Risk-Aware Multi-Armed Bandits

SPCOM 2022

Krishna Jagannathan[†] and Prashanth L. A.[♯]

† Department of Electrical Engineering, IIT Madras

♯ Department of Computer Science and Engineering, IIT Madras

12 Jul 2022

A Survey of Risk-Aware Multi-Armed Bandits

Vincent Y. F. Tan[†], **Prashanth L.A.**[*], **and Krishna Jagannathan**[*]

[†]National University of Singapore, [*]IIT Madras
vtan@nus.edu.sg, prashla@cse.iitm.ac.in, krishnaj@ee.iitm.ac.in

Abstract

In several applications such as clinical trials and financial portfolio optimization, the expected value (or the average reward) does not satisfactorily capture the merits of a drug or a portfolio. In such applications, *risk* plays a crucial role, and a risk-aware performance measure is preferable, so as to capture losses in the case of adverse events. This survey aims to consolidate and summarise the existing research on risk measures, specifically in the context of multi-armed bandits. We review various risk measures of interest, and comment on their properties. Next, we review existing concentration inequalities for various risk measures. Then, we proceed to defining risk-aware bandit problems, We consider algorithms for the regret minimization setting, where the exploration-exploitation trade-off manifests, as well as

player chooses or pulls one among several arms, each defined by a certain reward distribution. The player wishes to maximize his reward or find the best arm in the face of the uncertain environment the distributions are *a priori* unknown. There are two general sub-problems in the MAB literature, namely, regret minimization and best-arm identification (also called pure exploration). In the former, in which the exploration-exploitation trade-off manifests, the player wants to maximize his reward over a fixed time period. In the latter, the player simply wants to learn which arm is the best in either the quickest time possible with a given probability of success (the fixed-confidence setting) or he wants to do so with the highest probability of success given a fixed playing horizon (the fixed budget setting). In most of the MAB literature (see Lattimore and Szepesvári (2020) for an up-to-date survey), the metric of interest is defined simply as the mean of the reward distribution associated with the arm pulled.

However, in real-world applications, the mean

Now

Bandits 101: regret minimization, pure exploration, popular algorithms

Risk measures: common risk measures, risk estimation, concentration bounds

Now

Bandits 101: regret minimization, pure exploration, popular algorithms

Risk measures: common risk measures, risk estimation, concentration bounds

---

After coffee

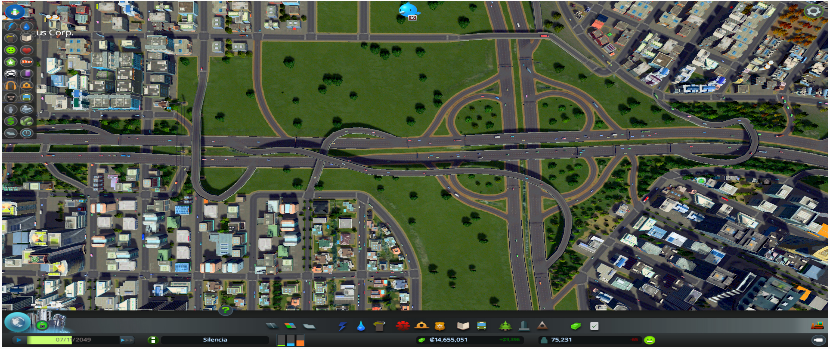
Risk-aware bandits for regret minimization: confidence bound and Thompson sampling based algorithms

Risk-aware bandits for pure exploration: Successive rejects, PAC algorithms


Leave, dad! Stop chilling with us all the day!
We need a new iPad and a PlayStation!
Go out and make some serious money! NOW!
Work-life balance is a challenge for most families
WUMO

Introduction

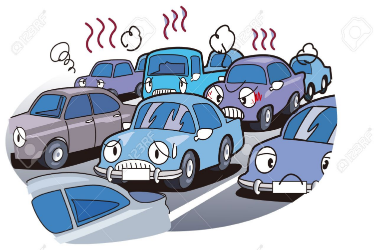
Going to office - bandit style



On every day

1. Pick a route to office
2. Reach office and record (suffered) delay

Bandit learning the best route



Delays are stochastic

Aim is to find the route that has the **lowest expected delay**

Complacs News Recommendation Platform

- NOAM database: 17 million articles from 2010

[1]Work done as a post-doc a long while ago

Complacs News Recommendation Platform

- NOAM database: 17 million articles from 2010
- Task: Find the best among 2000 news feeds

[1] Work done as a post-doc a long while ago

Complacs News Recommendation Platform

- NOAM database: 17 million articles from 2010
- Task: Find the best among 2000 news feeds
- Reward: Relevancy score of the article

[1] Work done as a post-doc a long while ago

Complacs News Recommendation Platform

- NOAM database: 17 million articles from 2010
- Task: Find the best among 2000 news feeds
- Reward: Relevancy score of the article
- Feature dimension: 80000 (approx)

[1]Work done as a post-doc a long while ago

More on relevancy score

Problem: Find the best news feed for Crime stories

Sample scores:

Five dead in Finnish mall shooting — Score: 1.93

More on relevancy score

Problem: Find the best news feed for Crime stories

Sample scores:

Five dead in Finnish mall shooting	Score: 1.93
Holidays provide more opportunities to drink	Score: -0.48

More on relevancy score

Problem: Find the best news feed for Crime stories

Sample scores:

Five dead in Finnish mall shooting	Score: 1.93
Holidays provide more opportunities to drink	Score: -0.48
Russia raises price of vodka	Score: 2.67

More on relevancy score

Problem: Find the best news feed for Crime stories

Sample scores:

Five dead in Finnish mall shooting	Score: 1.93
Holidays provide more opportunities to drink	Score: -0.48
Russia raises price of vodka	Score: 2.67
Why Obama Care Must Be Defeated	Score: 0.43

More on relevancy score

Problem: Find the best news feed for Crime stories

Sample scores:

Five dead in Finnish mall shooting	Score: 1.93
Holidays provide more opportunities to drink	Score: -0.48
Russia raises price of vodka	Score: 2.67
Why Obama Care Must Be Defeated	Score: 0.43
University closure due to weather	Score: -1.06

Figure 1: The *Featured* tab in Yahoo! Today module

[1] Yahoo User-Click Log Dataset given under the Webscope program

Two different frameworks

- Regret minimization: handles exploration-exploitation dilemma
- Best arm identification: a pure exploration
  - Fixed budget: identify best arm(s) with least probability of error in a given budget
  - Fixed confidence: identify best arm(s) with high probability with least expected sample complexity
    * skipped due to time constraints

Exploration exploitation dilemma

Exploitation:

Pull an arm that has the lowest estimated mean loss $\leftarrow$ best decision using historical information

Exploration:

Pull a (random) arm to estimate its mean loss $\leftarrow$ a decision to learn more about the environment

Regret formalizes this dilemma

Bandits: Regret minimization

Known # of arms K and horizon n

Unknown Distributions $F_i, i = 1, \ldots, K$,

Means : $\mu(1), \ldots, \mu(K)$

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

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

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

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

$$T_i(n) = \sum_{t=1}^{n} \mathbb{I}\{I_t = i\}, \quad \Delta_i = \mu(i) - \mu_*$$

Bandits: Regret minimization

Known # of arms K and horizon n

Unknown Distributions $F_i, i = 1, \ldots, K$,

Means : $\mu(1), \ldots, \mu(K)$

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

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

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

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

$$T_i(n) = \sum_{t=1}^{n} \mathbb{I}\{I_t = i\}, \quad \Delta_i = \mu(i) - \mu_*$$

Goal: Minimize expected regret $E(R_n)$

Performance measure

Best arm: $\mu_* = \min_{i=1,\ldots,K} \mu(i).$

Regret $R_n = \sum_{i=1}^{K} \mu(i) T_i(n) - n\mu_* = \sum_{i=1}^{K} T_i(n)\Delta_i$

Goal: ensure R_n grows sub-linearly with n

Performance measure

Best arm: $\mu_* = \min_{i=1,\dots,K} \mu(i).$

Regret $R_n = \sum_{i=1}^{K} \mu(i) T_i(n) - n\mu_* = \sum_{i=1}^{K} T_i(n)\Delta_i$

Goal: ensure R_n grows sub-linearly with n

Bandit algorithms ensure sub-linear regret!

Optimism in the face of uncertainty

LCB

Pull each arm once

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

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

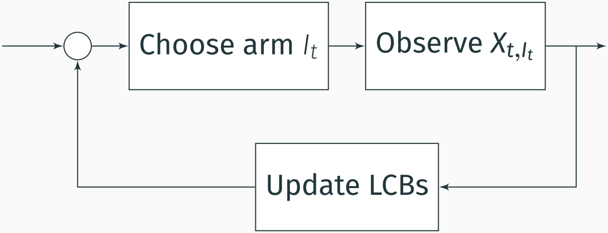
Compute an estimate $\mu_{i,T_i(t-1)}$ of $\mu(i)$

LCB index: $\mathrm{LCB}_t(i) = \mu_{i,T_i(t-1)} - w_{i,T_i(t-1)}$

Pull arm $I_t = \underset{i=1,\ldots,K}{\arg\min}\, \mathrm{LCB}_t(i).$

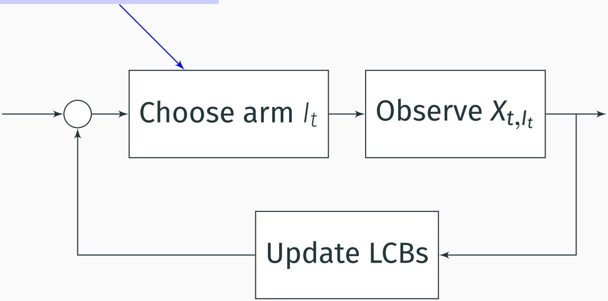
[1] Auer et al. (2002) Finite-time analysis of the multiarmed bandit problem. In: MLJ.

A bandit algorithm

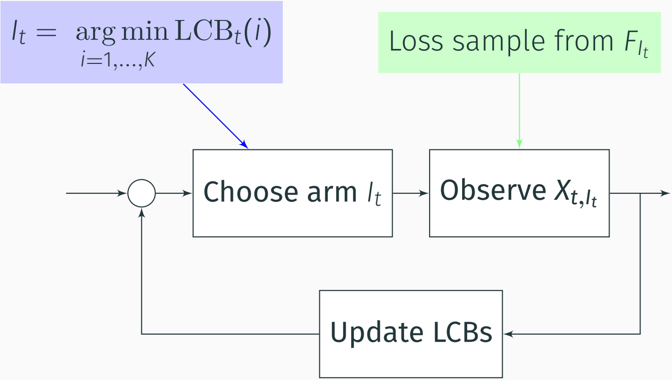


A bandit algorithm

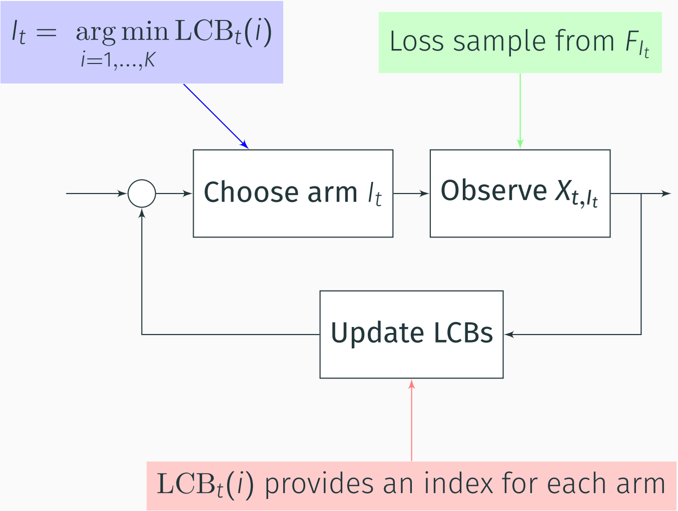
$$I_t = \arg\min_{i=1,\dots,K} \mathrm{LCB}_t(i)$$



A bandit algorithm



A bandit algorithm



Setting LCBs

- Mean-loss estimate

$$\mathrm{LCB}_t(i) = \mu_{i,T_i(t-1)} - w_{i,T_i(t-1)}$$

Setting LCBs

- Mean-loss estimate

$$\mathrm{LCB}_t(i) = \mu_{i,T_i(t-1)} - w_{i,T_i(t-1)}$$

- Confidence width

Setting LCBs

- Mean-loss estimate

$$\mathrm{LCB}_t(i) = \mu_{i,T_i(t-1)} - w_{i,T_i(t-1)}$$

- Confidence width

At each round t, select a tap.

Optimize the quality of n selected beers

Sub-Gaussian distributions

Assumptions on the tail of the distribution

- Need to put some restrictions on the tail distribution to obtain exponential concentration
- A common assumption:

(C1) X satisfies an exponential moment bound, i.e.,
There exist $\beta \geq 1$ and $\gamma > 0$ such that $\mathbb{E}\left(\exp\left(\gamma|X|^{\beta}\right)\right) < \top < \infty$.

Sub-Gaussian and sub-exponential r.v.s satisfy (C1)

We focus on sub-Gaussian distributions in this tutorial

Sub-Gaussian distributions

A random variable is X is sub-Gaussian if $\exists\ \sigma > 0$ s.t. $\forall \epsilon > 0$,

$$P(X \geq \epsilon) \leq \exp\left(-\frac{\epsilon^2}{2\sigma^2}\right)$$ ⟵ Tail dominated by a Gaussian

Or equivalently, $\mathbb{E}\left(\exp\left(\gamma X^2\right)\right) \leq 2$

γ is a constant multiple of σ

If $EX = 0$, then sub-Gaussianity is equivalent to
$\mathbb{E}\left[e^{\lambda X}\right] \leq e^{\frac{\sigma^2 \lambda^2}{2}},\ \forall \lambda \in \mathbb{R}$.

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 μ, and $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

When X is σ-sub-Gaussian:

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

Setting LCBs for sub-Gaussian arms

Assume arms' distributions are σ sub-Gaussian.

LCB index: $\mathrm{LCB}_t(i) = \mu_{i,T_i(t-1)} - w_{i,T_i(t-1)}$

$\mu_{i,T_i(t-1)}$: sample mean formed using $T_i(t-1)$ samples from arm i's distribution

$$w_{i,T_i(t-1)} = \sigma\sqrt{\frac{8\log(t)}{T_i(t-1)}}$$

On the confidence width

Recall: When X is σ-sub-Gaussian:

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

In high-confidence form,

$$\mathbb{P}\left[\mu \in \left[\hat{\mu}_n - \sigma\sqrt{\frac{2\log(\frac{1}{\delta})}{n}}, \hat{\mu}_n + \sigma\sqrt{\frac{2\log(\frac{1}{\delta})}{n}}\right]\right] \geq 1 - 2\delta.$$

On the confidence width

Recall: When X is σ-sub-Gaussian:

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

In high-confidence form,

$$\mathbb{P}\left[\mu \in \left[\hat{\mu}_n - \sigma\sqrt{\frac{2\log(\frac{1}{\delta})}{n}}, \hat{\mu}_n + \sigma\sqrt{\frac{2\log(\frac{1}{\delta})}{n}}\right]\right] \geq 1 - 2\delta.$$

Setting $\delta = \frac{1}{t^4}$, we obtain

$$\mathbb{P}\left[\mu_i \in \left[\mu_{i,T_i(t-1)} - \sigma\sqrt{\frac{8\log t}{T_i(t-1)}}, \mu_{i,T_i(t-1)} + \sqrt{\frac{8\log t}{T_k(t-1)}}\right]\right] \geq 1 - \frac{2}{t^4}$$

$\mu_{i,T_i(t-1)}$: sample mean formed using $T_i(t-1)$ samples from arm i's distribution

Gap-dependent regret upper bound

Gap-dependent:

$$\mathbb{E}(R_n) \leq \sum_{\{i:\Delta_i>0\}} \frac{32\sigma^2 \log n}{\Delta_i} + K\left(1 + \frac{\pi^2}{3}\right)\Delta_i$$

A regret bound that does not scale inversely with gaps:

$$\mathbb{E}(R_n) \leq \left(32K\sigma^2 \log n + K\Delta_i^2\left(\frac{\pi^2}{3} + 1\right)\right)^{\frac{1}{2}}\sqrt{n}.$$

The bound above matches the minimax lower bound on the regret up to constant factors

Bandit learning via simulation

Many practical stochastic optimization settings are difficult to optimize directly.

- Traffic signal control
- Portfolio optimization

Bandit learning via simulation

Many practical stochastic optimization settings are difficult to optimize directly.

- Traffic signal control
- Portfolio optimization

A good alternative of modeling and analysis is "Simulation"

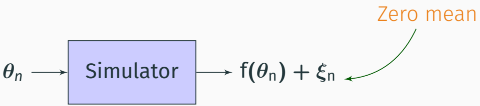


Figure 2: Simulation optimization

Best arm identification with a fixed budget

Known # of arms K and horizon n

Unknown Distributions $F_k, k = 1, \ldots, K$,

Means : $\mu(1), \mu(2), \ldots, \mu(K)$

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

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

Recommendation Arm J_n

Benchmark: $k^* = \underset{k=1,\ldots,K}{\arg\min}\, \mu(k)$.

Best arm identification with a fixed budget

Known # of arms K and horizon n

Unknown Distributions $F_k, k = 1, \ldots, K$,

Means : $\mu(1), \mu(2), \ldots, \mu(K)$

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

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

Recommendation Arm J_n

Benchmark: $k^* = \underset{k=1,\ldots,K}{\arg\min}\, \mu(k)$.

Goal: Minimize probability of erroneous recommendation

$$\mathbb{P}\left[J_n \neq k^*\right]$$

The Successive Rejects Algorithm[1]

- **Initialization**:
  $$A_1 = \{1, \dots, K\},$$
  $$n_k = \left\lceil \frac{1}{\overline{\log}(K)} \frac{n-K}{K+1-k} \right\rceil$$
- **Phase 1**: Play each arm $j \in A_1$, n_1 times; Set $A_2 = A_1 \setminus \arg\max_{j \in A_1} \hat{c}_{j,n_1}$
- **Phase 2**: Play each arm $j \in A_2$, $n_2 - n_1$ times; Eliminate …
- $\vdots$: $\vdots$
- **Phase $K-1$**: Play the remaining two arms $n_{K-1} - n_{K-2}$ times

- One arm played n_1 times, …, another played n_{K-2} times
- Two arms played n_{K-1} times
- $n_1 + \ldots + n_{K-1} + n_{K-1} \leq n$
- n_k increases with k
- Adaptive exploration: better than uniform (i.e., play each arm n/K times)

[1] Audibert et al., *Best Arm Identification in Multi-armed Bandits*, COLT 2010

Probability of error for Successive Rejects

- Suppose the arm distributions are all sub-Gaussian.
- Given a simulation budget n, the probability that the SR algorithm identifies a suboptimal arm as being optimal can be bounded as

$$\mathbb{P}\left[J_n \neq k^*\right] \leq \frac{K(K-1)}{2} \exp\left(-\frac{(n-K)}{H_2 \, \overline{\log}(K)}\right),$$

where

$$H_2 = \max_{k=1,2,\ldots,K} \frac{k}{\Delta_k^2} \leftarrow \text{ Hardness measure}$$

Notation: $\Delta_1 = \Delta_2$

Bottomline: SR needs $O(H_2)$ samples to identify the best arm w.h.p.
Uniform exploration would require $O\left(\frac{K}{\Delta_{\min}^2}\right)$ samples

Risk measures

Motivation

Risk is like fire: If controlled it will help you; if uncontrolled it will rise up and destroy you.

Theodore Roosevelt

Motivation

Risk is like fire: If controlled it will help you; if uncontrolled it will rise up and destroy you.

Theodore Roosevelt

The major difference between a thing that might go wrong and a thing that cannot possibly go wrong is that when a thing that cannot possibly go wrong goes wrong it usually turns out to be impossible to get at or repair.

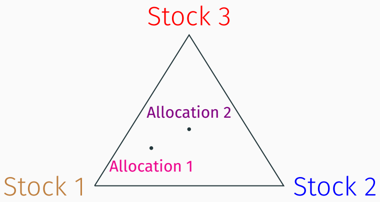
Douglas Adams

Money matters..

Portfolio composed of assets (e.g. stocks)

Stochastic gains for buying/selling assets

Aim find an investment strategy that maximizes the expected return

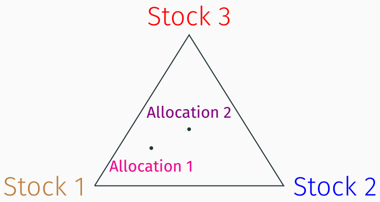


Money matters..

Portfolio composed of assets (e.g. stocks)

Stochastic gains for buying/selling assets

Aim find an investment strategy that maximizes the expected return



A *risk-averse* investor would prefer a strategy that

1. minimizes the risk(e.g. worst-case loss) of the portfolio, while
2. guaranteeing a minimum return

Poll #1

Choose between

A: a stock that gives 10000INR w.p. 0.001 and nothing otherwise

B: a stock that gives 10INR w.p. 1

Which of the two stocks would you favour?

Poll #2

Choose between

C: a stock that loses 10000INR w.p. 0.001 and nothing otherwise

D: a stock that loses 10INR w.p. 1

Which of the two stocks would you favour?

Poll #1

Choose between

A: a stock that gives 10000INR w.p. 0.001 and nothing otherwise

B: a stock that gives 10INR w.p. 1

Which of the two stocks would you favour?

People usually choose A over B

Poll #2

Choose between

C: a stock that loses 10000INR w.p. 0.001 and nothing otherwise

D: a stock that loses 10INR w.p. 1

Which of the two stocks would you favour?

People usually choose **D** over **C**

Message

Humans preferences can be explained using **distorted** probabilities!

People **overweight** extreme & unlikely events and **underweight** average events

Risk criteria

- Conditional Value-at-Risk *(Rockafellar, Ursayev 2000)*
- Spectral risk measures *(Acerbi 2002)*
- Utility-based shortfall risk *(Föllmer and Schied 2001)*
- Cumulative prospect theory *(Tversky,Kahnemann 1992)*
- Optimized certainty equivalent (OCE) risk *(Ben-Tal and Teboulle 2007)*
- Convex risk measure *(Föllmer and Schied 2001)*
- Coherent risk measures *(Artzner1 1999)*
- Rank-dependent expected utility *(Quiggin 2012)*

Mean-variance

Mean: $\mu = \mathbb{E}(X)$

Variance: $\sigma^2 = \mathbb{E}(X - \mu)^2$

Mean-Variance: $\mathrm{MV}(X) = \gamma\mu + \sigma^2$

$\gamma \to$ trade-off mean and variance
Border cases: $\gamma = 0$ and $\gamma \to -\infty$

Conditional Value-at-Risk

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 $\alpha = 0.95$)

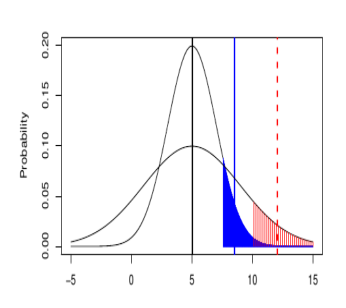
Value at Risk:

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

Conditional Value at Risk:

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

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



Defining CVaR

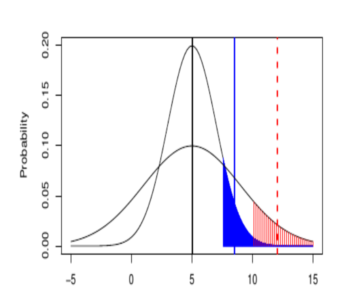
Value at Risk:

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

Conditional Value at Risk:

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

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



For a general r.v. X,

$$c_\alpha(X) = \inf_\xi \left\{\xi + \frac{1}{(1-\alpha)}\mathbb{E}\left(X - \xi\right)^+\right\}, \text{ where } (y)^+ = \max(y, 0)$$

CVaR is a *Coherent* Risk Metric

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

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

Note: VaR is not sub-additive[2]

[2] P. Artzner et al. "Coherent measures of risk." Mathematical finance 9.3 (1999).

Examples

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

Examples

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

Examples

1. Exponential Case: Suppose $X \sim \mathrm{Exp}(\mu)$
   - $v_\alpha(X) = \frac{1}{\mu} \ln \left(\frac{1}{1-\alpha} \right)$,
   - $c_\alpha(X) = v_\alpha(X) + \frac{1}{\mu}$ (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 μ and σ would do

Tea/Coffee

Spectral risk measures

Spectral Risk Measure

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

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

  - If ϕ is increasing and integrates to 1, then M_ϕ is a coherent risk measure
  - CVaR is a special case:

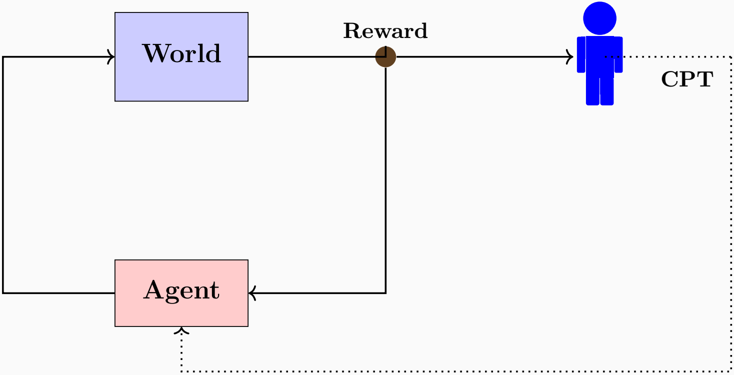
$$c_\alpha(X) = M_\phi \text{ for } \phi = (1-\alpha)^{-1} \mathbb{I}\{\beta \geq \alpha\}$$

- Using risk spectrum, one can assign higher weight to higher losses. In contrast, CVaR assigns same weight for all tail losses.

Cumulative prospect theory

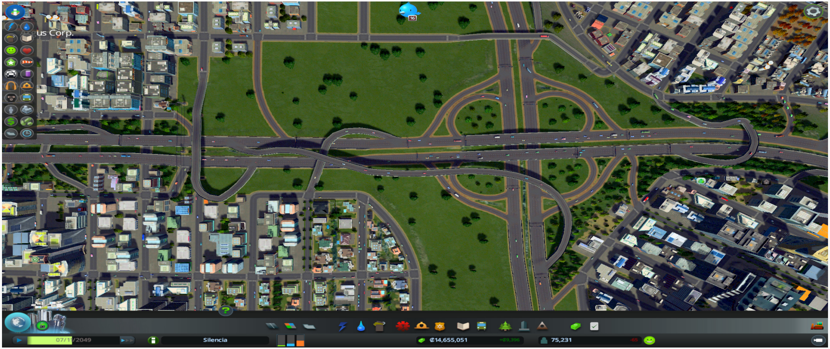
AI that benefits humans

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



Cumulative prospect theory (CPT) *captures* ***human preferences***

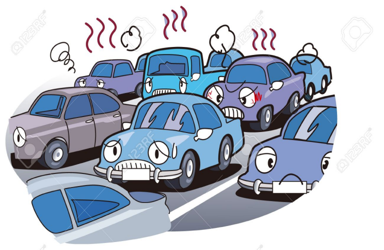
Going to office - bandit style



On every day

1. Pick a route to office
2. Reach office and record (suffered) delay

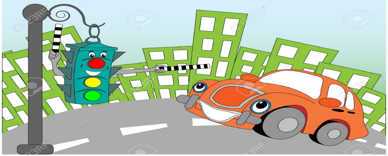
Why not distort?



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

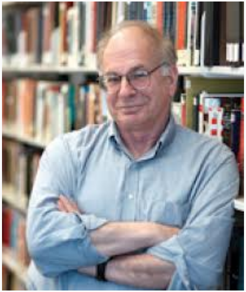
Route 2 has a *small* chance of *very* high delay, e.g. jammed traffic

I might prefer Route 1

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

Prospect Theory and its refinement (CPT)

Amos Tversky



Daniel Kahneman

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 $\mathcal{C}(X)$ is

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

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

CPT-value

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

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

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

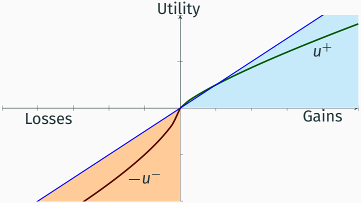
Connection to expected value:

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

$(a)^+ = \max(a, 0)$, $(a)^- = \max(-a, 0)$

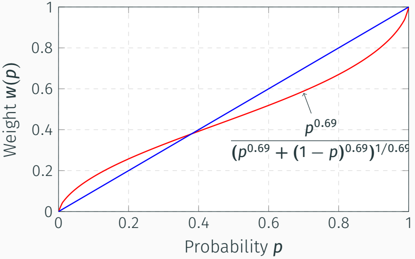
Utility and weight functions

Utility functions

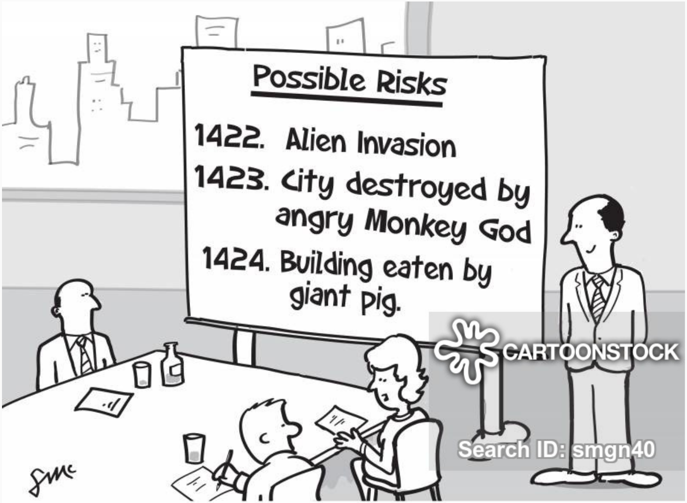


For losses, the disutility $-u^-$ is **convex**, for gains, the utility u^+ is **concave**

Weight function



Overweight low probabilities, **underweight** high probabilities



"Well he certainly does a very thorough risk analysis."

Risk estimation

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 these risk measures

Idea: Use finite sample bounds for Wasserstein distance
between empirical and true distribution

Empirical risk concentration: summary of contributions

Goal: Bound $\mathbb{P}\left[|\rho_n - \rho(X)| > \epsilon\right]$

$\rho(X) \to$ risk measure $\quad \rho_n \to$ estimate of $\rho(X)$ using n i.i.d. samples

Risk measure	Bounded support	Sub-Gaussian
Conditional Value-at-Risk	[Brown, 2007] [Gao et al. 2010]	[Bhat and L.A., 2019] [L.A. et al. 2020]
Spectral risk measure	[Bhat and L.A., 2019]	[Bhat and L.A., 2019]
Cumulative prospect theory	[Cheng et al. 2018]	[Bhat and L.A., 2019]

Wasserstein Distance

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

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

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 $\mathrm{d}x$ of x to the neighborhood $\mathrm{d}y$ of y is proportional to $\mathrm{d}F(x, y)$
- 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: Alternate Characterization

Suppose X and Y are r.v.s having CDFs F_1 and F_2, respectively. Then,

$$\sup |\mathbb{E}(f(X) - \mathbb{E}(f(Y))| = W_1(F_1, F_2)$$
$$= \int_{-\infty}^{\infty} |F_1(s) - F_2(s)| \mathrm{d}s = \int_0^1 |F_1^{-1}(\beta) - F_2^{-1}(\beta)| \mathrm{d}\beta,$$

where the supremum is over all functions $f : \mathbb{R} \to \mathbb{R}$ that are 1-Lipschitz

Wasserstein Distance: Concentration Bounds

$X \rightarrow$ r.v. with CDF F, $F_n \rightarrow$ empirical CDF formed using n i.i.d. samples.

Exponential moment bound:

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

Sub-Gaussian distributions satisfy this bound.

Empirical CDF concentration bound:[3]

$$\mathbb{P}\left(W_1(F_n, F) > \epsilon\right) \leq c_1 \left(\exp\left(-c_2 n \epsilon^2\right) \mathbb{I}\left\{\epsilon \leq 1\right\} + \exp\left(-c_3 n \epsilon^{\beta}\right) \mathbb{I}\left\{\epsilon > 1\right\}\right)$$

Note: The constants c_1, c_2, c_3 are some unknown functions of $\beta, \gamma, \top$.

[3] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. Probability Theory and Related Fields, 162(3-4):707–738, 2015.

Wasserstein Distance: Concentration Bounds

$X \rightarrow$ r.v. with CDF F, $\quad F_n \rightarrow$ empirical CDF from n i.i.d. samples.

Recall: A r.v. X is sub-Gaussian with parameter $\sigma > 0$ if

$$\mathbb{P}(X \geq \epsilon) \leq \exp\left(\frac{-\epsilon^2}{2\sigma^2}\right), \text{ and } \mathbb{P}(X \leq -\epsilon) \leq \exp\left(\frac{-\epsilon^2}{2\sigma^2}\right), \forall \epsilon > 0 \quad (1)$$

Empirical CDF concentration bound: [4] For a σ sub-Gaussian r.v. X

$$\mathbb{P}(W_1(F_n, F) > \epsilon) \leq \exp\left(-\frac{n}{256\sigma^2 e}\left(\epsilon - \frac{512\sigma}{\sqrt{n}}\right)^2\right),$$

for any $\frac{512\sigma}{\sqrt{n}} < \epsilon < \frac{512\sigma}{\sqrt{n}} + 16\sigma e$.

Note: The constants are explicit in this bound.

[4] J. Lei. Convergence and concentration of empirical measures under Wasserstein distance in unbounded functional spaces. Bernoulli, 26(1):767–798, 2020.
Prashanth L.A. and S.P.Bhat, A Wasserstein distance approach for concentration of empirical risk estimates, arXiv:1902.10709v4

CVaR estimation

CVaR estimation: The problem

Problem: Given i.i.d. samples $X_1, \ldots, X_n$ from the distribution F of r.v. X, estimate

$$c_\alpha(X) = \inf_\xi \left\{ \xi + \frac{1}{(1-\alpha)} \mathbb{E}\left(X - \xi\right)^+ \right\}$$

Nice to have: Sample complexity $O\left(1/\epsilon^2\right)$ for accuracy ϵ

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

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

Using EDF and the order statistics $X_{[1]} \leq X_{[2]} \leq \ldots, X_{[n]}$, form the following estimates[5]:

VaR estimate:

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

[5] Serfling, R. J. (2009). Approximation theorems of mathematical statistics, volume 162. John Wiley & Sons.

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[5]:

VaR estimate:

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

CVaR estimate:

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

[5] Serfling, R. J. (2009). Approximation theorems of mathematical statistics, volume 162. John Wiley & Sons.

A CVaR concentration result: sub-Gaussian case

When X is σ-sub-Gaussian,

$$\mathbb{P}\left[\left|\hat{C}_{n,\alpha} - C_\alpha\right| > \epsilon\right] \leq \exp\left(-\frac{n}{256\sigma^2 e}\left(\epsilon(1-\alpha) - \frac{512\sigma}{\sqrt{n}}\right)^2\right).$$

for any $\frac{512\sigma}{\sqrt{n}} < \epsilon(1-\alpha) < \frac{512\sigma}{\sqrt{n}} + 16\sigma e$.

Idea: Use a concentration result[6] for Wasserstein distance between EDF and CDF.

Note:

1) The dependence on n, ϵ cannot be improved

2) This bound allows a bandit application

[6] J. Lei. Convergence and concentration of empirical measures under Wasserstein distance in un- bounded functional spaces. Bernoulli, 26(1):767–798, 2020.

Proof Idea

We use the following alternative characterization of the Wasserstein distance

$$W_1(F_1, F_2) = \sup \left| \mathbb{E}(f(X)) - \mathbb{E}(f(Y)) \right|, \text{ where} \tag{2}$$

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 $\left| \hat{C}_{n,\alpha} - C_\alpha \right|$ 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 bound is invoked.

Spectral risk measure estimation

Estimating a Spectral Risk Measure

- Idea: apply M_ϕ to the empirical distribution F_n constructed from n i.i.d. samples of X

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

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

$$|M_\phi(X) - m_{n,\phi}| \leq K W_1(F, F_n)$$

- 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

$$W_1(F_1, F_2) = \int_0^1 |F_1^{-1}(\beta) - F_2^{-1}(\beta)| \mathrm{d}\beta, \text{ where} \tag{3}$$

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

The estimation error $|m_{n,\phi} - M_\phi(X)|$ 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.

Unification: (T1) risk measures

Hölder continuous Risk Measure[7]

A risk measure $\rho(\cdot)$ is Hölder continuous if $\exists\ \kappa \in (0,1]$ and $L > 0$ s.t. for any two distributions F, G,

$$|\rho(F) - \rho(G)| \leq L\left(W_1(F,G)\right)^{\kappa}.$$

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

Several popular risk measures are Hölder continuous

- CVaR $\kappa = 1$, $L = \dfrac{1}{1-\alpha}$
- Spectral risk measure $\kappa = 1$, $L = K$
- Utility-based shortfall risk $\kappa = 1$, for L, see the paper

Cumulative prospect theory is outside this class of risk measures

[7] P.L.A. and S.P. Bhat, "A Wasserstein distance approach for concentration of empirical risk estimates",2022

Estimating a Hölder continuous Risk Measure

Using EDF from an n-sample, form

$$\rho_n = \rho(F_n).$$

For CVaR, spectral risk measure and utility-based shortfall risk, ρ_n coincides with classic estimators.

A concentration bound for Hölder continuous risk measures

$$|\rho(F) - \rho(G)| \leq L\left(W_1(F,G)\right)^{\kappa}.$$

When X is sub-Gaussian with $\sigma > 0$,

$$\mathbb{P}\left[|\rho_n - \rho(X)| > \epsilon\right] \leq \exp\left(-\frac{n}{256\sigma^2 e}\left(\left(\frac{\epsilon}{L}\right)^{\frac{1}{\kappa}} - \frac{512\sigma}{\sqrt{n}}\right)^2\right),$$

$$\text{for any } \frac{512\sigma}{\sqrt{n}} < \left(\frac{\epsilon}{L}\right)^{\frac{1}{\kappa}} < \frac{512\sigma}{\sqrt{n}} + 16\sigma\sqrt{e}$$

Concentration bounds for CVaR, spectral risk measure and utility-based shortfall risk are corollaries to the result above.

CPT-value estimation

CPT-value estimation

Problem: Given samples $X_1, \ldots, X_n$ of X, estimate

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

Nice to have: Sample complexity $O\left(1/\epsilon^2\right)$ for accuracy ϵ

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 [8]

$$\overline{\mathcal{C}}_n = \underbrace{\int_0^{\infty} w^+(1-\hat{F}_n^+(x))dx}_{\text{Part (I)}} - \underbrace{\int_0^{\infty} w^-(1-\hat{F}_n^-(x))dx}_{\text{Part (II)}}$$

[8] Cheng et al. *Stochastic optimization in a cumulative prospect theory framework.* IEEE Transactions on Automatic Control, 2018.

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

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

Using EDFs, the CPT-value $\mathcal{C}(X)$ is estimated by [8]

$$\overline{\mathcal{C}}_n = \underbrace{\int_0^{\infty} w^+(1 - \hat{F}_n^+(x))dx}_{\textbf{Part (I)}} - \underbrace{\int_0^{\infty} w^-(1 - \hat{F}_n^-(x))dx}_{\textbf{Part (II)}}$$

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

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

[8] Cheng et al. *Stochastic optimization in a cumulative prospect theory framework*. IEEE Transactions on Automatic Control, 2018.

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). $X \in [0, B_1]$ a.s.

Concentration bound:

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

$$\mathbb{P}\left(\left|\overline{\mathcal{C}}_n - \mathcal{C}(X)\right| > \epsilon\right) \leq c_1 \exp\left(-2n\left[\frac{\epsilon}{LB_1}\right]^{\frac{2}{\alpha}}\right)$$

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). $X \in [0, B_1]$ a.s.

Concentration bound:

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

$$\mathbb{P}\left(\left|\overline{\mathcal{C}}_n - \mathcal{C}(X)\right| > \epsilon\right) \leq c_1 \exp\left(-2n\left[\frac{\epsilon}{LB_1}\right]^{\frac{2}{\alpha}}\right)$$

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

General $\alpha < 1$ case: *Sample complexity $O\left(1/\epsilon^{2/\alpha}\right)$ for accuracy* $\boldsymbol{\epsilon}$

CPT-value concentration: Sub-Gaussian case

Truncated estimator: [9]

$$\widetilde{C}_n = \int_0^{\tau_n} w^+(1-\hat{F}_n^+(z))\mathrm{d}z - \int_0^{\tau_n} w^-(1-\hat{F}_n^-(z))\mathrm{d}z, \text{ where } \tau_n = \Theta(\sqrt{\log n})$$

Assume: Weights w^+, w^- are Hölder continuous; u^+, u^- are differentiable, and their derivatives are bounded above and below by $K^+ > 0$ and $k^+ > 0$, and K^- and $k^- > 0$, respectively, in absolute value.

Concentration bound:

For any $\frac{512\sigma}{\sqrt{n}} < \left(\frac{\epsilon - c_3(n)}{c_4(n)}\right)^{\frac{1}{\alpha}} < \frac{512\sigma}{\sqrt{n}} + 16\sigma\sqrt{e}$,

$$\mathbb{P}\left(\left|\widetilde{C}_n - C(X)\right| > \epsilon\right) \leq \exp\left(-\frac{n}{256\sigma^2 e}\left(\left(\frac{\epsilon - c_3(n)}{c_4(n)}\right)^{\frac{1}{\alpha}} - \frac{512\sigma}{\sqrt{n}}\right)^2\right)$$

[9] P.L.A. and S.P.Bhat, *A Wasserstein distance approach for concentration of empirical risk estimates*. arXiv:1902.10709v4, 2022.

Proof Idea: Bounded case

We use the following alternative characterization of the Wasserstein distance

$$W_1(F_1, F_2) = \int_{-\infty}^{\infty} |F_1(s) - F_2(s)| \mathrm{d}s, \text{ where} \tag{4}$$

The estimation error $\left|\overline{\mathcal{C}}_n - \mathcal{C}(X)\right|$ is related to the Wasserstein distance in (4), 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.

Open Question ???

Given i.i.d. samples and an empirical version of a risk measure, for a sub-Gaussian distribution

Obtain concentration bounds for the given risk measure

Idea: Use a direct approach that is risk measure specific

Mean-variance estimation

Samples: $\{X_t\}_{t=1}^{n}$

Sample mean: $\hat{\mu}_n$

Sample variance: $\hat{\sigma}^2$

Mean-Variance: $\widehat{\mathrm{MV}}_n = \gamma \hat{\mu}_n + \hat{\sigma}_n^2$

A concentration bound for mean-variance

When X is sub-Gaussian with $\sigma > 0$,

$$\mathbb{P}\left[|\widehat{\mathrm{MV}}_n - \mathrm{MV}| > \epsilon\right] \leq 2\exp\left[-\frac{n\epsilon^2}{8\gamma^2\sigma^2}\right] + 2\exp\left(-\frac{n}{16}\min\left[\frac{\epsilon^2}{2\sigma^4}, \frac{\epsilon}{\sigma^2}\right]\right),$$

Proof uses sub-Gaussian and sub-exponential concentration bounds, cf. Wainwright's book.

VaR Concentration [10]

Assumption **(A1):** X is a continuous r.v. with a CDF F that satisfies a condition of *sufficient growth* around the VaR v_α: There exists constants $\delta, \eta > 0$ such that

$$\min\left(F\left(v_\alpha + \delta\right) - F\left(v_\alpha\right), F\left(v_\alpha\right) - F\left(v_\alpha - \delta\right)\right) \geq \eta\delta.$$

[10] Concentration bounds for empirical conditional value-at-risk: The unbounded case; R. Kolla, L.A. Prashanth, S. P. Bhat, K. Jagannathan; *Operations Research Letters*, 2019

VaR Concentration [10]

Assumption **(A1):** X is a continuous r.v. with a CDF F that satisfies a condition of *sufficient growth* around the VaR v_α: There exists constants $\delta, \eta > 0$ such that

$$\min\left(F\left(v_\alpha + \delta\right) - F\left(v_\alpha\right), F\left(v_\alpha\right) - F\left(v_\alpha - \delta\right)\right) \geq \eta\delta.$$

VaR concentration

For any $\epsilon \in (0, \delta)$ $\mathbb{P}\left[|\hat{v}_{n,\alpha} - v_\alpha| \geq \epsilon\right] \leq 2\exp\left(-2n\eta^2\epsilon^2\right)$

Proof uses DKW inequality; no tail assumptions required.

[10] Concentration bounds for empirical conditional value-at-risk: The unbounded case; R. Kolla, L.A. Prashanth, S. P. Bhat, K. Jagannathan; *Operations Research Letters*, 2019

CVaR concentration bound: sub-Gaussian case

Recall

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

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

CVaR concentration bound: sub-Gaussian case

Recall

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

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

Theorem (*CVaR concentration for sub-Gaussian*)
Assume (A1). Suppose that X_i, $i = 1, \ldots, n$ are σ-sub-Gaussian. Then, for any $\epsilon \in (0, \delta)$, we have

$$\mathbb{P}\left[|\hat{c}_{n,\alpha} - c_\alpha| > \epsilon\right] \leq 6 \exp\left[-n\psi_1(\epsilon)\right],$$

where $\psi_1(\epsilon) = \dfrac{\epsilon^2(1-\alpha)^2 \min\left(\eta^2, 1\right)}{8 \max\left(\sigma^2, 8\right)}$.

Risk-aware bandits: Regret minimization

Risk-aware bandits: Model

Known # of arms K and horizon n

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

Risk measure : $\rho(1), \ldots, \rho(K)$

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: $\rho_* = \min_{i=1,\ldots,K} \rho(i)$.

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

Risk-aware bandits: Model

Known # of arms K and horizon n

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

Risk measure : $\rho(1), \ldots, \rho(K)$

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: $\rho_* = \min_{i=1,\ldots,K} \rho(i)$.

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

---

Goal: Minimize expected regret $E(R_n)$

Optimizing risk with confidence[1]

Risk-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 $\rho_{i,T_i(t-1)}$ of $\rho(i)$

LCB index: $\mathrm{LCB}_t(i) = \rho_{i,T_i(t-1)} - w_{i,T_i(t-1)}$

Pull arm $I_t = \underset{i=1,\ldots,K}{\arg\min}\, \mathrm{LCB}_t(i).$

[1] Auer et al. (2002) Finite-time analysis of the multiarmed bandit problem. In: MLJ.

Confidence width for Hölder risk measures + sub-Gaussian arms

LCB index: $\mathrm{LCB}_t(i) = \rho_{i,T_i(t-1)} - w_{i,T_i(t-1)}$

$\rho_{i,T_i(t-1)}$: Formed by applying ρ to the EDF formed using $T_i(t-1)$ samples from arm i's distribution

$$w_{i,T_i(t-1)} = L^{\kappa}\left[\left(\frac{4\log(t)}{T_i(t-1)}\right)^{\frac{1}{2}} + \left(\frac{32\sigma^2}{T_i(t-1)}\right)^{\frac{1}{2}}\right]^{\kappa}$$

How I learn to stop regretting..

Upper bound for $\kappa = 1$

Gap-dependent:

$$\mathbb{E}(R_n) \leq \sum_{\{i:\Delta_i>0\}} \frac{(\sqrt{4\log n} + 32\sigma^2)^2 \; 4\,L^2}{\Delta_i} + K\left(1+\frac{\pi^2}{3}\right)\Delta_i$$

Worst-case bound:

$$\mathbb{E}(R_n) \leq \left(K(\sqrt{4\log(n)} + 32\sigma^2)^2 4L^2 + K\Delta_i^2\left(\frac{\pi^2}{3}+1\right)\right)^{\frac{1}{2}} \sqrt{n}.$$

The bound above matches the regular UCB upper bound (for optimizing expected value) up to constant factors

Confidence width for CPT risk measure + sub-Gaussian arms

Recall

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

Assume $w^\pm$ are Hölder with exponent $\alpha < 1$.

LCB index: $\mathrm{LCB}_t(i) = \rho_{i,T_i(t-1)} - w_{i,T_i(t-1)}$

$\rho_{i,T_i(t-1)}$: Truncated CPT estimator

$$w_{i,T_i(t-1)} = \left[L \max\left\{\frac{K^+}{k^+}, \frac{K^-}{k^-}\right\} \log T_i(t-1)\right]^\alpha \left[\left(\frac{4\log t}{T_i(t-1)}\right)^{\frac{1}{2}} + \left(\frac{32\sigma^2}{T_i(t-1)}\right)^{\frac{1}{2}}\right]^\alpha$$

Regret bound for CPT objective

Gap-dependent:

Upper bound

$$\mathbb{E}(R_n) \leq \sum_{\{i:\Delta_i>0\}} \frac{C_1 \log n}{\Delta_i^{\frac{2}{\alpha}-1}} + K\left(1+\frac{\pi^2}{3}\right)\Delta_i$$

Worst-case bound:

$$\mathbb{E}(R_n) \leq C_2 \left(K \log(n)\right)^{\frac{\alpha}{2}} n^{\frac{2-\alpha}{2}}.$$

For $\alpha < 1$, the bound above is worse than usual UCB upper bound of $O(\sqrt{n})$

A lower bound in [Gopalan et al. 2017] shows that the dependence on n and gaps Δ_i cannot be improved in a minimax sense.

Thompson Sampling for Risk-Regret Minimization

- Riou & Honda (ALT ’20): impactful paper showing asymptotic optimality of TS for expected regret, under (multinomial and) bounded distributions

Thompson Sampling for Risk-Regret Minimization

- Riou & Honda (ALT '20): impactful paper showing asymptotic optimality of TS for expected regret, under (multinomial and) bounded distributions
- Baudry, Gautron, Kaufmann, Maillard (ICML '21): TS is optimal for CVaR regret and bounded distributions

Thompson Sampling for Risk-Regret Minimization

- Riou & Honda (ALT '20): impactful paper showing asymptotic optimality of TS for expected regret, under (multinomial and) bounded distributions
- Baudry, Gautron, Kaufmann, Maillard (ICML '21): TS is optimal for CVaR regret and bounded distributions
- Zhu & Tan (ICML '20): TS for mean-variance under Gaussian losses.

Multinomial TS (Riou & Honda '20)

For multinomial distributions over $\{0, \frac{1}{M}, \frac{2}{M}, \dots, 1\}$

Input Horizon n, number of arms K and multinomial support size M.
Initialize $\alpha_k^m = 1$, $k \in [K]$, $m \in \{0, 1, \dots, M\}$

- Sample from Dirichlet distribution →

For $t = 1, 2, \dots, n$ **do**

For $k = 1$ **to** K
- Sample $L_k \sim \mathrm{Dir}(\alpha_0^k, \dots, \alpha_M^k)$

$I(t) = \arg\min_{k \in [K]} (0, \frac{1}{M}, \frac{2}{M}, \dots, 1)^\intercal L_k$

Play arm $I(t)$

Observe loss $\frac{m}{M}$.

Update parameter $\alpha_m^{I(t)} = \alpha_m^{I(t)} + 1$

Multinomial TS (Riou & Honda '20)

For multinomial distributions over $\{0, \frac{1}{M}, \frac{2}{M}, \dots, 1\}$

Input Horizon n, number of arms K and multinomial support size M.
Initialize $\alpha_k^m = 1$, $k \in [K], m \in \{0, 1, \dots, M\}$

- Sample from Dirichlet distribution →
- Arm index →

For $t = 1, 2, \dots, n$ **do**

For $k = 1$ **to** K
- Sample $L_k \sim \mathrm{Dir}(\alpha_0^k, \dots, \alpha_M^k)$

$I(t) = \arg\min_{k \in [K]} (0, \frac{1}{M}, \frac{2}{M}, \dots, 1)^\intercal L_k$

Play arm $I(t)$

Observe loss $\frac{m}{M}$.

Update parameter $\alpha_m^{I(t)} = \alpha_m^{I(t)} + 1$

Multinomial TS (Riou & Honda '20)

For multinomial distributions over $\{0, \frac{1}{M}, \frac{2}{M}, \dots, 1\}$

Input Horizon n, number of arms K and multinomial support size M.
Initialize $\alpha_k^m = 1$, $k \in [K], m \in \{0, 1, \dots, M\}$

- Sample from Dirichlet distribution
- Arm index
- Action and loss at time t

For $t = 1, 2, \dots, n$ **do**

For $k = 1$ **to** K
- Sample $L_k \sim \mathrm{Dir}(\alpha_0^k, \dots, \alpha_M^k)$

$$I(t) = \arg\min_{k \in [K]} (0, \frac{1}{M}, \frac{2}{M}, \dots, 1)^\intercal L_k$$

Play arm $I(t)$

Observe loss $\frac{m}{M}$.

Update parameter $\alpha_m^{I(t)} = \alpha_m^{I(t)} + 1$

Multinomial TS (Riou & Honda '20)

For multinomial distributions over $\{0, \frac{1}{M}, \frac{2}{M}, \ldots, 1\}$

Input Horizon n, number of arms K and multinomial support size M.
Initialize $\alpha_k^m = 1$, $k \in [K]$, $m \in \{0, 1, \ldots, M\}$

- Sample from Dirichlet distribution → For $t = 1, 2, \ldots, n$ **do**
  For $k = 1$ **to** K
  - Sample $L_k \sim \mathrm{Dir}(\alpha_0^k, \ldots, \alpha_M^k)$
- Arm index → $I(t) = \arg\min_{k \in [K]} (0, \frac{1}{M}, \frac{2}{M}, \ldots, 1)^\intercal L_k$
- Action and loss at time t → Play arm $I(t)$

  Observe loss $\frac{m}{M}$.
- Posterior update of Dirichlet parameters → Update parameter $\alpha_m^{I(t)} = \alpha_m^{I(t)} + 1$

Non-parametric TS (Riou & Honda '20)

Assume: arms' distributions support in $[0,1]$

Input Horizon n, number of arms K.
Initialize $X_k = 1, N_k = 1$ for each $k \in [K]$

- Sample from Dirichlet distribution of dimension N_k

For $t = 1, 2, \ldots, n$ **do**

For $k = 1$ **to** K
- Sample $p_k \sim \mathrm{Dir}(\mathbf{1}_{N_k})$

$I(t) = \arg\max_{k \in [K]} X_k^\intercal p_k$

Play arm $I(t)$

Observe loss $r_t^{I(t)}$.

Update $N_{I(t)} = N_{I(t)} + 1, X_{I(t)} = \begin{pmatrix} X_{I(t)} \\ r_t^{I(t)} \end{pmatrix}$

Non-parametric TS (Riou & Honda '20)

Assume: arms' distributions support in $[0,1]$

Input Horizon n, number of arms K.
Initialize $X_k = 1, N_k = 1$ for each $k \in [K]$

- Sample from Dirichlet distribution of dimension N_k →
- Arm index →

For $t = 1, 2, \ldots, n$ **do**

For $k = 1$ **to** K
- Sample $p_k \sim \mathrm{Dir}(\mathbf{1}_{N_k})$

$I(t) = \operatorname{arg\,max}_{k \in [K]} X_k^\intercal p_k$

Play arm $I(t)$

Observe loss $r_t^{I(t)}$.

Update $N_{I(t)} = N_{I(t)} + 1, X_{I(t)} = \begin{pmatrix} X_{I(t)} \\ r_t^{I(t)} \end{pmatrix}$

Non-parametric TS (Riou & Honda '20)

Assume: arms' distributions support in $[0,1]$

Input Horizon n, number of arms K.
Initialize $X_k = 1, N_k = 1$ for each $k \in [K]$

For $t = 1, 2, \ldots, n$ **do**

For $k = 1$ **to** K
- Sample $p_k \sim \mathrm{Dir}(\mathbf{1}_{N_k})$

$I(t) = \arg\max_{k \in [K]} X_k^\top p_k$

Play arm $I(t)$

Observe loss $r_t^{I(t)}$.

Update $N_{I(t)} = N_{I(t)} + 1$, $X_{I(t)} = \begin{pmatrix} X_{I(t)} \\ r_t^{I(t)} \end{pmatrix}$

- Sample from Dirichlet distribution of dimension N_k → (Sample p_k)
- Arm index → ($I(t)$)
- Action and loss at time t → (Play arm $I(t)$)

Non-parametric TS (Riou & Honda '20)

Assume: arms' distributions support in $[0, 1]$

Input Horizon n, number of arms K.
Initialize $X_k = 1, N_k = 1$ for each $k \in [K]$

- Sample from Dirichlet distribution of dimension N_k
- Arm index
- Action and loss at time t
- Update of loss vector X_k and number of pulls N_k

For $t = 1, 2, \ldots, n$ **do**

For $k = 1$ **to** K
- Sample $p_k \sim \mathrm{Dir}(\mathbf{1}_{N_k})$

$I(t) = \arg\max_{k \in [K]} X_k^\intercal p_k$

Play arm $I(t)$

Observe loss $r_t^{I(t)}$.

Update $N_{I(t)} = N_{I(t)} + 1$, $X_{I(t)} = \begin{pmatrix} X_{I(t)} \\ r_t^{I(t)} \end{pmatrix}$

Regret Optimality of TS (Riou & Honda '20)

Non-parametric TS achieves the optimal regret bound for bounded loss distributions:

$$\mathbb{E}(R_n) \leq \sum_{\{i:\Delta_i>0\}} \frac{\Delta_i \log n}{\mathcal{K}_{\text{inf}}(F_i, \mu^*)} + o(\log n)$$

- $\mathcal{K}_{\text{inf}}(F_i, \mu^*) = \inf_{G:\mathbb{E}[G]>\mu_1} \mathrm{KL}(F_i||G)$ denotes a 'disambiguation difficulty'
- Exactly matches the lower bound for regret (Burnetas & Katehakis '96)
- TS regret proofs typically follow pre-convergence and post-convergence analysis for the conjugate (Beta, Dirichlet etc.) parameters
- For non-parametric TS, 'convergence' is of the empirical distribution of the loss, in the sense of the Lévy distance

TS for minimizing CVaR regret (Baudry et. al. 2021)

For bounded distributions over $[0, B]$

Input Horizon n, number of arms K, bound B.
Initialize $t = 1$, $X_k = B$, $N_k = 1$ for each $k \in [K]$

- Sample uniformly over prob. simplex of dimension N_k →

For $t = 2, \ldots, n$ **do**

For $k = 1$ **to** K
- Sample $p_k \sim \mathrm{Unif}(\mathrm{Simp}_{N_k})$

$I(t) = \arg\max_{k \in [K]} C_\alpha(X_k, p_k)$

Play arm $I(t)$

Observe loss $r_t^{I(t)}$.

Update $N_{I(t)} = N_{I(t)} + 1$, $X_{I(t)} = \begin{pmatrix} X_{I(t)} \\ r_t^{I(t)} \end{pmatrix}$

TS for minimizing CVaR regret (Baudry et. al. 2021)

For bounded distributions over $[0, B]$

Input Horizon n, number of arms K, bound B.
Initialize $t = 1$, $X_k = B$, $N_k = 1$ for each $k \in [K]$

- Sample uniformly over prob. simplex of dimension N_k →
- Arm index →

For $t = 2, \ldots, n$ **do**

For $k = 1$ **to** K
 - Sample $p_k \sim \mathrm{Unif}(\mathrm{Simp}_{N_k})$

$I(t) = \arg\max_{k \in [K]} C_\alpha(X_k, p_k)$

Play arm $I(t)$

Observe loss $r_t^{I(t)}$.

Update $N_{I(t)} = N_{I(t)} + 1$, $X_{I(t)} = \begin{pmatrix} X_{I(t)} \\ r_t^{I(t)} \end{pmatrix}$

TS for minimizing CVaR regret (Baudry et. al. 2021)

For bounded distributions over $[0, B]$

Input Horizon n, number of arms K, bound B.
Initialize $t = 1$, $X_k = B$, $N_k = 1$ for each $k \in [K]$

For $t = 2, \ldots, n$ **do**

- Sample uniformly over prob. simplex of dimension N_k → For $k = 1$ **to** K - Sample $p_k \sim \mathrm{Unif}(\mathrm{Simp}_{N_k})$
- Arm index → $I(t) = \arg\max_{k \in [K]} C_\alpha(X_k, p_k)$
- Action and loss at time t → Play arm $I(t)$

Observe loss $r_t^{I(t)}$.

Update $N_{I(t)} = N_{I(t)} + 1$, $X_{I(t)} = \begin{pmatrix} X_{I(t)} \\ r_t^{I(t)} \end{pmatrix}$

TS for minimizing CVaR regret (Baudry et. al. 2021)

For bounded distributions over $[0, B]$

Input Horizon n, number of arms K, bound B.
Initialize $t = 1, X_k = B, N_k = 1$ for each $k \in [K]$

For $t = 2, \ldots, n$ **do**

- Sample uniformly over prob. simplex of dimension N_k → For $k = 1$ **to** K - Sample $p_k \sim \mathrm{Unif}(\mathrm{Simp}_{N_k})$
- Arm index → $I(t) = \arg\max_{k \in [K]} C_\alpha(X_k, p_k)$
- Action and loss at time t → Play arm $I(t)$

Observe loss $r_t^{I(t)}$.

- Update of loss vector X_k and number of pulls N_k → Update $N_{I(t)} = N_{I(t)} + 1$, $X_{I(t)} = \begin{pmatrix} X_{I(t)} \\ r_t^{I(t)} \end{pmatrix}$

CVaR Regret Optimality of TS (Baudry et.al. '21)

Non-parametric TS achieves the optimal regret bound for bounded loss distributions:

$$\mathbb{E}(R_n) \leq \sum_{\{i:\Delta_i^\alpha > 0\}} \frac{\Delta_i^\alpha \log n}{\mathcal{K}_{\inf}^\alpha(F_i, c_1^\alpha)} + o(\log n)$$

- $K_{\inf}^\alpha(F_i, c_1^\alpha) = \inf_{G:\mathrm{CVaR}_\alpha(G) > c_1^\alpha} \mathrm{KL}(F_i || G)$, and Δ_i^α are the CVaR gaps
- Exactly matches the lower bound for CVaR regret

TS algorithm for mean-variance learning

For Gaussian arms $\mathcal{N}(\mu_i, \sigma_i^2)$

Input Horizon n, number of arms K.

Initialize $\hat{\mu}_{i,0} = 0$ $T_{i,0} = 0, \alpha_{i,0} = 1/2, \beta_{i,0} = 1/2$, for each $i \in [K]$

- Precision sample from Gamma, mean sample from Gaussian

For $t = 1, 2, \ldots, n$ **do**

For $i = 1$ **to** K

Sample $\tau_i(t)$ from $\mathrm{Gamma}(\alpha_{i,t-1}, \beta_{i,t-1})$

Sample $\theta_i(t)$ from $\mathcal{N}(\hat{\mu}_{i,t-1}, 1/(T_{i,t-1} + 1))$

$$I(t) = \arg\min_{i \in [K]} \{\rho\theta_i(t) + 1/\tau_i(t)\}$$

Play arm $I(t)$

Observe loss $X_{I(t),t}$

$$(\hat{\mu}_{I(t),t}, T_{I(t),t}, \alpha_{I(t),t}, \beta_{I(t),t}) = \mathrm{Update}(\hat{\mu}_{I(t),t-1}, T_{I(t),t-1}, \alpha_{I(t),t-1}, \beta_{I(t),t-1}, X_{I(t),t})$$

TS algorithm for mean-variance learning

For Gaussian arms $\mathcal{N}(\mu_i, \sigma_i^2)$

Input Horizon n, number of arms K.
Initialize $\hat{\mu}_{i,0} = 0$ $T_{i,0} = 0, \alpha_{i,0} = 1/2, \beta_{i,0} = 1/2$, for each $i \in [K]$

For $t = 1, 2, \ldots, n$ **do**

- Precision sample from Gamma, mean sample from Gaussian

For $i = 1$ **to** K
Sample $\tau_i(t)$ from $\mathrm{Gamma}(\alpha_{i,t-1}, \beta_{i,t-1})$
Sample $\theta_i(t)$ from $\mathcal{N}(\hat{\mu}_{i,t-1}, 1/(T_{i,t-1}+1))$

- Arm index

$$I(t) = \operatorname*{arg\,min}_{i \in [K]} \{\rho\theta_i(t) + 1/\tau_i(t)\}$$

Play arm $I(t)$

Observe loss $X_{i(t),t}$

$$(\hat{\mu}_{i(t),t}, T_{i(t),t}, \alpha_{i(t),t}, \beta_{i(t),t}) = \mathrm{Update}(\hat{\mu}_{i(t),t-1}, T_{i(t),t-1}, \alpha_{i(t),t-1}, \beta_{i(t),t-1}, X_{i(t),t})$$

TS algorithm for mean-variance learning

For Gaussian arms $\mathcal{N}(\mu_i, \sigma_i^2)$

Input Horizon n, number of arms K.
Initialize $\hat{\mu}_{i,0} = 0$ $T_{i,0} = 0, \alpha_{i,0} = 1/2, \beta_{i,0} = 1/2$, for each $i \in [K]$

- Precision sample from Gamma, mean sample from Gaussian
- Arm index
- Action and loss at time t

For $t = 1, 2, \ldots, n$ **do**

For $i = 1$ **to** K
Sample $\tau_i(t)$ from $\mathrm{Gamma}(\alpha_{i,t-1}, \beta_{i,t-1})$
Sample $\theta_i(t)$ from $\mathcal{N}(\hat{\mu}_{i,t-1}, 1/(T_{i,t-1}+1))$

$$I(t) = \arg\min_{i\in[K]} \{\rho\theta_i(t) + 1/\tau_i(t)\}$$

Play arm $I(t)$

Observe loss $X_{i(t),t}$

$$(\hat{\mu}_{i(t),t}, T_{i(t),t}, \alpha_{i(t),t}, \beta_{i(t),t}) = \mathrm{Update}(\hat{\mu}_{i(t),t-1}, T_{i(t),t-1}, \alpha_{i(t),t-1}, \beta_{i(t),t-1}, X_{i(t),t})$$

TS algorithm for mean-variance learning

For Gaussian arms $\mathcal{N}(\mu_i, \sigma_i^2)$

Input Horizon n, number of arms K.

Initialize $\hat{\mu}_{i,0} = 0$ $T_{i,0} = 0, \alpha_{i,0} = 1/2, \beta_{i,0} = 1/2$, for each $i \in [K]$

For $t = 1, 2, \ldots, n$ **do**

- Precision sample from Gamma, mean sample from Gaussian

For $i = 1$ **to** K

Sample $\tau_i(t)$ from $\mathrm{Gamma}(\alpha_{i,t-1}, \beta_{i,t-1})$

Sample $\theta_i(t)$ from $\mathcal{N}(\hat{\mu}_{i,t-1}, 1/(T_{i,t-1} + 1))$

- Arm index

$$I(t) = \arg\min_{i \in [K]} \{\rho\theta_i(t) + 1/\tau_i(t)\}$$

- Action and loss at time t

Play arm $I(t)$

Observe loss $X_{i(t),t}$

- Posterior Update

$$(\hat{\mu}_{i(t),t}, T_{i(t),t}, \alpha_{i(t),t}, \beta_{i(t),t}) = \mathrm{Update}(\hat{\mu}_{i(t),t-1}, T_{i(t),t-1}, \alpha_{i(t),t-1}, \beta_{i(t),t-1}, X_{i(t),t})$$

Risk-aware bandits: Best arm identification

Fixed Budget, Risk- Aware BAI

Aim: Identify risk-optimal arm with least probability of error in a given budget n

- L.A. et.al. (ICML 2020): CVaR concentration for sG, sE, heavy-tailed cases, CVaR-SR algorithm for least CVaR arm
- Kagrecha et.al (NeurIPS 2019): Distribution oblivious setting, truncated version of SR to minimize a convex combination $\xi \mu_i + (1-\xi) c_\alpha^i$
- Zhang & Ong (ICML 2021): Quantile-SAR to identify m-best arms with highest VaR_α

Distribution oblivious, risk-aware BAI [11]

Known # of arms K and horizon n

Unknown Distributions $F_i, i = 1, \ldots, K,$

Risk measure : $\xi\mu(i) + (1-\xi)C_\alpha(i)$ for a given ξ

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

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

Recommendation Arm J_n

Benchmark: $k^* = \underset{k=1,\ldots,K}{\arg\min}\, \xi\mu(k) + (1-\xi)C_\alpha(k).$

Goal: Minimize probability of erroneous recommendation $p_e = \mathbb{P}[J_n \neq k^*]$

[11] A. Kagrecha, J. Nair, K. Jagannathan (NeurIPS '19)

Distribution oblivious, risk-aware BAI [12]

- Distribution oblivious: Nothing is known about the arm distributions
- Could be heavy tailed:
  $\mathbb{E}[X_i^p] < B$ for some $p > 1$, but B, p not known!

[12] A. Kagrecha, J. Nair, K. Jagannathan (NeurIPS '19)

Distribution oblivious, risk-aware BAI [12]

- Distribution oblivious: Nothing is known about the arm distributions
- Could be heavy tailed:
  $\mathbb{E}[X_i^p] < B$ for some $p > 1$, but B, p not known!
- Identify arm with the least $\xi\mu_i + (1-\xi)C_\alpha^i$,
- Here ξ decides the tradeoff between desire of average reward and risk appetite

[12] A. Kagrecha, J. Nair, K. Jagannathan (NeurIPS '19)

Distribution oblivious, risk-aware BAI

- Key challenge: empirical estimators for mean and CVaR lead to poor performance
- Key idea: work with projected samples $X_i^{(b)} = \min(\max(-b, X_i), b)$ to form mean and CVaR estimates
- Algorithm: Use SR or Uniform Exploration with projected samples, with $b = n^q$ for $q \in (0, 1)$

Distribution oblivious, risk-aware BAI

- Key challenge: empirical estimators for mean and CVaR lead to poor performance
- Key idea: work with projected samples $X_i^{(b)} = \min(\max(-b, X_i), b)$ to form mean and CVaR estimates
- Algorithm: Use SR or Uniform Exploration with projected samples, with $b = n^q$ for $q \in (0, 1)$
- Guarantee: $p_e = O(\exp(-\gamma n^{1-q}))$ for $n > Kn^*(q)$

Distribution oblivious, risk-aware BAI

- Key challenge: empirical estimators for mean and CVaR lead to poor performance
- Key idea: work with projected samples $X_i^{(b)} = \min(\max(-b, X_i), b)$ to form mean and CVaR estimates
- Algorithm: Use SR or Uniform Exploration with projected samples, with $b = n^q$ for $q \in (0, 1)$
- Guarantee: $p_e = O(\exp(-\gamma n^{1-q}))$ for $n > Kn^*(q)$
- A smaller q implies better asymptotic decay, but finite sample performance could be poorer

Distribution oblivious, risk-aware BAI

- Key challenge: empirical estimators for mean and CVaR lead to poor performance
- Key idea: work with projected samples $X_i^{(b)} = \min(\max(-b, X_i), b)$ to form mean and CVaR estimates
- Algorithm: Use SR or Uniform Exploration with projected samples, with $b = n^q$ for $q \in (0, 1)$
- Guarantee: $p_e = O(\exp(-\gamma n^{1-q}))$ for $n > Kn^*(q)$
- A smaller q implies better asymptotic decay, but finite sample performance could be poorer
- Lower bound (Kagrecha et.al., IEEE Trans. Info. Th. 2022): no consistent estimator can obtain exponential decay of p_e

Fixed Confidence, Risk- Aware BAI

Identify best arm(s) with high probability $1 - \delta$ with least expected sample complexity

- Szorenyi et.al. '15, David & Shimkin '16: PAC best-arm identification for VaR_α
- David et.al '18: find arm with best mean reward, subject to VaR_α risk constraint

- Acerbi, C. (2002). "Spectral measures of risk: A coherent representation of subjective risk aversion". In: *Journal of Banking & Finance* 26.7, pp. 1505–1518.
- Artzner, P. et al. (1999). "Coherent measures of risk". In: *Mathematical Finance* 9.3, pp. 203–228.
- Baudry, D. et al. (July 2021). "Optimal Thompson Sampling strategies for support-aware CVaR bandits". In: *Proceedings of the 38th International Conference on Machine Learning*. Vol. 139. PMLR, pp. 716–726.
- Bhat, S. P. and L. A. Prashanth (2019). "Concentration of risk measures: A Wasserstein distance approach". In: *Advances in Neural Information Processing Systems*, pp. 11739–11748.

Brown, D. B. (2007). "Large deviations bounds for estimating conditional value-at-risk". In: *Operations Research Letters* 35.6, pp. 722–730.

Cassel, A., S. Mannor, and A. Zeevi (2018). "A general approach to multi-armed bandits under risk criteria". In: *Proceedings of the 31st Conference On Learning Theory*, pp. 1295–1306.

Chang, J. Q. L. and V. Y. F. Tan (2022). "A Unifying Theory of Thompson Sampling for Continuous Risk-Averse Bandits". In: *Proc. of the 36th AAAI Conference on Artificial Intelligence*. AAAI Press.

David, Y. and N. Shimkin (2016). "Pure exploration for max-quantile bandits". In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, pp. 556–571.

Dembo, A. and O. Zeitouni (2009). *Large Deviations Techniques and Applications*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg. ISBN: 9783642033117. URL: *https://books.google.com.sg/books?id=iT9JRlGPx5gC*.

Föllmer, H. and A. Schied (2002). "Convex measures of risk and trading constraints". In: *Finance and Stochastics* 6.4, pp. 429–447.

- Gopalan, A. et al. (2017). "Weighted bandits or: How bandits learn distorted values that are not expected". In: *AAAI Conference on Artificial Intelligence*, pp. 1941–1947.
- Hou, Y., V. Y. F. Tan, and Z. Zhong (2022). *Almost Optimal Variance-Constrained Best Arm Identification*. arXiv 2201.10142. arXiv: `2201.10142 [cs.LG]`.
- Hu, Z. and D. Zhang (2018). "Utility-based shortfall risk: Efficient computations via Monte Carlo". In: *Naval Research Logistics (NRL)* 65.5, pp. 378–392.
- Jie, C. et al. (2018). "Stochastic Optimization in a Cumulative Prospect Theory Framework". In: *IEEE Transactions on Automatic Control* 63.9, pp. 2867–2882.

- Kagrecha, A., J. Nair, and K. Jagannathan (2019). "Distribution oblivious, risk-aware algorithms for multi-armed bandits with unbounded rewards". In: *Advances in Neural Information Processing Systems*, pp. 11269–11278.
- Kolla, R. K. et al. (2019). "Concentration bounds for empirical conditional value-at-risk: The unbounded case". In: *Operations Research Letters* 47.1, pp. 16–20.
- Markowitz, H. (1952). "Portfolio selection". In: *The Journal of Finance* 7.1, pp. 77–91.

- Pandey, A. K., L. A. Prashanth, and S. P. Bhat (May 2021). "Estimation of Spectral Risk Measures". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 35.13, pp. 12166–12173.
- Prashanth, L. A. and S. P. Bhat (2022). "A Wasserstein distance approach for concentration of empirical risk estimates". In: *arXiv preprint 1902.10709v4*.
- Prashanth, L. A., J. Cheng, et al. (2016). "Cumulative prospect theory meets reinforcement learning: prediction and control". In: *International Conference on Machine Learning*, pp. 1406–1415.

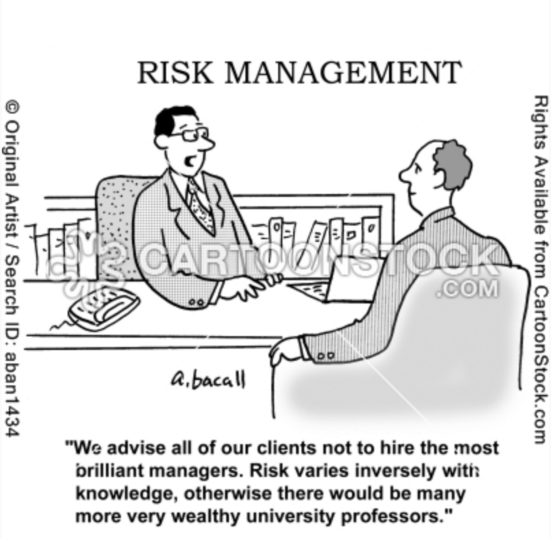
- Prashanth, L. A., K. Jagannathan, and R. K. Kolla (2020). "Concentration bounds for CVaR estimation: The cases of light-tailed and heavy-tailed distributions". In: *International Conference on Machine Learning*. Vol. 119. PMLR, pp. 5577–5586.
- Rockafellar, R. T. and S. Uryasev (2000). "Optimization of conditional value-at-risk". In: *Journal of Risk* 2.3, pp. 21–41.
- Sani, A., A. Lazaric, and R. Munos (2012). "Risk-aversion in multi-armed bandits". In: *Advances in Neural Information Processing Systems*, pp. 3275–3283.

- Szorenyi, B. et al. (2015). "Qualitative multi-armed bandits: A quantile-based approach". In: *International Conference on Machine Learning*. PMLR, pp. 1660–1668.
- Thomas, P. and E. Learned-Miller (2019). "Concentration Inequalities for Conditional Value at Risk". In: *International Conference on Machine Learning*, pp. 6225–6233.
- Tversky, Amos and Daniel Kahneman (1992). "Advances in prospect theory: Cumulative representation of uncertainty". In: *Journal of Risk and Uncertainty* 5.4, pp. 297–323.

- Vakili, S and Q. Zhao (2015). "Mean-variance and value at risk in multi-armed bandit problems". In: *Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1330–1335.
- Wang, Y. and F. Gao (2010). "Deviation inequalities for an estimator of the conditional value-at-risk". In: *Operations Research Letters* 38.3, pp. 236–239.
- Zhang, M. and C. S. Ong (2021). "Quantile Bandits for Best Arms Identification". In: *International Conference on Machine Learning*. PMLR, pp. 12513–12523.

Zhu, Q. and V. Y. F. Tan (2020). "Thompson Sampling Algorithms for Mean-Variance Bandits". In: *International Conference on Machine Learning*, pp. 2645–2654.

Will I make money?



Thank you