Importance Sampling for Chance Constrained Optimization*

Vyacheslav Gorchakov †

Abstract. Chance constrained problem is stated as minimizing a objective over solutions satisfying, with a given close to one probability, a system of convex constraints. These problems appears in many flavours of engineering. Probabilistic constraints, which appears at chance constrained problems, are not convex in general case. One of the main approach of solving these problems is to approximate the feasible set. Several state-of-the-art convex approximations are known: Markov, Chebyshev, Bernstein and sample average approximation (SAA). Moreover, we have the scenario approximation, which consists in sampling random points according to the nominal noise distribution and solving a deterministic problem afterwards. It is known to be one of the most accurate but time demanding approximation. Our goal is to propose a new way to reduce complexity of scenario approximation method for a linear objective and linear chance constraints with Gaussian noise by utilizing importance sampling.

Key words. chance constrained optimization, importance sampling, scenario approximation

AMS subject classifications. 68Q25, 68R10, 68U05

1. Introduction.

Chance constrained optimization appears in computer vision [7], robotics [11], gas and power systems [15], self-driving cars [14], district heating [16] and transportation networks[3]. The origins of these methods date back to the decision theory in 1950 [1] and, more recently, to robust optimization methods, where we are dealing with probabilistic and non-probabilistic models of robustness [4]. Since a robust solution is usually conservative, it is desired to have a compromise between the robustness and the risk of failure in many practical problems [2]. The robust optimization defines secure as ensuring feasibility for all realizations within a predefined uncertainty set, while chance-constrained optimization seeks to satisfy the constraints with a high probability [6]. Being significantly less conservative then the robust optimization methods, chance-constrained algorithms are widely used in engineering practice.

Being computationally intractable [9], the chance-constrained optimization admits efficient solutions only in a few cases. The case of Gaussian noise at constraints is the simplest approach. To this end, various approximate methods are used in practice. The convexity of constraints simplifies chance constrained problems and there are some state-of-the art approaches as Markov, Chebyshev, Bernstein approximations [12] and sample average approximation for mixed-integer stochastic problems [5]. These approximations find a convex set to approximate the feasible set.

Sample average approximation is a valuable alternative that theoretically can even lead to an optimal solution [10]. Unfortunately, this method requires a number of samples and often expensive for engineering practice. [13] In particular, if a standard deviation of the distribution is much less then a distance to the feasibility boundary the sample average approximation is inefficient as it requires significant efforts to generate even one infeasible point. Importance sampling allows to avoid this problem, by adjusting the distribution to sample from.

---

*Submitted to the editors DATE.

Funding: This work was funded by the Fog Research Institute under contract no. FRI-454.

† (ddoe@imag.com, http://www.imag.com/~ddoe/).

We investigate importance sampling approach for chance constrained optimization and evaluate its efficiency over simulated and real test cases coming from reliability analysis of power systems.

The paper is organized as follows. Our problem setup is in section 2, experiment design is in section 3, main results are in section 4, experimental results are in **??**, and the conclusions follow in section 8.

2. Background and Problem Setup.

Let $\{x_i\}_{i=1}^m, x_i \in \mathbb{R}^n$ be a sequence of points. Let $f(x) = c^\top x$ be a linear function. Consider linear probability constraints:

$$\text{Prob}(g(x,\xi) \geq 0) \leq \eta,$$

where ξ is scenario, η is confidence level, $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a function of x, ξ.

The optimization problem is then:

$$\begin{aligned} & c^\top x \to \min_{x \in \mathbb{R}^n} \\ & \text{s.t.:}\, \text{Prob}_\xi(g(x,\xi) \geq 0) \leq 1 - \eta \\ & \quad \xi \sim \mathcal{N}(\mu, \Sigma), \end{aligned}$$

where μ is expected value of gaussian noise and $Sigma$ is its covariance matrix.

Assumption 1. Hinge loss function. Let p, t be parameters. Then function

$$l(x) = \begin{cases} 0, x < p \\ ty + p, x \geq p \end{cases}$$

is called the Hinge loss function.

Firstly, let consider state-of-the-art methods. Our goal is to approximate the feasible set. It is known such approximations of feasible set as Bernshtein approximation, Markov approximation, Tchebyshev approximation and sample average approximation (SAA).

Convex approximations : Markov, Chebyshev, Bernstein. Assume a surrogate function $\psi(u)$ to replace and approximate the 0-1 loss function. $\psi(u) = \max\{1+u, 0\}$ for Markov approximation, $\psi(u) = (1+u)^2$ for Chebyshev approximation and $\psi(u) = e^u$ for Bernstein approximation. After convex approximation the original problem becomes problem with constraints as following:

$$\begin{aligned} & c^\top x \to \min_{x \in \mathbb{R}^n}, t \in \mathbb{R} \\ & \text{s.t.} t \cdot \mathbb{E}\left[\psi\left(\frac{g(x,\xi)}{t}\right)\right] - t\eta \geq 0 \\ & \quad \xi \sim \mathcal{N}(\mu, \Sigma) \end{aligned}$$

The constraints becomes a deterministic convex constraint.

State optimization problem:

$$
\begin{aligned}
-x &\to \min_{x \in \mathbb{R}^n} \\
\text{s.t.: } & \operatorname{Prob}_{\xi}(x\xi \leq 1) \leq 1 - \eta \\
& \xi \sim \mathcal{N}(\mu, \Sigma),
\end{aligned}
$$

Exact solution:

$$
\operatorname{Prob}_{\xi}(x\xi \leq 0) = \operatorname{Prob}\left(\frac{x\xi}{\sqrt{x\Sigma^2 x}} \leq \frac{1}{x\Sigma}\right) = F\left(\frac{1}{x\Sigma}\right)
$$

Then optimization problem is equivalent to:

$$
\begin{aligned}
-x &\to \min_{x \in \mathbb{R}^n} \\
\text{s.t.: } & F^{-1}(1-\eta)x\Sigma \leq 1
\end{aligned}
$$

Markov approximation.

Sample average approximation. Sample average approximation theoretically solves the problem, especially in cases $\eta \approx 1$, but requires a lot of samples. The main idea of this approach is to replace the probability constraints by sampling.

$$
\begin{aligned}
c^\top x &\to \min_{x \in \mathbb{R}^n} \\
\text{s.t.:} & \frac{1}{N}\sum_{i=1}^{N} I_{(0,+\infty)}(g(x,\xi_i)) \leq \gamma
\end{aligned}
$$

Notice that case $\gamma > \eta$ corresponds to the solution, which is a lower bound on the original problem. Case $\gamma = 0$ is called the scenario approximation.

3. Experiment design.

Competitors. We study empirical performance of the following algorithms:

1. Section 10.3.2 “Markov Approximation” from here, use two proxies for the indicator function, $f(x) = (1+x)_+$ and $f(x) = (1+x)_+^2$
2. Section 10.3.3 “Chebyshev Approximation” from here
3. Section 10.3.4 “Bernstein Approximation” here
4. Mixed-Integer programming + SAA approach from here

Place for importance sampling:

- First, estimate the “true” probability level by sampling. As all methods guarantees feasibility at least on level $1-\eta$, you can do conditional sampling beyond the boundary. See file “Drawing from..” at this folder.

- Second, one we have a probability of failure estimate π, and it is less then η at point $\bar{x}$ one can guarantee that the set

$$\{x : \{c^\top x \geq c^\top \bar{x}\} \wedge \{\sqrt{(x - x_0)^\top \Sigma^{-1}(x - x_0)} \leq -\log(\Phi^{-1}(1 - \pi))\}$$

So, we can try to use this to make a one-step improvement.

We will work with confidence bounds on the nest step.

First experiment. It is devoted to the simplest case with linear function $g(x, \xi)$:

$$\begin{aligned} &c^\top x \to \min \\ &\text{s.t.: } \mathrm{Prob}(a^\top (x + \xi) \leq b) \geq 1 - \eta, \end{aligned}$$

where $a \in \mathbb{R}^n$ is a fixed vector, x is a an optimization variable, and ξ is Gaussian uncertainty with known mean and covariance $\xi \in \mathcal{N}(\mu, \Sigma)$. In this case the result is known [8], while it is still interesting to understand how various methods perform. Our goal is to figure out which of the algorithms solve the problem exactly and which one solves approximately. Try various c ($c \| a$, $c \perp a$, different angles), and different dimensions (say 10, 100, 500, 1000, 2000, 5000). Try to play with Σ keeping $\mu = 0$, so that being outside of a hyperplane is a rare event (e.g. probability of being feasible at x_0 is 0.995 or less) or a common event (e.g. Prob $= 0.7 - 0.9$). Provide corresponding plots. (method/time/objective).

Second Experiment. We remain in the same setup: linear functions and Gaussian noise. Now use only 2 hyperplanes, so it is easy to visualize.

$$\begin{aligned} &c^\top x \to \min \\ &\text{s.t.: } \mathrm{Prob}(a_1^\top (x + \xi) \leq b_1 \wedge a_2^\top (x + \xi) \leq b_2) \geq 1 - \eta, \end{aligned}$$

Try various c, a_1 and a_2 and different dimensions (say 10, 100, 500, 1000, 2000, 5000). Try to play with Σ keeping $\mu = 0$, so that being outside of a hyperplane is a rare event (e.g. probability of being feasible at x_0 is 0.995 or less) or a common event (e.g. Prob $= 0.7 - 0.9$). Provide corresponding plots. Would be great to construct a table with various methods performance (method/time/objective).

Third experiment.. Now let us play with power grids. Use PowerModels.jl (video tutorial, package). We are looking for the DC grid, and DC Optimal power flow:

$$\begin{aligned} &c^\top p \to \min \\ &\text{s.t.:} p = B\theta \\ &\quad p_i^{\max} \geq p_i \geq p_i^{\min} \\ &\quad \theta_i^{\max} \geq \theta \geq \theta_i^{\min} \\ &\quad p \sim \mathcal{N}(\mu, \Sigma), \end{aligned}$$

where $\Sigma = \mathrm{Diag}(\Sigma)$, $\sqrt{\Sigma_{ii}} = 0.03 p_i$ (or $0.05 p_i$). In a chance constrained formulation, we should claim the probability of all inequality constraints (aka security constraints) to be satisfied together is at least 0.95, 0.99, or even 0.999 in a case of critical grids. Alternatively we can use PandaPower. I have a slight preference towards Julia, while it is still ok to use PandaPower.

Fourth experiment. Portfolio optimization. See this for the experiment setting. Please also take a look on this paper to have a nice modern competitor. Section 5.2. of this for the details regarding the data sets.

4. Main results.

5. Empirical Study.

6. Conclusion.
Summary + Extensions + Applications

7. Experimental results.

8. Conclusions.
Some conclusions here.

REFERENCES

[1] K. J. Arrow, *Decision theory and operations research.*

[2] J. W. Baker, M. Schubert, and M. H. Faber, *On the assessment of robustness*, Structural Safety, 30 (2008), pp. 253–267.

[3] Q. S. Z. Y. D. Bo Zhoua, Guo Chenb, *Robust chance-constrained programming approach for the planning of fastcharging stations in electrified transportation networks.*

[4] G. E. Box, *Robustness in the strategy of scientific model building*, in Robustness in statistics, Elsevier, 1979, pp. 201–236.

[5] M. Branda, *Sample approximation technique for mixed-integer stochastic programming problems with several chance constraints.*

[6] A. Charnes and W. W. Cooper, *Chance-constrained programming*, Management science, 6 (1959), pp. 73–79.

[7] S. M. Gal Dalal, Elad Gilboa and L. Wehenkel, *Chance-constrained outage scheduling using a machine learning proxy.*

[8] N. He, *Chance-constrained programming (ccps).*

[9] L. G. Khachiyan, *On the exact solution of systems of linear inequalities and linear programming problems.*

[10] A. J. Kleywegt, A. Shapiro, and T. Homem-de Mello, *The sample average approximation method for stochastic discrete optimization*, SIAM Journal on Optimization, 12 (2002), pp. 479–502.

[11] Y. K. J. B. Masahiro Ono, Marco Pavone, *Chance-constrained dynamic programming with application to risk-aware robotic space exploration.*

[12] A. Nemirovski and A. Shapiro, *Convex approximations of chance constrained programs.*

[13] B. K. Pagnoncelli, S. Ahmed, and A. Shapiro, *Sample average approximation method for chance constrained programming: theory and applications*, Journal of optimization theory and applications, 142 (2009), pp. 399–416.

[14] A. V. R.Austin Dollar, *Automated vehicles in hazardous merging traffic: A chance-constrained approach.*

[15] L. Roald and G. Andersson, *Chance-constrained ac optimal power flow: Reformulations and efficient algorithms.*

[16] Z. M. G. S. Zhinong Wei, Juan Sun, *Chance-constrained coordinated optimization for urban electricity and heat networks.*