantzig, Interfaces 20,1990 – nutrition problem

"When is an apple an apple and what do you mean by its cost and nutrition content? For example, when you say *apple* do you mean a Jonathan, or McIntosh, or ....? You see, it can make a difference, for the amount of ascorbic acid (vitamin C) can vary from 2.0 to 20.8 units per 100 grams depending upon the type of apple."

Uncertainties modeled as random $\longrightarrow$

STOCHASTIC PROGRAMMING

Multi-objective Optimization and Stochastic Programming Models

We shall consider now various approaches to mathematical formulation of stochastic programs.
Inspiration comes from multi-objective programming.

SP PROBLEM

Select the "best possible" decision which fulfills prescribed "hard" constraints, say, $\mathbf{x} \in \mathcal{X}$ where $\mathcal{X} \subset \mathbb{R}^n$ is a closed nonempty set.

Outcome of a decision $\mathbf{x}$ is influenced by a random element ω of a general nature whose realization is not known at the time of decision — **scenario**.

Random outcome of decision $\mathbf{x}$ is quantified by $f(\mathbf{x}, \omega)$ and different scenarios ω provide different optimal solutions,
$\mathbf{x}^*(\omega) \in \arg\min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \omega)$.

Scenario analysis

ASSUME:

Finite set of considered scenarios, $\{\omega^s,\ s = 1, \ldots, S\}$.

S optimal solutions $\mathbf{x}^s$ and S values $f(\mathbf{x}^s, \omega^s)$ of objective functions $f(\mathbf{x}, \omega^s)$, $s = 1, \ldots, S$.

The idea – analyse the values $f(\mathbf{x}^s, \omega^k)$ to see what happens when for decision $\mathbf{x}^s$ scenario ω^k occurs instead of ω^s.

EXAMPLE

Scenario-based SP I.

ASSUME AGAIN:

Finite set of considered scenarios, $\{\omega^s,\ s = 1, \ldots, S\}$.

Methods of multi-objective programming suggest to choose a solution efficient with respect to S objective functions $f(\mathbf{x}, \omega^s)$, $s = 1, \ldots, S$.

Such efficient solutions can be obtained, e.g., by minimization (or maximization) of a weighted sum of $f(\mathbf{x}, \omega^s)$, $s = 1, \ldots, S$.

In our case, it is natural to use **probabilities** p_s of scenarios ω^s at the place of weights t_s and the problem to be solved is

$$\min_{\mathbf{x} \in \mathcal{X}} \sum_{s=1}^{S} p_s f(\mathbf{x}, \omega^s).$$

The result is the widely used **expected value criterion**.

Scenario-based SP II.

Notice that we get efficient solutions regardless the origin of probabilities p_s, e.g., for p_s

- the true probabilites,
- subjective probabilities,
- probabilities offered by experts,
- equal probabilities obtained via simulation or coming from an empirical probability distribution.

Similarly, using **goal programming** approach we may get the **tracking model**, e.g.

$$\min_{\mathbf{x}\in\mathcal{X}} \sum_{s=1}^{S} p_s |f(\mathbf{x}, \omega^s) - f(\mathbf{x}^*(\omega^s), \omega^s)|.$$

Models of SP with non-stochastic constraints

ASSUME:

Known, general probability distribution P of ω, E_P denotes expectation.

Random elements appear only in the objective function $f(\mathbf{x}, \omega)$,

set $\mathcal{X}$ is fixed.

Similarly as in financial applications, choice among *various objective functions which depend only on the probability distribution P, NOT on observed scenarios.*

In case of minimization, we have for example

- $\min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \hat{\omega})$, with $\hat{\omega}$ a point estimate of ω
- $\min_{\mathbf{x} \in \mathcal{X}} E_P f(\mathbf{x}, \omega)$, risk neutral expected value criterium
- $\min_{\mathbf{x} \in \mathcal{X}} E_P u(f(\mathbf{x}, \omega))$, expected disutility criterium
- $\min_{\mathbf{x} \in \mathcal{X}} E_P f(\mathbf{x}, \omega) + \lambda R_P(f(\mathbf{x}, \omega))$ includes a risk measure R_P
- $\min_{\mathbf{x} \in \mathcal{X}} E_P f(\mathbf{x}, \omega) + \lambda var_P(f(\mathbf{x}, \omega))$, cf. Markowitz model

Random elements in constraints of (MP) I.

$$\mathcal{X}(\omega) = \{\mathbf{x} \in \mathcal{X}_0 : h_j(\mathbf{x}, \omega) = 0 \, \forall j, \, g_k(\mathbf{x}, \omega) \le 0 \, \forall k\}$$

WE CANNOT WAIT WHICH SCENARIO OCCURS!

How to choose SET OF FEASIBLE DECISIONS INDEPENDENT OF SCENARIOS?

AD HOC IDEA:

Replace ω by a fixed characteristics $\hat{\omega}$ which depends on P – point estimate, mostly expectation $E_P\omega$. Solve deterministic problem with set

$$\mathcal{X}(\hat{\omega}) = \{\mathbf{x} \in \mathcal{X}_0 : h_j(\mathbf{x}, \hat{\omega}) = 0 \, \forall j, \, g_k(\mathbf{x}, \hat{\omega}) \le 0 \, \forall k\}$$

NOT THE BEST IDEA — example

$$\min\{x_1 + x_2 : \alpha x_1 + x_2 \ge 7, \beta x_1 + x_2 \ge 4, x_1, x_2 \ge 0\}$$

(α, β) random, uniformly distributed on $[1, 4] \times [1/3, 1]$. Replace α, β by their expected values $5/2$ resp. $2/3$ and solve the LP.

Optimal solution $x_1^* = \frac{18}{11}$, $x_2^* = \frac{32}{11}$. Probability that (x_1^*, x_2^*) satisfies the random constraints is only $1/4$! Not acceptable.

Random elements in constraints of (MP) II.

PERMANENTLY FEASIBLE or FAT DECISIONS

$\mathbf{x} \in \mathcal{X}(\omega)$ for all scenarios ω, or $\mathbf{x} \in \mathcal{X}(\omega)$ almost surely.

$$P\{\omega : \mathbf{x} \in \mathcal{X}(\omega)\} = 1.$$

The set $\mathcal{X} = \cap_{\omega \in \Omega} \mathcal{X}(\omega)$ is small and it is often empty, e.g. for ω in equations and P absolutely continuous. BUT — Robust Optimization.

RELAX THE REQUIREMENT $\longrightarrow$ PROBABILITY or CHANCE CONSTRAINTS

JOINT PROBABILITY CONSTRAINT

$$P(\omega : \mathbf{x} \in \mathcal{X}(\omega)) \geq 1 - \epsilon,$$

with $0 \leq \epsilon \leq 1$ chosen by the decision maker.

Reliability type constraint; for absolutely continuous P can be used only for random parameters in inequality constraits.

PROBLEM: convexity of the resulting set of feasible decisions. Can be proved only for special structure of contraints and special probability distribution, e.g. normal.

Individual Probability Constraints

Given probability treasholds $\epsilon_1, \dots, \epsilon_m$ the feasible decisions are $\mathbf{x} \in \mathcal{X}_0$ that fulfil m INDIVIDUAL PROBABILITY CONSTRAINTS

$$P(\omega : g_k(\mathbf{x}, \omega) \leq 0) \geq 1 - \epsilon_k,\ k = 1, \dots, m \tag{1}$$

Easy structure of problem, namely, if ω are right-hand sides of contraints, i.e. $g_k(\mathbf{x}, \omega) = g_k(\mathbf{x}) - \omega_k$.

Denote F_k marginal d.f. of ω_k, $u_{\epsilon_k}(P)$ $100\epsilon_k\%$ quantile of $F_k \longrightarrow$ (1) reformulated

$$F_k(g_k(\mathbf{x})) \leq \epsilon_k \text{ i.e. } g_k(\mathbf{x}) \leq u_{\epsilon_k}(P)$$

For convex $g_k(\mathbf{x})$ — convex set of feasible decisions.

Notice: random rhs ω_k replaced by a specified QUANTILE of marginal pdf.

No more valid for joint probability constraints! Even for random rhs special requirement on P needed, cf. log-concave or quasiconcave probability distributions (Prékopa).

Joint Probability Constraints

THEOREM (Prékopa) Assume:

- $g_k(\mathbf{x}, \omega) \, \forall k$ are jointly convex in $\mathbf{x}$, ω on $\mathbb{R}^n \times \mathbb{R}^l$
- probability distribution P of ω is logarithmically concave on $\mathbb{R}^l$.

THEN is the function

$$h(\mathbf{x}) := P\{\omega : g_k(\mathbf{x}, \omega) \leq 0, \, \forall k\}$$

logarithmically concave on $\mathbb{R}^n$.

$\Longrightarrow$ convexity of the set $\mathcal{X}(P) := \{\mathbf{x} : h(\mathbf{x}) \geq \alpha\}$

GENERALIZATIONS: to quasiconcave probability distributions $\longrightarrow$
Applicable for a rich class of absolutely continuous probability distributions

PROBLEMS: Joint convexity of $g_k(\mathbf{x}, \omega)$ $\longrightarrow$
Theorem is applicable e.g. for $g_k(\mathbf{x}, \omega) = g_k(\mathbf{x}) - \omega_k$, i.e. for *separable* joint probabilistic constraints;
No direct application to discrete probability distributions

Random elements in constraints of (MP) — cont.

IDEA: Penalize discrepances and include the penalty into the objective function, cf. Problem of Private Investor. Recall

- Initial decision is independent of scenarios observed in future, *non-anticipativity*.
- It can be updated in dependence on observed scenarios later on to fit the requirements as much as possible.
- The cost of discrepace enters objective function; EXPECTED PENALTY TERM.

Exist various models. EXAMPLE — NEWSBOY PROBLEM

Illustrative examples – The newsboy problem

Newsboy sells newspapers for the cost c each. Before he starts selling, he has to buy the daily supply at the cost p a paper. The demand is random and the unsold newspapers are returned without refund at the end of the day. How many newspapers should he buy?

In the framework of stochastic programming, one assumes that the demand is random and the verbal description of the newsboy problem leads to the familiar mathematical formulation

$$\max_{x \geq 0} [(c - p)x - cE_P(x - \omega)^+] \tag{2}$$

where $c > p > 0$ and E_P denotes expectation with respect to a known probability distribution P of the random (nonnegative) demand ω. The optimal decision is then

$$x(P) = u_{1-\alpha}(P) \tag{3}$$

where $\alpha = p/c$ and $u_{1-\alpha}(P)$ is the 100 $(1 - \alpha)$% quantile of probability distribution P.

Newsboy problem cont.

In practice, the newsboy does not know probability distribution P. He may base his decision on historical records, or on a few expert forecasts - scenarios, he may use worst-case analysis

For instance, his decision based on independent identically distributed observed past realizations u^ν of $\omega, \nu = 1, \ldots, N$, can be obtained as

$$\arg\max_{x \geq 0} \left[(c-p)x - \frac{c}{N} \sum_{\nu=1}^{N} (x - u^\nu)^+\right] \quad (4)$$

with the optimal solution $x(P^N)$, the 100 $(1-\alpha)$% quantile of empirical distribution P^N. Sample from a censored distribution!

Applicability of this procedure depends on the available sample size N, in particular for α near to 0 or 1. With $0 < \alpha < 1$, empirical quantiles are asymptotically normal under quite general assumptions and the quantile process can be bootstrapped to obtain an estimate of the variance; consequently, for N large enough, asymptotical confidence intervals for $x(P)$ can be constructed. Any additional knowledge about the rules that influence the changes of demand can be in principle incorporated into this procedure.

Newsboy problem cont.

Alternatively, newsboy can confine himself to a parametric family of probability distributions, estimate its parameters from the sample and apply formula (2) for the obtained probability distribution $P^{\#}$. If he was right in his choice of the family (and this assumption seems to be the stumbling block of the approach), parametric analysis plus statistical inference or the worst case analysis with respect to the parameter values can be used to obtain a relationship between the true $x(P)$ and the obtained $x(P^{\#})$.

Known (or estimated) range and moments of P in connection with the minimax approach can be used to obtain lower and upper bounds L and U on the optimal value of the objective function in (1): one considers a family of probability distributions, say $\mathcal{P}$ described by the moments values and solves the problem for the "worst" and the "best" distribution of the family. $\rightarrow$ bounds

$$L(\mathcal{P}) = \max_{x \geq 0} \min_{P \in \mathcal{P}} [(c-p)x - cE_P(x-\omega)^+] \tag{5}$$

$$U(\mathcal{P}) = \max_{x \geq 0} \max_{P \in \mathcal{P}} [(c-p)x - cE_P(x-\omega)^+] \tag{6}$$

such that

Newsboy problem cont.

$$L(\mathcal{P}) \leq \max_{x \geq 0}[(c-p)x - cE_P(x-\omega)^+] \leq U(\mathcal{P}) \quad \forall P \in \mathcal{P}$$

and, wrt. family $\mathcal{P}$, they are tight. The result depends, of course, on the choice of $\mathcal{P}$. If (for the chosen family $\mathcal{P}$) the difference between $L(\mathcal{P})$ and $U(\mathcal{P})$ is too large, the newsboy should try to collect an additional information about the distribution of demand.

Without any historical records, the newsboy might base his decision on experts' estimates of "low" and "high" demand (we can relate these values to the given range of P) augmented perhaps by subjective probabilities of these outcomes or by a qualitative information such as ranking probabilities of the outcomes. In the former case, he solves (1) for the corresponding discrete distribution P and, naturally, he gets interested in the robustness of the obtained decision, its sensitivity on the occurrence of another outcome (scenario), etc. In the later case, the available qualitative information can be used to define the family $\mathcal{P}$ needed for the worst case analysis.

Newsboy problem cont.

Finally, the newsboy may prefer ANOTHER MODEL

$$\max_{x \geq 0} (c - p)x$$

subject to a reliability type constraint

$$P(x \leq \omega) \geq 1 - \varepsilon.$$

– individual probability constraint which may be written as

$$F(x) \leq \varepsilon \text{ or } x \leq u_\varepsilon(P)$$

with $u_\varepsilon(P)$ the $100\varepsilon\%$ quantile of P. The newsboy chooses the value of $\varepsilon \in (0, 1)$ and according the problem formulation, it will be a value close to 0. The optimal decision is

$$x(P) = u_\varepsilon(P).$$

Notice, that in both cases, the resulting optimal decision can be obtained by solving simple optimization problem

$$\max_{x \geq 0} (c - p)x \text{ subject to } x \leq \hat{\omega}$$

obtained by replacing the random demand by $\hat{\omega}$ — a quantile of its probability distribution P; NOT ecpectation $E_P\omega$!