Concentration of risk measures: A Wasserstein distance approach

Prashanth L. A.‡

Joint work with Sanjay P. Bhat†

‡ IIT Madras
† TCS Research

Introduction
Risk criteria

- Conditional Value-at-Risk *(Rockafellar, Ursayev 2000)*
- Spectral risk measures *(Acerbi 2002)*
- Cumulative prospect theory *(Tversky, Kahnemann 1992)*
Open Question ???

Given i.i.d. samples and an empirical version of the risk measure, for a distribution with unbounded support

Obtain concentration bounds for each of the three risk measures

Idea: Use finite sample bounds for Wasserstein distance between empirical and true distributions
Empirical risk concentration: summary of contributions

**Goal:** Bound \( \mathbb{P} [ |\hat{r}_n - r(X) | > \epsilon ] \)

\( \hat{r}_n \rightarrow \) empirical risk using \( n \) i.i.d. samples,
\( r(X) \rightarrow \) true risk

<table>
<thead>
<tr>
<th>Risk measure</th>
<th>Bounded support</th>
<th>Sub-Gaussian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral risk measures</td>
<td>Our work</td>
<td>Our work</td>
</tr>
<tr>
<td>Cumulative prospect theory</td>
<td>[Cheng et al. 2018]</td>
<td>Our work</td>
</tr>
</tbody>
</table>

**Unified approach:** For each bound, the estimation error is related to Wasserstein distance between empirical and true distributions\(^1\)

Wasserstein Distance
The Wasserstein distance between two CDFs $F_1$ and $F_2$ on $\mathbb{R}$ is

$$W_1(F_1, F_2) = \left[ \inf \int_{\mathbb{R}^2} |x - y| dF(x, y) \right],$$

where the infimum is over all joint distributions having marginals $F_1$ and $F_2$.

Related to the **Kantorovich mass transference** problem

- **Ship** masses around so that the initial mass distribution $F_1$ changes into $F_2$
- **Shipping plan**: given by joint distribution $F$ with marginals $F_1$ and $F_2$ such that the amount of mass shipped from a neighborhood $dx$ of $x$ to the neighborhood $dy$ of $y$ is proportional to $dF(x, y)$
- The integral above is then the total transportation distance under the shipping plan $F$
- **Wasserstein distance** between $F_1$ and $F_2$ is the transportation distance under the optimal shipping plan
Wasserstein Distance: Concentration Bounds

\( X \rightarrow \text{r.v. with CDF } F, \quad F_n \rightarrow \text{empirical CDF formed using } n \text{ i.i.d. samples.} \) Then\(^2\),

\[
\mathbb{P} \left( W_1(F_n, F) > \epsilon \right) \leq B(n, \epsilon), \text{ for any } \epsilon > 0,
\]

Exponential moment bound:

If \( \exists \beta > 1 \) and \( \gamma > 0 \) such that \( \mathbb{E} \left( \exp \left( \gamma |X - \mathbb{E}(X)|^\beta \right) \right) < \infty \), then

\[
B(n, \epsilon) = C \left( \exp (-cn\epsilon^2) \mathbb{1} \{\epsilon \leq 1\} + \exp (-cn\epsilon^\beta) \mathbb{1} \{\epsilon > 1\} \right)
\]

Higher moment bound:

If \( \exists \beta > 2 \) such that \( \mathbb{E} \left( |X - \mathbb{E}(X)|^\beta \right) < \infty \), then, for any \( \eta \in (0, \beta) \),

\[
B(n, \epsilon) = C \left( \exp (-cn\epsilon^2) \mathbb{1} \{\epsilon \leq 1\} + n (n\epsilon)^{-(\beta-\eta)/p} \mathbb{1} \{\epsilon > 1\} \right)
\]

Conditional Value-at-Risk
VaR and CVaR are Risk-Sensitive Metrics

- Widely used in financial portfolio optimization, credit risk assessment and insurance
VaR and CVaR are Risk-Sensitive Metrics

• Widely used in financial portfolio optimization, credit risk assessment and insurance

• Let $X$ be a continuous random variable

• Fix a ‘risk level’ $\alpha \in (0, 1)$
VaR and CVaR are Risk-Sensitive Metrics

- Widely used in financial portfolio optimization, credit risk assessment and insurance
- Let $X$ be a continuous random variable
- Fix a ‘risk level’ $\alpha \in (0, 1)$ (say $\alpha = 0.95$)
VaR and CVaR are Risk-Sensitive Metrics

- Widely used in financial portfolio optimization, credit risk assessment and insurance
- Let $X$ be a continuous random variable
- Fix a ‘risk level’ $\alpha \in (0, 1)$ (say $\alpha = 0.95$)

Value at Risk:

$$V_\alpha(X) = F_X^{-1}(\alpha)$$
VaR and CVaR are Risk-Sensitive Metrics

• Widely used in financial portfolio optimization, credit risk assessment and insurance
• Let $X$ be a continuous random variable
• Fix a ‘risk level’ $\alpha \in (0, 1)$ (say $\alpha = 0.95$)

Value at Risk:

$$v_\alpha(X) = F_X^{-1}(\alpha)$$

Conditional Value at Risk:

$$c_\alpha(X) = \mathbb{E} [X | X > v_\alpha(X)]$$

$$= v_\alpha(X) + \frac{1}{1 - \alpha} \mathbb{E} [X - v_\alpha(X)]^+$$
Defining CVaR

Value at Risk:

\[ v_\alpha(X) = F_X^{-1}(\alpha) \]

Conditional Value at Risk:

\[ c_\alpha(X) = \mathbb{E}[X | X > v_\alpha(X)] \]

\[ = v_\alpha(X) + \frac{1}{1 - \alpha} \mathbb{E} [(X - v_\alpha(X))^+] \]

For a general r.v. \( X \),

\[ c_\alpha(X) = \inf_\xi \left\{ \xi + \frac{1}{1 - \alpha} \mathbb{E} (X - \xi)^+ \right\}, \quad \text{where} \ (y)^+ = \max(y, 0) \]
CVaR is a Coherent Risk Metric

• **Monotonicity**: If $X \leq Y$, then $c(X) \leq c(Y)$

• **Sub-additivity**: $c(X + Y) \leq c(X) + c(Y)$, i.e., diversification cannot lead to increased risk.

• **Positive Homogeneity**: $c(\lambda X) = \lambda c(X)$ for any $\lambda \geq 0$.

• **Translation Invariance**: For deterministic $a > 0$, $c(X + a) = c(X) - a$.

---

CVaR is a *Coherent* Risk Metric

- **Monotonicity**: If $X \leq Y$, then $c(X) \leq c(Y)$
- **Sub-additivity**: $c(X + Y) \leq c(X) + c(Y)$, i.e., diversification cannot lead to increased risk.
- **Positive Homogeneity**: $c(\lambda X) = \lambda c(X)$ for any $\lambda \geq 0$.
- **Translation Invariance**: For deterministic $a > 0$, $c(X + a) = c(X) - a$.

Note: VaR is not sub-additive\(^3\)

1. **Exponential Case:** Suppose $X \sim \text{Exp}(\mu)$

   - $v_\alpha(X) = \frac{1}{\mu} \ln \left( \frac{1}{1 - \alpha} \right)$,
   
   - $c_\alpha(X) = v_\alpha(X) + \frac{1}{\mu}$ (memoryless!)
Examples

1. **Exponential Case:** Suppose $X \sim \text{Exp} (\mu)$
   
   - $v_\alpha (X) = \frac{1}{\mu} \ln \left( \frac{1}{1 - \alpha} \right),$
   
   - $c_\alpha (X) = v_\alpha (X) + \frac{1}{\mu}$ (memoryless!)

2. **Gaussian Case:** Suppose $X \sim \mathcal{N} (\mu, \sigma^2)$
   
   - $v_\alpha (X) = \mu - \sigma Q^{-1} (\alpha)$
   
   - $c_\alpha (X) = \mu + \sigma c_\alpha (Z), \quad Z \sim \mathcal{N} (0, 1)$
1. **Exponential Case:** Suppose $X \sim \text{Exp}(\mu)$
   
   - $v_\alpha(X) = \frac{1}{\mu} \ln \left( \frac{1}{1 - \alpha} \right)$,
   
   - $c_\alpha(X) = v_\alpha(X) + \frac{1}{\mu}$ (memoryless!)

2. **Gaussian Case:** Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$
   
   - $v_\alpha(X) = \mu - \sigma Q^{-1}(\alpha)$
   
   - $c_\alpha(X) = \mu + \sigma c_\alpha(Z)$, $Z \sim \mathcal{N}(0, 1)$

For these distributions, no separate CVaR estimate is necessary – estimating $\mu$ and $\sigma$ would do.
Problem: Given i.i.d. samples $X_1, \ldots, X_n$ from the distribution $F$ of r.v. $X$, estimate

$$c_\alpha(X) = \mathbb{E}[X | X > v_\alpha(X)]$$

Nice to have: Sample complexity $O\left(1/\epsilon^2\right)$ for accuracy $\epsilon$
Empirical distribution function (EDF): Given samples $X_1, \ldots, X_n$ from distribution $F$,

\[
\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I} \{X_i \leq x\}, \ x \in \mathbb{R}
\]

Using EDF and the order statistics $X_{[1]} \leq X_{[2]} \leq \ldots, X_{[n]}$, form the following estimates\(^4\):

**VaR estimate:**

\[
\hat{\nu}_{n,\alpha} = \inf\{x : \hat{F}_n(x) \geq \alpha\} = X_{[\lceil n\alpha \rceil]}.
\]

---

Empirical distribution function (EDF): Given samples $X_1, \ldots, X_n$ from distribution $F$,

\[
\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{X_i \leq x\}, \ x \in \mathbb{R}
\]

Using EDF and the order statistics $X_{[1]} \leq X_{[2]} \leq \ldots, X_{[n]}$, form the following estimates$^4$:

**VaR estimate:**

\[
\hat{\nu}_{n,\alpha} = \inf\{x : \hat{F}_n(x) \geq \alpha\} = X_{[\lceil n\alpha \rceil]}.
\]

**CVaR estimate:**

\[
\hat{C}_{n,\alpha} = \hat{\nu}_{n,\alpha} + \frac{1}{n(1-\alpha)} \sum_{i=1}^{n} (X_i - \hat{\nu}_{n,\alpha})^+
\]

Concentration bounds for CVaR Estimation

- Need to put some restrictions on the tail distribution to obtain exponential concentration
- Our assumptions:
  - (C1) $X$ satisfies an exponential moment bound, i.e.,
    $\exists \beta > 0$ and $\gamma > 0$ s.t. $\mathbb{E} \left( \exp \left( \gamma |X - \mu|^\beta \right) \right) < T < \infty$, where $\mu = \mathbb{E}(X)$
  
  or
  
  - (C2) $X$ satisfies a higher-moment bound, i.e.,
    $\beta > 0$ such that $\mathbb{E} \left( |X - \mu|^\beta \right) < T < \infty$

Sub-Gaussian r.v.s satisfy (C1), while sub-exponential r.v.s satisfy (C2)
A random variable is \( X \) is sub-Gaussian if \( \exists \sigma > 0 \) s.t.

\[
\mathbb{E} \left[ e^{\lambda X} \right] \leq e^{\frac{\sigma^2 \lambda^2}{2}}, \quad \forall \lambda \in \mathbb{R}.
\]

Or equivalently, letting \( Z \sim \mathcal{N}(0, \sigma^2) \),

\[
\mathbb{P} [X > \epsilon] \leq c \mathbb{P} [Z > \epsilon], \quad \forall \epsilon > 0.
\]

Tail dominated by a Gaussian
A random variable is $X$ is sub-Gaussian if $\exists \sigma > 0$ s.t.

$$\mathbb{E} \left[ e^{\lambda X} \right] \leq e^{\frac{\sigma^2 \lambda^2}{2}}, \; \forall \lambda \in \mathbb{R}. $$

Or equivalently, letting $Z \sim \mathcal{N}(0, \sigma^2)$,

$$\mathbb{P}[X > \epsilon] \leq c \mathbb{P}[Z > \epsilon], \; \forall \epsilon > 0. $$

Tail dominated by a Gaussian

A random variable is $X$ is sub-exponential if $\exists c_0 > 0$ s.t.

$$\mathbb{E} \left[ e^{\lambda X} \right] < \infty, \; \forall |\lambda| < c_0. $$

Or equivalently, $\exists \sigma, b > 0$ s.t.

$$\mathbb{E} \left[ e^{\lambda X} \right] \leq e^{\frac{\sigma^2 \lambda^2}{2}}, \; \forall |\lambda| \leq \frac{1}{b}. $$

Or

$$\mathbb{P}[X > \epsilon] \leq c_1 \exp(-c_2 \epsilon), \; \forall \epsilon > 0. $$

Tail dominated by an exponential r.v
A few well-known concentration inequalities

Let $X_1, \ldots, X_n$ be i.i.d. samples from the distribution of r.v. $X$ with mean $\mu$, and $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

When $X$ is $\sigma$-sub-Gaussian:

$$
\mathbb{P} \left[ |\hat{\mu}_n - \mu| > \epsilon \right] \leq 2 \exp \left( -\frac{n\epsilon^2}{2\sigma^2} \right)
$$
A few well-known concentration inequalities

Let $X_1, \ldots, X_n$ be i.i.d. samples from the distribution of r.v. $X$ with mean $\mu$, and $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

When $X$ is $\sigma$-sub-Gaussian:

$$
P [ |\hat{\mu}_n - \mu| > \epsilon ] \leq 2 \exp \left( -\frac{n \epsilon^2}{2 \sigma^2} \right)
$$

When $X$ is $(\sigma, b)$-sub-exponential:

$$
P [ |\hat{\mu}_n - \mu| > \epsilon ] \leq \begin{cases} 
2 \exp \left( -\frac{n \epsilon^2}{2 \sigma^2} \right), & 0 \leq \epsilon \leq \frac{\sigma^2}{b}, \\
2 \exp \left( -\frac{n \epsilon}{2b} \right), & \epsilon > \frac{\sigma^2}{b}.
\end{cases}
$$
A CVaR concentration result using Wasserstein distance: sub-Gaussian case

When $X$ is $\sigma$-sub-Gaussian,

$$\mathbb{P} [ |\hat{c}_{n,\alpha} - c_\alpha | > \epsilon ] \leq 2C \exp \left( -cn(1 - \alpha)^2 \epsilon^2 \right), \text{ for any } \epsilon \geq 0,$$

where $C, c$ are constants that depend on $\sigma$.

Idea: Use a concentration result\(^5\) for Wasserstein distance between EDF and CDF.

**Note:**

1) The dependence on $n, \epsilon$ cannot be improved

2) Our bound allows a bandit application, as $C, c$ depend on $\sigma$ (assumed to be known in bandit settings)

---

When \( X \) is sub-exponential, for any \( \epsilon \geq 0 \),

\[
\mathbb{P} \left[ |\hat{c}_{n,\alpha} - c_{\alpha}| > \epsilon \right] \leq \begin{cases} 
C \exp \left[ -cn(1 - \alpha)^2\epsilon^2 \right], & 0 \leq \epsilon \leq 1, \\
C n [n(1 - \alpha)\epsilon]^{\eta - 3}, & \epsilon > 1
\end{cases},
\]

where \( C, c \) are universal constants, and \( \eta \) is chosen arbitrarily from \((0, \beta)\).

**Note:**

For \( \epsilon \leq 1 \), the bound above is satisfactory.

For large \( \epsilon \), the second term exhibits polynomial decay, and this is not an artifact of our analysis. Instead, it relates to the sub-optimal rate obtained in [Fourner-Guillin, 2015].

Recent work in [Prashanth et al. 2019] has closed this gap, using a different proof technique.
Proof Idea

We use the following alternative characterization of the Wasserstein distance

\[ W_1(F_1, F_2) = \sup \left| \mathbb{E}(f(X)) - \mathbb{E}(f(Y)) \right|, \]

where (1)

\[ X \] and \( Y \) are random variables having CDFs \( F_1 \) and \( F_2 \), respectively, and supremum is over all 1-Lipschitz functions \( f : \mathbb{R} \to \mathbb{R} \).

The estimation error \( |\hat{c}_{n,\alpha} - c_\alpha| \) is related to the Wasserstein distance in (1), with EDF \( F_n \) as \( F_1 \) and the true distribution \( F \) as \( F_2 \), and Wasserstein distance concentration bounds from [Fournier and Guillin. 2015] are invoked.
Spectral risk measures
Spectral Risk Measure

• A risk spectrum $\phi : [0, 1] \rightarrow [0, \infty)$, defines a risk measure

$$M_\phi(X) = \int_0^1 \phi(\beta) F^{-1}(\beta) d\beta$$

• If $\phi$ is increasing and integrates to 1, then $M_\phi$ is a coherent risk measure
• CVaR is a special case:

$$c_\alpha(X) = M_\phi \text{ for } \phi = (1 - \alpha)^{-1} \mathbb{I}\{\beta \geq \alpha\}$$

• Using risk spectrum, one can assign higher weight to higher losses. In contrast, CVaR assigns same weight for all tail losses.
Estimating a Spectral Risk Measure

- Idea: apply $M_\phi$ to the empirical distribution $F_n$ constructed from $n$ i.i.d. samples of $X$

$$m_{n,\phi} = \int_0^1 \phi(\beta)F_n^{-1}(\beta)d\beta$$

- If $|\phi(\cdot)|$ is bounded above by $K$, then

$$|M_\phi(X) - m_{n,\phi}| \leq KW_1(F,F_n)$$

- Bounds on $W_1(F, F_n)$ immediately yield concentration bounds for the estimator $m_{n,\phi}$
Proof Idea

We use the following alternative characterization of the Wasserstein distance:

$$W_1(F_1, F_2) = \int_0^1 |F_1^{-1}(\beta) - F_2^{-1}(\beta)| \, d\beta, \text{ where}$$

where $F_i^{-1}(\beta) = \inf\{x \in \mathbb{R} : F_i(x) \geq \beta\}$ is the $\beta$-quantile under $F_i$.

The estimation error $|m_{n,\phi} - M_\phi(X)|$ is related to the Wasserstein distance in (2), with EDF $F_n$ as $F_1$ and the true distribution $F$ as $F_2$, and Wasserstein distance concentration bounds from [Fournier and Guillin. 2015] are invoked.
Cumulative prospect theory
AI that benefits humans

Sequential decision making (RL/bandits) setting with rewards evaluated by humans

Cumulative prospect theory (CPT) captures human preferences
Going to office - bandit style

On every day

1. Pick a route to office
2. Reach office and record (suffered) delay
Why not distort?

Delays are stochastic

In choosing between routes, humans *need not* minimize expected delay
Why not distort?

Two-route scenario: Average delay(Route 2) slightly below that of Route 1

Route 2 has a *small* chance of *very* high delay, e.g. jammed traffic

I might prefer Route 1

In choosing between routes, humans *need not* minimize expected delay
Prospect Theory and its refinement (CPT)

Kahneman & Tversky (1979) “Prospect Theory: An analysis of decision under risk” is the second most cited paper in economics during the period, 1975-2000

Cumulative prospect theory - Tversky & Kahneman (1992)
Rank-dependent expected utility - Quiggin (1982)
CPT-value

For a given r.v. $X$, CPT-value $C(X)$ is

$$C(X) := \int_{0}^{\infty} w^+ \left( \mathbb{P} \left( u^+(X) > z \right) \right) dz - \int_{0}^{\infty} w^- \left( \mathbb{P} \left( u^-(X) > z \right) \right) dz$$

Utility functions $u^+, u^- : \mathbb{R} \rightarrow \mathbb{R}_+$, $u^+(x) = 0$ when $x \leq 0$, $u^-(x) = 0$ when $x \geq 0$

Weight functions $w^+, w^- : [0, 1] \rightarrow [0, 1]$ with $w(0) = 0$, $w(1) = 1$
CPT-value

For a given r.v. $X$, CPT-value $C(X)$ is

$$C(X) := \int_0^{\infty} w^+ (\mathbb{P}(u^+(X) > z)) \, dz - \int_0^{\infty} w^- (\mathbb{P}(u^-(X) > z)) \, dz$$

Utility functions $u^+, u^- : \mathbb{R} \to \mathbb{R}_+$, $u^+(x) = 0$ when $x \leq 0$, $u^-(x) = 0$ when $x \geq 0$

Weight functions $w^+, w^- : [0, 1] \to [0, 1]$ with $w(0) = 0$, $w(1) = 1$

Connection to expected value:

$$C(X) = \int_0^{\infty} \mathbb{P}(X > z) \, dz - \int_0^{\infty} \mathbb{P}(-X > z) \, dz$$

$$= \mathbb{E}(X^+) - \mathbb{E}(X^-)$$

$(a)^+ = \max(a, 0)$, $(a)^- = \max(-a, 0)$
Utility and weight functions

Utility functions

For losses, the disutility $-u^-$ is **convex**, for gains, the utility $u^+$ is **concave**.

Weight function

*Overweight* low probabilities, *underweight* high probabilities.
Problem: Given samples \( X_1, \ldots, X_n \) of \( X \), estimate

\[
C(X) := \int_0^\infty w^+ \left( \Pr(u^+(X) > z) \right) \, dz - \int_0^\infty w^- \left( \Pr(u^-(X) > z) \right) \, dz
\]

Nice to have: Sample complexity \( O\left(1/\epsilon^2\right) \) for accuracy \( \epsilon \)
Empirical distribution function (EDF): Given samples $X_1, \ldots, X_n$ of $X$,

$$
\hat{F}^+_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{(u^+(X_i) \leq x)}, \quad \text{and} \quad \hat{F}^-_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{(u^-(X_i) \leq x)}
$$

Using EDFs, the CPT-value $\mathcal{C}(X)$ is estimated by

$$
\bar{C}_n = \int_0^\infty w^+ (1 - \hat{F}^+_n(x)) \, dx - \int_0^\infty w^- (1 - \hat{F}^-_n(x)) \, dx
$$

\hspace{1cm}
Part (I) \hspace{1cm} \text{Part (II)}

---

Empirical distribution function (EDF): Given samples $X_1, \ldots, X_n$ of $X$,

$$
\hat{F}_n^+(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(u^+(X_i) \leq x)}, \quad \text{and} \quad \hat{F}_n^-(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(u^-(X_i) \leq x)}
$$

Using EDFs, the CPT-value $C(X)$ is estimated by \(^6\)

$$
\bar{C}_n = \int_0^\infty w^+ (1 - \hat{F}_n^+ (x)) \, dx - \int_0^\infty w^- (1 - \hat{F}_n^- (x)) \, dx
$$

\begin{align*}
\text{Part (I)} & = \int_0^\infty w^+ (1 - \hat{F}_n^+ (x)) \, dx \\
\text{Part (II)} & = \int_0^\infty w^- (1 - \hat{F}_n^- (x)) \, dx
\end{align*}

Computing Part (I): Let $X_{[1]}, X_{[2]}, \ldots, X_{[n]}$ denote the order-statistics

$$
\text{Part (I)} = \sum_{i=1}^{n} u^+(X_{[i]}) \left( w^+ \left( \frac{n + 1 - i}{n} \right) - w^+ \left( \frac{n - i}{n} \right) \right),
$$

\(^6\text{Cheng et al. Stochastic optimization in a cumulative prospect theory framework. IEEE Transactions on Automatic Control, 2018.}\)
(A1). Weights $w^+, w^-$ are Hölder continuous, i.e.,
$|w^+(x) - w^+(y)| \leq L|x - y|^\alpha, \forall x, y \in [0, 1]$

(A2). Utilities $u^+(X)$ and $u^-(X)$ are bounded above by $M < \infty$

Concentration bound:
Under (A1) and (A2), for any $\epsilon > 0$, we have

$$
\mathbb{P} \left( |\overline{C}_n - C(X)| > \epsilon \right) \leq 2C \exp \left( -\frac{cn\epsilon^{2/\alpha}}{(2LM)^{2/\alpha}} \right)
$$
(A1). Weights $w^+, w^-$ are Hölder continuous, i.e.,
$|w^+(x) - w^+(y)| \leq L|x - y|^\alpha$, $\forall x, y \in [0, 1]$

(A2). Utilities $u^+(X)$ and $u^-(X)$ are bounded above by $M < \infty$

Concentration bound:
Under (A1) and (A2), for any $\epsilon > 0$, we have

$$
\mathbb{P} \left( |\bar{C}_n - C(X)| > \epsilon \right) \leq 2C \exp \left( - \frac{cn\epsilon^{2/\alpha}}{(2LM)^{2/\alpha}} \right)
$$

Lipschitz weights ($\alpha = 1$): Sample complexity $O \left( 1/\epsilon^2 \right)$ for accuracy $\epsilon$

General $\alpha < 1$ case: Sample complexity $O \left( 1/\epsilon^{2/\alpha} \right)$ for accuracy $\epsilon$
CPT-value concentration: Sub-Gaussian case

Truncated estimator:

\[
\tilde{C}_n = \int_0^{\tau_n} w^+(1 - \hat{F}^+_n(z))dz - \int_0^{\tau_n} w^-(1 - \hat{F}^-_n(z))dz, \quad \text{where}
\]

\[
\tau_n = \sigma \left( \sqrt{\log n} + \sqrt{\log \log n} \right)
\]

(A1). Weights \( w^+, w^- \) are Hölder continuous

(A2). Utilities \( u^+(X) \) and \( u^-(X) \) are sub-Gaussian with parameter \( \sigma \)

Concentration bound:

For any \( \epsilon > \frac{8L\sigma^2}{\alpha n^{\alpha/2}} \), and for \( n \) s.t. \( \sigma \sqrt{\log \log n} > \max (\mathbb{E}(u^+(X)), \mathbb{E}(u^-(X))) + 1 \),

\[
\mathbb{P} \left( \left| \tilde{C}_n - C(X) \right| > \epsilon \right) \leq 2C \exp \left( -cn \left( \frac{\epsilon - \frac{8L\sigma^2}{\alpha n^{\alpha/2}}}{L\sqrt{\log n}} \right)^{\frac{2}{\alpha}} \right)
\]
Proof Idea: Bounded case

We use the following alternative characterization of the Wasserstein distance

$$W_1(F_1, F_2) = \int_{-\infty}^{\infty} |F_1(s) - F_2(s)| ds$$

where

The estimation error $|\bar{C}_n - C(X)|$ is related to the Wasserstein distance in (3), with EDF $F_n$ as $F_1$ and the true distribution $F$ as $F_2$, and Wasserstein distance concentration bounds from [Fournier and Guillin. 2015] are invoked.
CVaR bandits
CVaR-aware bandits: Model

**Known**  # of arms $K$ and horizon $n$

**Unknown** Distributions $P_i, i = 1, \ldots, K$,

**CVaR-values** (at fixed risk level $\alpha$): $C_\alpha(1), \ldots, C_\alpha(K)$

**Interaction** In each round $t = 1, \ldots, n$

- pull arm $I_t \in \{1, \ldots, K\}$
- observe a sample loss from $P_{I_t}$

**Benchmark:** $C_* = \min_{i=1,\ldots,K} C_\alpha(i)$.

**Regret** $R_n = \sum_{i=1}^{K} C_\alpha(i)T_i(n) - nC_* = \sum_{i=1}^{K} T_i(n)\Delta_i,$
CVaR-aware bandits: Model

**Known** # of arms $K$ and horizon $n$

**Unknown** Distributions $P_i, i = 1, \ldots, K$,

CVaR-values (at fixed risk level $\alpha$): $C_\alpha(1), \ldots, C_\alpha(K)$

**Interaction** In each round $t = 1, \ldots, n$

- pull arm $I_t \in \{1, \ldots, K\}$
- observe a sample loss from $P_{I_t}$

**Benchmark:** $C_* = \min_{i=1,\ldots,K} C_\alpha(i)$.

**Regret** $R_n = \sum_{i=1}^{K} C_\alpha(i)T_i(n) - nC_* = \sum_{i=1}^{K} T_i(n)\Delta_i,$

**Goal:** Minimize expected regret $E(R_n)$
Optimizing CVaR using confidence bounds

CVaR-LCB

Pull each arm once

For each round $t = 1, 2, \ldots, n$ do

For each arm $i = 1, \ldots, K$ do

Compute an estimate $c_{i,T_i(t-1)}$ of CVaR value $C_{\alpha}(i)$

LCB index: $LCB_t(i) = c_{i,T_i(t-1)} - \frac{2}{1 - \alpha} \sqrt{\frac{\log(C_t)}{c_{T_i(t-1)}}}$

Pull arm $I_t = \arg \min_{i=1,\ldots,K} LCB_t(i)$.

How I learn to stop regretting..

Upper bound

Gap-dependent:

\[ \mathbb{E}(R_n) \leq \sum_{i: \Delta_i > 0} \frac{16 \log(Cn)}{(1 - \alpha)^2 \Delta_i} + K \left( 1 + \frac{\pi^2}{3} \right) \Delta_i \]

Worst-case bound:

\[ \mathbb{E}(R_n) \leq \frac{8}{(1 - \alpha)} \sqrt{Kn \log(Cn)} + \left( \frac{\pi^2}{3} + 1 \right) \sum_i \Delta_i \]

The bound above matches the regular UCB upper bound (for optimizing expected value) up to constant factors.
Sanjay P. Bhat and Prashanth L.A. (2019),
*Concentration of risk measures: A Wasserstein distance approach*,

Prashanth L.A., Krishna Jagannathan and Ravi Kumar Kolla, (2019),
*Concentration bounds for CVaR estimation: The cases of light-tailed and heavy-tailed distributions*,

C. Acerbi (2002),
*Spectral measures of risk: A coherent representation of subjective risk aversion*,
Journal of Banking and Finance.

A. Tversky and D. Kahneman (1992)
*Advances in prospect theory: Cumulative representation of uncertainty*,
Journal of Risk and Uncertainty.

*Deviation inequalities for an estimator of the conditional value-at-risk*,

*Large deviations bounds for estimating conditional value-at-risk*,
Thank you