Answers and questions in selected topics of probabilistic programming

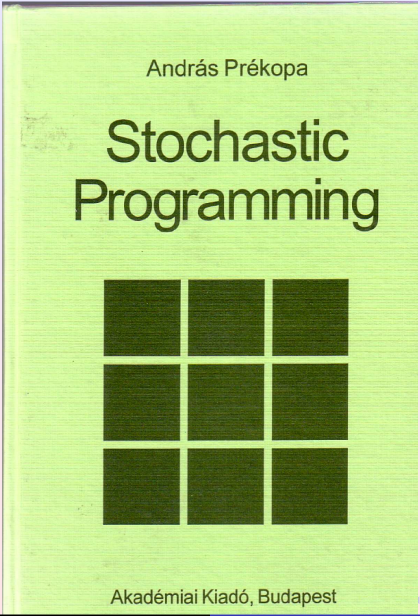
René Henrion, Weierstrass Institute Berlin

International Colloquium on Stochastic Modeling and Optimization
Dedicated to the 80th birthday of Professor András Prékopa

RUTCOR, Rutgers University

November 30, 2009

The Book of Answers



The Big Answer

The Big Answer

Theorem (Prékopa 1971)

If a random vector ξ has a log-concave density, then its law is log-concave.

The Big Answer

Theorem (Prékopa 1971)

If a random vector ξ has a log-concave density, then its law is log-concave.

What makes the beauty of this Theorem?

The Big Answer

Theorem (Prékopa 1971)

If a random vector ξ has a log-concave density, then its law is log-concave.

What makes the beauty of this Theorem?

- The statement is very simple.

The Big Answer

Theorem (Prékopa 1971)

If a random vector ξ has a log-concave density, then its law is log-concave.

What makes the beauty of this Theorem?

- The statement is very simple.
- The result is powerful.

▸ Insertion

The Big Answer

Theorem (Prékopa 1971)

If a random vector ξ has a log-concave density, then its law is log-concave.

What makes the beauty of this Theorem?

- The statement is very simple.
- The result is powerful.
- The consequences are manifold

▸ Insertion

The Big Answer

Theorem (Prékopa 1971)

If a random vector ξ has a log-concave density, then its law is log-concave.

What makes the beauty of this Theorem?

- The statement is very simple.
- The result is powerful.
- The consequences are manifold

Insertion

Key for convexity theory in probabilistic programming. Impact on

The Big Answer

Theorem (Prékopa 1971)

If a random vector ξ has a log-concave density, then its law is log-concave.

What makes the beauty of this Theorem?

- The statement is very simple.
- The result is powerful. ▸ Insertion
- The consequences are manifold

Key for convexity theory in probabilistic programming. Impact on

- Numerics ▸ Insertion

The Big Answer

Theorem (Prékopa 1971)

If a random vector ξ has a log-concave density, then its law is log-concave.

What makes the beauty of this Theorem?

- The statement is very simple.
- The result is powerful. ▸ Insertion
- The consequences are manifold

Key for convexity theory in probabilistic programming. Impact on

- Numerics ▸ Insertion
- Stability
- Structure

Stability

Optimization problem: $\min\{f(x) \mid x \in C,\ \mathbb{P}(\xi \leq Ax) \geq p\}$

Distribution of ξ rarely known $\Longrightarrow$ Approximation by some $\eta \Longrightarrow$ Stability?

Solution set mapping: $\Psi(\eta) := \operatorname{argmin}\{f(x) \mid x \in C,\ \mathbb{P}(\eta \leq Ax) \geq p\}$

Stability

Optimization problem: $\min\{f(x) \mid x \in C, \mathbb{P}(\xi \leq Ax) \geq p\}$

Distribution of ξ rarely known $\Longrightarrow$ Approximation by some $\eta \Longrightarrow$ Stability?

Solution set mapping: $\Psi(\eta) := \operatorname{argmin}\{f(x) \mid x \in C, \mathbb{P}(\eta \leq Ax) \geq p\}$

Theorem (R.H./W.Römisch 2004)

- *f convex, C convex, closed, ξ has log-concave distribution function*
- *$\Psi(\xi)$ nonempty and bounded*
- *$\exists\, x \in C: \quad \mathbb{P}(\xi \leq Ax) > p$ (Slater point)*

Then, Ψ is upper semicontinuous at ξ:

$$\Psi(\eta) \subseteq \Psi(\xi) + \varepsilon\mathbb{B} \quad \text{for} \quad d_K(\mathbb{P}\circ\eta^{-1}, \mathbb{P}\circ\xi^{-1}) < \delta$$

Stability

Optimization problem: $\min\{f(x) \mid x \in C, \mathbb{P}(\xi \le Ax) \ge p\}$

Distribution of ξ rarely known $\Longrightarrow$ Approximation by some $\eta \Longrightarrow$ Stability?

Solution set mapping: $\Psi(\eta) := \operatorname{argmin}\{f(x) \mid x \in C, \mathbb{P}(\eta \le Ax) \ge p\}$

Theorem (R.H./W.Römisch 2004)

- *f convex, C convex, closed, ξ has log-concave distribution function*
- *$\Psi(\xi)$ nonempty and bounded*
- *$\exists x \in C: \quad \mathbb{P}(\xi \le Ax) > p$ (Slater point)*

Then, Ψ is upper semicontinuous at ξ:

$$\Psi(\eta) \subseteq \Psi(\xi) + \varepsilon\mathbb{B} \quad \text{for} \quad d_K(\mathbb{P}\circ\eta^{-1}, \mathbb{P}\circ\xi^{-1}) < \delta$$

If in addition

- *f convex-quadratic, C polyhedron,*
- *ξ has strongly log-concave distribution function,*

then Ψ is locally Hausdorff-Hölder continuous at ξ:

$$d_{\mathrm{Haus}}(\Psi(\eta), \Psi(\xi)) \le \sqrt{d_K(\mathbb{P}\circ\eta^{-1}, \mathbb{P}\circ\xi^{-1})} \quad \text{(locally around } \xi)$$

Strongly log-concave distribution functions

When is a distribution function F_ξ strongly log-concave?

$$\log F_\xi(\lambda x + (1-\lambda)y) \geq \lambda \log F_\xi(x) + (1-\lambda)\log F_\xi(y) + \kappa\lambda(1-\lambda)\|x-y\|^2$$

Strongly log-concave distribution functions

When is a distribution function F_ξ strongly log-concave?

$$\log F_\xi(\lambda x + (1-\lambda)y) \geq \lambda \log F_\xi(x) + (1-\lambda)\log F_\xi(y) + \kappa\lambda(1-\lambda)\|x-y\|^2$$

Proposition

ξ_i independent, F_{ξ_i} strongly log-concave $\Longrightarrow$ F_ξ strongly log-concave.

Example

- The multivariate normal distribution function with independent components is strongly log-concave on bounded convex sets.
- The uniform distribution on multivariate intervals $[a,b]$ is strongly log-concave on int $[a,b]$.

Strongly log-concave distribution functions

When is a distribution function F_ξ strongly log-concave?

$$\log F_\xi(\lambda x + (1-\lambda)y) \geq \lambda \log F_\xi(x) + (1-\lambda)\log F_\xi(y) + \kappa\lambda(1-\lambda)\|x-y\|^2$$

Proposition

ξ_i independent, F_{ξ_i} strongly log-concave $\Longrightarrow$ F_ξ strongly log-concave.

Example

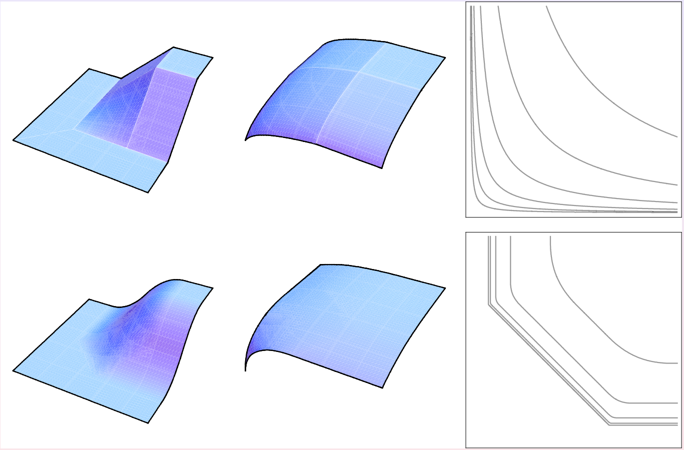
- The multivariate normal distribution function with independent components is strongly log-concave on bounded convex sets.
- The uniform distribution on multivariate intervals $[a, b]$ is strongly log-concave on int $[a, b]$.

Question 1: Density strongly log-concave $\Longrightarrow$ Distribution function strongly log-concave?

Question 1': Multivariate normal distribution function strongly log-concave?

Failure of rotation invariance of strong log-concavity

uniform distribution on unit square and log



45^o rotation of unit square

Lipschitz continuity of distribution functions of quasi-concave probability measures

Definition

A probability measure $\mathbb{P}$ on $\mathbb{R}^n$ is *quasi-concave* if

$$\mathbb{P}(\lambda A + (1-\lambda)B) \geq \min\{\mathbb{P}(A), \mathbb{P}(B)\}$$

for all $\lambda \in [0,1], \;\; A, B \in \mathcal{B}(\mathbb{R}^n)$.

Log-concavity (and α-concavity) implies quasi-concavity.

Theorem (R.H./W.Römisch 2005)

Let ξ have a quasi-concave law $\mathbb{P}$ and denote its distribution function by F_ξ.

$$F_\xi \text{ is Lipschitz} \iff F_\xi \text{ is continuous}$$
$$\iff \operatorname{supp}\mathbb{P} \not\in \text{ canonic hyperplane} \iff \operatorname{Var}\xi_i \neq 0 \;\forall i$$

Lipschitz continuity of distribution functions of quasi-concave probability measures

Definition

A probability measure $\mathbb{P}$ on $\mathbb{R}^n$ is *quasi-concave* if

$$\mathbb{P}(\lambda A + (1-\lambda)B) \geq \min\{\mathbb{P}(A), \mathbb{P}(B)\}$$

for all $\lambda \in [0,1]$, $A, B \in \mathcal{B}(\mathbb{R}^n)$.

Log-concavity (and α-concavity) implies quasi-concavity.

Theorem (R.H./W.Römisch 2005)

Let ξ have a quasi-concave law $\mathbb{P}$ and denote its distribution function by F_ξ.

$$F_\xi \text{ is Lipschitz} \iff F_\xi \text{ is continuous}$$
$$\iff \operatorname{supp} \mathbb{P} \not\subseteq \text{ canonic hyperplane} \iff \operatorname{Var} \xi_i \neq 0 \ \forall i$$

Corollary

If the s-dimensional random vector has a density f_ξ such that $f_\xi^{-1/s}$ is convex, then its distribution function F_ξ is Lipschitz continuous.

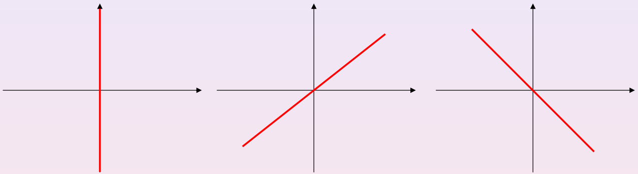
Lipschitz continuity of singular normal distributions

$$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

Lipschitz continuity of singular normal distributions

$$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

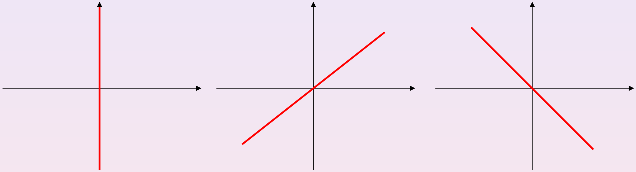
supp P

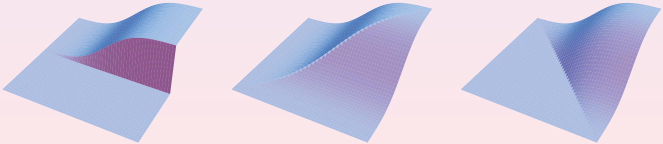


Lipschitz continuity of singular normal distributions

$$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

supp P





Linear probabilistic constraints with Gaussian coefficient matrix: structural properties

Feasible Set: $M_p := \{x \in \mathbb{R}^n \mid \mathbb{P}(\Xi x \leq a) \geq p\}$

Denoting by ξ_i the rows of Ξ, we assume that $\xi_i \sim \mathcal{N}(\mu_i, \Sigma_i) \quad \forall i$.

Theorem (R.H. 2007)

- *M_p is compact if $p > \min_i \{\Phi(\|\mu_i\|_{\Sigma_i^{-1}})\}$*

  (Φ = one-dimensional standard normal distribution function)
- *M_p is empty if $p \geq \min_{a_i < 0} \{\Phi(\|\mu_i\|_{\Sigma_i^{-1}})\}$*
- *If $a \geq 0$ then M_p is a nonempty and star-shaped ($\Longrightarrow$ connected) set.*

Linear probabilistic constraints with Gaussian coefficient matrix: convexity

$$\boxed{M_p := \{x \in \mathbb{R}^n \mid \mathbb{P}(\Xi x \leq a) \geq p\}} \quad \mathrm{vec}\,\Xi \sim \mathcal{N}(\mu, \Sigma) \quad (\xi_i = \text{rows of } \Xi)$$

Theorem (Prékopa 1974)

If all cross covariance matrices $\mathrm{Cov}\,(\xi_i, \xi_j)$ *are proportional, then* M_p *is convex for* $p \geq 0.5$. *The same holds true, if the* ξ_i *refer to the columns of* Σ.

Linear probabilistic constraints with Gaussian coefficient matrix: convexity

$$\boxed{M_p := \{x \in \mathbb{R}^n \mid \mathbb{P}(\Xi x \leq a) \geq p\}} \quad \mathrm{vec}\,\Xi \sim \mathcal{N}(\mu, \Sigma) \quad (\xi_i = \text{rows of } \Xi)$$

Theorem (Prékopa 1974)

If all cross covariance matrices $\mathrm{Cov}\,(\xi_i, \xi_j)$ *are proportional, then* M_p *is convex for* $p \geq 0.5$. *The same holds true, if the* ξ_i *refer to the columns of* Σ.

Theorem (R.H./C. Strugarek 2008)

If all ξ_i *are pairwise independent (not the components of* ξ_i*!), then* M_p *is convex for* $p \geq p^* := \Phi(\max\{\sqrt{3}, \tau\})$ *with*

$$\tau := \max_i \lambda_{max}^{(i)} [\lambda_{min}^{(i)}]^{-3/2} \|\mu_i\|$$

$$\lambda_{max}^{(i)}, \lambda_{min}^{(i)} := \text{largest and smallest eigenvalue of } \Sigma_i.$$

Linear probabilistic constraints with Gaussian coefficient matrix: convexity

$$M_p := \{x \in \mathbb{R}^n \mid \mathbb{P}(\Xi x \leq a) \geq p\} \quad \mathrm{vec}\,\Xi \sim \mathcal{N}(\mu, \Sigma) \quad (\xi_i = \text{rows of } \Xi)$$

Theorem (Prékopa 1974)

If all cross covariance matrices $\mathrm{Cov}\,(\xi_i, \xi_j)$ are proportional, then M_p is convex for $p \geq 0.5$. The same holds true, if the ξ_i refer to the columns of Σ.

Theorem (R.H./C. Strugarek 2008)

If all ξ_i are pairwise independent (not the components of ξ_i!), then M_p is convex for $p \geq p^* := \Phi(\max\{\sqrt{3}, \tau\})$ with

$$\tau := \max_i \lambda^{(i)}_{max} [\lambda^{(i)}_{min}]^{-3/2} \|\mu_i\|$$

$$\lambda^{(i)}_{max}, \lambda^{(i)}_{min} := \text{largest and smallest eigenvalue of } \Sigma_i.$$

Question: Does it hold true that M_p is convex for $p \geq 0.5$ (or: $p \geq \bar{p}$) for any multivariate normal distribution of the elements of Ξ?

Linear probabilistic constraints with a random coefficient row vector

$$M_p^\alpha := \{x \in \mathbb{R}^n \mid \mathbb{P}(\langle \xi, x \rangle \leq \alpha) \geq p\}$$

Theorem (R.H. 2007)

Let ξ have a density. Then, there exists some d such that

$$M_p^\alpha = \{d\} + \overline{\left(M_{1-p}^{-\alpha}\right)^c} \quad \forall \alpha \neq 0 \ \forall p \in (0,1)$$

Result has an impact on properties which are not affected by translation and closure (e.g., convexity, boundedness, nontriviality).

Linear probabilistic constraints with Gaussian coefficient matrix: Gradients

$$\boxed{M_p := \{x \in \mathbb{R}^n \mid \mathbb{P}(\Xi x \leq a) \geq p\}} \quad \mathrm{vec}\,\Xi \sim \mathcal{N}(\mu, \Sigma) \quad (\xi_i = \text{rows of } \Xi)$$

Linear probabilistic constraints with Gaussian coefficient matrix: Gradients

$$\boxed{M_p := \{x \in \mathbb{R}^n \mid \mathbb{P}(\Xi x \leq a) \geq p\}} \quad \mathrm{vec}\,\Xi \sim \mathcal{N}(\mu, \Sigma) \quad (\xi_i = \text{rows of } \Xi)$$

$$\begin{aligned} \mathbb{P}(\Xi x \leq a) \geq p\} &= \Phi^{0,\Sigma(x)}(\alpha(x)) \\ (\Sigma(x))_{i,j} &:= \langle x, \mathrm{Cov}\,(\xi_i, \xi_j) x \rangle \\ \alpha_i(x) &:= a_i - \langle \mu_i, x \rangle \end{aligned}$$

$\Longrightarrow$ for fixed x: value of a multivariate normal distribution function

Linear probabilistic constraints with Gaussian coefficient matrix: Gradients

$$\boxed{M_p := \{x \in \mathbb{R}^n \mid \mathbb{P}(\Xi x \le a) \ge p\}} \quad \mathrm{vec}\,\Xi \sim \mathcal{N}(\mu, \Sigma) \quad (\xi_i = \text{rows of } \Xi)$$

$$\begin{aligned} \mathbb{P}(\Xi x \le a) \ge p\} &= \Phi^{0,\Sigma(x)}(\alpha(x)) \\ (\Sigma(x))_{i,j} &:= \langle x, \mathrm{Cov}\,(\xi_i, \xi_j) x\rangle \\ \alpha_i(x) &:= a_i - \langle \mu_i, x\rangle \end{aligned}$$

$\Longrightarrow$ for fixed x: value of a multivariate normal distribution function Question:

Gradients?

Sensitivities of normal distribution functions w.r.t. correlations?

Boldog éveket,

jó egészséget

és még sok uj tételt

kivánok önnek!

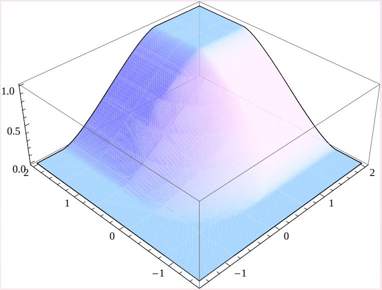
Uniform Distribution on the Disk

Formula for the distribution function:

$$F_\xi(x,y) = \begin{cases} \varphi(x,y) & \text{if } x \le 0,\ y \le 0 \\ 2\varphi(0,y) - \varphi(-x,y) & \text{if } y \le 0,\ x \ge 0 \\ 2\varphi(x,0) - \varphi(x,-y) & \text{if } y \ge 0,\ x \le 0 \\ 1 - 2\varphi(0,-y) - 2\varphi(-x,0) + \varphi(-x,-y) & \text{if } y \ge 0,\ x \ge 0 \end{cases},$$

where

$$\varphi(x,y) := \begin{cases} \frac{1}{2\pi}\left(x\sqrt{1-x^2} + y\sqrt{1-y^2} + 2yx + \arcsin x - \arcsin\left(-\sqrt{1-y^2}\right)\right) & \text{if } x^2+y^2 \le 1 \\ 0 & \text{else} \end{cases}$$



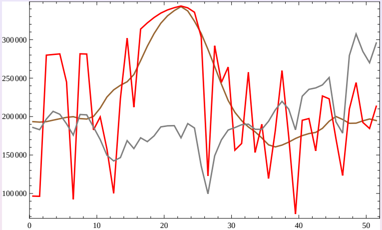
Check log-concavity!

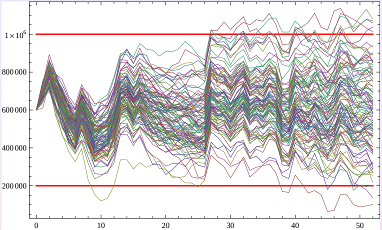
Density is a constant $\Longrightarrow$ log-concave.

back

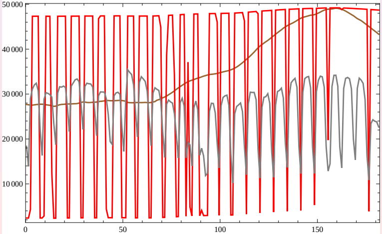
Probabilistic constraints in hydro power management

Numerical experiments for Electricité de France (A. Möller, WIAS Berlin)
Profit maximization in hydro power production, reservoir with random inflow and prices, probabilistic filling level constraints.
Optimal release policy with weekly decisions over one year (n=52):





Optimal release policy with daily decisions over half a year (n=182):



back