Selection of the Best System using Large Deviations, and Multi-Arm Bandits

Sandeep Juneja, TIFR

joint work with Peter Glynn, Stanford

Large Deviations Theory, ICTS

August 23, 2017

Selection of the Best System

- d different populations or probability distributions are compared.

Selection of the Best System

- d different populations or probability distributions are compared.

- Do not know the underlying distributions but can generate samples from them.

Selection of the Best System

- d different populations or probability distributions are compared.
- Do not know the underlying distributions but can generate samples from them.
- Goal is only to identify the population with smallest mean and not to actually estimate the means.

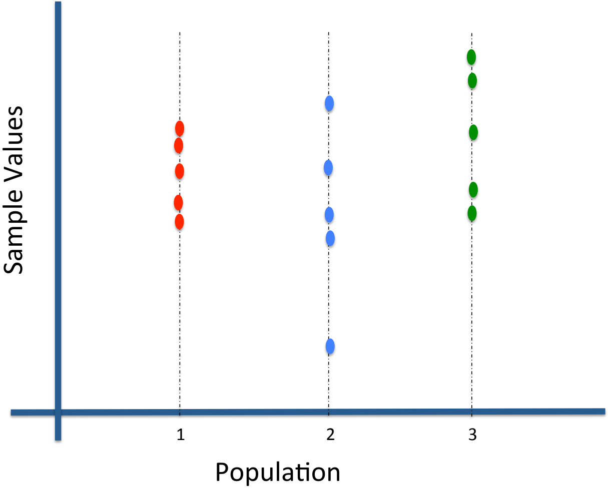
Selection of the Best System

- d different populations or probability distributions are compared.
- Do not know the underlying distributions but can generate samples from them.
- Goal is only to identify the population with smallest mean and not to actually estimate the means.
- For random variables $X(i), i \leq d$, the goal is to identify

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

in minimum number of samples while controlling the probability of false selection.

Determining the smallest mean population: Discrete stochastic optimization



Some applications

- Given stochastic models of different road network designs, finding the one with least average congestion.

Some applications

- Given stochastic models of different road network designs, finding the one with least average congestion.

- Given a manufacturing system evaluating the best maintenance strategy.

Some applications

- Given stochastic models of different road network designs, finding the one with least average congestion.

- Given a manufacturing system evaluating the best maintenance strategy.

- Given many medicinal treatments for a given disease, finding the one that causes maximum benefit on average.

Historical review

- Traditionally, this and related problems studied by statisticians, operations researchers and lately computer scientists. Some names - Bechhofer (54, 58), Rinott 78, Nelson and Goldsman (80's, 90's), Bubeck (2008 +).

Historical review

- Traditionally, this and related problems studied by statisticians, operations researchers and lately computer scientists. Some names - Bechhofer (54, 58), Rinott 78, Nelson and Goldsman (80's, 90's), Bubeck (2008 $+$).

- Underlying distributions are typically assumed to be Gaussian (asymptotically valid), Bernoulli (verifiable).

Historical review

- Traditionally, this and related problems studied by statisticians, operations researchers and lately computer scientists. Some names - Bechhofer (54, 58), Rinott 78, Nelson and Goldsman (80's, 90's), Bubeck (2008 +).
- Underlying distributions are typically assumed to be Gaussian (asymptotically valid), Bernoulli (verifiable).
- Underlying analysis relies on the central limit theorem

$$\left(\frac{X_1 + \cdots + X_n}{n} - EX_i\right)\frac{\sqrt{n}}{\sigma} \Rightarrow N(0,1)$$

- Typically, the means are assumed to be separated by $\epsilon > 0$, so that

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

for all suboptimal j.

- Typically, the means are assumed to be separated by $\epsilon > 0$, so that

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

for all suboptimal j.

- For fixed $\delta > 0$, asymptotic results:

$$\limsup_{\epsilon \to 0} P(FS) \le \delta$$

using $O(\epsilon^{-2})$ samples.

- Typically, the means are assumed to be separated by $\epsilon > 0$, so that

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

for all suboptimal j.

- For fixed $\delta > 0$, asymptotic results:

$$\limsup_{\epsilon \to 0} P(FS) \le \delta$$

using $O(\epsilon^{-2})$ samples.

- This talk focusses instead on keeping ϵ fixed and letting $\delta \to 0$.

Historical review ....

- Ho et al. (1990) observed that probability of false selection decays at an exponential rate for light-tailed distributions.

Historical review ....

- Ho et al. (1990) observed that probability of false selection decays at an exponential rate for light-tailed distributions.

- L. Dai (1996) showed using large deviation methods that if

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

then

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

for large deviations rate function $\mathcal{I} > 0$.

- 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.

HOPE

- If $P(FS) \le e^{-n\mathcal{I}}$, for some $\mathcal{I} > 0$, then

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

HOPE

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

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

- However this relies on estimating $\mathcal{I}$ from the samples generated.

HOPE

- If $P(FS) \le e^{-n\mathcal{I}}$, for some $\mathcal{I} > 0$, then
$$n = \frac{1}{\mathcal{I}} \log(1/\delta) \text{ ensures } P(FS) \le \delta.$$

- However this relies on estimating $\mathcal{I}$ from the samples generated.

- Even if one could get a lower bound on $\mathcal{I}$ in order $\log(1/\delta)$ samples, correct with probability $1 - \delta/2$, that would work.

Asymptotic HOPE

- In spirit of earlier CLT based asymptotic analysis, one hopes for algorithms 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.

Observations

- $O(\log(1/\delta))$ 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 $\delta \to 0$.

Observations

- $O(\log(1/\delta))$ 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 $\delta \to 0$.

- $O(\log(1/\delta)^{1+\epsilon})$ is sufficient

$$\delta^{-1} P(FS) \leq \delta^{-1} e^{-n\mathcal{I}} = \delta^{-1} e^{-\log(1/\delta)^{1+\epsilon}\mathcal{I}} = \delta^{\log(1/\delta)^{\epsilon}\mathcal{I}-1}$$

which goes to zero as $\delta \to 0$.

Contributions

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

Contributions

- We argue through a popular implementation 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 in these big data times.

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 (mutually absolutely continuous) with unbounded support and finite mean, the expected number of samples cannot be $O(\log(1/\delta))$.

Positive contribution

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

Positive contribution

- 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*.

Basic large deviations theory

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

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

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

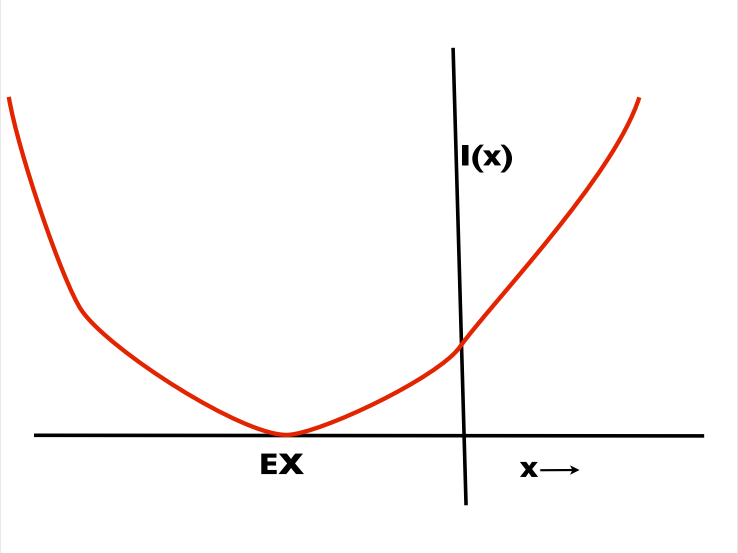
- Cramer's Theorem

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

where, the large deviations rate function

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

The rate function



A simple setting of $d = 2$

- Consider a single rv X with unknown mean EX. Need to decide whether $EX > 0$ or $EX < 0$ with error probability $\leq \delta$. Decision based on whether $\bar{X}_n > 0$ or $\bar{X}_n < 0$.

A simple setting of $d = 2$

- Consider a single rv X with unknown mean EX. Need to decide whether $EX > 0$ or $EX < 0$ with error probability $\leq \delta$. Decision based on whether $\bar{X}_n > 0$ or $\bar{X}_n < 0$.

- Then if $EX < 0$, probability of false selection $P(\bar{X}_n > 0)$ is approximated by
$$\exp(-nI(0)).$$

A simple setting of $d = 2$

- Consider a single rv X with unknown mean EX. Need to decide whether $EX > 0$ or $EX < 0$ with error probability $\leq \delta$. Decision based on whether $\bar{X}_n > 0$ or $\bar{X}_n < 0$.

- Then if $EX < 0$, probability of false selection $P(\bar{X}_n > 0)$ is approximated by

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

- If $EX > 0$, again probability of false selection $P(\bar{X}_n < 0)$ is approximated by

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

Two phase implementation

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

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

- 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 $\hat{I}_m(0)$.

- 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.

- 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 $\bar{X}_n \leq 0$.

- 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 $I(0)$.

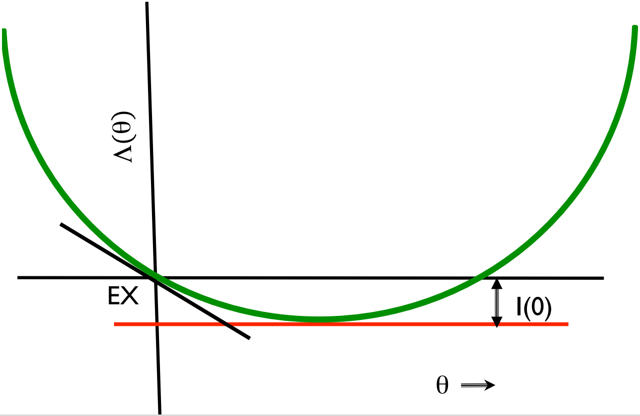
Estimating rate function

Graphic view of $I(0) = -\inf_\theta \Lambda(\theta)$

- The log-moment generating function of X

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

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



Estimating rate function $I(0)$

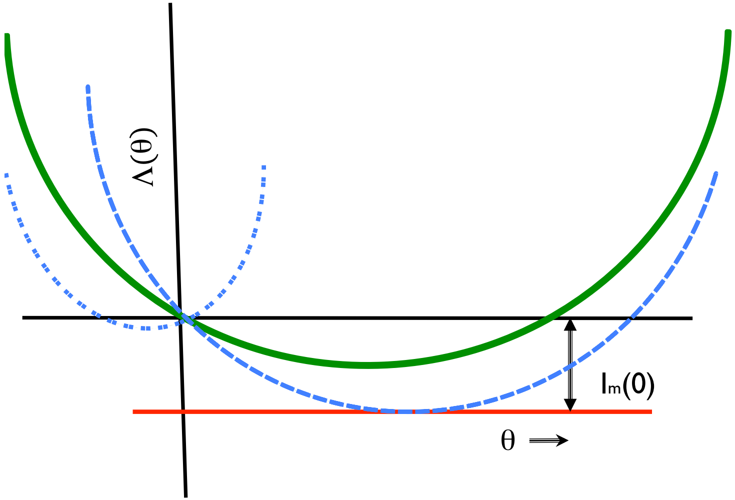
- A natural estimator for $I(0)$ based on samples $(X_i : 1 \leq i \leq m)$ is

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

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 $\hat{I}_m(0)$

- *Theorem:* 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 $\hat{I}_m(0)$

- *Theorem:* 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 $\hat{I}_m(0)$

- *Theorem:* 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 $a < I(0)$,

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

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.

Lower Bound for P(FS)

- *Theorem:*

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

Lower Bound for P(FS)

- *Theorem:*

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

- In particular,

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

Key Negative Result

- Let $\mathcal{L}$ denote a collection of probability distributions with finite mean, unbounded positive support, that are equivalent to each other.

- Let $\mathcal{L}$ denote a collection of probability distributions with finite mean, unbounded positive support, that are equivalent to each other.

- Let $\mathcal{P}(\epsilon, \delta)$ denote a policy that can adaptively sample from any two distributions in $\mathcal{L}$ and select one 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, that are equivalent to each other.

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

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

- **Theorem - 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 $\tilde{G} > \mu_F + \epsilon$.
- *Theorem:* 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,

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

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 $\epsilon > 0$, and $K > \mu_G$ there exists a distribution $\tilde{G}$ such that

$$\mathcal{KL}(G, \tilde{G}) \leq \epsilon$$

and

$$\mu_{\tilde{G}} \geq K.$$

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.

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

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

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

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

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

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

$\mathcal{P}(\epsilon, \delta)$ 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, $X \geq 0$. Then,

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, $X \geq 0$. Then,

$$\max_X EXI(X \geq u) \leq u \left(\frac{a - f(0)}{f(u) - f(0)} \right).$$

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, $X \geq 0$. Then,

$$\max_X EXI(X \geq u) \leq u \left(\frac{a - f(0)}{f(u) - f(0)} \right).$$

- 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 $Ef(X) \le a$. Further, $X \ge 0$. Then,

$$\max_X EXI(X \ge u) \le u \left(\frac{a - f(0)}{f(u) - f(0)} \right).$$

- This follows from the optimisation problem

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

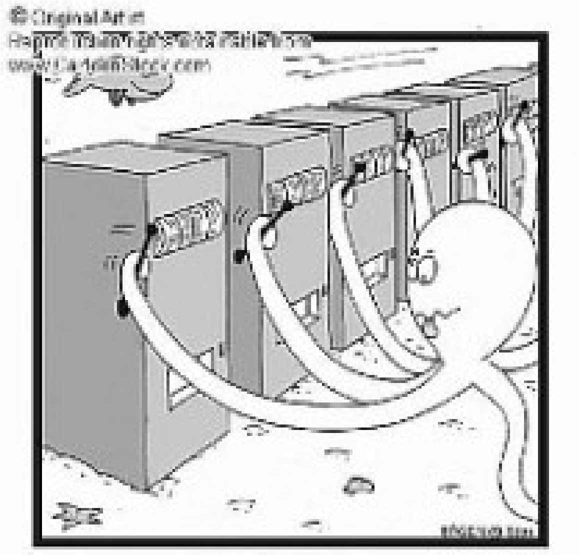
$$\text{such that } Ef(X) \le a.$$

- This has a two point solution relying on observation that

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

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

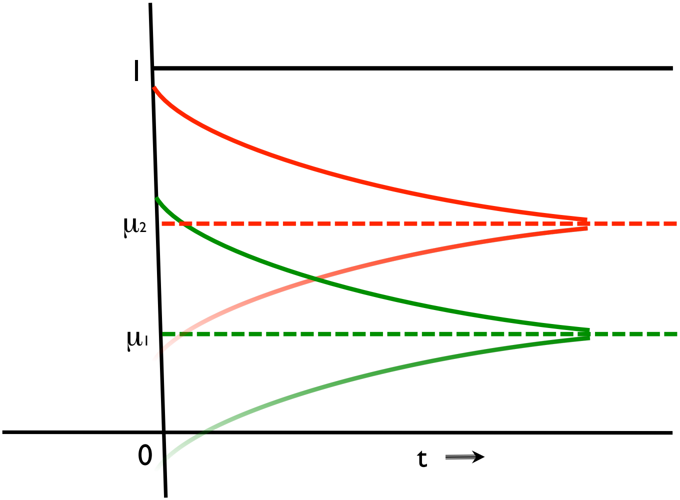
Popular successive rejection algorithm

- 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

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

  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 in light-tailed settings, the probability of false selection decays at an exponential rate, suggesting that order $\log(1/\delta)$ computational algorithms that upper bound this probability by δ may be feasible.

In conclusion

- We discussed that in light-tailed settings, the probability of false selection decays at an exponential rate, suggesting that order $\log(1/\delta)$ computational algorithms that upper bound this probability by δ may be feasible.

- 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 in light-tailed settings, the probability of false selection decays at an exponential rate, suggesting that order $\log(1/\delta)$ computational algorithms that upper bound this probability by δ may be feasible.
- 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 in light-tailed settings, the probability of false selection decays at an exponential rate, suggesting that order $\log(1/\delta)$ computational algorithms that upper bound this probability by δ may be feasible.
- 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.