Concentration of risk measures: A Wasserstein distance approach

Sanjay P. Bhat† and Prashanth L. A.*

† TCS Research
* IIT Madras
Outline

Introduction

CVaR estimation

Cumulative prospect theory

CVaR bandits
Risk criteria

- Conditional Value-at-Risk \((\text{Rockafellar, Ursayev 2000})\)
- Cumulative prospect theory \((\text{Tversky, Kahnemann 1992})\)
Open Question ???

Given i.i.d. samples and an empirical version of the risk measure

Obtain concentration bounds for each of the 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>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\(^1\)

---

Outline

Introduction

CVaR estimation

Cumulative prospect theory

CVaR bandits
Concentration of risk measures

1) 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:
$$c_\alpha(X) = \mathbb{E}[X|X > v_\alpha(X)]$$
$$= v_\alpha(X) + \frac{1}{1 - \alpha} \mathbb{E}[X - v_\alpha(X)]^+$$
CVaR is a \textit{Coherent} Risk Metric

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

\footnote{P. Artzner et al. "Coherent measures of risk." Mathematical finance 9.3 (1999).}
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\(^2\)

---

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 $X \sim \text{Exp}(\mu)$
   
   \begin{itemize}
   \item $v_\alpha(X) = \frac{1}{\mu} \ln \left( \frac{1}{1 - \alpha} \right)$,
   \item $c_\alpha(X) = v_\alpha(X) + \frac{1}{\mu}$ (memoryless!)
   \end{itemize}

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

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

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 $O\left(\frac{1}{\epsilon^2}\right)$ for accuracy $\epsilon$
Empirical distribution function (EDF): Given samples $X_1, \ldots, X_n$ from distribution $F$,

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

Using EDF and the order statistics $X_{[1]} \leq X_{[2]} \leq \ldots, X_{[n]}$, form the following estimates$^3$:

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$^3$:

**VaR estimate:**

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

**CVaR estimate:**

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

---

Concentration for CVaR$_\alpha$ Estimator

• Need to put some restrictions on the tail distribution to obtain exponential concentration

• Our assumptions:

(C1) $X$ satisfies an exponential moment bound, i.e.,
\[ \exists \beta > 0 \text{ and } \gamma > 0 \text{ s.t. } \mathbb{E} \left( \exp \left( \gamma |X - \mu|^\beta \right) \right) < T < \infty, \text{ where } \mu = \mathbb{E}(X) \]

or

(C2) $X$ satisfies a higher-moment bound, i.e.,
\[ \beta > 0 \text{ such that } \mathbb{E} \left( |X - \mu|^\beta \right) < T < \infty \]

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

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

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

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

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

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:

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

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

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

$$\mathbb{P} [ |\hat{\mu}_n - \mu| > \epsilon ] \leq \begin{cases} 2 \exp \left( -\frac{n\epsilon^2}{2\sigma^2} \right), & 0 \leq \epsilon \leq \frac{\sigma^2}{b}, \\ 2 \exp \left( -\frac{n\epsilon}{2b} \right), & \epsilon > \frac{\sigma^2}{b}. \end{cases}$$
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\(^4\) 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)

A CVaR concentration result using Wasserstein distance: sub-exponential case

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

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

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.
Outline

Introduction

CVaR estimation

Cumulative prospect theory

CVaR bandits
Concentration of risk measures

2) Cumulative prospect theory
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
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)
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 $w(0) = 0$, $w(1) = 1$
For a given r.v. $X$, CPT-value $C(X)$ is

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

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 $w(0) = 0$, $w(1) = 1$

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

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

Nice to have: Sample complexity $O(1/\epsilon^2)$ 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 $C(X)$ is estimated by

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

Part (I) \hspace{2cm} Part (II)

---

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

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

Using EDFs, the CPT-value $C(X)$ is estimated by 5

$$\overline{C}_n = \int_{0}^{\infty} w^+(1 - \hat{F}_n^+(x))dx - \int_{0}^{\infty} w^-(1 - \hat{F}_n^-(x))dx$$

\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

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

CPT-value concentration: Bounded case

(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

$$\mathbb{P} \left( |\overline{C}_n - \mathcal{C}(X)| > \epsilon \right) \leq 2C \exp \left( -\frac{cn\epsilon^{2/\alpha}}{(2LM)^{2/\alpha}} \right)$$
CPT-value concentration: Bounded case

(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:

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

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

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

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

Concentration bound:

For any \( \epsilon > 0 \), and for all \( n \) s.t. \( \sigma \sqrt{\log \log n} > \max(\mathbb{E}(u^+(X)), \mathbb{E}(u^-(X))) \),

\[ \mathbb{P} \left( |\tilde{C}_n - C(X)| > \epsilon \right) \leq 2C \exp \left( -cn \left( \frac{\alpha \epsilon}{L \sqrt{\log n}} \right)^\frac{2}{\alpha} \right) \]
CVaR optimization in a multi-armed bandit framework
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:** $C_\star = \min_{i=1,\ldots,K} C_\alpha(i)$.

**Regret** $R_n = \sum_{i=1}^{K} C_\alpha(i)T_i(n) - nC_\star = \sum_{i=1}^{K} T_i(n)\Delta_i,$
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:**  $C_\ast = \min_{i=1,\ldots,K} C_\alpha(i)$.

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

**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 \( l_t = \arg \min_{i=1,\ldots,K} LCB_t(i) \).

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.
Thank you