Notations:

- We use $\mathbb{R}$ to denote real space, and $P$ to denote probability measure.
- We use $I_A(x)$ to denote an indicator function, as $I_A(x) = 1$ when $x \in A$ and $I_A(x) = 0$ when $x \notin A$.
- We use $x^\top$ to denote the transpose of $x$.

### 10.1 Problem Formulation of CCPs

In this lecture, we mainly concern the so-called **chance-constrained programming** problem (CCPs), which can be dated back to the work in 1950s [CC59].

**Problem 10.1 (P$\epsilon$)** The main problem concerned in this lecture is

$$
\min_x f(x) \\
\text{s.t. } P(g(x, \xi) \leq 0) \geq 1 - \epsilon,
$$

where $\xi$ is a random variable following some distribution, while $g(x, \xi) : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$.

This is a totally different kind of problem because the constraint is in a probabilistic form, rather than the deterministic form, i.e. $g(x, \xi) \leq 0$, as we learned in stochastic optimization. Intuitively, the constraint here can be interpreted as a “reliability constraint” (or “risk constraint”). There are many applications that CCPs can fit in, for example, financial risk management, weather forecast [Gel12].

To facilitate the understanding of chance constraints, let’s consider the following toy example.

**Example 10.2 (Random Linear Program [Lin02])** Assume $w_1 \sim \text{Unif}[1, 4]$, $w_2 \sim \text{Unif}[1/3, 1]$, consider the following problem

$$
\min_x x_1 + x_2 \\
\text{s.t. } w_1 x_1 + x_2 \geq 7 \\
\quad w_2 x_1 + x_2 \geq 4 \\
\quad x_1, x_2 \geq 0
$$

As shown in Figure 10.1, the randomness of $w_1$ and $w_2$ will vary the feasible region of the problem. With the changing feasible region, correspondingly the optimal point will also change, so probabilistic uncertainty will be incurred into the problem. Chance constraint is an option to handle this uncertainty, from a probabilistic view [Nem12].

**Remark 10.3 (Challenges with Chance Constraints)** But at the same time, the change in the constraints will bring new challenges when we solve the problems.
Evaluation issue: to check the feasibility of a point \( x \in \mathbb{R}^n \), we need to compute the following integral (denoted as \( Pr(x) \)):

\[
Pr(x) := P(g(x, \xi) \leq 0) = \int_{\xi \in \mathbb{R}^d} I_{g(x, \xi) \leq 0}(\xi) P(\xi) d\xi
\]  

(10.3)

where \( P(\xi) \) is the probability density function of \( \xi \). But generally the integral is difficult to compute.

(Non)convexity issue: sometimes the constraints will lead the feasible set to be a nonconvex set, here we still use the random linear program above for illustration.

Example 10.4 (Nonconvex Feasible Set [Lin02]) Based on the original random problem, we can formulate the following chance-constrained problem:

\[
\begin{align*}
\min \quad & x_1 + x_2 \\
\text{s.t.} \quad & P(w_1x_1 + x_2 \geq 7) \geq 1 - \epsilon_1 \\
& P(w_2x_1 + x_2 \geq 4) \geq 1 - \epsilon_2 \\
& x_1, x_2 \geq 0
\end{align*}
\]  

(10.4)

with \( w = (w_1, w_2) \) following a discrete distribution as

\[
w = (w_1, w_2) = \begin{cases} 
(1, 1) & \text{w.p. 0.1 (Case 1)} \\
(2, 5/9) & \text{w.p. 0.4 (Case 2)} \\
(3, 7/9) & \text{w.p. 0.4 (Case 3)} \\
(4, 1/3) & \text{w.p. 0.1 (Case 4)} 
\end{cases}
\]  

(10.5)

assume \( w_1 \) and \( w_2 \) are independent, consider the joint chance constraint, i.e.

\[
P(w_1x_1 + x_2 \geq 7, w_2x_1 + x_2 \geq 4) \geq 1 - \epsilon
\]  

(10.6)

let \( \epsilon = 0.1 \), then we can have that feasible \( x = (x_1, x_2) \) should satisfy the inequalities in the chance constraint simultaneously when \( w \) falls into case 1, 2, 3 or case 2, 3, 4. Informally speaking,

Feasible set = (Case 1, 2, 3) \( \cup \) (Case 2, 3, 4)

as shown in Figure 10.2, the set is a nonconvex set, which makes the optimization problem more difficult.
Figure 10.2: Illustration of nonconvex feasible region [Lin02]. The eight straight lines correspond to all possible constraints with different $w_1$ and $w_2$, respectively; the highlighted green broken-line corresponds to the boundary of the feasible set when $\epsilon = 0.1$, in which inequalities in case 1, 2, 3 or case 2, 3, 4 should be satisfied, clearly it shows that the feasible set is a nonconvex set.

Generally there are three scenarios to consider in CCPs:

- **Case 1**: The evaluation of the integral is easy, also the constraint is convex. In this case, the problem will be easier to solve.

- **Case 2**: The evaluation part is easy, but the constraint is not convex. As shown in the example above, the distribution of $\xi$ is discrete, so the evaluation will be easy, but the resulting constraint may be nonconvex. Generally in this case, mixed-integer programming (MIP) or other integer programming-based approach will be applied.

- **Case 3**: The evaluation part is hard and the constraint is not convex. Generally the approach will be approximating the constraint to some other constraints that are easy to deal with, also it is called convex approximation approach.

**Example 10.5 (Case 1 of CCPs) Consider the following constraint ($N$ is the normal distribution)**

\[
P(\xi^\top x \leq b) \geq 1 - \epsilon, \quad \xi \sim \mathcal{N}(\mu, \Sigma)
\]  

(10.7)

where $\xi, x \in \mathbb{R}^n$, so $\xi^\top x \sim \mathcal{N}(\mu^\top x, x^\top \Sigma x)$, we have it is equivalent to

\[
P(\xi^\top x \leq b) = P\left(\frac{\xi^\top x - \mu^\top x}{\sqrt{x^\top \Sigma x}} \leq \frac{b - \mu^\top x}{\sqrt{x^\top \Sigma x}}\right)
= F\left(\frac{b - \mu^\top x}{\sqrt{x^\top \Sigma x}}\right) \geq 1 - \epsilon
\]

(10.8)

where $F(\cdot)$ is the cumulative distribution function (CDF) of Gaussian distribution, so we have

\[
P(\xi^\top x \leq b) \geq 1 - \epsilon \iff F^{-1}(1 - \epsilon)\sqrt{x^\top \Sigma x} + \mu^\top x - b \leq 0
\]

(10.9)

where the RHS is analytic to evaluate.
10.2 Structural Properties

Here we will introduce several important definitions and conclusions in CCPs, one main concern is that in which scenarios the constraint will be convex. We first assume $g(x, \xi)$ is convex in $(x, \xi)$. For simplicity, we denote $\Phi(x) := P(g(x, \xi) \leq 0)$, $X := \{x : \Phi(x) \geq 1 - \epsilon\}$ (10.10)

**Proposition 10.6** If $g$ is lower semi-continuous (l.s.c., or equivalently, closed [Bec17]), then $X$ is closed.

**Proof:** Refer to the proof of Lemma 2.2.1 in [VA13]

**Example 10.7** Let’s consider the following simple example 

$$g(x, \xi) = \xi - h(x)$$ (10.11)

so it is linear in $\xi$ and separable for $(x, \xi)$, assume we know the CDF of $\xi$ denoted as $F_\xi(\cdot)$, then we have 

$$\Phi(x) = P(\xi \leq h(x)) = F_\xi(h(x))$$ (10.12)

Now we want $X$ to be convex, then $\Phi(\cdot)$ should be concave, then a sufficient condition to attain the concavity is that (refer to Section 3.2.4 of [BV04])

- $h(\cdot)$ is concave, which can be ensured by choosing appropriate functions.
- $F_\xi(\cdot)$ is non-decreasing, which naturally holds because of the CDF.
- $F_\xi(\cdot)$ is concave, which is what we need to concern.

In another way, recall that $\Phi(\cdot)$, as a probability, will always be non-negative, so we can have the following reformulation

$$X = \{x : \log(\Phi(x)) \geq \log(1 - \epsilon)\}$$ (10.13)

so we can hope that $\log(\Phi(\cdot))$, i.e., $\log(F_\xi(h(x)))$ is concave, then a sufficient condition to attain the concavity is that

- $h(\cdot)$ is concave, which can be ensured by choosing appropriate functions.
- $\log(F_\xi(\cdot))$ is non-decreasing, which naturally holds.
- $\log(F_\xi(\cdot))$ is concave.

the last condition inspires a very important concept called log-concavity, which holds in many distributions, for example, Gaussian distribution.

**Definition 10.8** (Log-concave [BV04]) A function $f : \mathbb{R}^n \to \mathbb{R}$ is log-concave is equivalent to that $f(x) > 0$ and $\log f$ is concave, or formally, 

$$f(\lambda x + (1 - \lambda)y) \geq [f(x)]^\lambda[f(y)]^{1-\lambda}, \ \forall (x, y) \in \text{dom}(f), \ \lambda \in [0, 1]$$ (10.14)

**Proposition 10.9** If $\xi$ has a log-concave distribution (density) function, then $F_\xi$ is log-concave.
Proof: By definition, for any two realizations $A$ and $B$, define $P(A) := P(w : \xi(w) \in A)$, we have

$$P(\lambda A + (1 - \lambda)B) \geq [P(A)]^\lambda [P(B)]^{1-\lambda}, \quad \lambda \in [0, 1] \quad (10.15)$$

which immediately implies that $F_\xi(\cdot)$ is log-concave.

(For another approach of the proof, refer to Theorem 1 in [BB05].) \hfill \blacksquare

Proposition 10.10 (Prékopa [Pré71]) For a continuous random variable $\xi$, if the density function $f_\xi(\cdot)$ is log-concave, then $\xi$ has a log-concave distribution.

Intuitively, with the above two propositions, a distribution is log-concave if and only if its density is also log-concave, which can further imply the convexity of the feasible set. We can further relax the convexity assumption of $g(\cdot, \cdot)$, by introducing quasi-convexity.

Definition 10.11 (Quasi-convex [BV04]) A function $f : \mathbb{R}^n \to \mathbb{R}$ is quasi-convex (or unimodal) if $\text{dom}(f)$ and all its sublevel sets

$$S_\alpha := \{ x \in \text{dom}(f) | f(x) \leq \alpha \} \quad (10.16)$$

are convex for all $\alpha \in \mathbb{R}$. Also $f$ is quasi-convex if and only if $\text{dom}(f)$ is convex and

$$f(\lambda x + (1 - \lambda)y) \leq \max\{f(x), f(y)\}, \quad \forall (x, y) \in \text{dom}(f), \quad \lambda \in [0, 1] \quad (10.17)$$

Theorem 10.12 (Convexity of Feasible Set) If $g(x, \xi)$ is quasi-convex in $(x, \xi)$, $\xi$ has a log-concave distribution function, then $X$ is convex.

Proof: By definition, we have for any $x, y \in X$ and $\lambda \in [0, 1]$,

$$P(g(\lambda x + (1 - \lambda)y, \xi) \leq 0) = \Phi(\lambda x + (1 - \lambda)y) \quad (10.18)$$

denote $A_x := \{ \xi : g(x, \xi) \leq 0 \}$, $A_y := \{ \xi : g(y, \xi) \leq 0 \}$, then for any $\xi \in A_{\lambda x+(1-\lambda)y}$, we have

$$g(\lambda x + (1 - \lambda)y, \xi) \leq 0 \quad (10.19)$$

then note that

$$\Phi(\lambda x + (1 - \lambda)y) = P(g(\lambda x + (1 - \lambda)y, \xi) \leq 0)$$

$$\geq P(\lambda A_x + (1 - \lambda)A_y)$$

$$\geq [P(A_x)]^\lambda \cdot [P(A_y)]^{1-\lambda}$$

$$= [\Phi(x)]^\lambda \cdot [\Phi(y)]^{1-\lambda} \quad (10.20)$$

so we have

$$\Phi(\lambda x + (1 - \lambda)y) \geq [\Phi(x)]^\lambda \cdot [\Phi(y)]^{1-\lambda} \geq 1 - \epsilon \quad (10.21)$$

so $\lambda x + (1 - \lambda)y \in X$, which concludes the proof. \hfill \blacksquare

Remark 10.13 The above theorem looks general and elegant, but in practice, we should know that

- Log-concave assumption on the distribution is not satisfied in many cases, or it is very hard to verify.
- Besides the quasi-convex assumption, another potential assumption on $g(\cdot, \cdot)$ is jointly quasi-convex. But some simple functions like $g(x, \xi) = \xi^T x$ does not satisfy this assumption. So in summary, we can see that the theoretical guarantees in this section is restricted to some extent.
10.3 Safe Convex Approximation

Based on previous discussions, we found that it needs pretty strong assumptions to ensure the convexity of the constraint, generally the feasible set will be nonconvex. So now we turn to the next important approach to solve CCPs: find a convex set $C \subseteq X$ to approximate the feasible set (not relax).

10.3.1 Basic Framework

To start, let’s assume $g(\cdot, \cdot)$ is convex in $(x, \xi)$, recall that

$$
\Phi(x) := P(g(x, \xi) \geq 0) = E[I_{++}(g(x, \xi))]
$$

where $I_{++}(x) = 1$ if $x > 0$ or 0 if otherwise, so the feasible set can be rewritten as $X := \{x : \Phi(x) \leq \epsilon\}$.

Now we consider to find a convex surrogate function $\psi(\cdot)$ to replace and approximate the 0-1 loss function. We hope the surrogate $\psi(\cdot)$ comes with the following properties:

- Convex
- Non-negative
- Non-decreasing
- For normalization, we require $\psi(0) = 1$

For example, $\psi(u) = (1 + u)^2$ or $\psi(u) = \max(1 + u, 0)$ satisfy the requirements.

**Proposition 10.14 (Perspective Function [He16])** If $f(x)$ is convex, then its perspective $h(x, t) := tf(x/t)$ is convex in $\text{dom}(h) := \{(x, t) | x/t \in \text{dom}(f), t > 0\}$.

Then we know that

$$
\Phi(x) \leq \epsilon \iff E[\psi\left(\frac{g(x, \xi)}{t}\right)] \leq \epsilon \quad \forall t \geq 0
$$

$$
\iff tE\left[\psi\left(\frac{g(x, \xi)}{t}\right)\right] \leq t\epsilon
$$

$$
\iff \inf_{t} \left\{ tE\left[\psi\left(\frac{g(x, \xi)}{t}\right)\right] - t\epsilon \right\} \leq 0
$$

so we can define the following set as an approximation of $X$:

$$
C := \left\{ x : \inf_{t} \left\{ tE\left[\psi\left(\frac{g(x, \xi)}{t}\right)\right] - t\epsilon \right\} \leq 0 \right\} \subseteq X
$$

for the convexity of $C$, we argue that the convexity of $g$ and property of perspective functions implies the joint convexity of $tE\left[\psi\left(\frac{g(x, \xi)}{t}\right)\right]$, then take the infimum over $t$ will result in a convex function, so its level set, i.e. $C$, is convex. So now we can transform the original problem into the following approximation.

**Problem 10.15 (P')** The original problem (P$_\epsilon$), after convex approximation, becomes problem (P') as following

$$
\min_{x,t} \quad f(x)
$$

s.t. \quad tE\left[\psi\left(\frac{g(x, \xi)}{t}\right)\right] - t\epsilon \leq 0

(10.25)
10.3.2 Markov Approximation

Intuitively, the tightest or best approximation of the 0-1 loss is (in fact we can prove it [NS07])
\[
\psi(u) = \max(1 + u, 0) = [1 + u]^+
\]
which is also called Markov approximation. So we have
\[
\inf_{t \geq 0} \left\{ t \mathbb{E} \left[ \psi \left( \frac{g(x, \xi)}{t} \right) \right] - t \epsilon \right\} \leq 0
\]
\[
\Leftrightarrow \inf_{t \geq 0} \left\{ \mathbb{E} \left[ \left( t + g(x, \xi) \right)^+_+ - t \epsilon \right] \leq 0 \right. \right.
\]
\[
\Leftrightarrow \inf_{t \geq 0} \left\{ \frac{1}{\epsilon} \mathbb{E} \left[ \left( t + g(x, \xi) \right)^+_+ - t \right] \leq 0 \right. \right.
\]
\[
\Leftrightarrow \inf_{t} \left\{ \frac{1}{\epsilon} \mathbb{E} \left[ \left( g(x, \xi) - t \right)^+_+ + t \right] \leq 0 \right. \right.
\]

The third implication holds by replacing \( t \) with \(-t\) and ignore the negativity constraint [Duc18]. The expression in the second line is related to the value-at-risk (VaR), and the fourth line corresponds to the conditional value-at-risk (CVaR).

Definition 10.16 (VaR and CVaR) For a random variable \( Z \), given a scale \( 1 - \epsilon \in [0, 1] \), its value-at-risk ((VaR)) at level \( 1 - \epsilon \) is defined as
\[
\text{VaR}_{1-\epsilon}(Z) := \inf \left\{ t : P(Z \leq t) \geq 1 - \epsilon \right\}
\]
and its conditional value-at-risk ((CVaR)) at level \( 1 - \epsilon \) is defined as
\[
\text{CVaR}_{1-\epsilon}(Z) := \mathbb{E}[Z|Z \geq \text{VaR}_{1-\epsilon}(Z)] = \inf \left\{ t + \frac{1}{\epsilon} \mathbb{E}[z - t]^+_+ \right\}
\]

VaR and CVaR are concepts commonly used in the field of financial risk management, for more details about them in stochastic optimization, we refer the interested readers to materials like [Duc18] (Section 6) and references therein. Now we can conclude that in Markov approximation, for the original constraint
\[
\inf_{t \geq 0} \left\{ t \mathbb{E} \left[ \psi \left( \frac{g(x, \xi)}{t} \right) \right] - t \epsilon \right\} \leq 0 \quad \Leftrightarrow \quad \text{CVaR}_{1-\epsilon}(g(x, \xi)) \leq 0
\]
In fact \( \text{CVaR}_{1-\epsilon}(g(x, \xi)) \) is convex in \( x \) if \( g \) is also convex in \( x \), but the expectation will incur difficulties into the calculation in real applications, Monte Carlo method may be an option to approximate.

10.3.3 Tchebyshev Approximation

Another convex approximation method, i.e., Tchebyshev approximation, is letting
\[
\psi(u) = (1 + u)^2,
\]
and this converts problem (10.25) to the following problem:
\[
\begin{align*}
\min_{x, t} & \quad f(x) \\
\text{s.t.} & \quad \frac{\mathbb{E} [g(x, \xi)^2]}{t} + 2\mathbb{E} [g(x, \xi)] + (1 - \epsilon) \leq 0.
\end{align*}
\]
Supposing \( g(x, \xi) = \xi^\top x + b \) is a linear function, where \( \mathbb{E}[\xi] = \mu \) and \( \mathbb{E}[\xi^\top \xi] = \Sigma + \mu \mu^\top \), the constraint in (10.32) then becomes

\[
\frac{\mathbb{E}[g(x, \xi)^2]}{t} + 2\mathbb{E}[g(x, \xi)] + (1 - \epsilon) \leq 0 \tag{10.33}
\]

\[
\Rightarrow \quad \mathbb{E}[g(x, \xi)] + \sqrt{1 - \epsilon} \mathbb{E}[g(x, \xi)^2] \leq 0 \tag{10.34}
\]

\[
\Rightarrow \quad \mu^\top x + b + \sqrt{1 - \epsilon} \sqrt{x^\top (\Sigma + \mu \mu^\top) x + 2b \mu^\top + b^2} \leq 0, \tag{10.35}
\]

which is a second order conic programming problem.

### 10.3.4 Bernstein Approximation

We could also let

\[
\psi(u) = e^u, \tag{10.36}
\]

which is what being called Bernstein approximation.

Supposing \( g(x, \xi) = \xi^\top x \), where \( \xi_1, \ldots, \xi_n \) are independent, the constraint in (10.32) then becomes

\[
t \log \mathbb{E} \left[ e^\frac{g(x, \xi)}{t} \right] - t \log(\epsilon) \leq 0 \tag{10.37}
\]

\[
\Rightarrow \quad t \log \mathbb{E} \left[ e^{\sum_{i=1}^{n} \xi_i x_i / t} \right] - t \log(\epsilon) \leq 0 \tag{10.38}
\]

\[
\Rightarrow \quad t \log \mathbb{E} \left[ \prod_{i=1}^{n} e^{\xi_i x_i / t} \right] - t \log(\epsilon) \leq 0 \tag{10.39}
\]

\[
\Rightarrow \quad t \log \prod_{i=1}^{n} \mathbb{E} \left[ e^{\xi_i x_i / t} \right] - t \log(\epsilon) \leq 0 \tag{10.40}
\]

\[
\Rightarrow \quad t \left( \sum_{i=1}^{n} \log \mathbb{E} \left[ e^{\xi_i x_i / t} \right] \right) - t \log(\epsilon) \leq 0 \tag{10.41}
\]

\[
\Rightarrow \quad \left( \sum_{i=1}^{n} \Lambda_i \left( \frac{x_i}{t} \right) \right) - \log(\epsilon) \leq 0, \tag{10.42}
\]

where \( \Lambda_i(u) = \log \mathbb{E} \left[ e^{\xi_i u} \right] \) is logarithmic moment generating function of \( \xi_i \).

Therefore, the constraint becomes a deterministic convex constraint.

### 10.4 Sample Average Approximation

The three convex approximations, i.e., Markov approximation, Tchebyshev approximation and Bernstein approximation, definitely offer a convex set inside the original non-convex constraints. However, we can hardly have idea about how good the approximations are.

To have a sense of how good our approximations are, a simple yet decent way is to replace the probability
constraint in (10.57) by sampling, as in the following problem (P_{\gamma N}) in (10.43).

\[
\begin{align*}
\min_x & \quad f(x) \\
\text{s.t.} & \quad \frac{1}{N} \sum_{i=1}^{N} I_{(0, +\infty)}(g(x, \xi)) \leq \gamma,
\end{align*}
\]  

(10.43)

where \(\xi_i\) is the \(i^{th}\) sample that independently drawn from the distribution of \(\xi\).

Clearly, a solution that satisfy constraint (10.43) can not guarantee to satisfy the original chance constraint in (10.57). Instead, the likelihood of a solution that satisfying the original chance constraint depends of our choice of \(\gamma\). If \(\gamma > \epsilon\), the solution can serve as a lower bound on the original problem. If \(\gamma \leq \epsilon\), then it is somewhat likely we will get a feasible solution. Particularly, if \(\gamma = 0\), it is requiring that \(g(x, \gamma_i) \leq 0\) for all \(i = 1 \ldots N\), which is called scenario approximation.

The key question is, if a solution satisfying (P_{\gamma N}), then how likely will it be feasible for (P_{\epsilon})? Denoting

\[
X_{\epsilon} = \{ x \mid P(g(x, \xi) \leq 0) \geq 1 - \epsilon \}
\]

(10.44)

\[
X_{\gamma} = \{ x \mid \frac{1}{N} \sum_{i=1}^{N} I_{(0, +\infty)}(g(x, \xi_i)) \leq \gamma \},
\]

(10.45)

we have the following proposition.

**Proposition 10.17** If ground set \(X\) is a finite set, and \(0 < \gamma \leq \epsilon\), it holds that

\[
P(X_{\gamma} \subseteq X_{\epsilon}) \geq 1 - |X|e^{-2N(\epsilon-\gamma)^2}.
\]

(10.46)

**Proof:** If \(x \in X \setminus X_{\epsilon}\), which means \(P(g(x, \xi) \leq 0) < 1 - \epsilon\), then denoting \(y = I_{(0, +\infty)}(g(x, \xi))\), we have

\[
\mathbb{E}[y] \geq \epsilon.
\]

Let \(y_i = I_{0, +\infty}(g(x, \xi_i))\), it holds that

\[
P(x \in X_{\gamma}^N) = P\left(\frac{1}{N} \sum_{i=1}^{N} y_i \leq \gamma\right)
\]

\[
\leq P\left(\frac{1}{N} \sum_{i=1}^{N} y_i - \mathbb{E}[y] \leq \gamma - \epsilon\right)
\]

\[
\leq e^{-2N(\epsilon-\gamma)^2},
\]

(10.48)

(10.49)

where the last inequality is due to Hoeffding’s inequality.

Denoting event \(A_x = \{ x \in X_{\gamma}^N \mid x \in X \setminus X_{\epsilon}\}\), we have shown that \(P(A_x) \leq e^{-2N(\epsilon-\gamma)^2}\). Therefore, applying union bound, it can be derived that

\[
P(X_{\gamma}^N \nsubseteq X_{\epsilon}) = P(\cup_{x \in X \setminus X_{\epsilon}} A_x)
\]

\[
\leq \sum_{x \in X \setminus X_{\epsilon}} P(A_x) \leq |X|P(A_x)
\]

\[
\leq |X|e^{-2N(\epsilon-\gamma)^2}
\]

(10.50)

(10.51)

(10.52)
That is being said,

\[ P(X^\gamma_N \subseteq X) = 1 - P(X^\gamma_N \nsubseteq X) \]

\[ \geq 1 - |X|e^{-2N(\epsilon-\gamma)^2} \]  

(10.53) \hspace{1cm} (10.54)  

Feasibility of approximated problem \((P^\gamma_N)\) have shown in Proposition 10.17, the following question of how can we solve \((P^\gamma_N)\) naturally arises.

As we can see, \((P^\gamma_N)\) is equivalent to a mixed integer programming as shown in the following.

\[ (P^\gamma_N) \iff \min_{x,z} f(x) \]

\[ \text{s.t.} \quad \frac{1}{N} \sum_{i=1}^{N} z_i \leq \gamma \]

\[ g(x,\xi) \leq M_i z_i \quad i = 1, \ldots, N \]  

(10.55)  

where \(z_i \in \{0,1\}\) and \(M_i = \max_x g(x,\xi_i)\).

However, integer variables can make an optimization problem non-convex, and therefore, generally speaking, the problem is not easy to solve.

### 10.5 Extensions

Besides the basic Chance Constrained Programming (CCPs) we have introduced, there are many other extensions.

#### 10.5.1 Ambiguous Chance Constrained Programs

In the original CCPs \((P_\epsilon)\), it is implicitly assumed that the probability measure \(P\) associated with the random variable \(\xi\) is fixed. However, in practice, the underlying probability measure \(P\) is rarely known exactly, but is estimated from data. Therefore, it is usual that the underlying distribution is ambiguous with some error, i.e., the probability measure \(P \in \mathcal{P}\) where \(\mathcal{P}\) is a set of probability measure.

The model considering the above-mentioned ambiguity is

\[ \min_{x} f(x) \]

\[ \text{s.t.} \quad P(g(x,\xi) \leq 0) \geq 1 - \epsilon, \quad \forall P \in \mathcal{P}. \]  

(10.56)

#### 10.5.2 Robust Optimization Approximation

Instead of considering probability, from a perspective of robustness, we can form the original CCPs \((P_\epsilon)\) as the following robust optimization problem.

\[ \min_{x} f(x) \]

\[ \text{s.t.} \quad g(x,\xi) \leq 0, \quad \forall \xi \in \mathcal{U}(\epsilon), \]  

(10.57)
where \( \mathcal{U}(\epsilon) \) is an uncertainty set of \( \xi \) with error \( \epsilon \).

We note that it is similar and related to problem \((P')\), depending on which \(\psi(\cdot)\) we choose.

References


[He16] Niao He. Ie 598: Big data optimization, 2016.


