Ordinal optimization - Empirical large deviations rate estimators, and multi-armed bandit methods

Sandeep Juneja

Tata Institute of Fundamental Research
Mumbai, India

joint work with Peter Glynn

Applied Probability Frontiers - Banff event

June 1, 2015

The ordinal optimization problem

- d different designs are compared on the basis of random performance measures $X(i), i \leq d$, and the goal is to identify

$$i^* = \arg \min_{1 \leq j \leq d} EX(j).$$

The ordinal optimization problem

- d different designs are compared on the basis of random performance measures $X(i), i \leq d$, and the goal is to identify

$$i^* = \arg\min_{1 \leq j \leq d} EX(j).$$

- Goal is only to identify the best design and not to actually estimate the performance.

The ordinal optimization problem

- d different designs are compared on the basis of random performance measures $X(i), i \leq d$, and the goal is to identify

$$i^* = \arg \min_{1 \leq j \leq d} EX(j).$$

- Goal is only to identify the best design and not to actually estimate the performance.

- We have the ability to generate iid realizations of each of the d random variables.

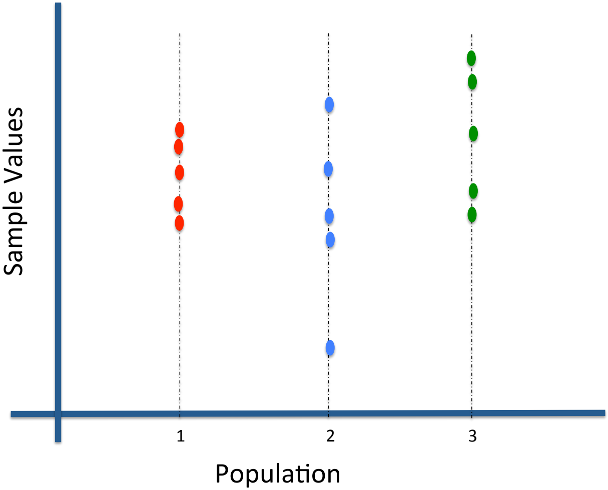
The ordinal optimization problem

- d different designs are compared on the basis of random performance measures $X(i), i \leq d$, and the goal is to identify

$$i^* = \arg \min_{1 \leq j \leq d} EX(j).$$

- Goal is only to identify the best design and not to actually estimate the performance.
- We have the ability to generate iid realizations of each of the d random variables.
- We focus primarily on $d = 2$, so given independent samples of X we want to find if $EX > 0$ or $EX < 0$.

Determining the smallest mean population



Talk Overview

- Estimating difference of mean values relies on central limit theorem with an associated slow $n^{-1/2}$ convergence rate.

Talk Overview

- Estimating difference of mean values relies on central limit theorem with an associated slow $n^{-1/2}$ convergence rate.
- Ho et al. (1990) observed that identifying the best system typically has a faster convergence rate.

Talk Overview

- Estimating difference of mean values relies on central limit theorem with an associated slow $n^{-1/2}$ convergence rate.

- Ho et al. (1990) observed that identifying the best system typically has a faster convergence rate.

$$\text{Finding } EX < EY, \text{ or } EX > EY$$

easier than determining the value

$$EX - EY.$$

- L. Dai (1996) showed using large deviation methods that the probability of false selection decays at an exponential rate under mild light tailed assumptions. Thus, if

$$EX_1 < \min_{i \geq 2} EX_i$$

then

$$\lim_{n \to \infty} \frac{1}{n} \log P(\bar{X}_1(n) > \min_{i \geq 2} \bar{X}_i(n)) = I$$

for large deviations rate function $I > 0$.

- L. Dai (1996) showed using large deviation methods that the probability of false selection decays at an exponential rate under mild light tailed assumptions. Thus, if

$$EX_1 < \min_{i \geq 2} EX_i$$

then

$$\lim_{n \to \infty} \frac{1}{n} \log P(\bar{X}_1(n) > \min_{i \geq 2} \bar{X}_i(n)) = I$$

for large deviations rate function $I > 0$.

- In a series of papers by Chen, Yucesan, L Dai, Chick and others (2000, 200*) attempted to optimize the budget allocated to each population under Normality assumption.

- L. Dai (1996) showed using large deviation methods that the probability of false selection decays at an exponential rate under mild light tailed assumptions. Thus, if

$$EX_1 < \min_{i \geq 2} EX_i$$

then

$$\lim_{n \to \infty} \frac{1}{n} \log P(\bar{X}_1(n) > \min_{i \geq 2} \bar{X}_i(n)) = I$$

for large deviations rate function $I > 0$.

- In a series of papers by Chen, Yucesan, L Dai, Chick and others (2000, 200*) attempted to optimize the budget allocated to each population under Normality assumption.

- Substantial literature on selecting the best system amongst many alternatives using ranking/selection procedures. Gaussian assumption is critical to most of the analysis (Nelson and others)

- Glynn and J (2004) observed that if

$$EX_1 < \min_{i \geq 2} EX_i$$

then for $p_i > 0$, $\sum_{i=1}^{d} p_i = 1$

$$P(\bar{X}_1(p_1 n) > \min_{i \geq 2} \bar{X}_i(p_i n)) \approx e^{-nH(p_1,\ldots,p_d)}$$

so that $H(p_1, \ldots, p_d)$ can be optimised to determine optimal allocations.

- Glynn and J (2004) observed that if

$$EX_1 < \min_{i \geq 2} EX_i$$

then for $p_i > 0$, $\sum_{i=1}^{d} p_i = 1$

$$P(\bar{X}_1(p_1 n) > \min_{i \geq 2} \bar{X}_i(p_i n)) \approx e^{-nH(p_1,\ldots,p_d)}$$

so that $H(p_1, \ldots, p_d)$ can be optimised to determine optimal allocations.

- Significant literature since then relying on large deviations analysis (E.g., Hunter and Pasupathy 2013, Szechtman and Yucesan 2008, Broadie, Han Zeevi 2007, Blanchet, Liu, Zwart 2008).

Some observations

- If $P(FS) \leq e^{-nI}$, for some $I > 0$, then

$$n = \frac{1}{I}\log(1/\delta)) \text{ ensures } P(FS) \leq \delta.$$

Some observations

- If $P(FS) \leq e^{-nI}$, for some $I > 0$, then

$$n = \frac{1}{I}\log(1/\delta)) \text{ ensures } P(FS) \leq \delta.$$

- $O(\log(1/\delta))$ effort is necessary. If $\log(1/\delta)^{1-\epsilon}$ samples are generated, then

$$P(X_i \in A)^{\log(1/\delta)^{1-\epsilon}} = \delta^{\frac{\text{positive no.}}{\log(1/\delta)^{\epsilon}}} >> \delta$$

as $\delta \to 0$.

HOPE

- However methods relying on $P(FS) \leq e^{-nI}$, rely on estimating the large deviations rate function I from the samples generated.

HOPE

- However methods relying on $P(FS) \leq e^{-nI}$, rely on estimating the large deviations rate function I from the samples generated.

- One hopes for algorihms that for $n = O(\log(1/\delta))$ ensure that at least asymptotically $P(FS) \leq \delta$, that is,

$$\limsup_{\delta \to 0} P(FS)\delta^{-1} \leq 1$$

even when the means are separated by a fixed and known $\epsilon > 0$. So that

$$\min_{1 \leq i \leq d} EX_i < EX_j - \epsilon$$

for all suboptimal j.

Contributions

- We argue through two popular implementations that these rate functions are difficult to estimate accurately using $O(\log(1/\delta))$ samples

Contributions

- We argue through two popular implementations that these rate functions are difficult to estimate accurately using $O(\log(1/\delta))$ samples

- Enroute, we identify the large deviations rate function of the empirically estimated rate function. This may be of independent interest.

Key negative result

- Given any (ϵ, δ) algorithm - one that correctly separates designs with mean difference at least ϵ with

$$\limsup_{\delta \to 0} P(FS)\delta^{-1} \leq 1,$$

Key negative result

- Given any (ϵ, δ) algorithm - one that correctly separates designs with mean difference at least ϵ with

$$\limsup_{\delta \to 0} P(FS)\delta^{-1} \leq 1,$$

- We prove that for populations with unbounded support under mild restrictions, the expected number of samples cannot be $O(\log(1/\delta))$.

Positive contributions

- Under explicitly available moment upper bounds, we develop truncation based $O(\log(1/\delta))$ computation time (ϵ, δ) algorithms.

Positive contributions

- Under explicitly available moment upper bounds, we develop truncation based $O(\log(1/\delta))$ computation time (ϵ, δ) algorithms.

- We also adapt the recently proposed sequential algorithms in multi-armed bandit regret setting to this *pure exploration setting*.

Two phase ordinal optimisation implementation

Basic large deviations theory

- Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. samples of X and $a > EX$.

Basic large deviations theory

- Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. samples of X and $a > EX$.
- Then, for $\theta > 0$, Cramer's bound

$$P(\bar{X}_n \geq a) \leq e^{-n(\theta a - \Lambda(\theta))}$$

where $\Lambda(\theta) = \log Ee^{\theta X}$.

Basic large deviations theory

- Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. samples of X and $a > EX$.
- Then, for $\theta > 0$, Cramer's bound

$$P(\bar{X}_n \geq a) \leq e^{-n(\theta a - \Lambda(\theta))}$$

where $\Lambda(\theta) = \log Ee^{\theta X}$.

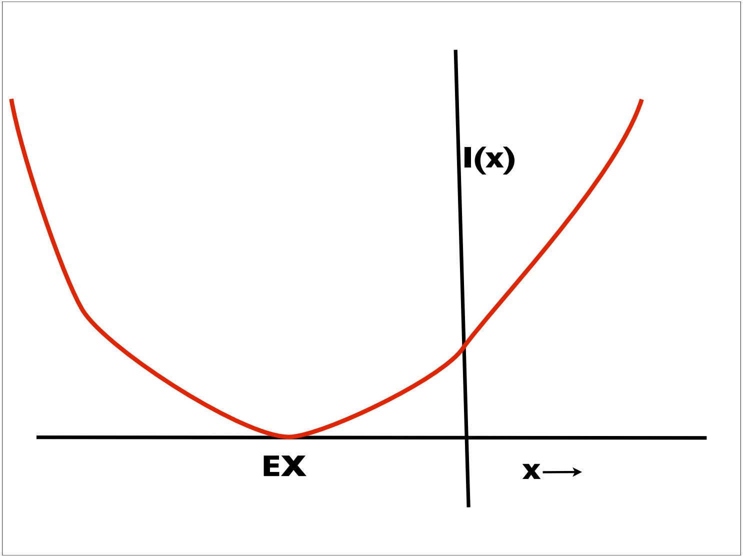
- Cramer's Theorem

$$P(\bar{X}_n \geq a) = e^{-nI(a)(1+o(1))}$$

where, the large deviations rate function

$$I(a) = \sup_{\theta \in \Re} \left(\theta a - \Lambda(\theta)\right).$$

The rate function



A simple setting

- Consider a single rv X with unknown mean EX. Need to decide whether $EX > 0$ or $EX < 0$ with error probability $\leq \delta$.

A simple setting

- Consider a single rv X with unknown mean EX. Need to decide whether $EX > 0$ or $EX < 0$ with error probability $\leq \delta$.

- Then if $EX < 0$, we may take

$$\exp(-nI(0))$$

as a proxy for $P(\bar{X}_n \geq 0)$, the probability of false selection.

A simple setting

- Consider a single rv X with unknown mean EX. Need to decide whether $EX > 0$ or $EX < 0$ with error probability $\leq \delta$.

- Then if $EX < 0$, we may take

$$\exp(-nI(0))$$

as a proxy for $P(\bar{X}_n \geq 0)$, the probability of false selection.

- If $EX > 0$, we may again take

$$\exp(-nI(0))$$

as a proxy for $P(\bar{X}_n \leq 0)$, the probability of false selection.

A two phase implementation

- Thus, $\frac{\log(1/\delta)}{I(0)}$ samples ensure that $P(FS) \leq \delta$.

A two phase implementation

- Thus, $\frac{\log(1/\delta)}{I(0)}$ samples ensure that $P(FS) \leq \delta$.
- Hence, one reasonable estimation procedure is

A two phase implementation

- Thus, $\frac{\log(1/\delta)}{I(0)}$ samples ensure that $P(FS) \leq \delta$.
- Hence, one reasonable estimation procedure is
  - First phase - Generate $m = \log(1/\delta)$ samples to estimate $I(0)$ by $\hat{I}_m(0)$.

A two phase implementation

- Thus, $\frac{\log(1/\delta)}{I(0)}$ samples ensure that $P(FS) \leq \delta$.
- Hence, one reasonable estimation procedure is
  - First phase - Generate $m = \log(1/\delta)$ samples to estimate $I(0)$ by $\hat{I}_m(0)$.
  - Second phase - Generate

$$\log(1/\delta)/\hat{I}_m(0) = m/\hat{I}_m(0) \triangleq n$$

samples of X.

A two phase implementation

- Thus, $\frac{\log(1/\delta)}{I(0)}$ samples ensure that $P(FS) \leq \delta$.
- Hence, one reasonable estimation procedure is
  - First phase - Generate $m = \log(1/\delta)$ samples to estimate $I(0)$ by $\hat{I}_m(0)$.
  - Second phase - Generate

    $$\log(1/\delta)/\hat{I}_m(0) = m/\hat{I}_m(0) \triangleq n$$

    samples of X.
  - Decide the sign of EX based on whether $\bar{X}_n > 0$ or $\bar{X}_n \leq 0$.

A two phase implementation

- Thus, $\frac{\log(1/\delta)}{I(0)}$ samples ensure that $P(FS) \leq \delta$.
- Hence, one reasonable estimation procedure is
  - First phase - Generate $m = \log(1/\delta)$ samples to estimate $I(0)$ by $\hat{I}_m(0)$.
  - Second phase - Generate

$$\log(1/\delta)/\hat{I}_m(0) = m/\hat{I}_m(0) \triangleq n$$

    samples of X.
  - Decide the sign of EX based on whether $\bar{X}_n > 0$ or $\bar{X}_n \leq 0$.
- We now discuss estimation of $I(0)$.

Estimating rate function

Graphic view of I(0)

- The log-moment generating function of X

$$\Lambda(\theta) = \log E \exp(\theta X)$$

is convex with $\Lambda(0) = 0$ and $\Lambda'(0) = EX$.

Graphic view of I(0)

- The log-moment generating function of X

$$\Lambda(\theta) = \log E \exp(\theta X)$$

  is convex with $\Lambda(0) = 0$ and $\Lambda'(0) = EX$.

- Then, $I(0) = -\inf_\theta \Lambda(\theta)$.

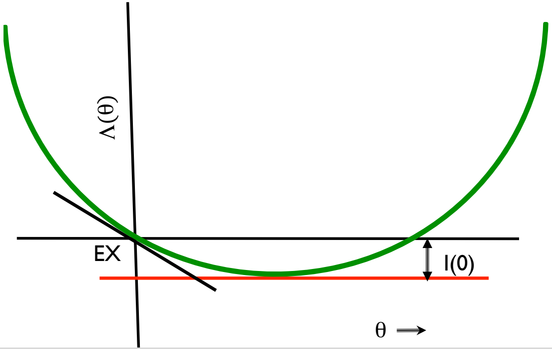
Graphic view of I(0)

- The log-moment generating function of X

$$\Lambda(\theta) = \log E \exp(\theta X)$$

  is convex with $\Lambda(0) = 0$ and $\Lambda'(0) = EX$.

- Then, $I(0) = -\inf_\theta \Lambda(\theta)$.



Estimating rate function $I(0)$

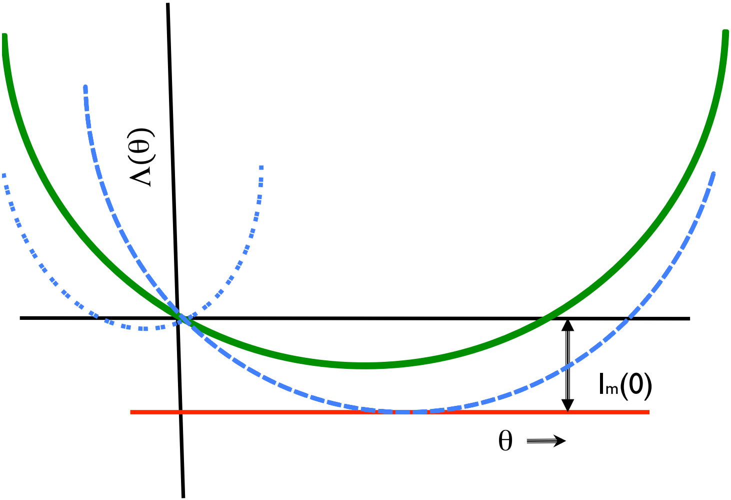
- A natural estimator for $I(0)$ based on samples $(X_i : 1 \leq i \leq m)$ is

$$\hat{I}_m(0) = -\inf_{\theta \in \Re} \hat{\Lambda}_m(\theta)$$

where

$$\hat{\Lambda}_m(\theta) = \log\left(\frac{1}{m}\sum_{i=1}^{m} \exp(\theta X_i)\right)$$

Graphic view of estimated log moment generating function



Large deviations rate function of $\hat{I}_m(0)$

- Can show that for $a > I(0)$,

$$\lim_{m\to\infty} \frac{1}{m} \log P(\hat{I}_m(0) \geq a), \text{ equals}$$

Large deviations rate function of $\hat{I}_m(0)$

- Can show that for $a > I(0)$,

$$\lim_{m\to\infty} \frac{1}{m} \log P(\hat{I}_m(0) \geq a), \text{ equals}$$

$$\lim_{m\to\infty} \frac{1}{m} \log P\left(\inf_{\theta} \frac{1}{m} \sum_{i=1}^{m} e^{\theta X_i} \leq e^{-a} \right) = - \inf_{\theta \in \Re} \mathcal{I}_\theta(e^{-a}),$$

where

$$\mathcal{I}_\theta(\nu) = \sup_{\alpha}(\alpha\nu - \log E \exp(\alpha e^{\theta X})).$$

Large deviations rate function of $\hat{I}_m(0)$

- Can show that for $a > I(0)$,

$$\lim_{m\to\infty} \frac{1}{m} \log P(\hat{I}_m(0) \geq a), \text{ equals}$$

$$\lim_{m\to\infty} \frac{1}{m} \log P\left(\inf_{\theta} \frac{1}{m} \sum_{i=1}^{m} e^{\theta X_i} \leq e^{-a}\right) = -\inf_{\theta\in\Re} \mathcal{I}_\theta(e^{-a}),$$

where

$$\mathcal{I}_\theta(\nu) = \sup_{\alpha}(\alpha\nu - \log E \exp(\alpha e^{\theta X})).$$

- Further, for $a < I(0)$,

$$\lim_{m\to\infty} \frac{1}{m} \log P(\hat{I}_m(0) \leq a) = -\sup_{\theta\in\Re} \mathcal{I}_\theta(e^{-a})$$

- Also, the natural estimator for $I(x)$ is

$$\hat{I}_m(x) = \sup_{\theta \in \Re} (\theta x - \hat{\Lambda}_m(\theta)).$$

- Also, the natural estimator for $I(x)$ is

$$\hat{I}_m(x) = \sup_{\theta \in \Re}(\theta x - \hat{\Lambda}_m(\theta)).$$

- it can be seen that for $a > I(x)$, the rate function for $\hat{I}_m(x)$, equals

$$\inf_{\theta \in \Re} \mathcal{I}_\theta(e^{-a+\theta x})$$

for $a > I(x)$, and equals

$$\sup_{\theta \in \Re} \mathcal{I}_\theta(e^{-a+\theta x})$$

$a < I(x)$.

Negative Result 1

Failure of the Naive

Returning to two phase procedure

- We generate samples $X_1, \ldots, X_m$ for $m = \log(1/\delta)$ and set

$$\hat{I}_m(0) = -\inf_{\theta} \hat{\Lambda}_m(\theta).$$

Returning to two phase procedure

- We generate samples $X_1, \ldots, X_m$ for $m = \log(1/\delta)$ and set

$$\hat{I}_m(0) = -\inf_{\theta} \hat{\Lambda}_m(\theta).$$

- Then generate $\log(1/\delta)/\hat{I}_m(0) = m/\hat{I}_m(0)$ samples of X in the second phase.

Returning to two phase procedure

- We generate samples $X_1, \ldots, X_m$ for $m = \log(1/\delta)$ and set

$$\hat{I}_m(0) = -\inf_{\theta} \hat{\Lambda}_m(\theta).$$

- Then generate $\log(1/\delta)/\hat{I}_m(0) = m/\hat{I}_m(0)$ samples of X in the second phase.

$$P(FS) \approx E \exp\left(-\frac{m}{\hat{I}_m(0)} I(0)\right)$$

- Errors due to large values of $\hat{I}_m(0)$ that lead to under sampling in second phase - Due to conspiratorial large deviations behaviour of all the terms.

Lower Bound for P(FS)

- Can show

$$\lim_{m} \frac{1}{m} \log P(FS) = \sup_{a>0} \sup_{\theta} \left(-\frac{I(0)}{a} - \mathcal{I}_{\theta}(e^{-a}) \right).$$

Lower Bound for P(FS)

- Can show

$$\lim_{m} \frac{1}{m} \log P(FS) = \sup_{a>0} \sup_{\theta} \left(-\frac{I(0)}{a} - \mathcal{I}_\theta(e^{-a}) \right).$$

- In particular,

$$\liminf_{\delta \to 0} P(FS)\delta^{-1} > 1.$$

Negative Result 2

Fall of the Sophisticated

Another common estimation method

- Generate $m = c\log(1/\delta)$ samples in the first phase to estimate $I(0)$ by $\hat{I}_m(0)$.

Another common estimation method

- Generate $m = c \log(1/\delta)$ samples in the first phase to estimate $I(0)$ by $\hat{I}_m(0)$.

- If $\exp(-m\hat{I}_m(0)) \leq \delta$, stop.

Another common estimation method

- Generate $m = c\log(1/\delta)$ samples in the first phase to estimate $I(0)$ by $\hat{I}_m(0)$.

- If $\exp(-m\hat{I}_m(0)) \leq \delta$, stop.

- Else, provide another $c\log(1/\delta)$ of computational budget and so on.

Another common estimation method

- Generate $m = c\log(1/\delta)$ samples in the first phase to estimate $I(0)$ by $\hat{I}_m(0)$.

- If $\exp(-m\hat{I}_m(0)) \leq \delta$, stop.

- Else, provide another $c\log(1/\delta)$ of computational budget and so on.

Can identify distributions for which this would not be accurate.

- Need to find X with $EX < 0$ so that

$$P(\bar{X}_m \geq 0 \text{ and } \exp(-m\hat{I}_m(0)) \leq \delta) > \delta.$$

- Need to find X with $EX < 0$ so that

$$P(\bar{X}_m \geq 0 \text{ and } \exp(-m\hat{I}_m(0)) \leq \delta) > \delta.$$

- This follows if for some $\theta < 0$,

$$\mathcal{I}_\theta(e^{-1/c}) < 1/c.$$

- Need to find X with $EX < 0$ so that

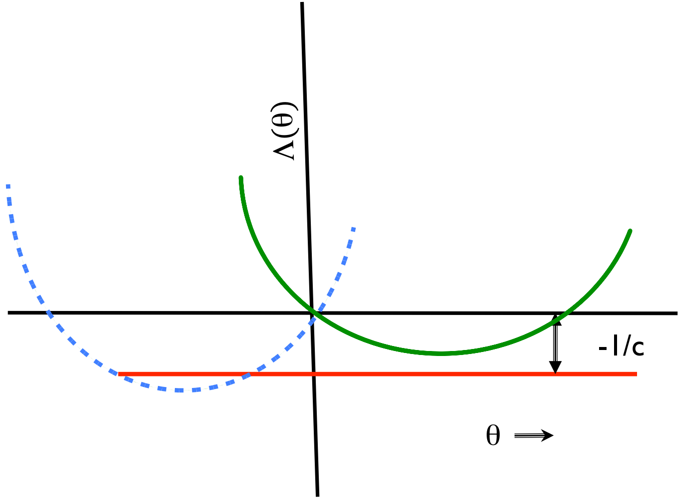
$$P(\bar{X}_m \geq 0 \text{ and } \exp(-m\hat{I}_m(0)) \leq \delta) > \delta.$$

- This follows if for some $\theta < 0$,

$$\mathcal{I}_\theta(e^{-1/c}) < 1/c.$$

- Can be shown for some light tailed distributions even for large c.

Graphic view: $I(0) < 1/c$ and $\hat{I}_m(0) > 1/c$



Negative Result 3

Perdition

- Let $\mathcal{L}$ denote a collection of probability distributions with finite mean, unbounded positive support, open under Kullback-Leibler distance.

- Let $\mathcal{L}$ denote a collection of probability distributions with finite mean, unbounded positive support, open under Kullback-Leibler distance.

- Let $\mathcal{P}(\epsilon, \delta)$ denote a policy that can adaptively sample from any two distributions in $\mathcal{L}$ with

$$\limsup_{\delta \to 0} P(FS)\delta^{-1} \leq 1.$$

- Let $\mathcal{L}$ denote a collection of probability distributions with finite mean, unbounded positive support, open under Kullback-Leibler distance.

- Let $\mathcal{P}(\epsilon, \delta)$ denote a policy that can adaptively sample from any two distributions in $\mathcal{L}$ with

$$\limsup_{\delta \to 0} P(FS)\delta^{-1} \leq 1.$$

- **Result - For any two such distributions in $\mathcal{L}$ with arbitrarily apart mean, $\mathcal{P}(\epsilon, \delta)$ policy on average requires more than $O(\log(1/\delta))$ samples.**

Details

- Under probability P_a,
  - $\{X_i\}$ has distribution F, mean μ_F,
  - $\{Y_i\}$ has distribution G, mean $\mu_G < \mu_F - \epsilon$.

Details

- Under probability P_a,
  - $\{X_i\}$ has distribution F, mean μ_F,
  - $\{Y_i\}$ has distribution G, mean $\mu_G < \mu_F - \epsilon$.
- Under probability P_b
  - $\{X_i\}$ has distribution F,
  - $\{Y_i\}$ has distribution $\tilde{G} > \mu_F + \epsilon$.

Details

- Under probability P_a,
  - $\{X_i\}$ has distribution F, mean μ_F,
  - $\{Y_i\}$ has distribution G, mean $\mu_G < \mu_F - \epsilon$.
- Under probability P_b
  - $\{X_i\}$ has distribution F,
  - $\{Y_i\}$ has distribution $\tilde{G} > \mu_F + \epsilon$.
- Under $\mathcal{P}(\epsilon, \delta)$,

$$\liminf_{\delta \to 0} \frac{E_a T_G}{\log(1/\delta)} \geq \frac{1}{3\,\mathcal{KL}(G, \tilde{G})}.$$

where $\mathcal{KL}(G, \tilde{G}) = \int_{x \in \Re} \left(\log \frac{dG}{d\tilde{G}}(x) \right) dG(x))$.

Proof similar to Lai and Robbins 1985, Mannor and Tsitsiklis 2004

Asymptotically,

$$P_a(\text{ algorithm selects F}) \geq 1 - \tilde{\delta}$$

Proof similar to Lai and Robbins 1985, Mannor and Tsitsiklis 2004

Asymptotically,

$$P_a(\text{ algorithm selects F}) \geq 1 - \tilde{\delta}$$

$$P_b(\text{ algorithm selects F}) \leq \tilde{\delta}.$$

Asymptotically,

$$P_a(\text{ algorithm selects F}) \geq 1 - \tilde{\delta}$$

$$P_b(\text{ algorithm selects F}) \leq \tilde{\delta}.$$

$$\begin{aligned} P_b(\text{ algo. selects F}) &= E_a\left(\prod_{i=1}^{T_G} \frac{d\tilde{G}}{dG}(Y_i) I(\text{ algo. selects F})\right) \\ &\approx E_a\left(e^{-T_G \times \mathcal{KL}(G,\tilde{G})} I(\text{ algo. selects F})\right) \\ &\approx\geq e^{-2E_a(T_G) \times \mathcal{KL}(G,\tilde{G})} P_a(\text{ algo. selects F}) \end{aligned}$$

and the result is easily deduced.

Result

- Given G with finite mean and unbounded positive support, for any $\alpha > 0$, and $k > \mu_G$ there exists a distribution G_k such that
$$\mathcal{KL}(G, G_k) \leq \alpha$$
and
$$\mu_{G_k} \geq k.$$

Way forward

- Additional information needed to attain $\log(1/\delta)$ convergence rates.

- Additional information needed to attain $\log(1/\delta)$ convergence rates.

- Often upper bounds on moments may be available in simulation models.

- Additional information needed to attain $\log(1/\delta)$ convergence rates.

- Often upper bounds on moments may be available in simulation models.

- Easy to develop such bounds once suitable Lyapunov functions can be identified (not to be discussed here)

- Additional information needed to attain $\log(1/\delta)$ convergence rates.

- Often upper bounds on moments may be available in simulation models.

- Easy to develop such bounds once suitable Lyapunov functions can be identified (not to be discussed here)

- Use such bounds to develop (ϵ, δ) strategies by truncating random variables while controlling the error to be less than ϵ. Then use Hoeffding.

- Additional information needed to attain $\log(1/\delta)$ convergence rates.

- Often upper bounds on moments may be available in simulation models.

- Easy to develop such bounds once suitable Lyapunov functions can be identified (not to be discussed here)

- Use such bounds to develop (ϵ, δ) strategies by truncating random variables while controlling the error to be less than ϵ. Then use Hoeffding.

- Recent multi-armed-bandits methods do this in a sequential and adaptive manner.

δ guarantees using $\log(1/\delta)$ samples: Static approach

$\mathcal{P}(\epsilon, \delta)$ policy for bounded random variables

- Consider $\mathcal{X}_\epsilon = \{X : |EX| > \epsilon, a \leq X \leq b\}$. A reasonable algorithm on $\mathcal{X}_\epsilon$ is:

$\mathcal{P}(\epsilon, \delta)$ policy for bounded random variables

- Consider $\mathcal{X}_\epsilon = \{X : |EX| > \epsilon, a \leq X \leq b\}$. A reasonable algorithm on $\mathcal{X}_\epsilon$ is:
  - Generate iid samples $X_1, X_2, \ldots, X_n$ of X.

$\mathcal{P}(\epsilon, \delta)$ policy for bounded random variables

- Consider $\mathcal{X}_\epsilon = \{X : |EX| > \epsilon, a \leq X \leq b\}$. A reasonable algorithm on $\mathcal{X}_\epsilon$ is:
  - Generate iid samples $X_1, X_2, \ldots, X_n$ of X.
  - If $\bar{X}_n \geq 0$ declare, $EX > 0$.

$\mathcal{P}(\epsilon, \delta)$ policy for bounded random variables

- Consider $\mathcal{X}_\epsilon = \{X : |EX| > \epsilon, a \leq X \leq b\}$. A reasonable algorithm on $\mathcal{X}_\epsilon$ is:
  - Generate iid samples $X_1, X_2, \ldots, X_n$ of X.
  - If $\bar{X}_n \geq 0$ declare, $EX > 0$.
  - If $\bar{X}_n < 0$, declare, $EX < 0$.

$\mathcal{P}(\epsilon, \delta)$ policy for bounded random variables

- Consider $\mathcal{X}_\epsilon = \{X : |EX| > \epsilon, a \leq X \leq b\}$. A reasonable algorithm on $\mathcal{X}_\epsilon$ is:
  - Generate iid samples $X_1, X_2, \ldots, X_n$ of X.
  - If $\bar{X}_n \geq 0$ declare, $EX > 0$.
  - If $\bar{X}_n < 0$, declare, $EX < 0$.
- Hoeffding's inequality can be used to bound probability of false selection. Suppose, $EX < -\epsilon$,

$$P(\bar{X}_n \geq 0) \leq P(\bar{X}_n - EX \geq \epsilon) \leq \exp(-2n\epsilon^2/(b-a)^2)$$

$\mathcal{P}(\epsilon, \delta)$ policy for bounded random variables

- Consider $\mathcal{X}_\epsilon = \{X : |EX| > \epsilon, a \leq X \leq b\}$. A reasonable algorithm on $\mathcal{X}_\epsilon$ is:
  - Generate iid samples $X_1, X_2, \ldots, X_n$ of X.
  - If $\bar{X}_n \geq 0$ declare, $EX > 0$.
  - If $\bar{X}_n < 0$, declare, $EX < 0$.

- Hoeffding's inequality can be used to bound probability of false selection. Suppose, $EX < -\epsilon$,

$$P(\bar{X}_n \geq 0) \leq P(\bar{X}_n - EX \geq \epsilon) \leq \exp(-2n\epsilon^2/(b-a)^2)$$

- Thus, $n = \frac{(b-a)^2}{2\epsilon^2} \log(1/\delta)$ provides the desired $\mathcal{P}(\epsilon, \delta)$ policy.

Truncation when explicit bounds on function of rv known

- Suppose f is a strictly increasing convex function and we know that $Ef(X) \leq a$. Further, $X \geq b$. Then,

Truncation when explicit bounds on function of rv known

- Suppose f is a strictly increasing convex function and we know that $Ef(X) \le a$. Further, $X \ge b$. Then,

$$\max_X EXI(X \ge u) \le u \left(\frac{a - f(b)}{f(x) - f(b)} \right).$$

Truncation when explicit bounds on function of rv known

- Suppose f is a strictly increasing convex function and we know that $Ef(X) \leq a$. Further, $X \geq b$. Then,

$$\max_X EXI(X \geq u) \leq u \left(\frac{a - f(b)}{f(x) - f(b)} \right).$$

- This follows from the optimisation problem

$$\max_X E[XI(X \geq u)]$$

$$\text{such that } Ef(X) \leq a.$$

Truncation when explicit bounds on function of rv known

- Suppose f is a strictly increasing convex function and we know that $Ef(X) \leq a$. Further, $X \geq b$. Then,

$$\max_X EXI(X \geq u) \leq u \left(\frac{a - f(b)}{f(x) - f(b)} \right).$$

- This follows from the optimisation problem

$$\max_X E[XI(X \geq u)]$$

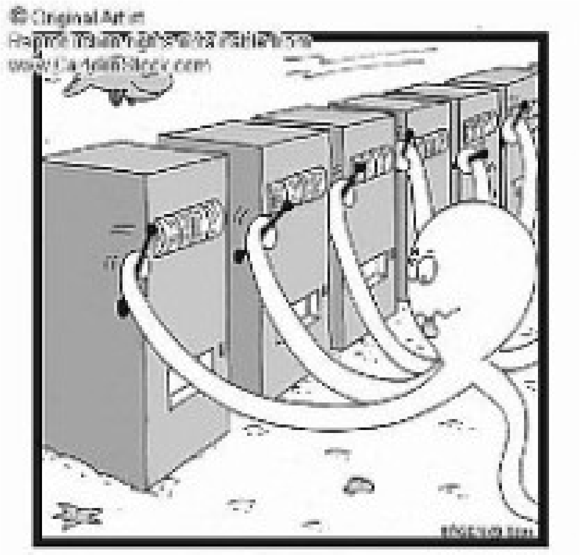
$$\text{such that } Ef(X) \leq a.$$

- This has a two point solution relying on observation that

$$Y = E[X|X < u]I(X < u) + E[X|X \geq u]I(X \geq u)$$

is better than X

Pure Exploration Multi-Armed Bandit Approach



Pure exploration bandit algorithms

- Total n arms. Each arm a when sampled gives a Bernoulli reward with mean p_a.

Pure exploration bandit algorithms

- Total n arms. Each arm a when sampled gives a Bernoulli reward with mean p_a.
- Let $a^* = \arg\max_{a \in A} p_a$ and let $\Delta_a = p_{a^*} - p_a$.

Pure exploration bandit algorithms

- Total n arms. Each arm a when sampled gives a Bernoulli reward with mean p_a.
- Let $a^* = \arg\max_{a \in A} p_a$ and let $\Delta_a = p_{a^*} - p_a$.
- Even Dar et al. 2006 devise a sequential sampling strategy to find a^* with probability at least $1 - \delta$.

Pure exploration bandit algorithms

- Total n arms. Each arm a when sampled gives a Bernoulli reward with mean p_a.
- Let $a^* = \arg\max_{a \in A} p_a$ and let $\Delta_a = p_{a^*} - p_a$.
- Even Dar et al. 2006 devise a sequential sampling strategy to find a^* with probability at least $1 - \delta$.
- Expected computational effort

$$O\left(\sum_{a \neq a^*} \frac{\ln(n/\delta)}{\Delta_a^2}\right).$$

Popular successive rejection algorithm

- Sample every arm a once and let $\hat{\mu}_a^t$ be the average reward of arm a by time t;

Popular successive rejection algorithm

- Sample every arm a once and let $\hat{\mu}_a^t$ be the average reward of arm a by time t;
- Each arm a such that

$$\hat{\mu}_{\max}^t - \hat{\mu}_a^t \geq 2\alpha_t$$

is removed from consideration. $\alpha_t = \sqrt{\log(5nt^2/\delta)/t}$;

Popular successive rejection algorithm

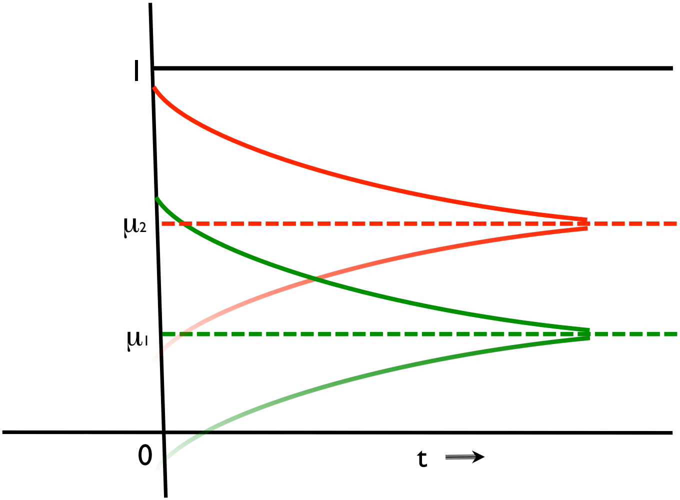
- Sample every arm a once and let $\hat{\mu}_a^t$ be the average reward of arm a by time t;
- Each arm a such that

$$\hat{\mu}_{\text{max}}^t - \hat{\mu}_a^t \geq 2\alpha_t$$

is removed from consideration. $\alpha_t = \sqrt{\log(5nt^2/\delta)/t}$;

- $t = t + 1$; Repeat till one arm left.

Key idea



Generalizing to heavy tails

- In Bubeck, Cesa-Bianchi, Lugosi 2013, they develop $\log(1/\delta)$ algorithms in regret settings when $1+\epsilon$ moments of each arm output are available.

Generalizing to heavy tails

- In Bubeck, Cesa-Bianchi, Lugosi 2013, they develop $\log(1/\delta)$ algorithms in regret settings when $1+\epsilon$ moments of each arm output are available.
- Analysis again relies on forming a cone, which they do through truncation and clever usage of Bernstein inequality.

Generalizing to heavy tails

- In Bubeck, Cesa-Bianchi, Lugosi 2013, they develop $\log(1/\delta)$ algorithms in regret settings when $1+\epsilon$ moments of each arm output are available.
- Analysis again relies on forming a cone, which they do through truncation and clever usage of Bernstein inequality.
- We adapt these algorithms to pure exploration settings.

In conclusion

- We discussed that the ordinal optimisation method relies on exponential convergence rate of the the probability of false selection

In conclusion

- We discussed that the ordinal optimisation method relies on exponential convergence rate of the the probability of false selection
- However, we show through a series of negative results that this convergence rate, or equivalently $O(\log(1/\delta))$ computation algorithms, are not possible for unbounded support distributions without further restrictions.

In conclusion

- We discussed that the ordinal optimisation method relies on exponential convergence rate of the the probability of false selection

- However, we show through a series of negative results that this convergence rate, or equivalently $O(\log(1/\delta))$ computation algorithms, are not possible for unbounded support distributions without further restrictions.

- Under explicit restrictions on moments of underlying random variables, we devise $O(\log(1/\delta))$ algorithms.

In conclusion

- We discussed that the ordinal optimisation method relies on exponential convergence rate of the the probability of false selection
- However, we show through a series of negative results that this convergence rate, or equivalently $O(\log(1/\delta))$ computation algorithms, are not possible for unbounded support distributions without further restrictions.
- Under explicit restrictions on moments of underlying random variables, we devise $O(\log(1/\delta))$ algorithms.
- These are closely related to evolving multi-arm bandit literature on pure exploration methods.