Uncertainty and Variability Modeling via Data-Driven Chance Constraints

Bruno A. Calfa, Ignacio E. Grossmann
Department of Chemical Engineering
Carnegie Mellon University
Pittsburgh, PA 15213

Anshul Agarwal, John M. Wassick, Scott J. Bury
The Dow Chemical Company
Midland, MI

September 17th – 18th, 2013

Modeling Uncertainty in Optimization

- Optimization models for real-world applications are expected to generate "robust" decisions in the face of uncertainty.
- Models use historical and predicted data subject to uncertainty.
- Examples of data variability and uncertainty:
  - Product demand and selling price;
  - Raw material supply;
  - Production rates etc.
- Some approaches of optimization under uncertainty:
  - Stochastic Programming (Birge & Louveaux, 2011);
  - Robust Optimization (Ben-Tal, Ghaoui, & Nemirovski, 2009);
  - Approximate Dynamic Programming (Powell, 2011) etc.
- **This work:** Chance-Constrained Optimization (Charnes & Cooper, 1959).

Chance-Constrained Optimization

- Chance constrains (CCs) are also known as probabilistic constraints.
- Optimization model with joint chance constraint (JCC)

$$\max_x \ f(x)$$
$$\text{s.t.} \quad x \in X$$
$$\mathbb{P}\left\{g_j(x, \tilde{\xi}_j) \geq 0,\ j = 1,\ \ldots,\ m\right\} \geq 1 - \alpha$$

- Optimization model with individual (or disjoint) chance constraint (ICC)

$$\max_x \ f(x)$$
$$\text{s.t.} \quad x \in X$$
$$\mathbb{P}\left\{g_j(x, \tilde{\xi}_j) \geq 0\right\} \geq 1 - \alpha_j \qquad j = 1,\ \ldots,\ m$$

where $X \subset \mathbb{R}^{n_x}$ determines the feasible region (e.g., polyhedron)

- **This work:** right-hand side (RHS) uncertainty only.

Classical Chance Constraints

- CCs with RHS uncertainty are very common in engineering applications
  - Chemical Engineering: maximum production rates, demand satisfaction, setpoint tracking in dynamic optimization
  - Electrical Engineering: power supply for demand satisfaction
  - Hydrology: reservoir storage levels (water availability)
- The general chance constraints

$$\mathbb{P}\left\{g_j(x) \geq \tilde{\xi}_j\right\} \geq 1 - \alpha_j, \qquad j = 1, \ldots, m \quad \text{(ICC)}$$

$$\mathbb{P}\left\{g_j(x) \geq \tilde{\xi}_j,\ j = 1, \ldots, m\right\} \geq 1 - \alpha \quad \text{(JCC)}$$

can be reformulated as follows

$$g_j(x) \geq F_{\xi_j}^{-1}(1 - \alpha_j), \qquad \text{for } j = 1, \ldots, m \quad \text{(ICC)}$$

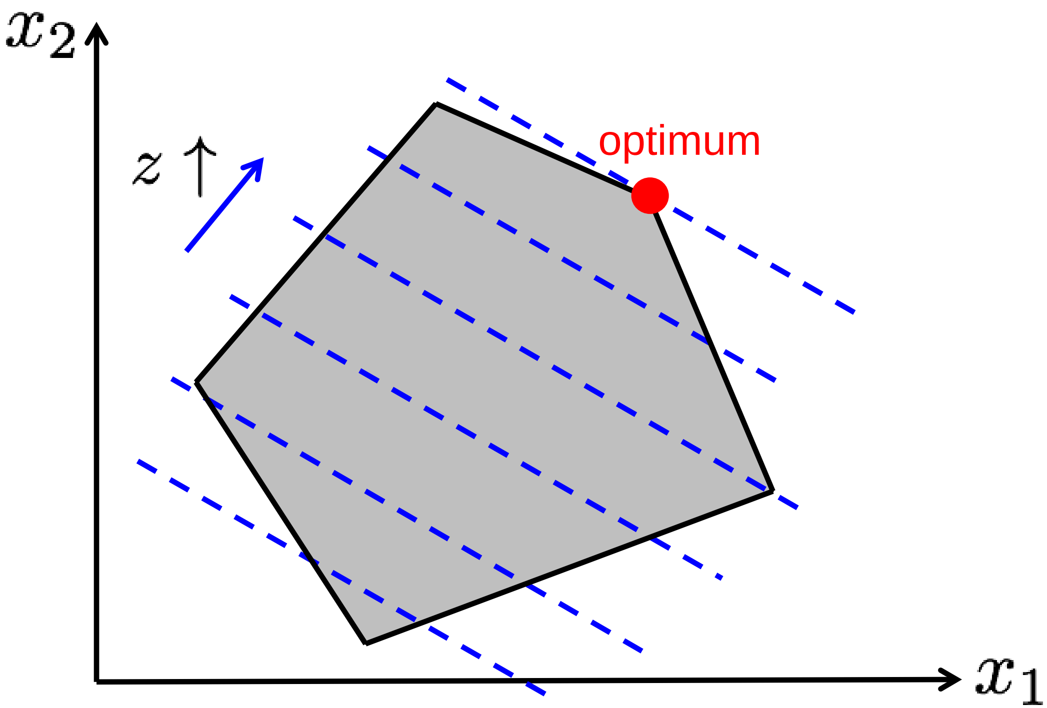
$$F_{\tilde{\xi}_{j=1}^{m}}(g_j(x), \ldots, g_m(x)) \geq 1 - \alpha \quad \text{(JCC)}$$

where $F_{\xi_j}^{-1}(\cdot)$ is the inverse (univariate) CDF (or quantile function) and $F_{\tilde{\xi}_{j=1}^{m}}(\cdot, \ldots, \cdot)$ is the (multivariate) joint CDF.

Illustration: Deterministic Model

- For simplicity, consider an LP model in two dimensions

$$\begin{aligned} \max_{x_1, x_2} \quad & z = c_1 x_1 + c_2 x_2 \\ \text{s.t.} \quad & A_{i,1} x_1 + A_{i,2} x_2 \leq b_i \\ & \qquad \forall\, i = 1, \ldots, 5 \\ & x_1,\ x_2 \geq 0 \end{aligned}$$



Individual Chance-Constrained Model

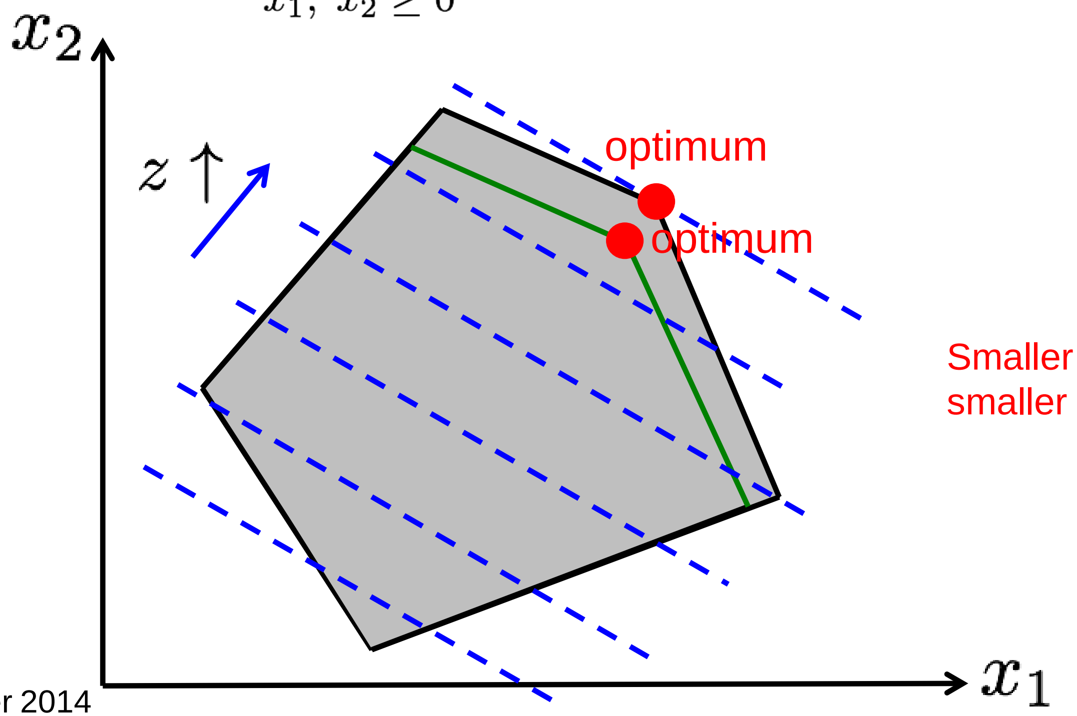
- Individual chance constraints (model remains an LP)

$$\max_{x_1, x_2} \; z = c_1 x_1 + c_2 x_2$$

$$\text{s.t.} \quad A_{i,1} x_1 + A_{i,2} x_2 \leq b_i \quad \forall\, i = 1,\ 2,\ 3$$

$$\mathbb{P}\left\{ A_{i,1} x_1 + A_{i,2} x_2 \leq \tilde{b}_i \right\} \geq 1 - \alpha_i \quad \forall\, i = 4,\ 5$$

$$x_1,\ x_2 \geq 0$$



Smaller feasible region for smaller risk level (α)

Individual Chance-Constrained Model

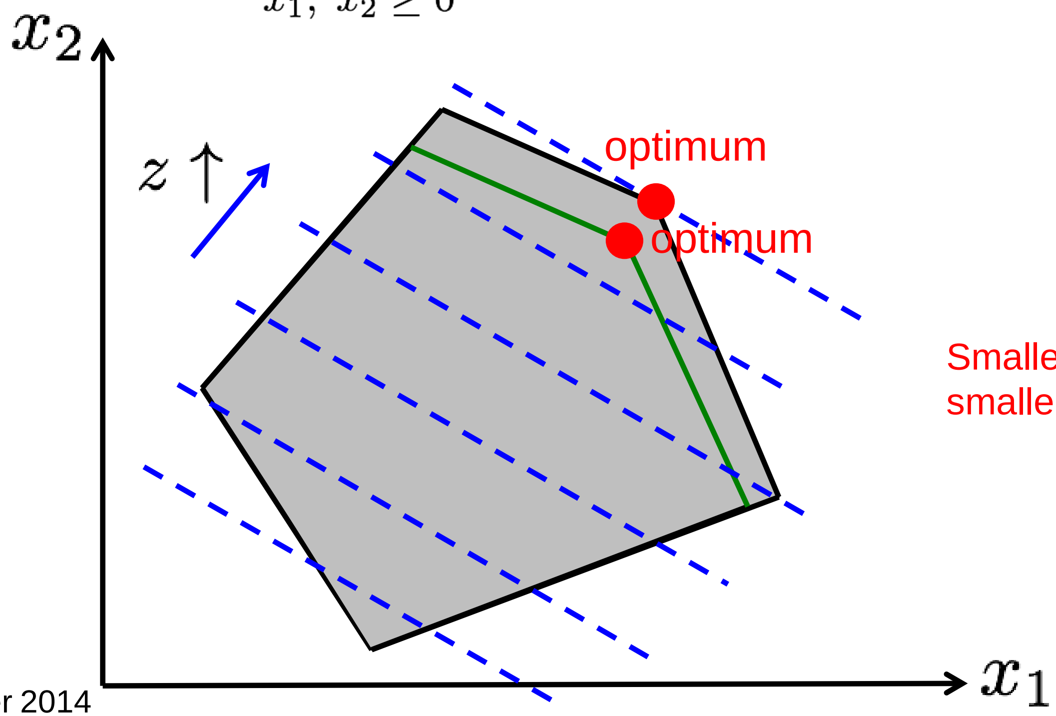
- Individual chance constraints (model remains an LP)

$$\max_{x_1,x_2} \; z = c_1 x_1 + c_2 x_2$$

$$\text{s.t.} \quad A_{i,1}x_1 + A_{i,2}x_2 \le b_i \quad \forall\, i = 1,\ 2,\ 3$$

$$A_{i,1}x_1 + A_{i,2}x_2 \le F_{\tilde{b}_i}^{-1}(\alpha_i) \quad \forall\, i = 4,\ 5$$

$$x_1,\ x_2 \ge 0$$



Smaller feasible region for smaller risk level (α)

Joint Chance-Constrained Model

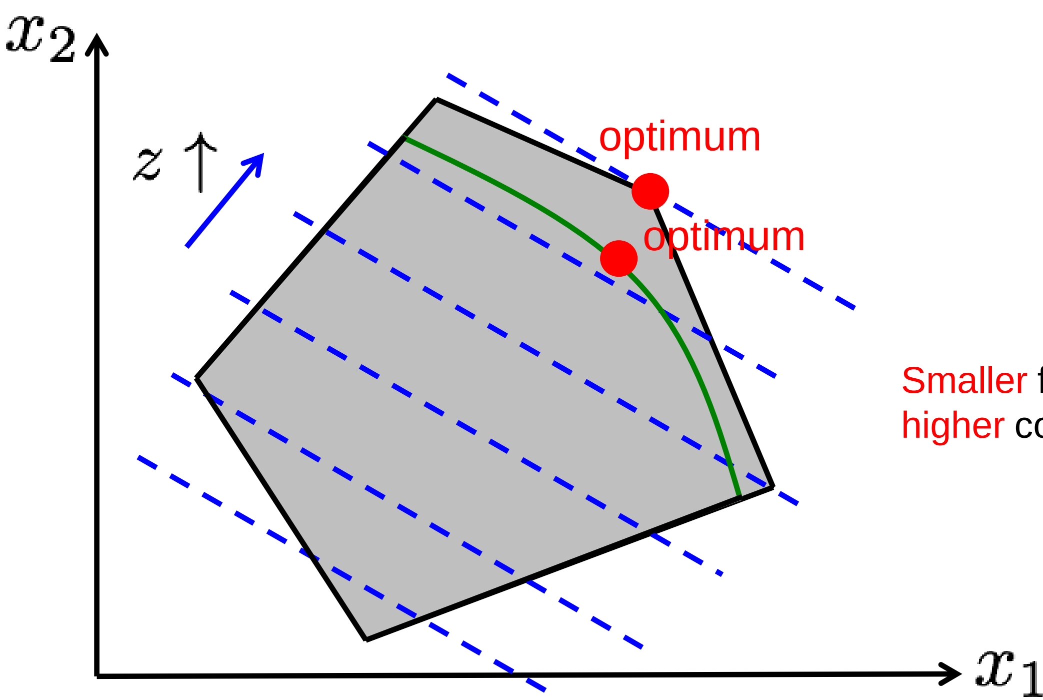
- Joint chance constraints (model is an NLP)

$$\max_{x_1,x_2} \; z = c_1x_1 + c_2x_2$$

$$\text{s.t.} \quad A_{i,1}x_1 + A_{i,2}x_2 \le b_i \qquad \forall\, i = 1,\ 2,\ 3$$

$$\mathbb{P}\left\{A_{i,1}x_1 + A_{i,2}x_2 \le \tilde{b}_i,\ \ i = 4, 5\right\} \ge 1 - \alpha$$

$$x_1,\ x_2 \ge 0$$



Smaller feasible region for higher confidence level $(1 - \alpha)$

Joint Chance-Constrained Model

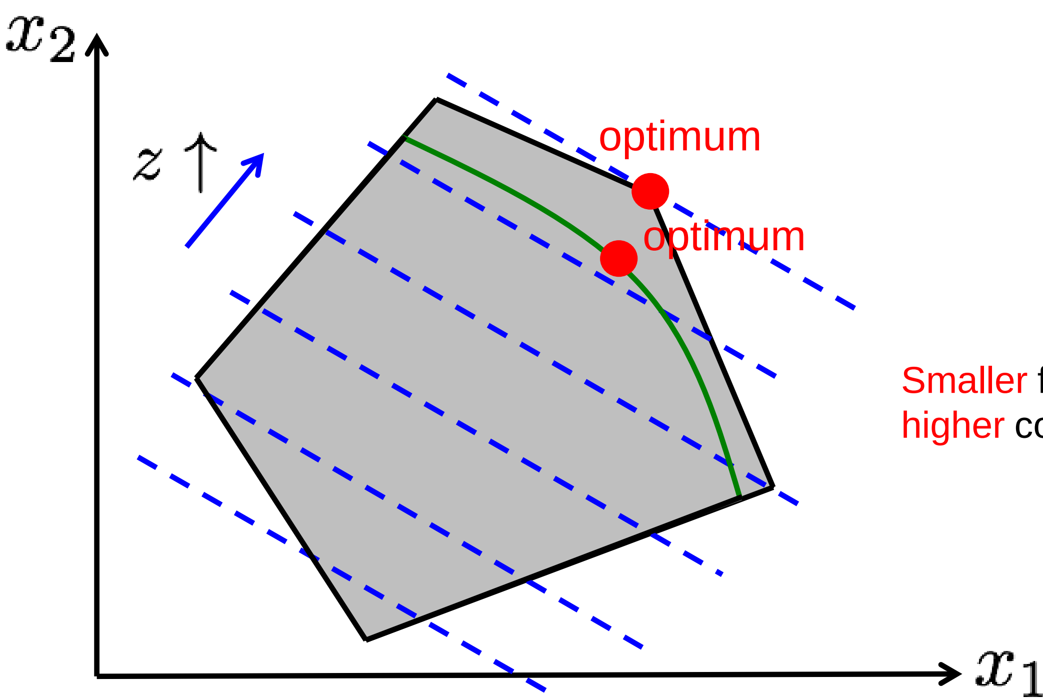
- Joint chance constraints (model is an NLP)

$$\max_{x_1, x_2} \; z = c_1 x_1 + c_2 x_2$$

$$\text{s.t.} \quad A_{i,1} x_1 + A_{i,2} x_2 \leq b_i \qquad \forall\, i = 1,\ 2,\ 3$$

$$\int_{A_{4,1}x_1 + A_{4,2}x_2}^{\infty} \int_{A_{5,1}x_1 + A_{5,2}x_2}^{\infty} f_{\tilde{b}_4, \tilde{b}_5}(b_4, b_5)\,\mathrm{d}b_4 \mathrm{d}b_5 \geq 1 - \alpha$$

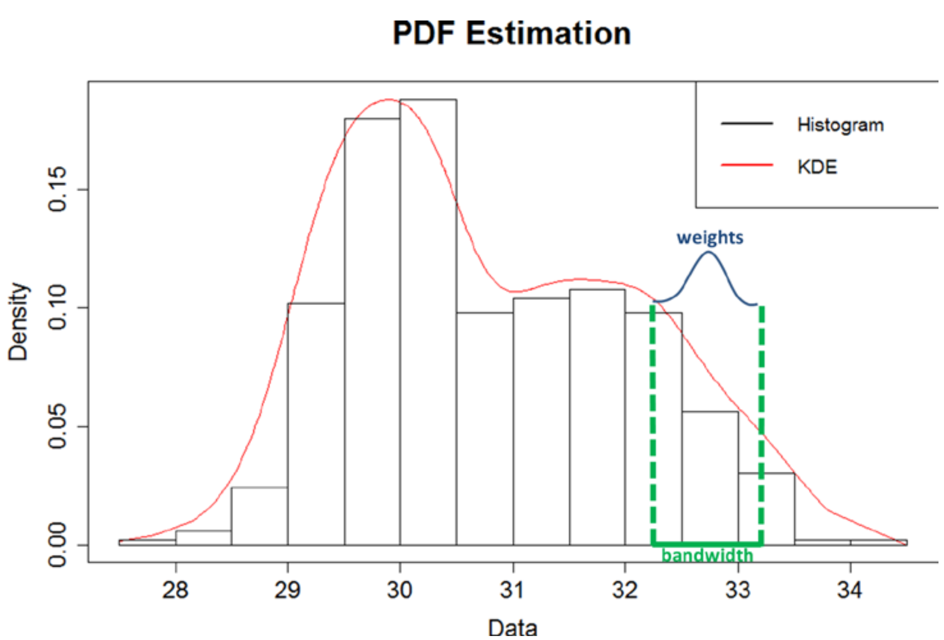
$$x_1,\ x_2 \geq 0$$



Smaller feasible region for higher confidence level (1 − α)

Data-Driven Chance-Constrained Optimization

- Very often in reality, distribution function is unknown or ambiguous.
- Nonparametric techniques: weaker model assumptions.
- Main result (Jiang and Guan, 2013)
  - Replace assumed "true" distribution $\mathbb{P}$ with its estimate $\hat{\mathbb{P}}$.
  - Risk level α is decreased to α'.



Kernel Density Estimation (KDE)

- Given i.i.d. data
- Estimate the PDF/CDF with a smooth curve
- Can estimate quantiles
- Decisions
  - Kernel function (typically Gaussian)
  - Bandwidth (optimization and cross-validation)
- Rigorous routines in R np package (Hayfield and Racine, 2008)

KDE Formulas

- Mathematical expressions for the univariate case (1-D):

PDF: $$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

CDF: $$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_h(x - x_i) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}\left(\frac{x - x_i}{h}\right)$$

where $\mathcal{K}(\cdot)$ is the integrated kernel

- For multivariate, use *product kernels* (standard technique):

PDF: $$\hat{f}(x_1, \ldots, x_m) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{m} K_{h,j}(x_j - x_{i,j}) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{m} \frac{1}{h_j} K_j\left(\frac{x_j - x_{i,j}}{h_j}\right)$$

CDF: $$\hat{F}(x_1, \ldots, x_m) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{m} \mathcal{K}_{h,j}(x_j - x_{i,j}) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{m} \mathcal{K}_j\left(\frac{x_j - x_{i,j}}{h_j}\right)$$

- If using Gaussian kernel, then

$$K(u) = \phi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \qquad \mathcal{K}(u) = \Phi(u) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{u}{\sqrt{2}}\right)\right]$$

CarnegieMellon ENGINEERING

CAPD CENTER

Kernel-Based Reformulation

- Original reformulated CCs:

$$g_j(x) \geq F_{\xi_j}^{-1}(1-\alpha_j), \qquad j = 1, \ldots, m \qquad \text{(ICC)}$$

$$F_{\tilde{\xi}_{j=1}^m}(g_1(x), \ldots, g_m(x)) \geq 1-\alpha \qquad \text{(JCC)}$$

- Kernel-based reformulation:

$$g_j(x) \geq \hat{F}_{\xi_j}^{-1}(1-\alpha'_{j,+}), \qquad j = 1, \ldots, m \qquad \text{(ICC)}$$

$$\hat{F}_{\tilde{\xi}_{j=1}^m}(g_1(x), \ldots, g_m(x)) \geq 1-\alpha'_+ \qquad \text{(JCC)}$$

or

$$\frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{m}\mathcal{K}_j\left(\frac{g_j(x)-x_{i,j}}{h_j}\right) \geq 1-\alpha'_+ \qquad \text{(JCC)}$$

where α' is the decreased risk level due to the estimation process.

- **This work's contribution:** use of point-wise standard errors of the KDE process to compute α' using ϕ-divergences.

Calculating ϕ-Divergence Tolerance

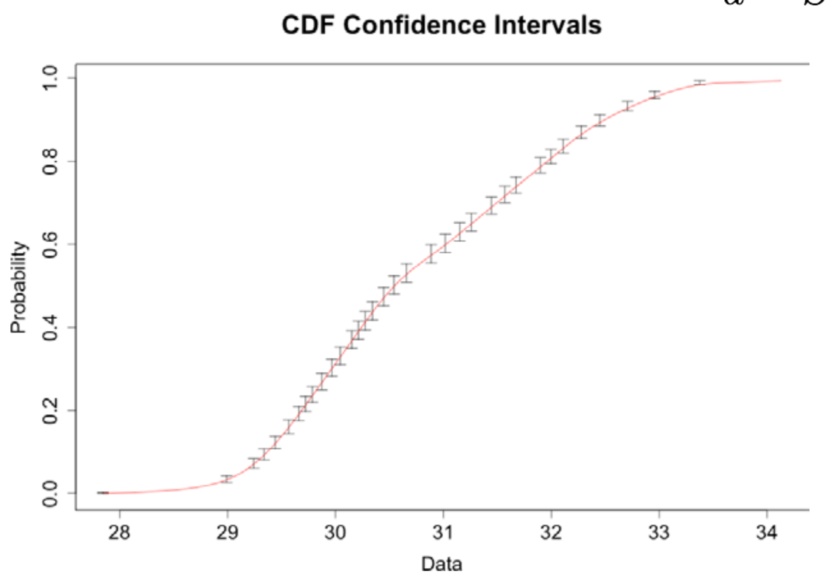
- Density-based confidence set and ϕ-divergence (distance between distributions):

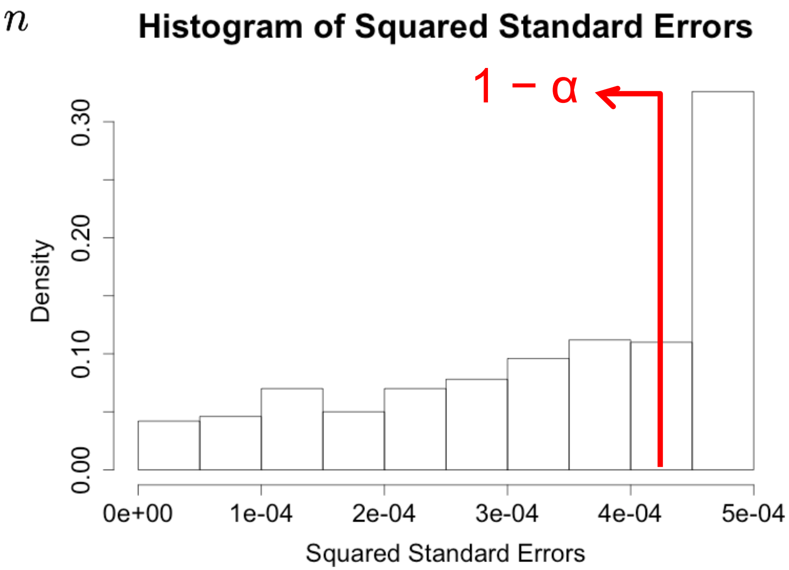
$$\mathcal{D} = \left\{ \mathbb{P} \in \mathcal{M}_+ : \; D_\phi(f||\hat{f}) \leq d \;\; f = \mathrm{d}\mathbb{P}/\mathrm{d}\tilde{\xi} \right\} \qquad D_\phi(f||\hat{f}) = \int_{\mathbb{R}^m} \phi\left(\frac{f(\tilde{\xi})}{\hat{f}(\tilde{\xi})}\right) \hat{f}(\tilde{\xi})\mathrm{d}\tilde{\xi}$$

e.g., Kullback-Leibler (K-L) divergence, $\phi(t) = t\log t - t + 1$.

- Proposed approach for calculating the divergence tolerance *d*
  - Square of point-wise standard errors from the quantile/distribution estimation via KDE. Bootstrapping or formulas based on the asymptotic normality of kernel estimators.

$$d = SE^2_{1-\alpha} \propto \frac{1}{n}$$

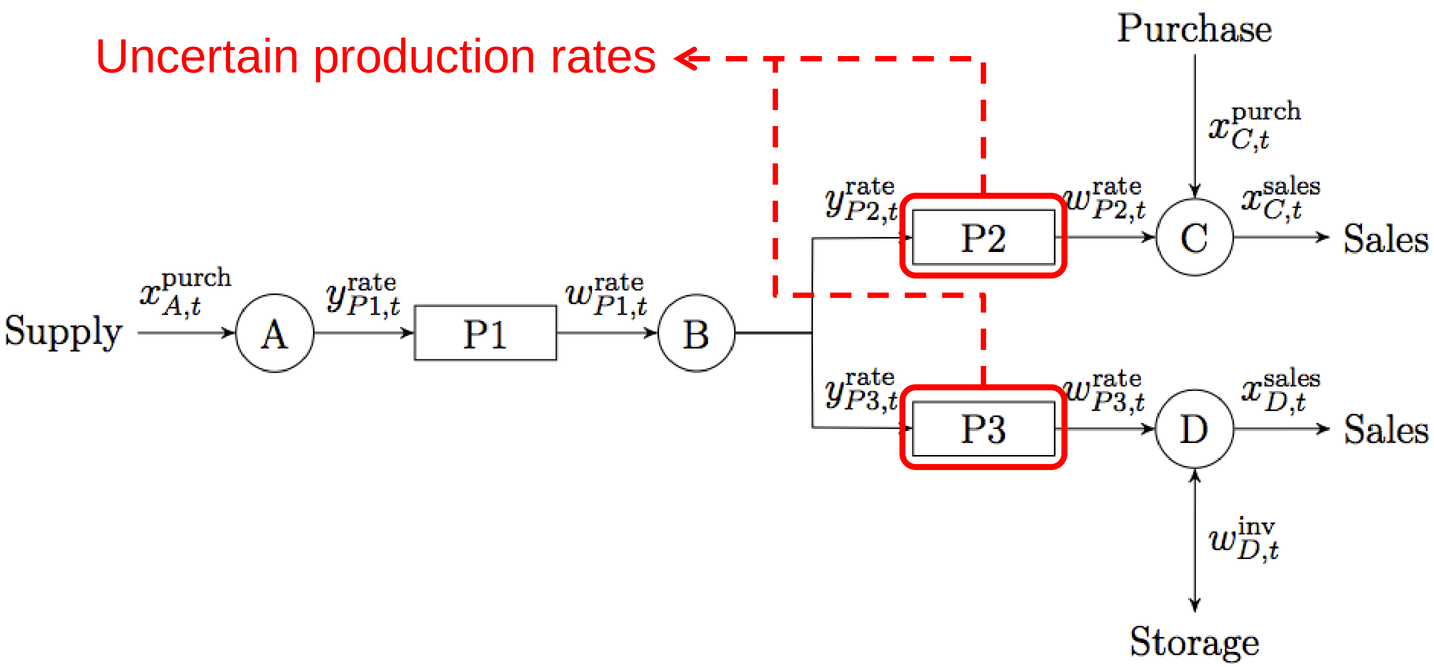




(Ben-Tal *et al.*, 2011; Jiang and Guan, 2013)

Motivating Example

- Network of chemical plants



- 1 raw material (A), 1 intermediate product (B), two finished products (C and D), 1 site
- Only D can be stored and C can be purchased from elsewhere (may simulate inter-site transfers)

Modeling Chance Constraints

- Uncertain production rates of plants *P2* and *P3*.
- ICC approach yields and LP (PPICC):

$$\mathbb{P}\left\{w^{\text{rate}}_{P2,t} \leq \tilde{\mathrm{w}}^{\text{rate,max}}_{P2}\right\} \geq 1-\alpha_{P2} \qquad \forall\, t \in T$$
$$\mathbb{P}\left\{w^{\text{rate}}_{P3,t} \leq \tilde{\mathrm{w}}^{\text{rate,max}}_{P3}\right\} \geq 1-\alpha_{P3} \qquad \forall\, t \in T$$

$$w^{\text{rate}}_{P2,t} \leq \hat{F}^{-1}_{\tilde{\mathrm{w}}^{\text{rate,max}}_{P2}}(\alpha'_{P2,+}) \qquad \forall\, t \in T$$
$$w^{\text{rate}}_{P3,t} \leq \hat{F}^{-1}_{\tilde{\mathrm{w}}^{\text{rate,max}}_{P3}}(\alpha'_{P3,+}) \qquad \forall\, t \in T$$

- JCC approach yields an NLP (PPJCC):

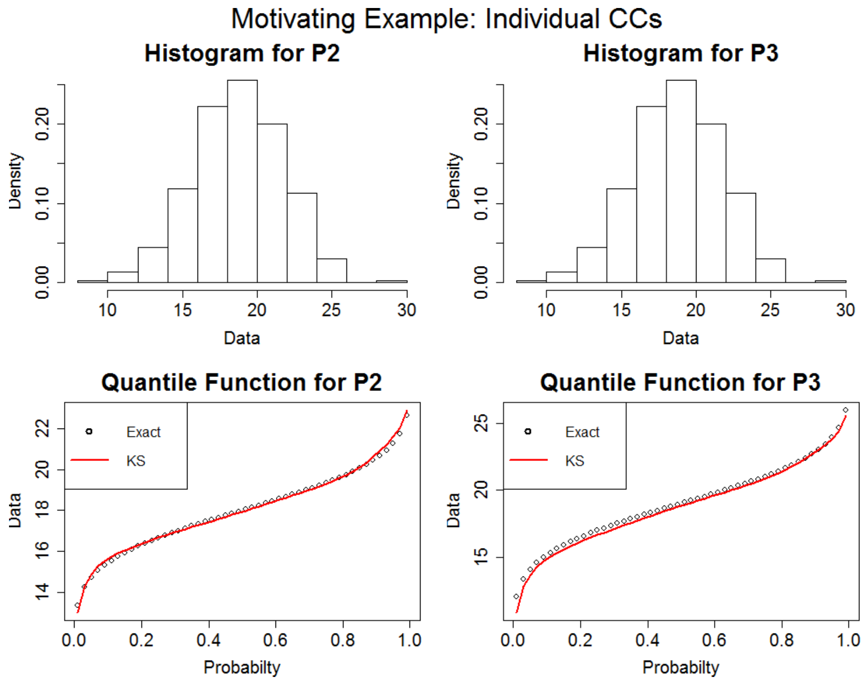
$$\mathbb{P}\left\{\begin{matrix} w^{\text{rate}}_{P2,t} \leq \tilde{\mathrm{w}}^{\text{rate,max}}_{P2} \\ w^{\text{rate}}_{P3,t} \leq \tilde{\mathrm{w}}^{\text{rate,max}}_{P3} \end{matrix}\right\} \geq 1-\alpha \qquad \forall\, t \in T$$

$$\frac{1}{n}\sum_{i=1}^{n}\left[1-\mathcal{K}_{P2}\left(\frac{w^{\text{rate}}_{P2,t}-\mathrm{w}^{\text{rate,max}}_{P2,i}}{h_{P2}}\right)\right]\left[1-\mathcal{K}_{P3}\left(\frac{w^{\text{rate}}_{P3,t}-\mathrm{w}^{\text{rate,max}}_{P3,i}}{h_{P3}}\right)\right] \geq 1-\alpha'_{+} \qquad \forall\, t \in T$$

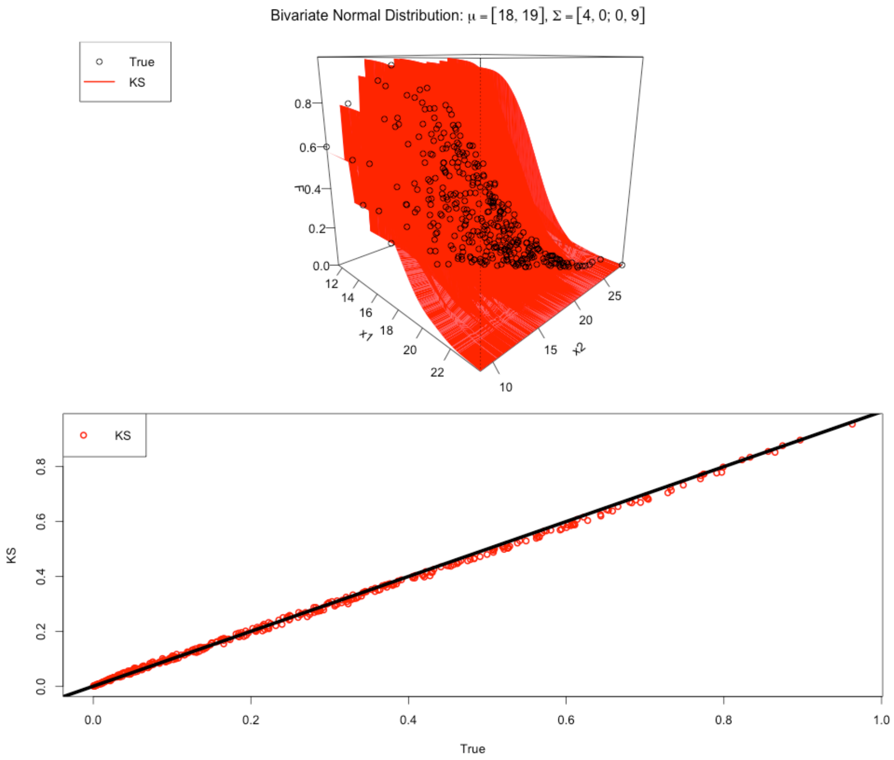
- Artificially generated historical data
  - Normal distributions with μ = [18; 19] and Σ = [4, 0; 0, 9].
- Compared KDE-based reformulation solutions with Exact reformulation
- Modeled in AIMMS 3.13.5 with GUROBI 5.1 (LP) and IPOPT 3.10.1 (NLP).

Artificial Historical Data

Data and Quantile Estimation for (PPICC)

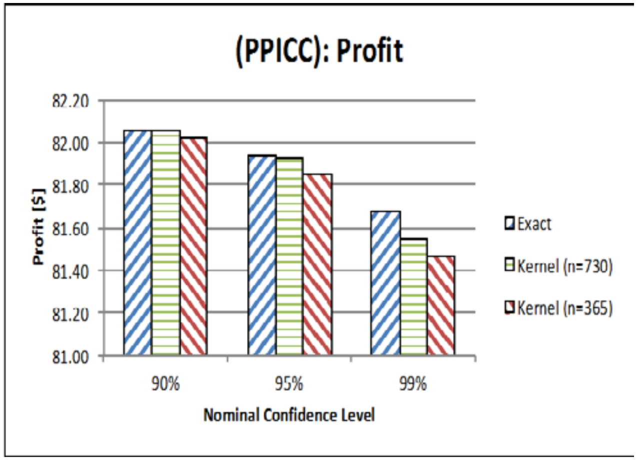


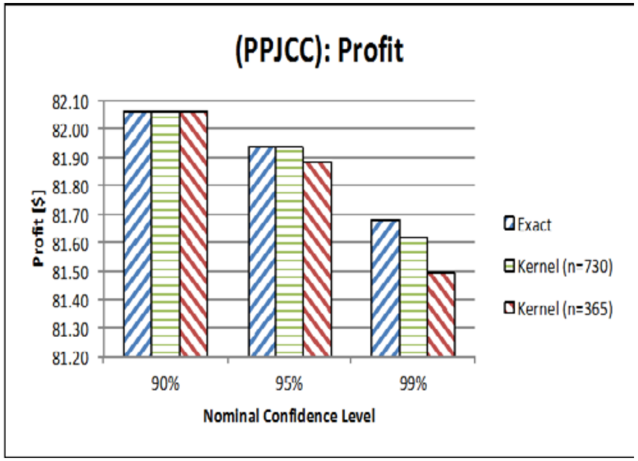
Data and Joint CDF Estimation for (PPJCC)



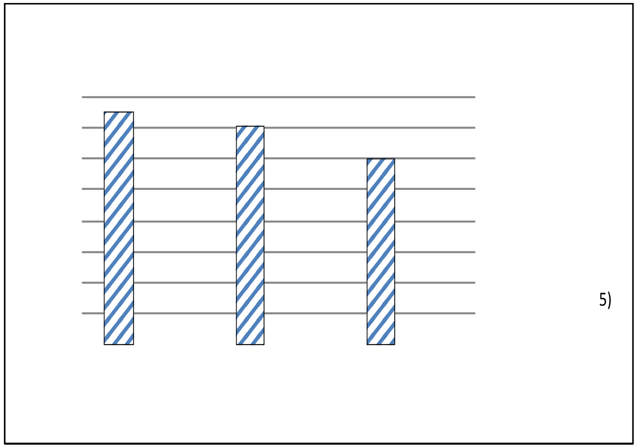
Results

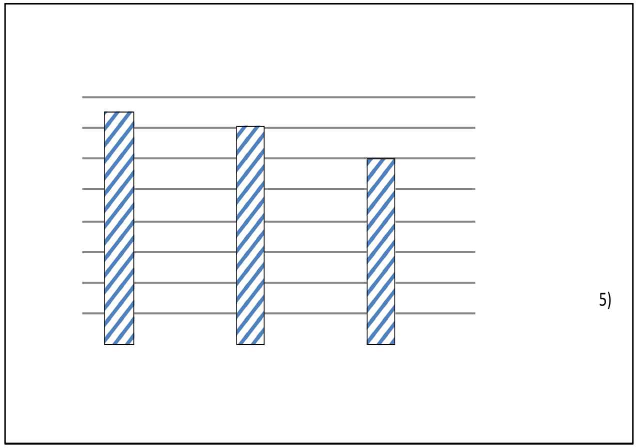
- Profits





- Flowrates out of *P3*





Conclusions

- Kernel smoothing: data-driven statistical method for reformulating chance constraints with RHS uncertainty.
- Use of density-based confidence set with ϕ-divergence. Specification of the divergence tolerance *d* from distribution of squared point-wise standard errors estimated from the KDE process.
- Sample size influences the accuracy of the estimation, and ultimately, the quality of solution.
- ICC problem is generally easier to solve. Can use its solution to initialize the JCC problem.

Acknowledgments

Collaboration with The Dow Chemical Company, and kind technical support on the R np package and kernel smoothing by Dr. Jeffrey S. Racine (Department of Economics at McMaster University).

References

- Ben-Tal, A., Ghaoui, L. E., & Nemirovski, A. (2009). *Robust Optimization*. Princeton Series in Applied Mathematics. New Jersey, USA.
- Birge, J. R., & Louveaux, F. (2011). *Introduction to Stochastic Programming*. Springer, second edition.
- Charnes, A., & Cooper, W. W. (1959). Chance-Constrained Programming. *Management Science.* 6(1):73–79.
- Hayfield, T., & Racine, J. S. (2008). Nonparametric Econometrics: The np Package. *Journal of Statistical Software* 27(5).
- Jiang, R., and Guan, Y. 2013. *Data-driven Chance Constrained Stochastic Program*. **Optimization Online**. Retrieved from http://www.optimization-online.org/DB_FILE/2013/09/4044.pdf on Feb 24, 2014.
- Powell, W. B. (2011). *Approximate Dynamic Programming: Solving the Curses of Dimension- ality*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., second edition. Hoboken, NJ. USA.
- R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.