Chance Constrained Programs

Theory and Solution Methods

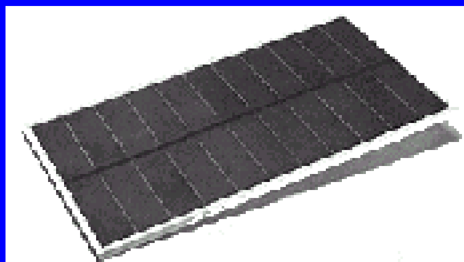
R. Henrion

Weierstrass Institute Berlin

9th International Conference on Stochastic Programming

Humboldt University, Berlin

An Example



stochastic inflow

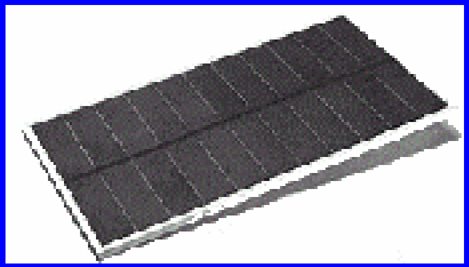
solar panel:

- stochastic output according to daylight

electrical device:

- maximize output over 1 day
- minimum performance $\geq \epsilon > 0$

An Example



solar panel:

- stochastic output according to daylight

stochastic inflow

controlled extraction

electrical device:

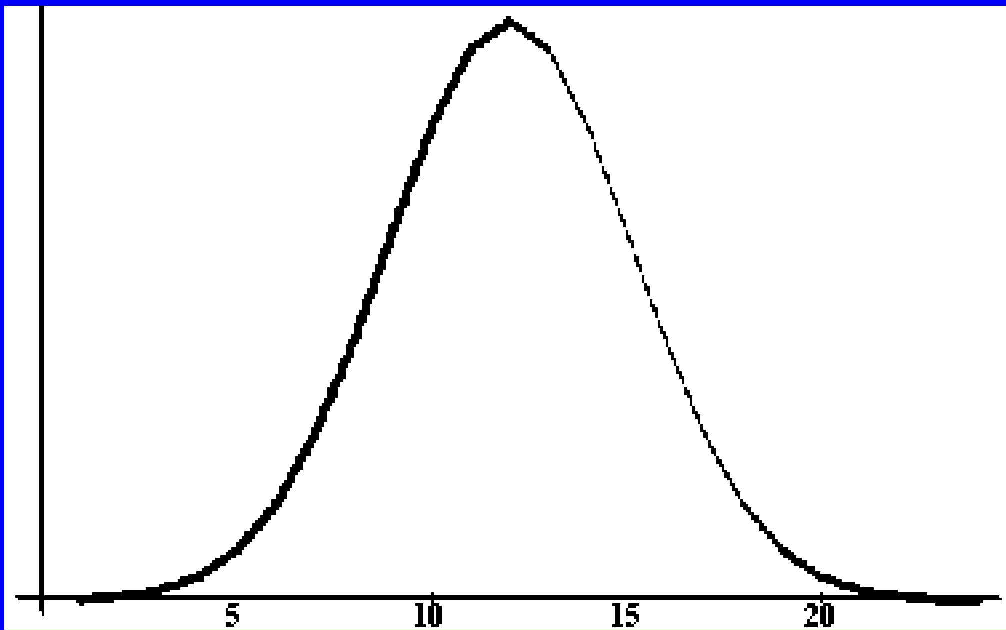
- maximize output over 1 day
- minimum performance $\geq \epsilon > 0$
- hourly programmable output

battery:

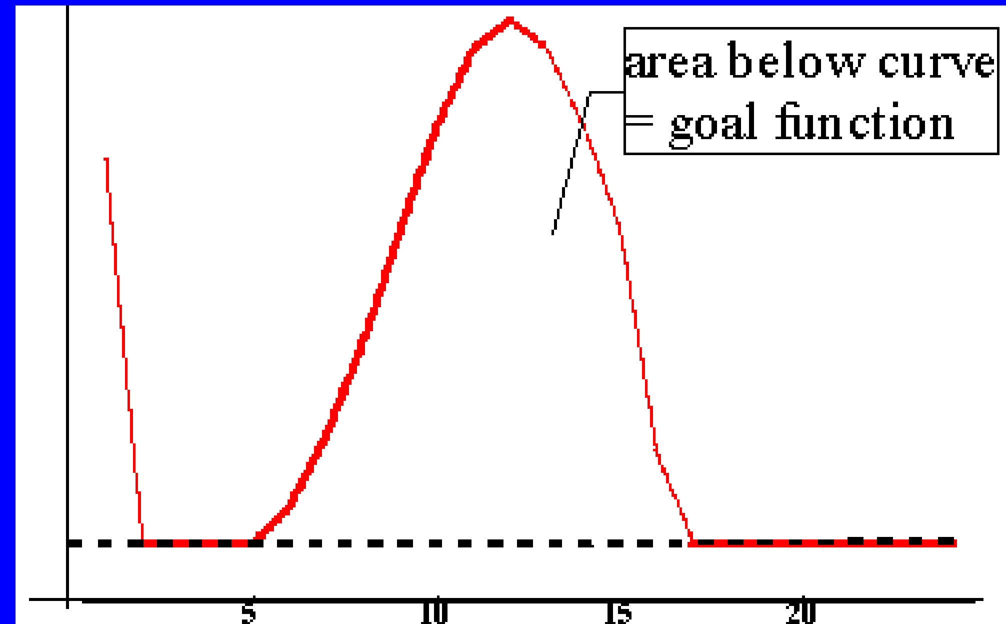
- minimum filling level (capacity) $L > 0$

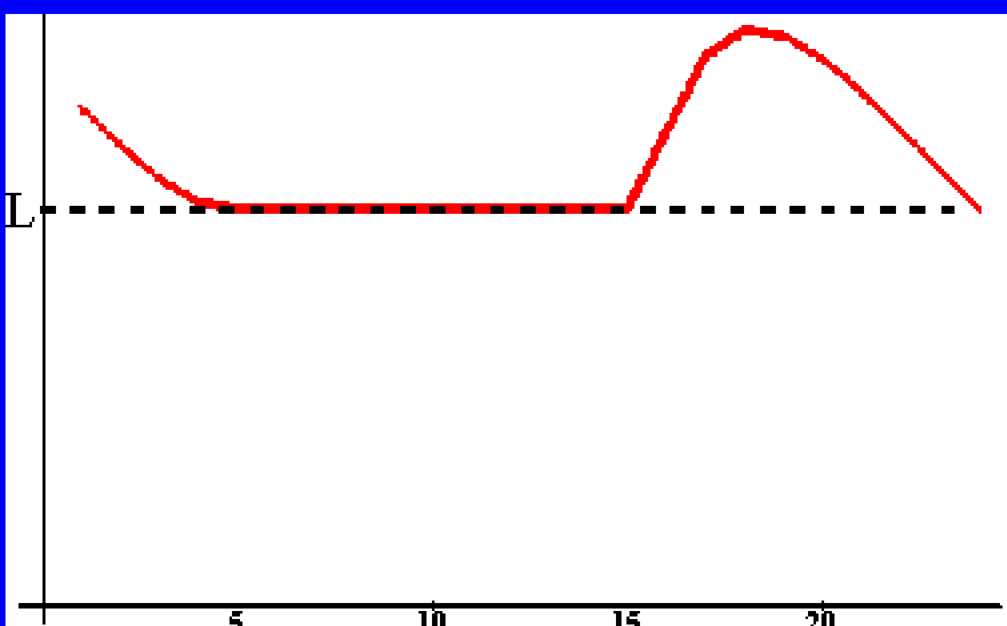
  (risk of damage by low capacity)

solar panel: typical profile of energy output



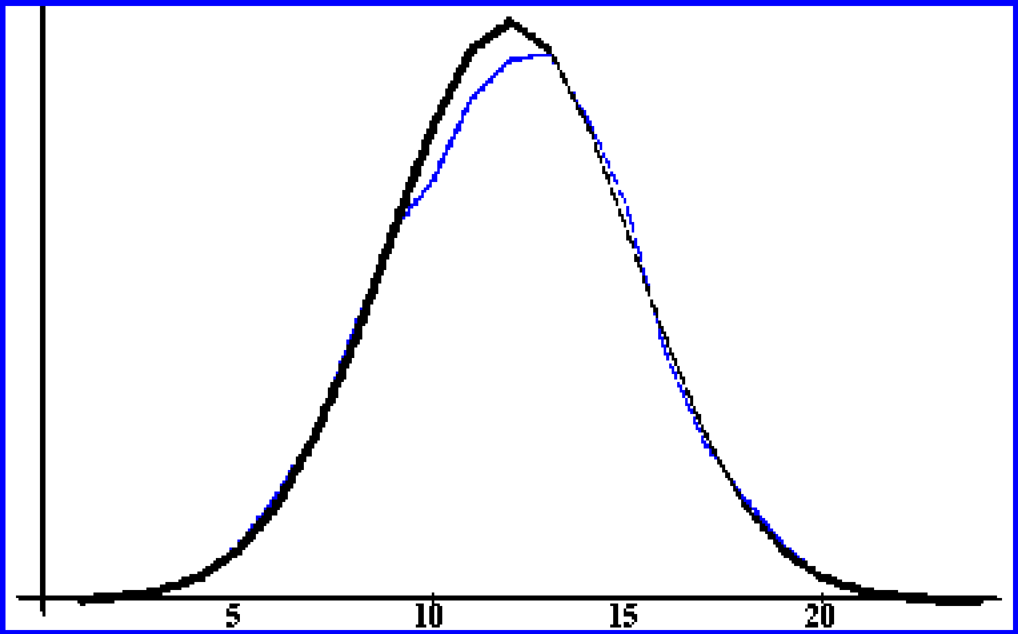
electrical device: optimal consumption profile



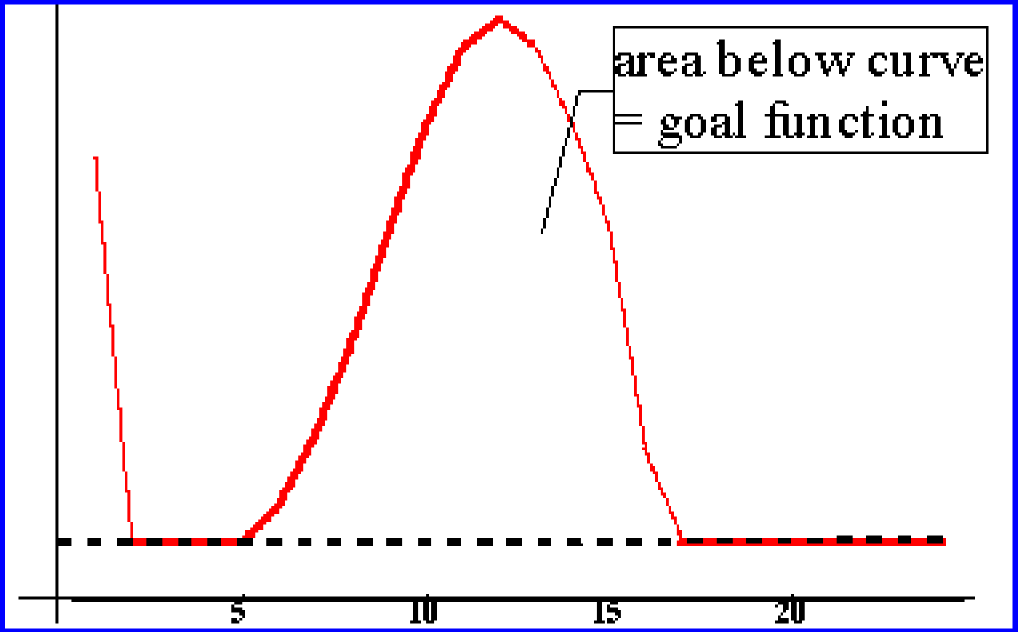


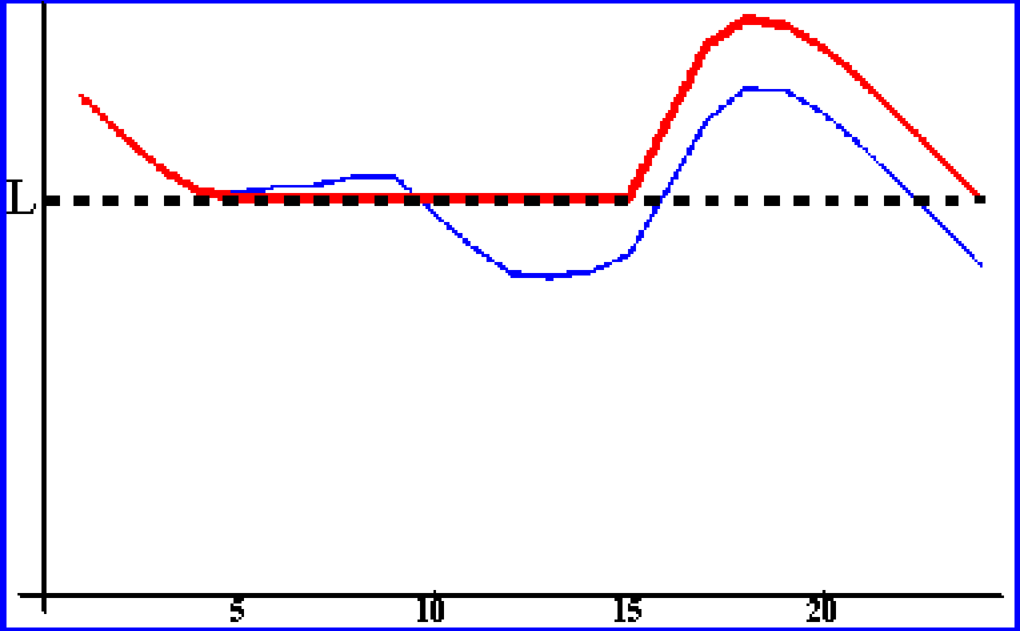
battery: profile of filling level (capacity)

solar panel: sample profile



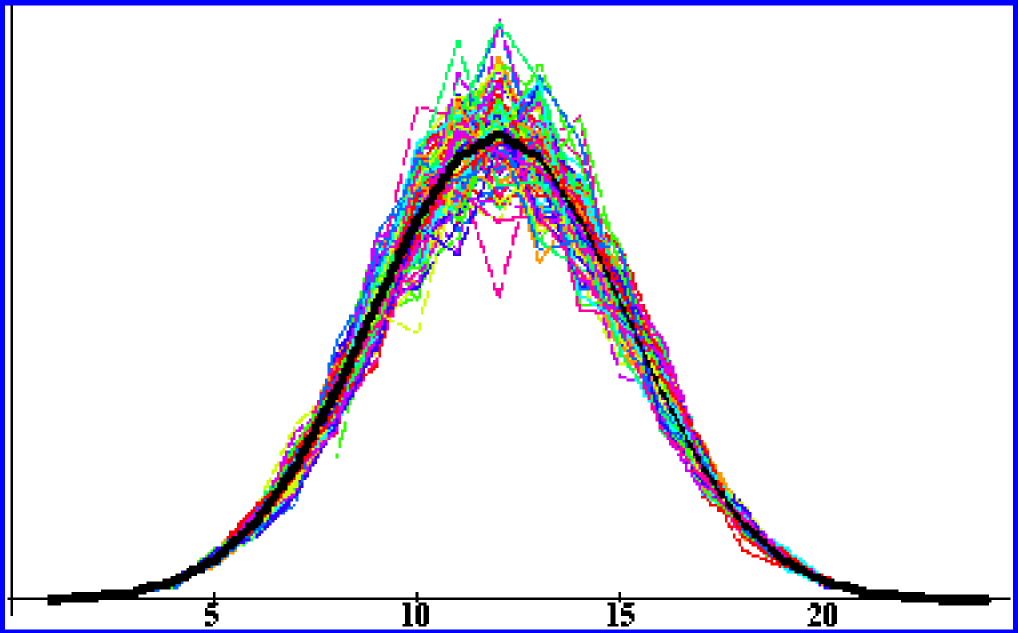
electrical device: optimal consumption profile



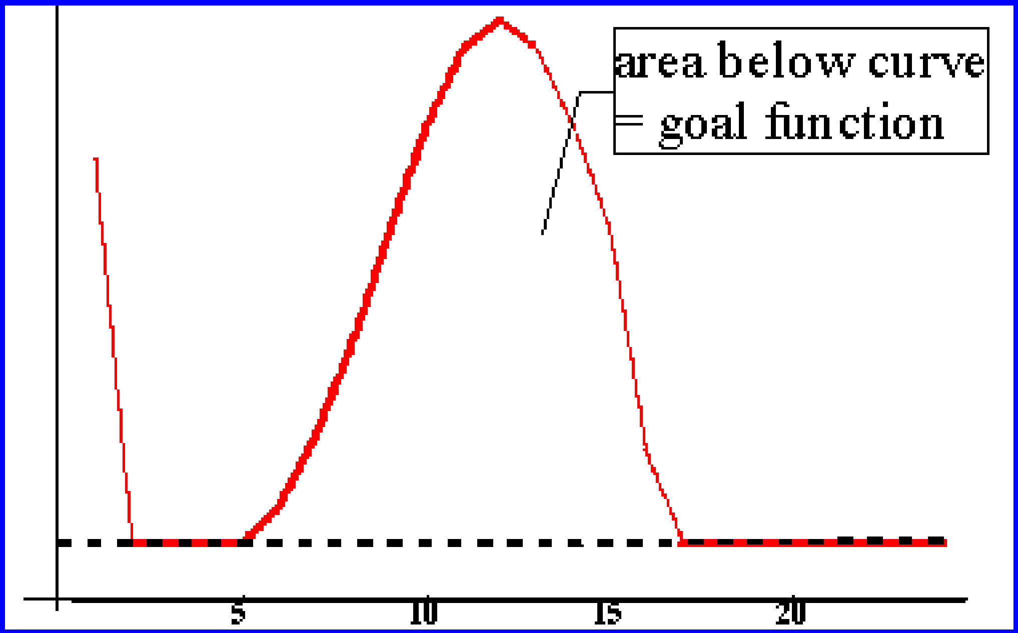


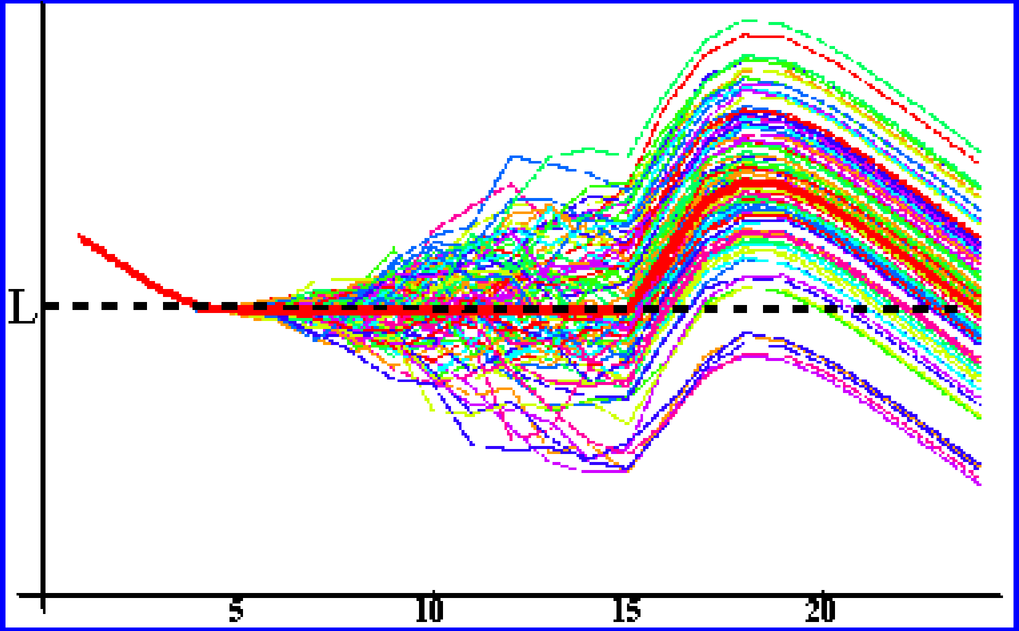
battery: profile of filling level (capacity)

solar panel: 100 sample profiles

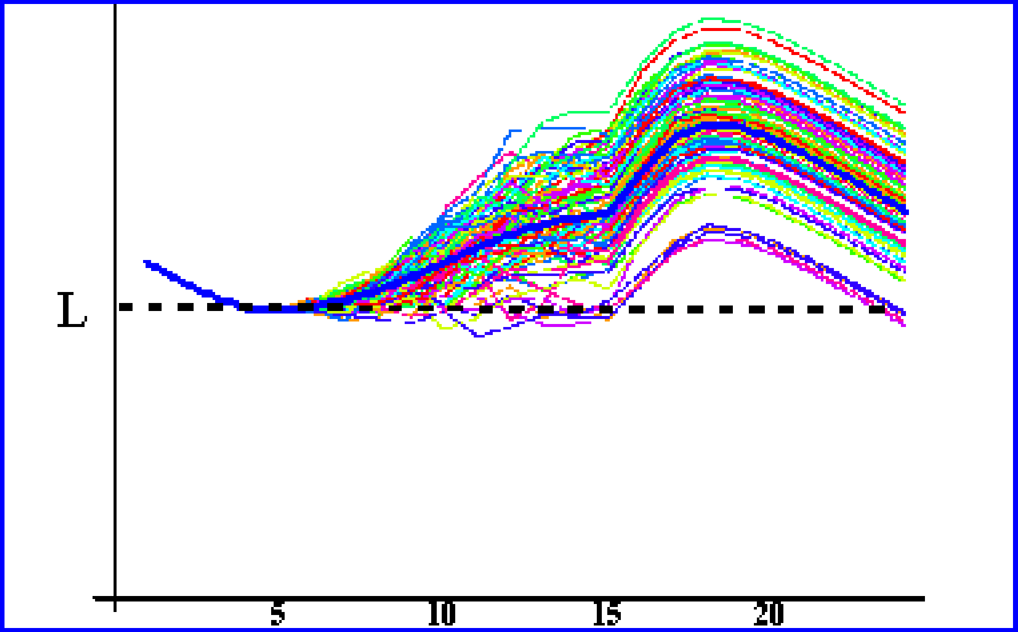


electrical device: optimal consumption profile



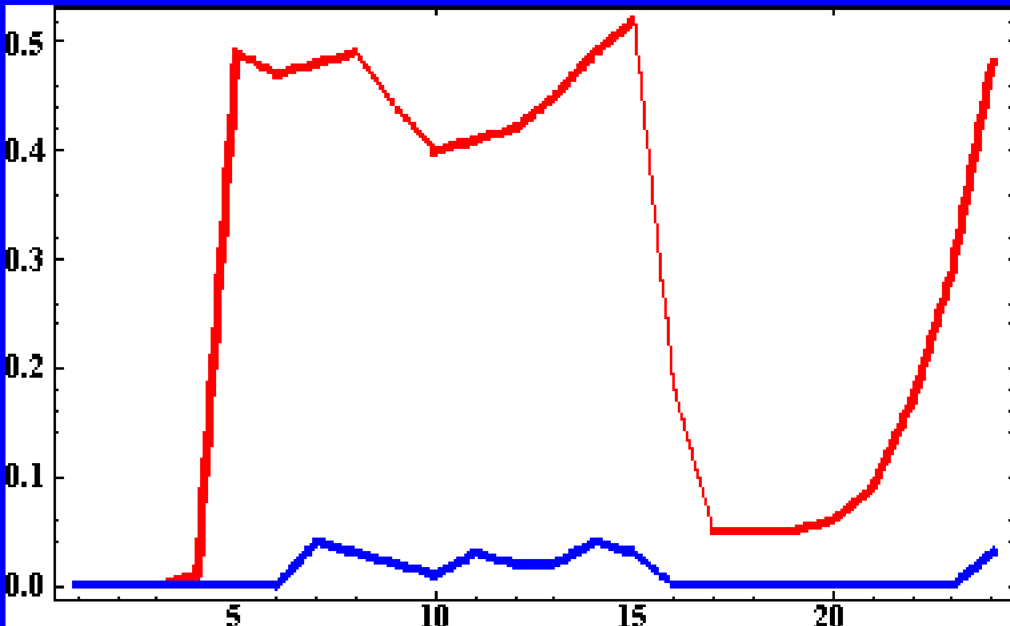


battery: profile of filling level (capacity)

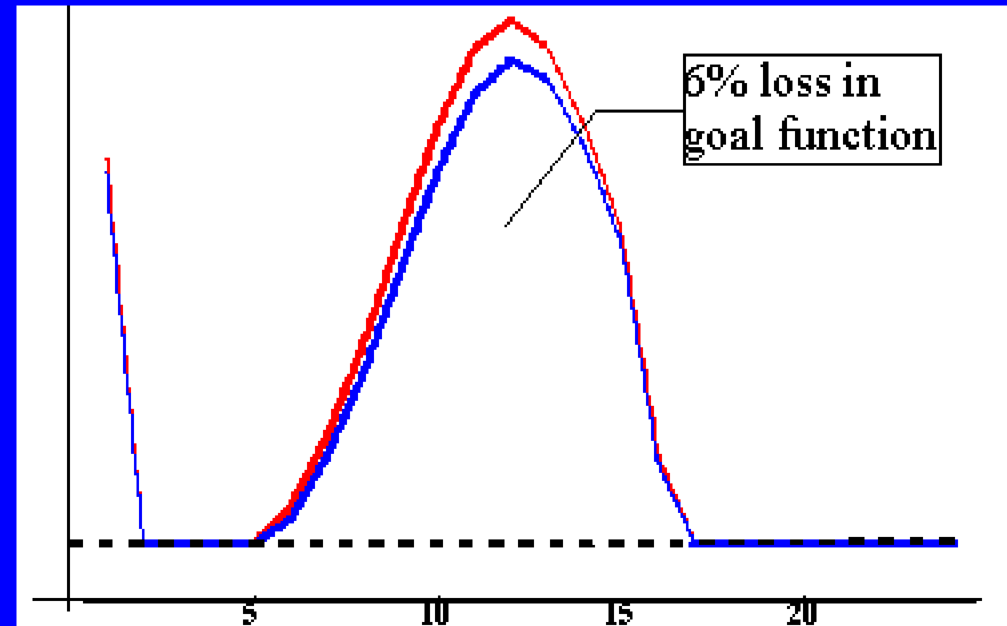


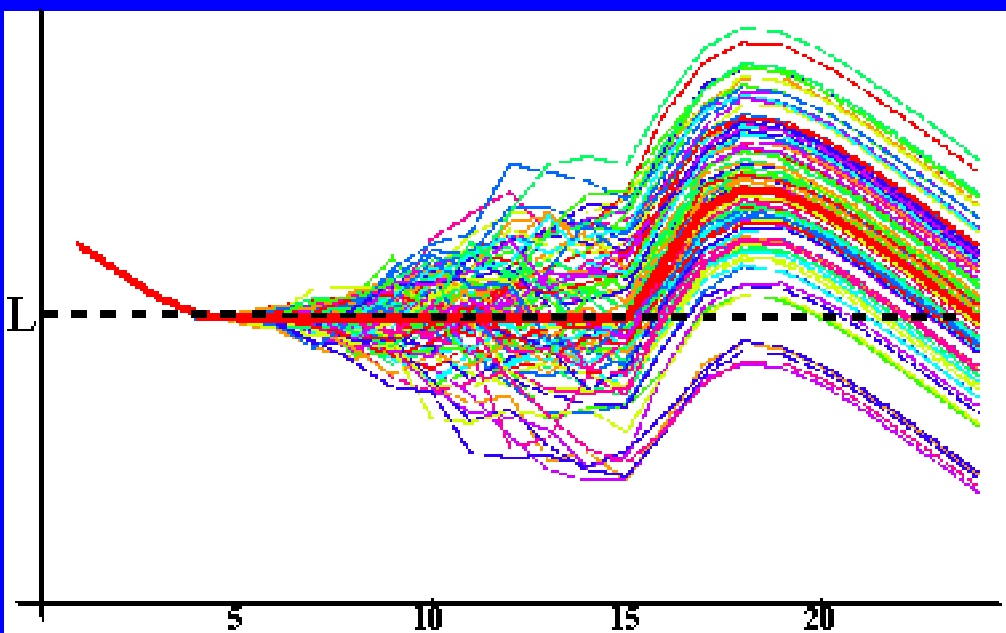
battery: profile of filling level (capacity)

Probability of violating the constraint

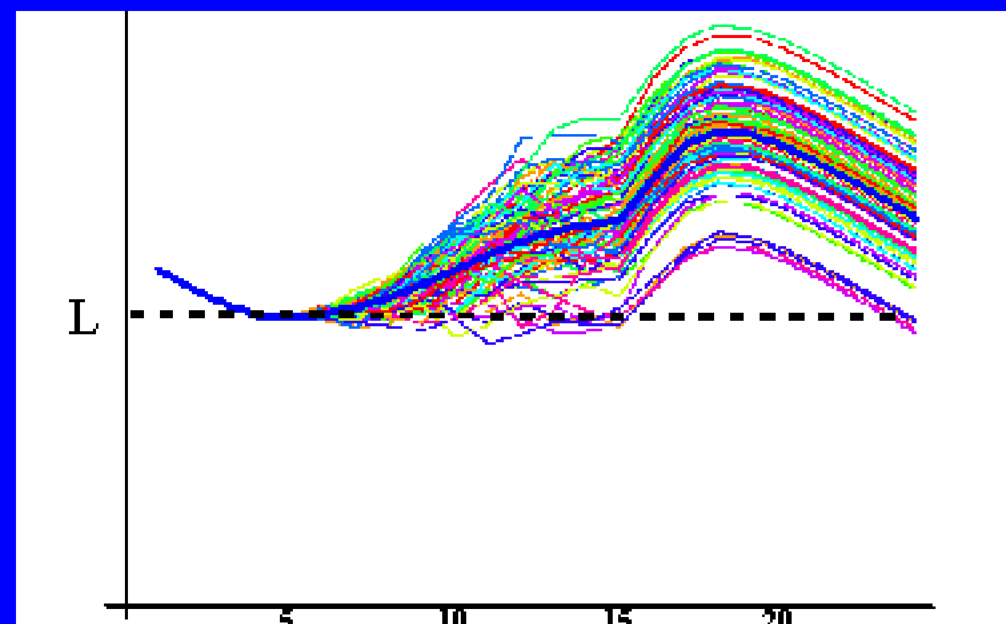


electrical device: optimal consumption profile



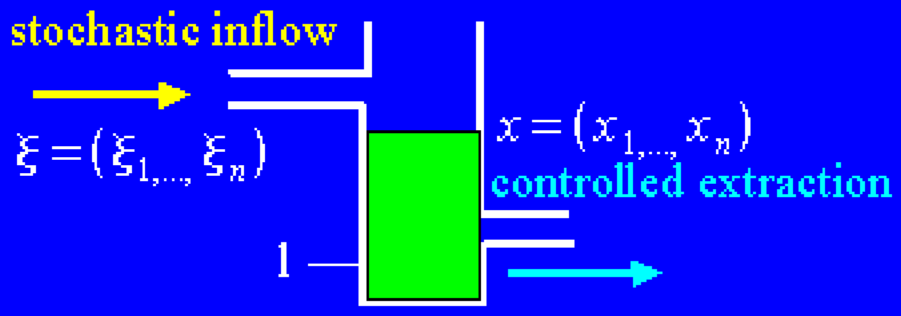


battery: profile of filling level (capacity)



battery: profile of filling level (capacity)

Storage Level Constraints



filling level at time i:

$$l_0 + \xi_1 \cdots + \xi_i - x_1 \cdots - x_i$$

↑ Initial filling level

Assumption: decision on x to be made before observation of ξ

Feasibility: decision x should be chosen such that with high probability the critical lower level l is not violated.

1. Variant: $P(l_0 + \xi_1 \cdots + \xi_i - x_1 \cdots - x_i - l \geq 0) \geq p \quad (i = 1,\dots,n)$ 'pointwise constraints'

↑ probability measure ↑ probability level

2. Variant: $P(l_0 + \xi_1 \cdots + \xi_i - x_1 \cdots - x_i - l \geq 0 \quad (i = 1,\dots,n)) \geq p$ 'simultaneous constraints'

short: $P(L\xi - Lx - \tilde{l} \geq 0) \geq p$ with

$$L = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

Chance Constraints

Deterministic system of inequalities: $h(x) \geq 0 \quad (h: \mathbb{R}^n \to \mathbb{R}^m)$

Stochastic system of inequalities: $h(x, \xi) \geq 0 \quad (h: \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^m)$

$\xi \sim s-dimensional\ random vector\ on\ (\Omega, A, P)$

x ⟶ ξ ('here and now')

Chance constraints: $P(h(x, \xi) \geq 0) \geq p \quad (p \in (0,1))$

⇕

$\phi(x) \geq p$

(ordinary nonlinear inequality)

Major challenges:

1. Structural properties: Impact of h, ξ on properties of ϕ (convexity, differentiability) and of the feasible set (polyhedrality, convexity, connectedness)

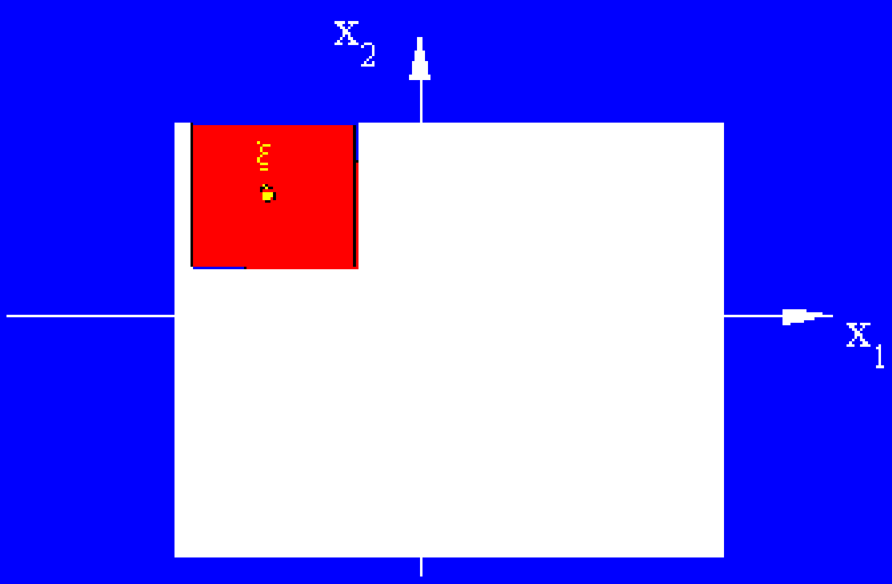
2. Numerical evaluation: How to calculate ϕ and $\nabla \phi$?

Random Sets and Chance Constraints

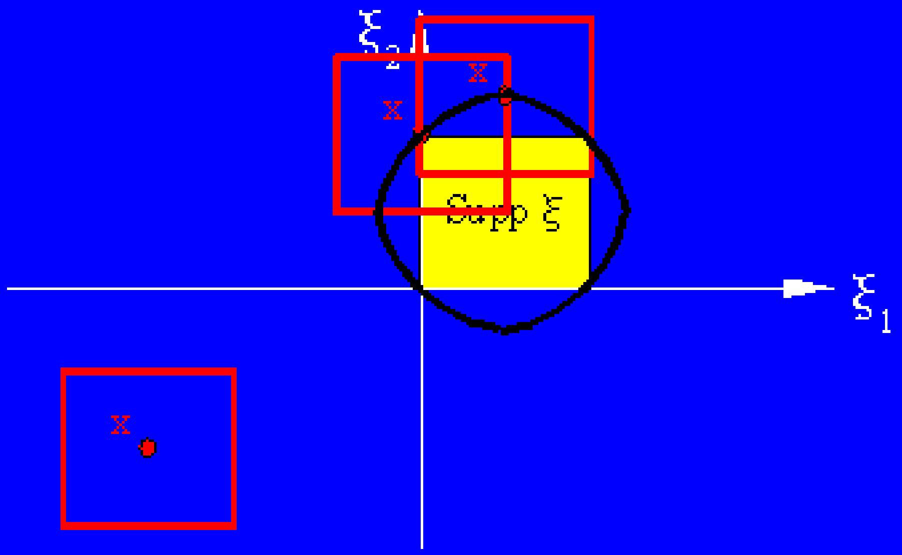
Example: $P(\ |x_1-\xi_1|\le 1/2,\ |x_2-\xi_2|\le 1/2\ )\ \ge\ 1/4$

$$\xi \sim U([0,1]\times[0,1])$$

x- space



ξ - space



Linear chance constraints

1. Type: $P(Ax \geq \xi) \geq p$, i.e., $h(x,\xi) = Ax - \xi$

→ linear chance constraints with random right-hand side

Example: storage level constraints $P(L\xi - Lx - \tilde{l} \geq 0) \geq p$

$$\Updownarrow$$

$$P(Ax \geq \tilde{\xi}) \geq p \quad (A := -L,\ \tilde{\xi} := \tilde{l} - L\xi)$$

2. Type[1]: $P(\Xi x \geq \xi) \geq p$, i.e., $h(x,\tilde{\xi}) = \Xi x - \xi$ for $\tilde{\xi} = (\Xi, \xi)$

→ linear chance constraints with random technology matrix

1: van de Panne and Popp (1963)

Special cases and simplifications

Joint chance constraints[1]: $P(h(x,\xi)\geq 0)\geq p$

Individual chance constraints[2]: $P(h_i(x,\xi)\geq 0)\geq p_i \quad (i=1,\dots,m)$

↕ 1-dimensional random variables

Chance constraints with independent components[3]: $P(h_1(x,\xi)\geq 0)\cdots P(h_m(x,\xi)\geq 0)\geq p$

Explicit chance constraints: $P(h(x)\geq \xi)\geq p$

↕ composition of distribution and constraint function

$F_\xi(h(x))\geq p$

1: Prékopa (1970) 2: Charnes, Cooper, Symonds (1958) 3: Miller, Wagner (1965)

Individual explicit chance constraints

$$P(h_i(x)\geq \xi_i)\geq p_i \quad (i=1,\dots,m) \longleftrightarrow F_{\xi_i}(h_i(x))\geq p_i \quad (i=1,\dots,m)$$

$$\longleftrightarrow h_i(x)\geq \alpha_i \quad \left(\alpha_i=\left(F_{\xi_i}\right)^{-1}(p_i)\right) \quad (i=1,\dots,m)$$

These chance constraints have the same structure as the original constraints!

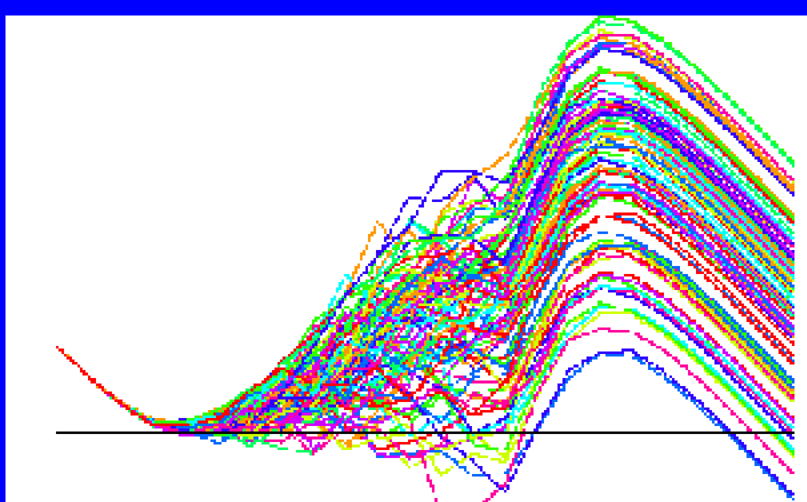
Example: pointwise storage level constraints.

$$P(l_0+\xi_1\cdots+\xi_i-x_1\cdots-x_i-l\geq 0)\geq p \quad (i=1,\dots,n) \quad (h_i(x)=-x_1\cdots-x_i;\ \tilde{\xi}_i=l-l_0-\xi_1-\cdots-\xi_i)$$

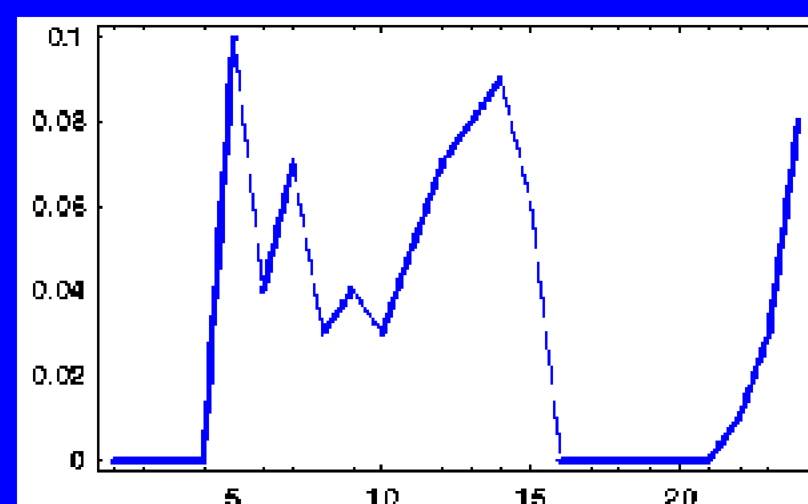
h_i linear $\longrightarrow$ feasible set defined by chance constraints is a polyhedron.

Individual chance constraints are weaker than joint chance constraints.

pointwise probability of violation: 5%



empirical probabilities



Individual explicit chance constraints

$$P(h_i(x) \geq \xi_i) \geq p_i \quad (i = 1,\ldots,m) \longleftrightarrow F_{\xi_i}(h_i(x)) \geq p_i \quad (i = 1,\ldots,m)$$

$$\longleftrightarrow h_i(x) \geq \alpha_i \quad (\alpha_i = \left(F_{\xi_i}\right)^{-1}(p_i)) \quad (i = 1,\ldots,m)$$

These chance constraints have the same structure as the original constraints!

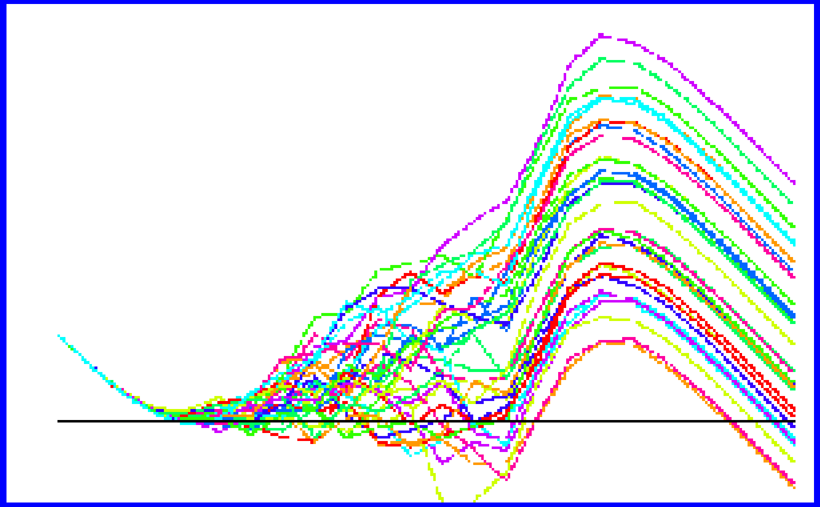
Example: pointwise storage level constraints.

$$P(l_0 + \xi_1 \cdots + \xi_i - x_1 \cdots - x_i - l \geq 0) \geq p \quad (i = 1,\ldots,n) \quad (h_i(x) = -x_1 \cdots - x_i;\ \tilde{\xi}_i = l - l_0 - \xi_1 - \cdots - \xi_i)$$

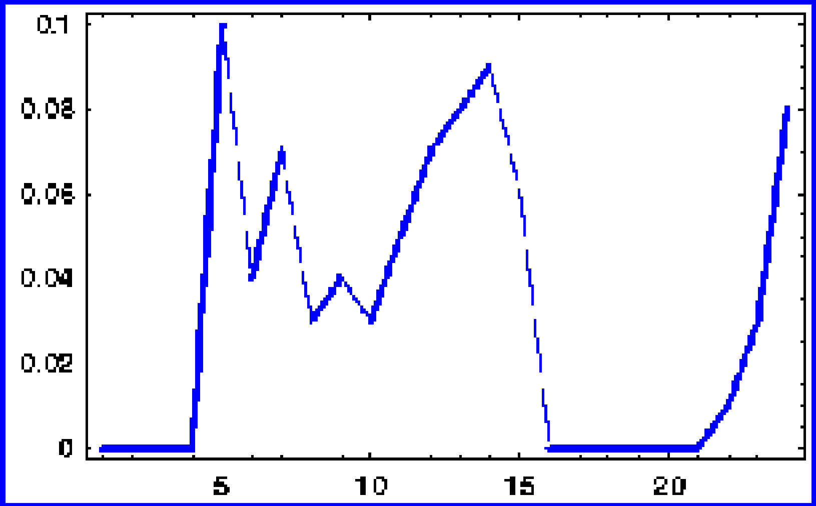
h_i linear ⟶ feasible set defined by chance constraints is a polyhedron.

Individual chance constraints are weaker than joint chance constraints.

probability of simultaneous violation: 35%



empirical probabilities



Convexity of chance constraints

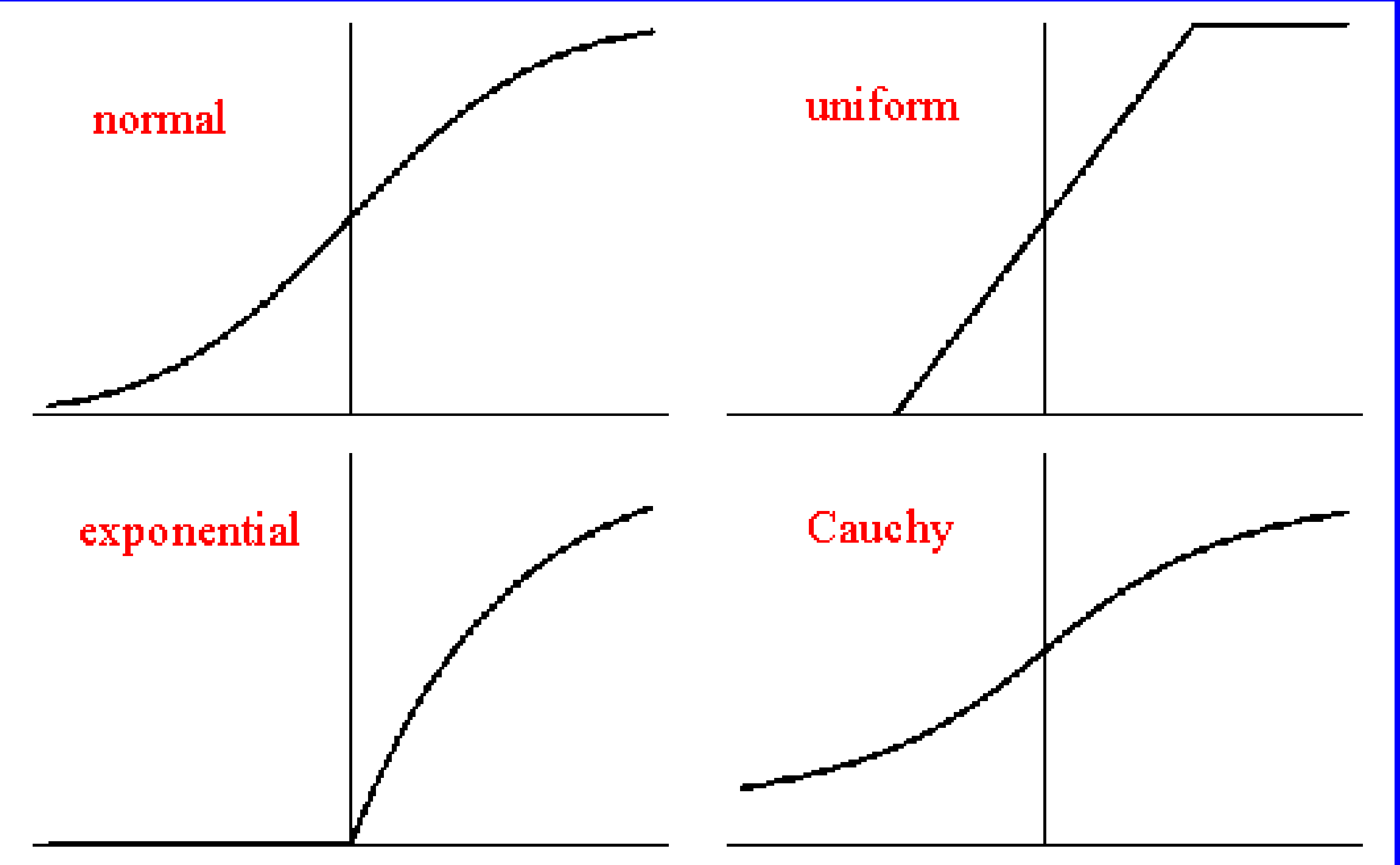
When is the set $\{x \mid P(h(x,\xi)\geq 0)\geq p\}$ convex ?

Conditions on the constraint function h and the probability distribution $P\circ\xi^{-1}$?

Convexity (concavity) of h may be easily checked, but:

Distribution functions are neither convex nor concave!

F_ξ



Convexity of chance constraints

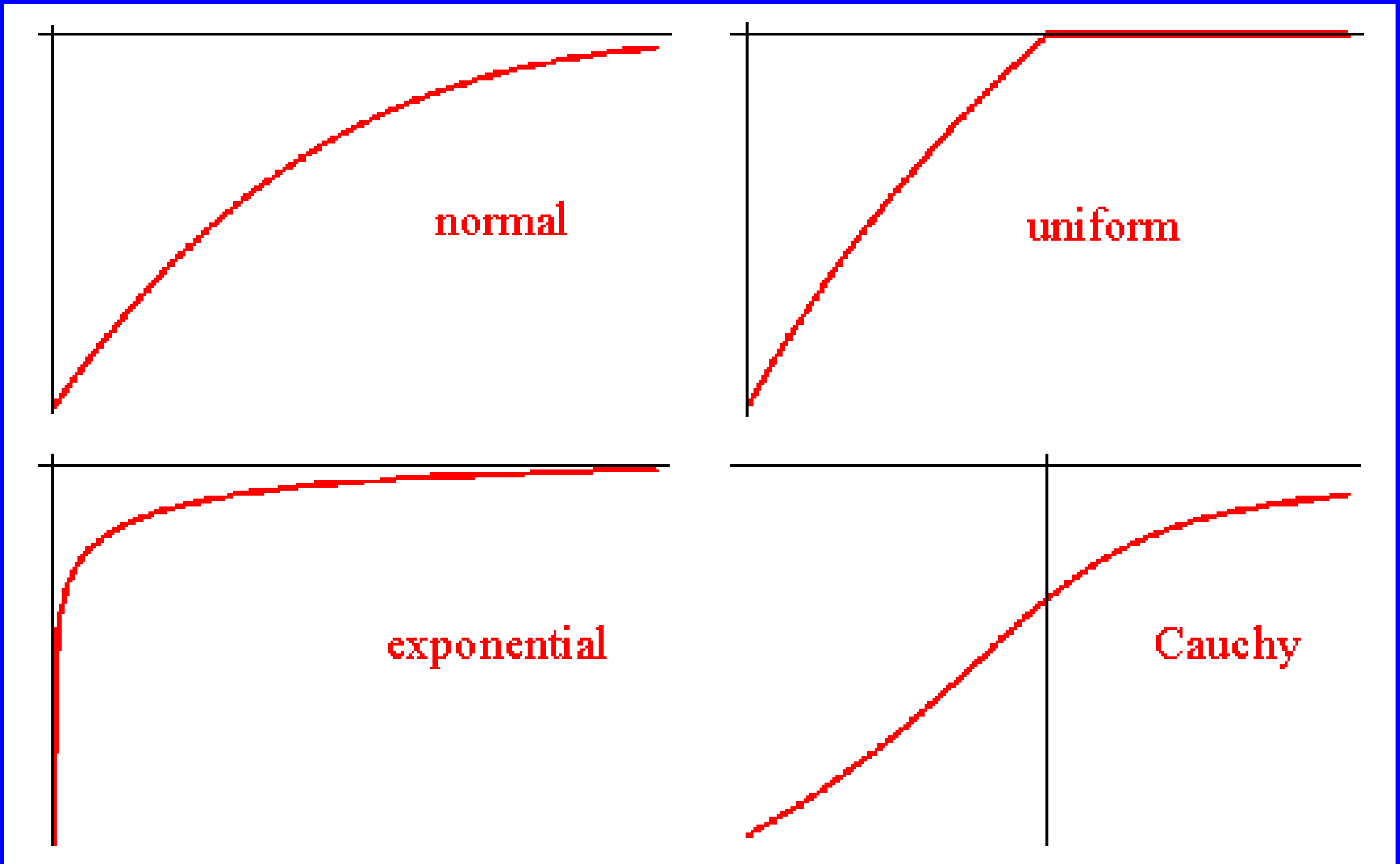
When is the set $\{x \mid P(h(x,\xi)\geq 0)\geq p\}$ convex ?

Conditions on the constraint function h and the probability distribution $P\circ\xi^{-1}$?

Convexity (concavity) of h may be easily checked, but:

Distribution functions are neither convex nor concave!

$\log(F_\xi)$



Convexity of chance constraints

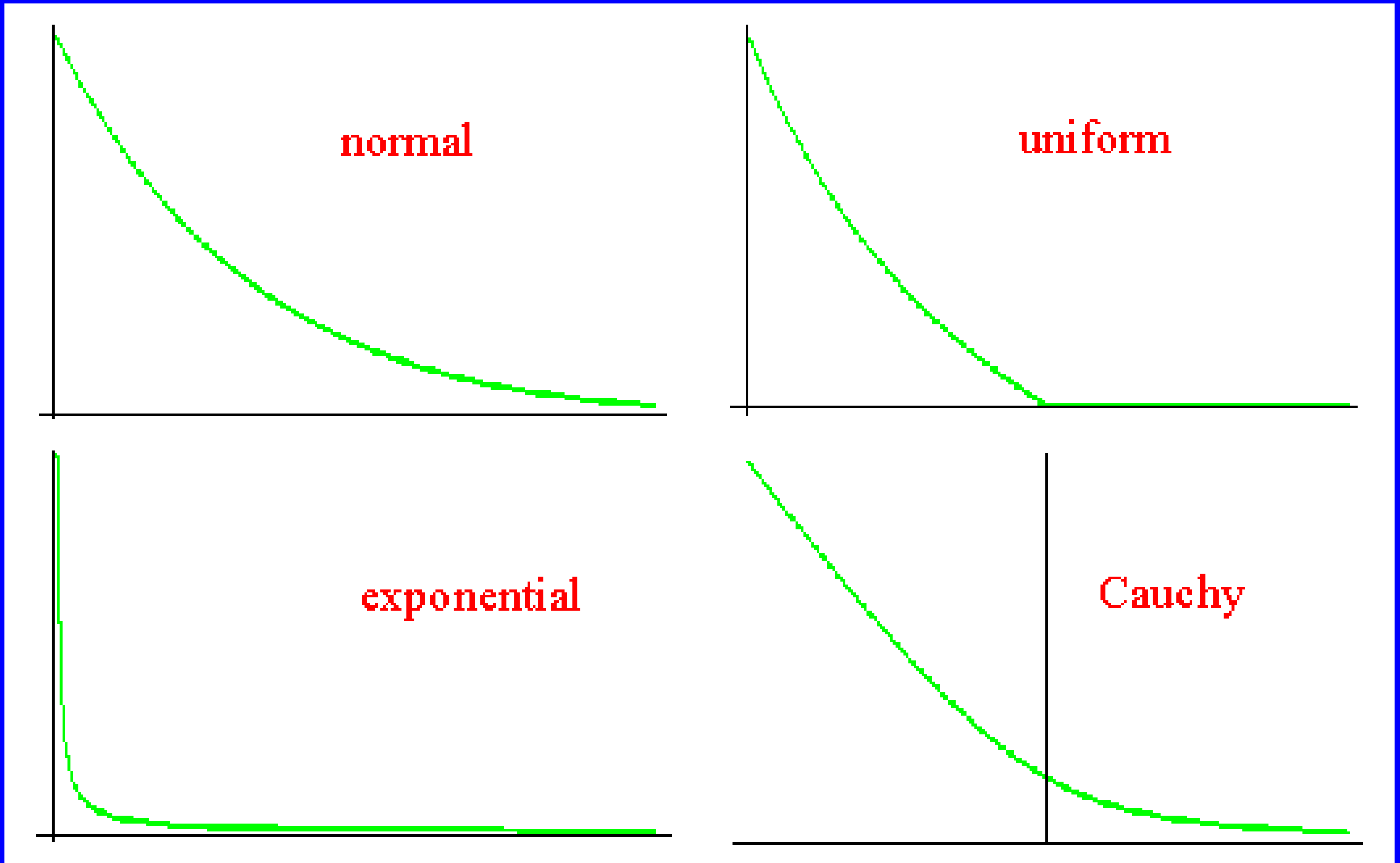
When is the set $\{x \mid P(h(x,\xi)\geq 0)\geq p\}$ convex ?

Conditions on the constraint function h and the probability distribution $P\circ\xi^{-1}$?

Convexity (concavity) of h may be easily checked, but:

Distribution functions are neither convex nor concave!

$$(F_{\xi})^{-1}$$



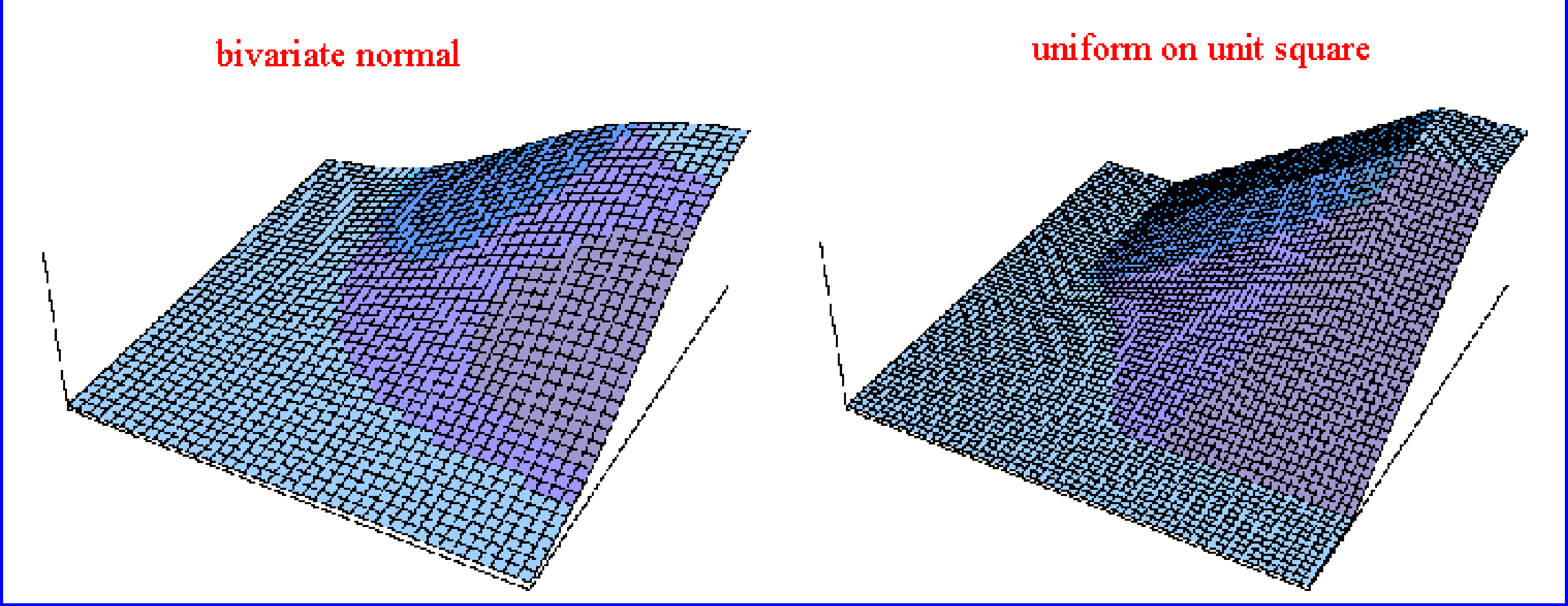
Convexity of chance constraints

When is the set $\{x \mid P(h(x,\xi)\geq 0)\geq p\}$ convex ?

Conditions on the constraint function h and the probability distribution $P\circ\xi^{-1}$?

Convexity (concavity) of h may be easily checked, but:

Distribution functions are neither convex nor concave!



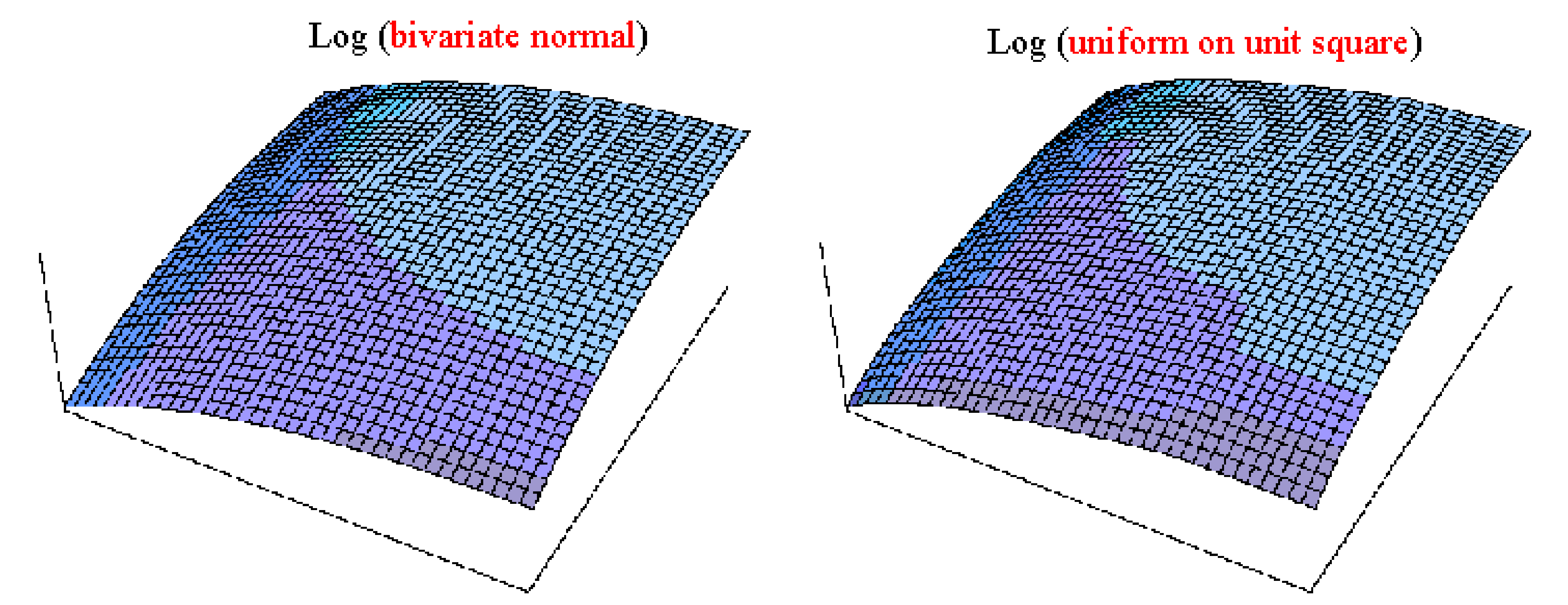
Convexity of chance constraints

When is the set $\{x \mid P(h(x,\xi)\geq 0)\geq p\}$ convex ?

Conditions on the constraint function h and the probability distribution $P\circ\xi^{-1}$?

Convexity (concavity) of h may be easily checked, but:

Distribution functions are neither convex nor concave!



Log- and r-concave probability measures

A probability measure μ on $\mathbb{R}^s$ is called Log-concave or r-concave if

$$\log(\mu(\lambda A+(1-\lambda)B))\geq \lambda \log \mu(A)+(1-\lambda)\log \mu(B)$$

or

$$\mu^r(\lambda A+(1-\lambda)B)\leq \lambda\, \mu^r(A)+(1-\lambda)\mu^r(B)$$

$$\forall A,B\subseteq \mathbb{R}^s (\textit{Borel measurable, convex})\,\forall \lambda\in[0,1]$$

i.e., $\log \mu$ is a concave set function and μ^r is a convex set function.

Lemma: If in the chance constrained set $M = \{x \mid P(h(x,\xi)\geq 0)\geq p\}$

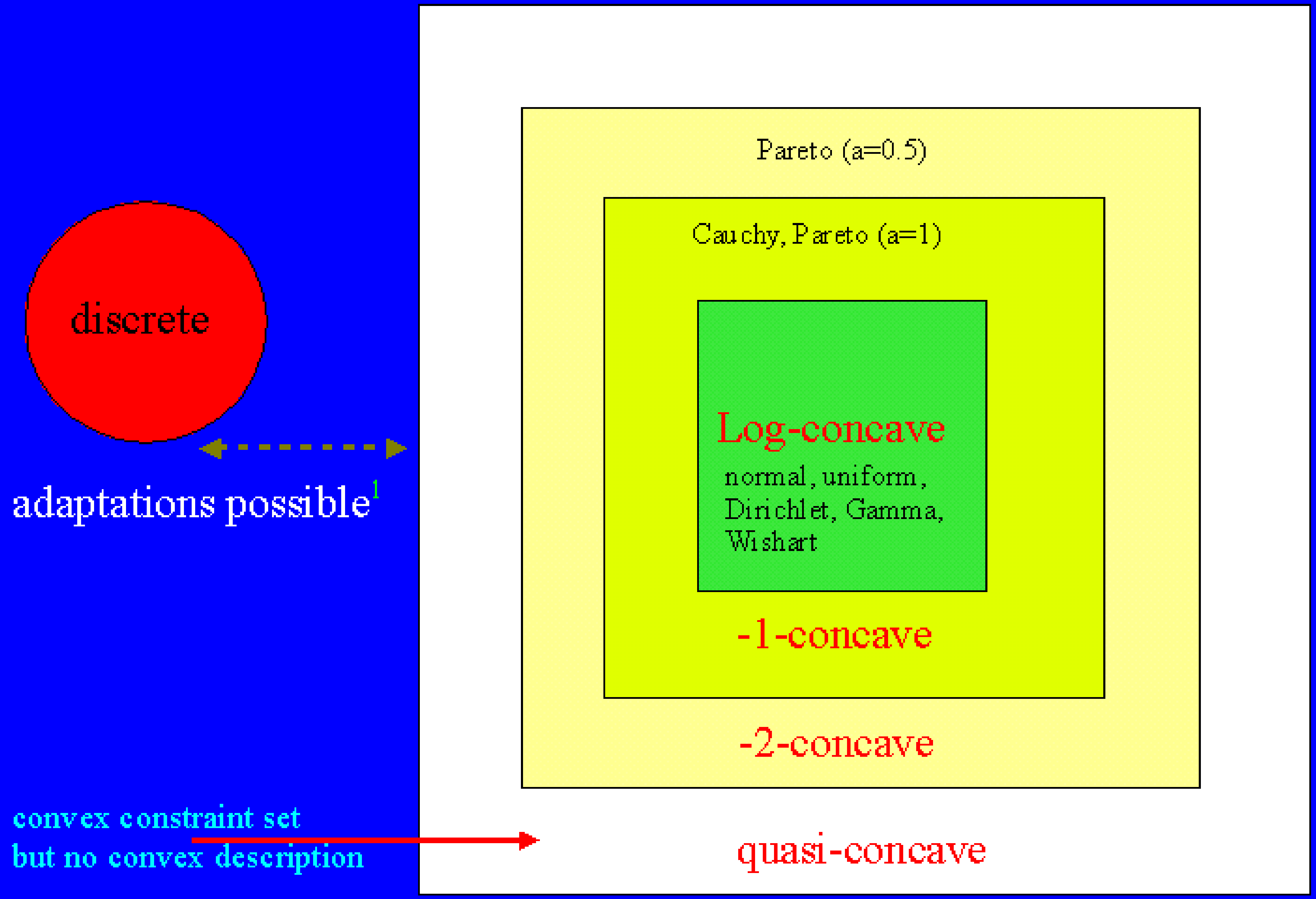
- h is quasi-concave (e.g.: concave)
- $\mu := P\circ\xi^{-1}$ is log-concave or r-concave for some $r < 0$

then M is a convex set and gets the convex description

$$-\log P(h(x,\xi)\geq 0)\leq -\log p \qquad \text{or} \qquad P^r(h(x,\xi))\leq p^r$$

convex inequalities in x

Hierarchy of r-concave measures



1: Dentcheva, Prékopa, Ruszczynski (1998)

Characterization of r-concavity from density functions

Assume that the probability measure μ on $\mathbb{R}^s$ is induced by a density f.

Theorem[1]: If log f is concave then μ is a logconcave measure.

Theorem[2]: If f^r is convex for some $r \in [-s^{-1}, 0)$ then μ is $r(1+sr)^{-1}$ – concave.

Examples:

- multivariate normal: $f(x) = K \exp(-(x-m)^T C^{-1}(x-m)/2)$

  → $\log f(x) = \log K - (x-m)^T C^{-1}(x-m)/2$ → concave → μ log-concave

- Cauchy: $f(x) = [\pi(1+x^2)]^{-1}$ $s = 1,\ r = -0.5$

  → $f^{-0.5}(x) = \sqrt{\pi(1+x^2)}$ → μ (-1) concave

1: Prékopa (1973) 2: Borell (1975)

Convexity of linear chance constraints

For sets $M = \{x \mid P(Ax \geq \xi) \geq p\}$ defined by linear chance constraints with random right hand side we have the following corollary:

Corollary: If $P \circ \xi^{-1}$ is quasi-concave (e.g. log-concave or r-concave for some r<0) then M is convex.

However, if also the technology matrix is random, then the constraint set $M = \{x \mid P(\Xi x \geq \xi) \geq p\}$ is not necessarily convex.

Theorem[1]: Let $(\Xi, \xi) \sim N(m, C)$. Then, the constraint set

$$M = \{x \mid P([\Xi x]_i \geq \xi_i) \geq p_i \ \ (i = 1, \ldots, k)\}$$

defined by individual linear chance constraints is convex if

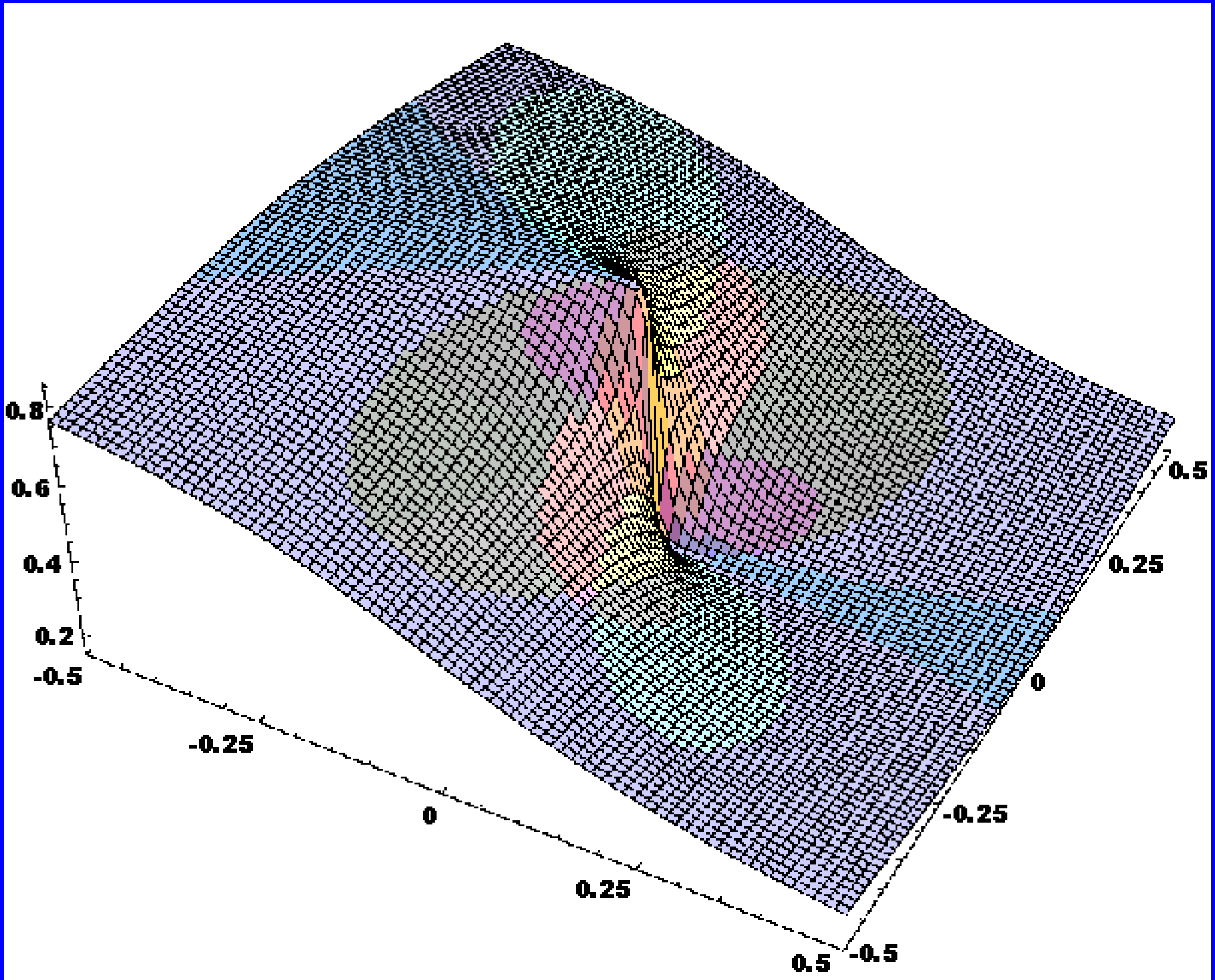
$$p_i \in (-\infty, \Phi(-\lambda^{-1/2} \|m\|)) \cup [0.5, \infty) \quad (i = 1, \ldots k).$$

Φ = standard normal distribution; λ = smallest eigenvalue of C.

1: van de Panne and Popp (1963)

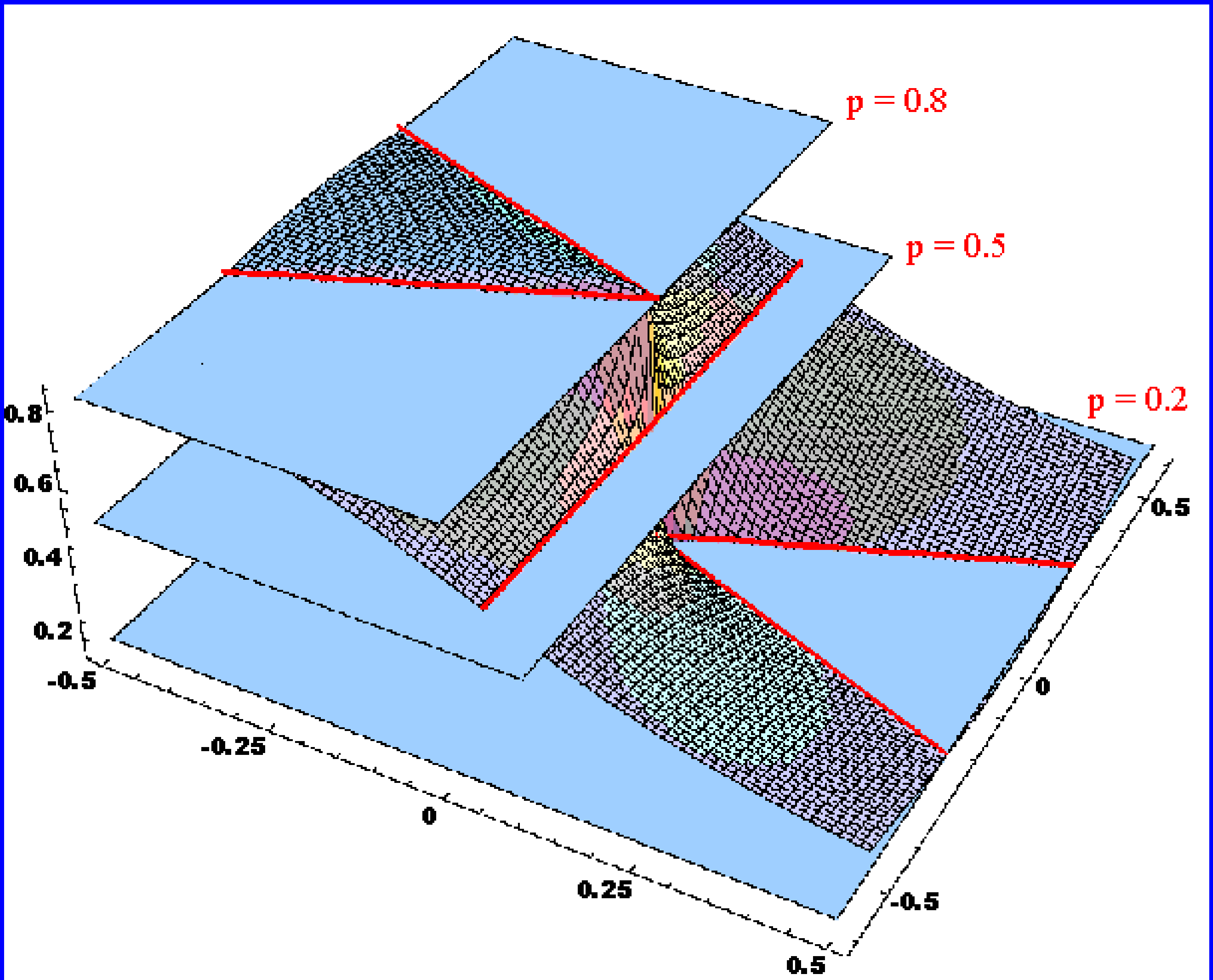
Example: $M=\{x\in\mathbb{R}^2 \mid P(\langle x,\xi\rangle\leq 0)\geq p\}\quad \xi\sim N((1,0),I_2)$

Equivalent description: $M=\{x\in\mathbb{R}^2 \mid \Phi(-x_1\|x\|^{-1})\geq p\}$



Example: $M=\{x\in\mathbb{R}^2 \mid P(\langle x,\xi\rangle\le 0)\ge p\}\quad \xi\sim N((1,0),\boldsymbol{I}_2)$

Equivalent description: $M=\{x\in\mathbb{R}^2 \mid \Phi(-x_1\|x\|^{-1})\ge p\}$



Connectedness of chance constraints

Beyond convexity, the weaker property of connectedness of chance constraints may be of theoretical and numerical interest (e.g., homotopy methods).

Theorem[1]: The set $M = \{x \mid P(Ax \geq \xi) \geq p\}$ is (path-) connected provided that the rows of A are positively linear independent.

No assumption on the probability distribution!

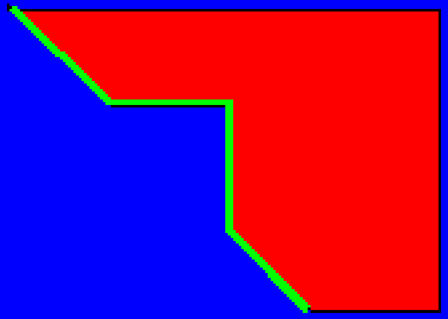
→ Result also applies to discrete distributions.

Example:

$$M = \{(x,y) \in \mathbb{R}^2 \mid P(x \geq \xi_1, y \geq \xi_2, x + y \geq \xi_3) \geq 0.5\}$$

$$\xi \sim (\delta_{e_1} + \delta_{e_2} + \delta_{e_3})/3$$

$$\rightarrow \quad A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$$



not convex

but connected

Connectedness of chance constraints

Beyond convexity, the weaker property of connectedness of chance constraints may be of theoretical and numerical interest (e.g., homotopy methods).

Theorem[1]: The set $M=\{x \mid P(Ax \geq \xi) \geq p\}$ is (path-) connected provided that the rows of A are positively linear independent.

No assumption on the probability distribution!

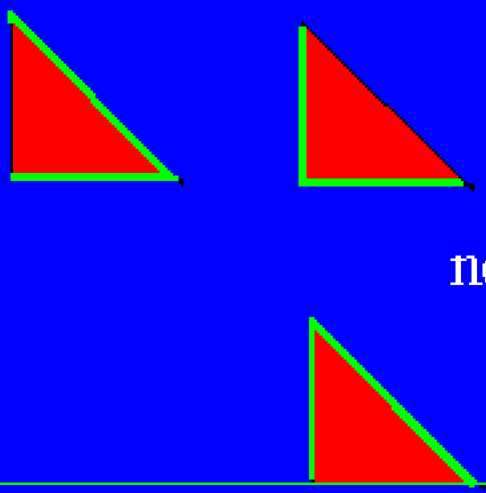
→ Result also applies to discrete distributions.

Example:

$$M=\{(x,y)\in \mathbb{R}^2 \mid P(x\geq \xi_1, y \geq \xi_2, -x-y\geq \xi_3)\geq 0.5\}$$

$$\xi \sim (\delta_{e_1}+\delta_{e_2}+\delta_{e_3})/3$$

→ $A=\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix}$ rows not positively independent !



not connected !

1: Henrion (2001)

Calculating chance constraints

Main job in numerical treatment of chance constraints:

Calculation of $\phi(x) := P(h(x,\xi) \geq 0)$ and of $\nabla \phi(x)$ $(\nabla^2 \phi(x))$

We content ourselves with explicit joint chance constraints under multivariate normal distribution:

$$\phi(x) := P(h(x) \geq \xi) = F_\xi(h(x)) \quad (\xi \sim N(m, C))$$

distribution function

Since h is analytically given, it suffices to calculate F_ξ, ∇F_ξ .

Standardization: With $\tilde{\xi}$ defined by $\tilde{\xi}_i := c_{ii}^{-1/2}(\xi_i - m_i)$ it follows that

$\tilde{\xi} \sim N(0, R)$ (R = correlation matrix) and

$P(h(x) \geq \xi) = P(\tilde{h}(x) \geq \tilde{\xi}) = \Phi_R(x)$ with $\tilde{h}_i := c_{ii}^{-1/2}(h_i - m_i)$

distribution function of standard normal distribution

Reduction of derivatives: Partial derivatives of $\Phi_R(x)$ can be analytically to values of $\Phi_R(x)$ itself.

Calculating the standard normal distribution

Main approaches for calculating $\Phi_R(x)$ in s dimensions:

- 'exact' calculation for s = 1,2[1,2]
- numerical integration[3,4] (up to s ~ 10)
- Monte Carlo simulation
- bounds and simulation (up to s~50)

Distribution function from the probability of the union of events:

$$\Phi_R(x) = P(x_i \geq \xi_i \ \ (i = 1,\ldots,s)) = 1 - P(A_1 \cup \cdots \cup A_s)$$

where $A_i = (\xi_i > x_i)$

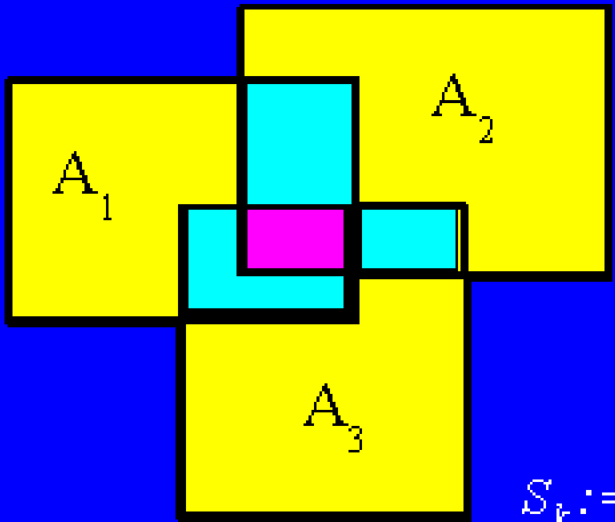
→ bounds for $P(A_1 \cup \cdots \cup A_s)$

Test example: $R = I_4$ (independent components), $x = (1.28, 1.28, 1.28, 0)$

$$\Phi_R(x) = [\Phi_1(1.28)]^3 \cdot \Phi_1(0) = 0.9^3 \cdot 0.5 = 0.3645$$

1: Donnely (1973) 2: Drezner, Wesolowsky (1990) 3: Schervish (1984) 4: Genz (1992)

Bonferroni bounds



$$P(A_1 \cup A_2 \cup A_3) \le P(A_1) + P(A_2) + P(A_3) =: S_1$$

$$P(A_1 \cup A_2 \cup A_3) \ge S_1 - \underbrace{\left[P(A_1 \cap A_2) + P(A_2 \cap A_3) + P(A_1 \cap A_3) \right]}_{S_2}$$

$$P(A_1 \cup A_2 \cup A_3) = S_1 - S_2 + P(A_1 \cap A_2 \cap A_3)$$

$$S_k := \sum_{i_1 < \cdots < i_k} P(A_{i_1} \cap \cdots \cap A_{i_k})$$ (sums of k-th order intersections)

Bonferroni: $P(A_1 \cup \cdots \cup A_s) = S_1 - S_2 + S_3 \cdots - / + \cdots (-1)^{s-1} S_s$

For calculation of $\Phi_R(x)$

$$P(A_i) = P(\xi_i > x_i) = \int_{x_i}^{\infty} f_1(t)\,dt \qquad P(A_i \cap A_j) = P(\xi_i > x_i, \xi_j > x_j) = \int_{x_i}^{\infty} \int_{x_j}^{\infty} f_2(t_1, t_2)\,dt_1\,dt_2$$

1- and 2-dimensional integration

no problem to calculate S_1, S_2 Difficult: $S_3, S_4, \cdots$

Bounds based on S_1, S_2: $\Phi_R(x) = 1 - P(A_1 \cup \cdots \cup A_s) \in \left[1 - S_1, 1 - S_1 + S_2 \right]$

Example: $\Phi_R(\ 1.28,\ 1.28,\ 1.28,\ 0\ )$

$P(A_1) = P(A_2) = P(A_3) = 0.1;\ \ P(A_4) = 0.5 \quad \Rightarrow S_1 = 0.8 \ \Rightarrow 1 - S_1 = \ 0.2$ lower bound

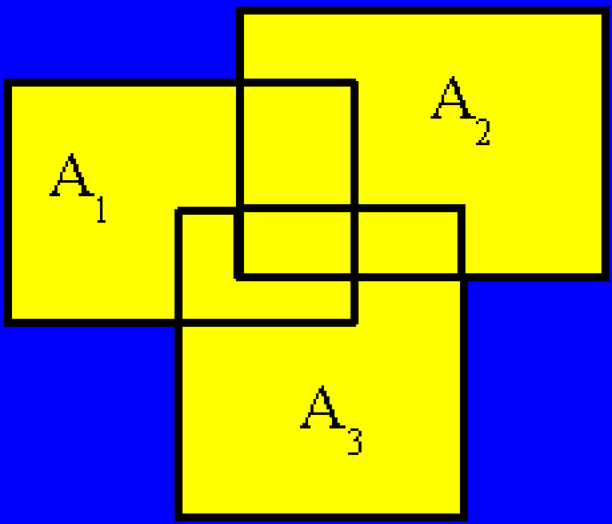
$P(A_1 \cap A_2) = P(A_2 \cap A_3) = P(A_1 \cap A_3) = 0.01$

$P(A_1 \cap A_4) = P(A_2 \cap A_4) = P(A_3 \cap A_4) = 0.05 \quad \Rightarrow S_2 = 0.18 \ \Rightarrow 1 - S_1 + S_2 = \ 0.38$

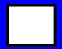
upper bound

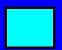
True value: 0.3645

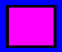
Bounds by linear programming[1]



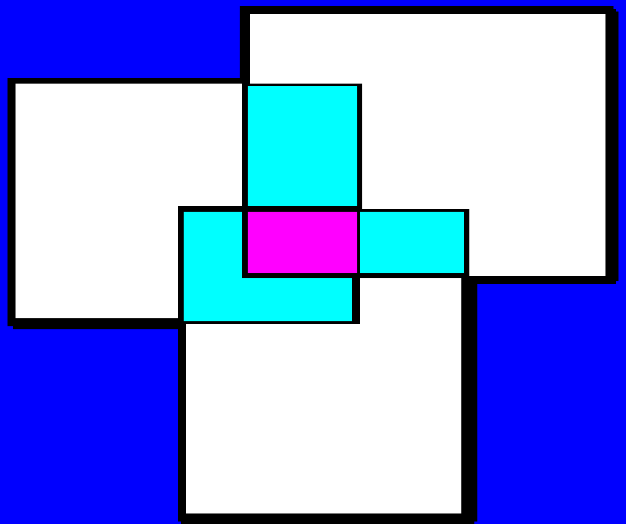
Disjoint decomposition:

$p_1 :=$ $P($  $)$

$p_2 :=$ $P($  $)$

$p_3 :=$ $P($  $)$

$$P(A_1 \cup A_2 \cup A_3) = p_1 + p_2 + p_3$$



Relation between the p_i and S_i:

$$p_1 + 2p_2 + 3p_3 = S_1$$
$$p_2 + 3p_3 = S_2$$
~~$p_3 = S_3$~~

Assume that S_3 is not known.

→ Solve the linear problem

$$\min(\max)\; p_1 + p_2 + p_3$$
$$p_1 + 2p_2 + 3p_3 = S_1$$
$$p_2 + 3p_3 = S_2$$
$$p_1, p_2, p_3 \geq 0; \quad p_1 + p_2 + p_3 \leq 1$$

Explicit bound: $P(A_1 \cup \cdots \cup A_s) \leq S_1 - \frac{2}{s} S_2$

→ lower bound: $\Phi_R(x) \geq 1 - S_1 + \frac{2}{s} S_2$

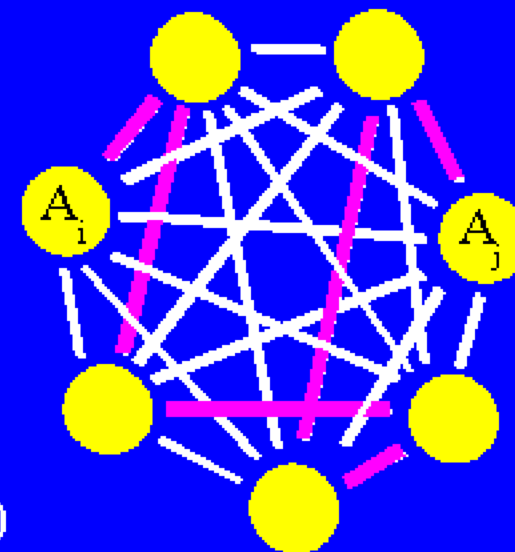
Example: $\Phi_R(x) \geq 1 - 0.8 + 0.5 \cdot 0.18 = 0.29$

Bonferroni: 0.2; true: 0.3645

1: Prékopa (1988)

Graph-theoretical bounds

For a bound of $P(A_1 \cup \cdots \cup A_s)$ consider a graph with nodes A_i and edges (A_i, A_j) weighted by $P(A_i \cap A_j)$



For any spanning tree T denote its weight by

$$|T| := \sum_{(i,j) \in T} P(A_i \cap A_j)$$

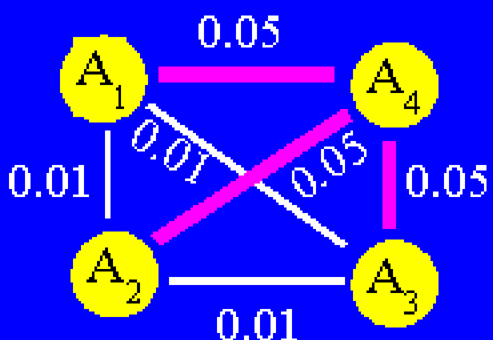
Find a spanning tree T^* with maximum weight (Kruskal)

→ upper bound[1]: $P(A_1 \cup \cdots \cup A_s) \leq S_1 - |T^*|$

corresponds to lower bound: $\Phi_R(x) \geq 1 - S_1 + |T^*|$

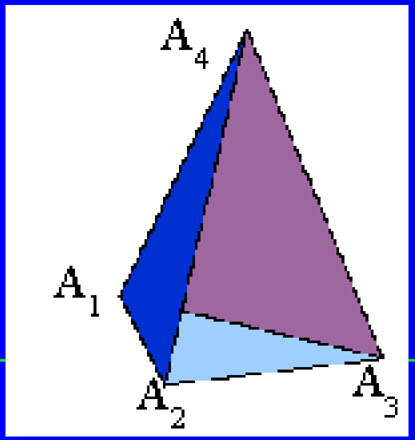
Example:

→ $|T^*| = 0.15, \quad \Phi_R(x) \geq 1 - 0.8 + 0.15 = 0.35$



Bonferroni: 0.2; linear programming bound: 0.29; true = 0.3645; upper = 0.38

Modern improvements:
hypergraphs and hypertrees[2,3]



1: Hunter (1976) 2:Tomescu (1986) 3: Bukszár, Prékopa (2001)

Simulation scheme with variance reduction

Crude Monte Carlo estimator for $\Phi_R(x)$:

Generate sample $\{z^k\}_{k=1,\ldots,n} \subseteq \mathbb{R}^s$ of size n according to distribution $\mathcal{N}(0,R)$.

Then, $\Phi_R(x) \approx n^{-1} \cdot card\{k | z^k \leq x\}$.

Put $l^k := card\{i | z_i^k \leq x_i\}$ $(k=1,\ldots,n)$. Then, the following are estimators of $\Phi_R(x)$[1]

P_1: $n^{-1} \sum_{k=1}^{n} max\{l^k - s + 1, 0\}$

(crude Monte Carlo)

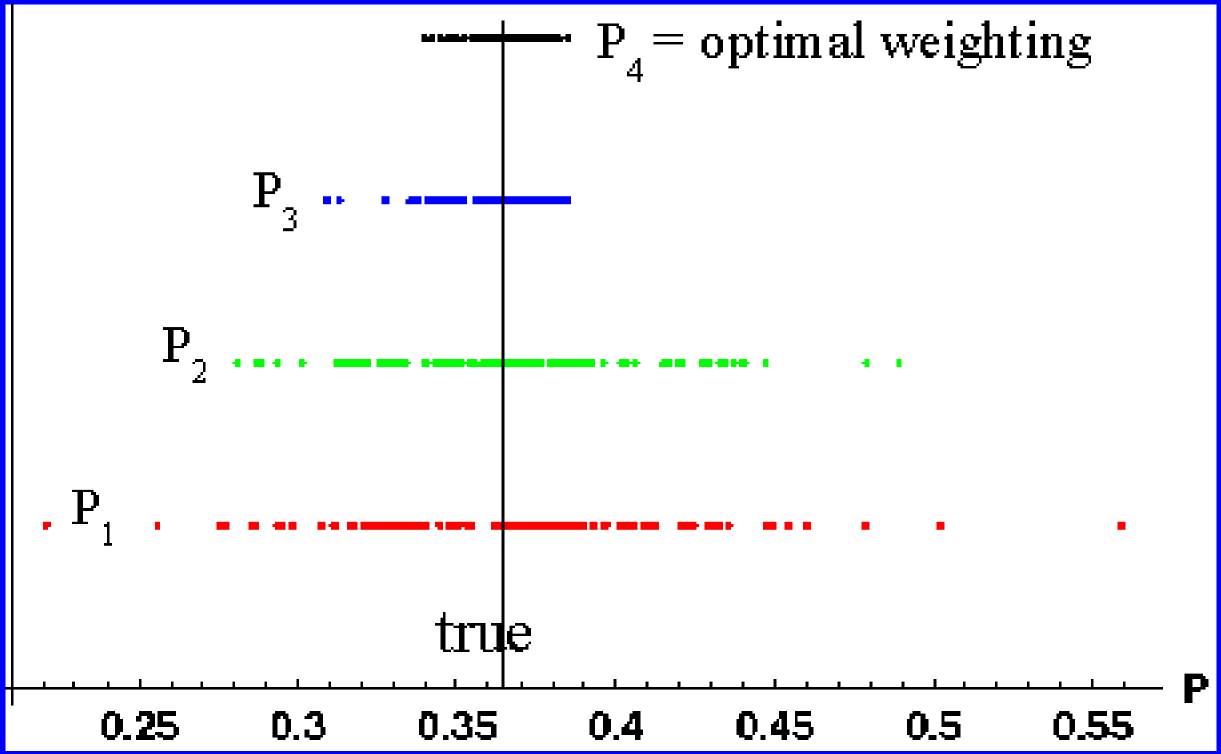
P_2: $1 - S_1 + n^{-1} \sum_{k=1}^{n} max\{s - l^k - 1, 0\}$

(Monte Carlo based on S_1)

P_3: $1 - S_1 + S_2 - n^{-1} \sum_{k=1}^{n} \binom{s - l^k - 2}{2}$

(Monte Carlo based on S_2)

Example: n = 100



Chance constraints in a solution method

Chance constraints in cutting plane methods[1]:

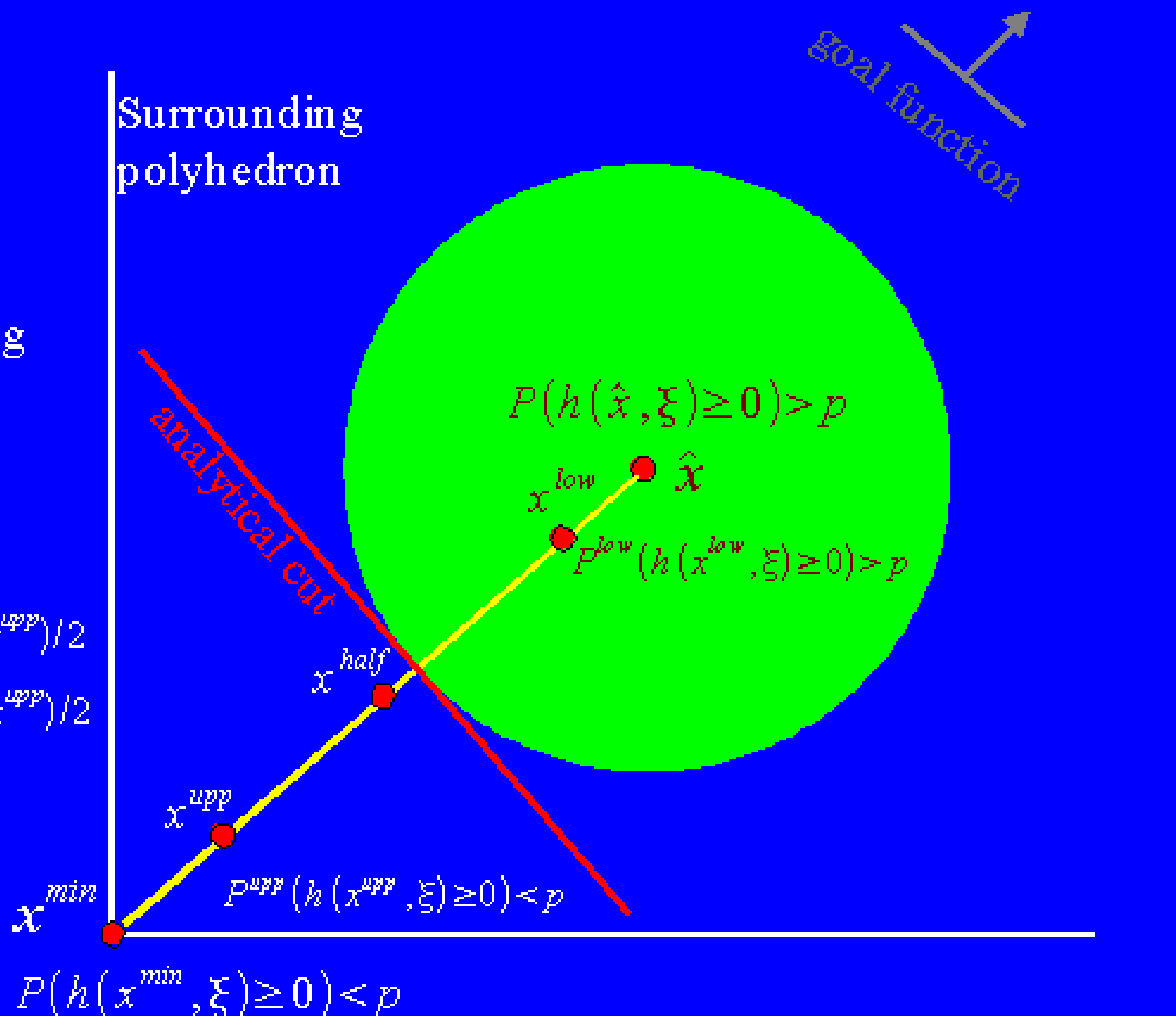
- Determine an approximate cutting point x^{half} by bisection based on probability bounds
- If $P(h(x^{half},\xi)\geq 0)$

$\geq p+\epsilon \longrightarrow x^{low}:=x^{half}\ x^{half}:=(x^{low}+x^{upp})/2$

$\leq p-2\epsilon \longrightarrow x^{upp}:=x^{half}\ x^{half}:=(x^{low}+x^{upp})/2$

$\in(p-2\epsilon,p-\epsilon) \longrightarrow$ Stop

$\in(p-\epsilon,p+\epsilon) \longrightarrow$ Increase accuracy for P



Related Method: Reduced Gradients[2]

1: Szántai (1988) 2: Mayer (1988)