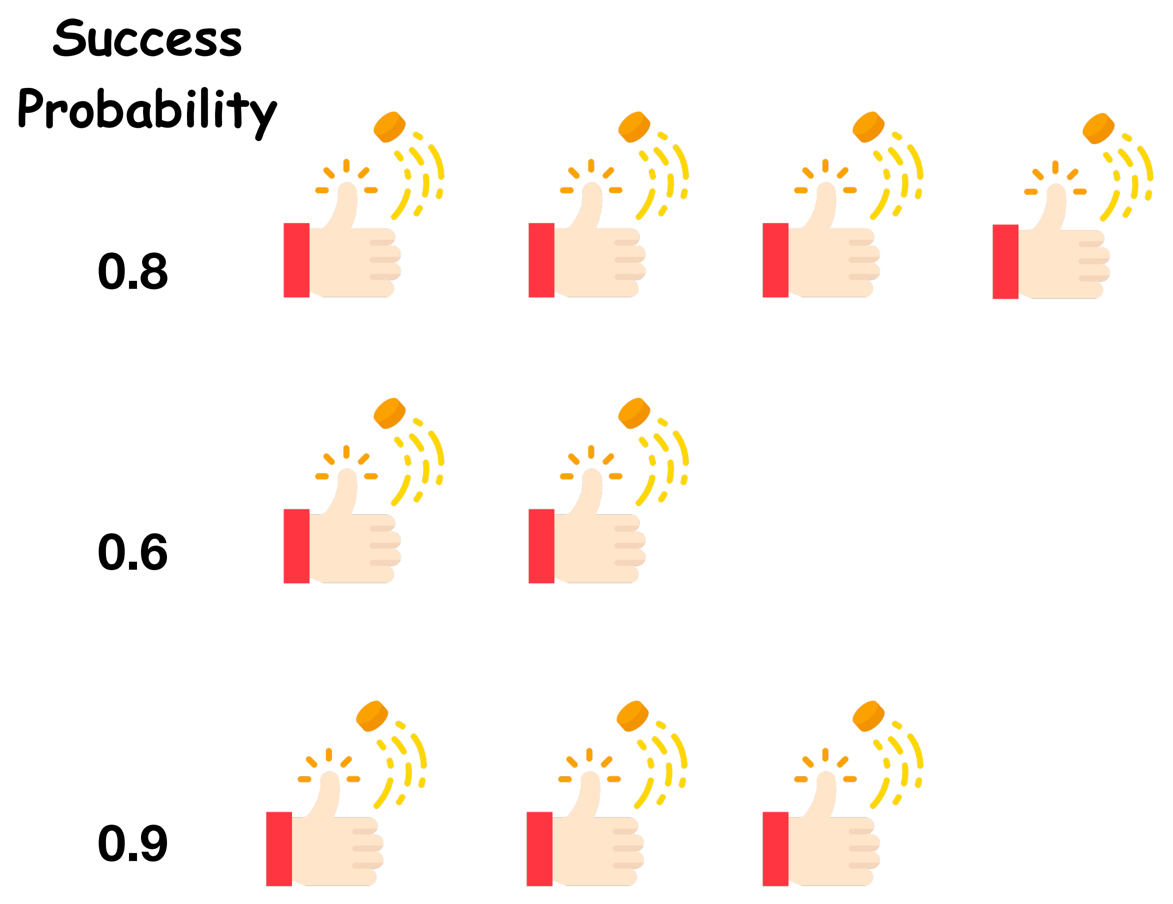
Best arm identification, through fluid based methods

Sandeep Juneja, TIFR,
Networks, ICTS Feb 1, 2024

Joint work with Agniv Bandyopadhyay,
Shubhada Agrawal, Karthikeyan Shanmugam and Arun Suggala

Best arm selection problem

Given K unknown probability distributions that can be sampled from, find the distribution with the largest mean, using fewest samples while keeping the probability of false selection to $\leq \delta$

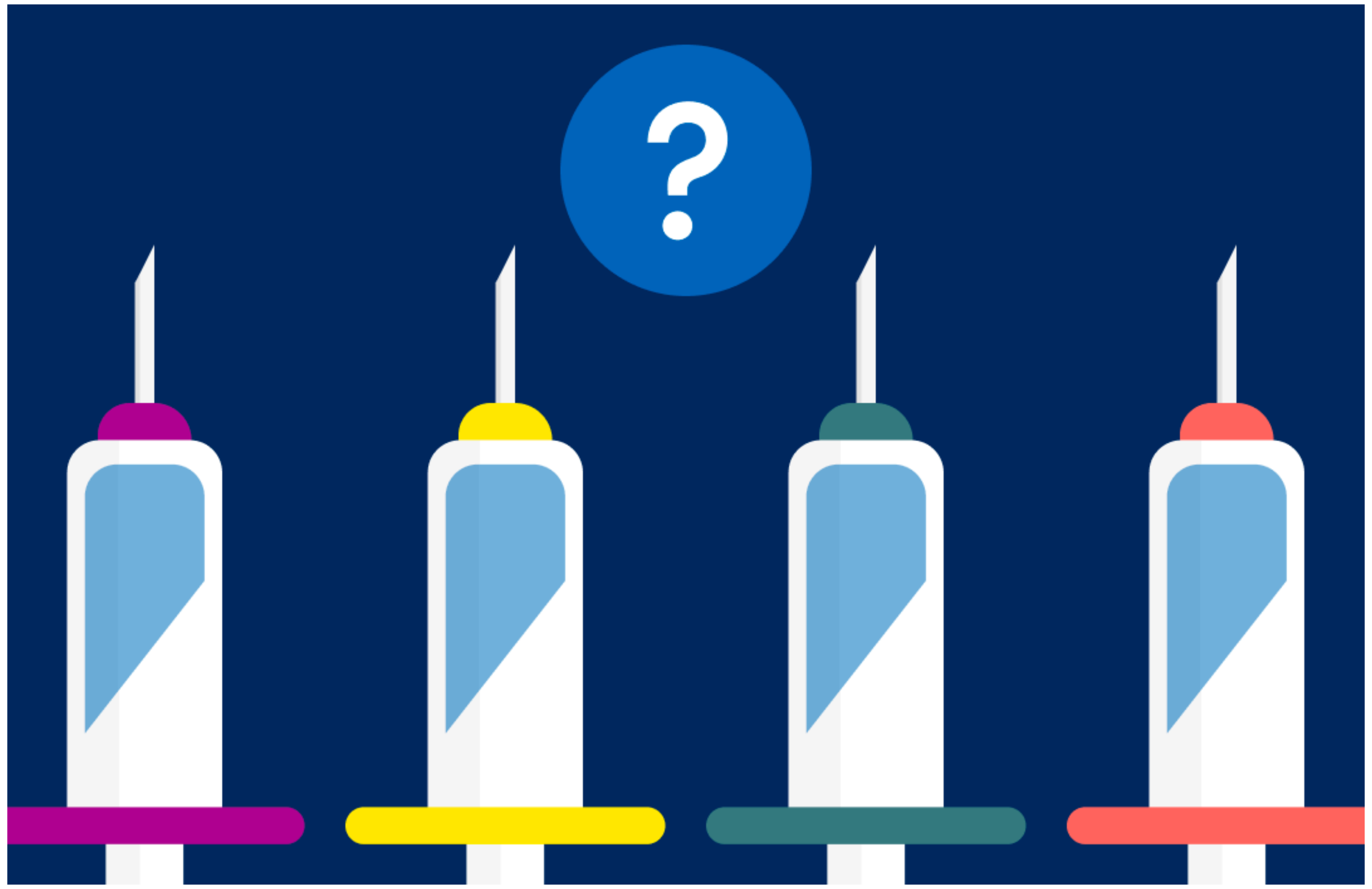


- An intuitive overview
- Optimal top 2 algorithm - chasing the fluid limit

Which coin has the highest probability of heads?

Stop sampling when you are 95% sure

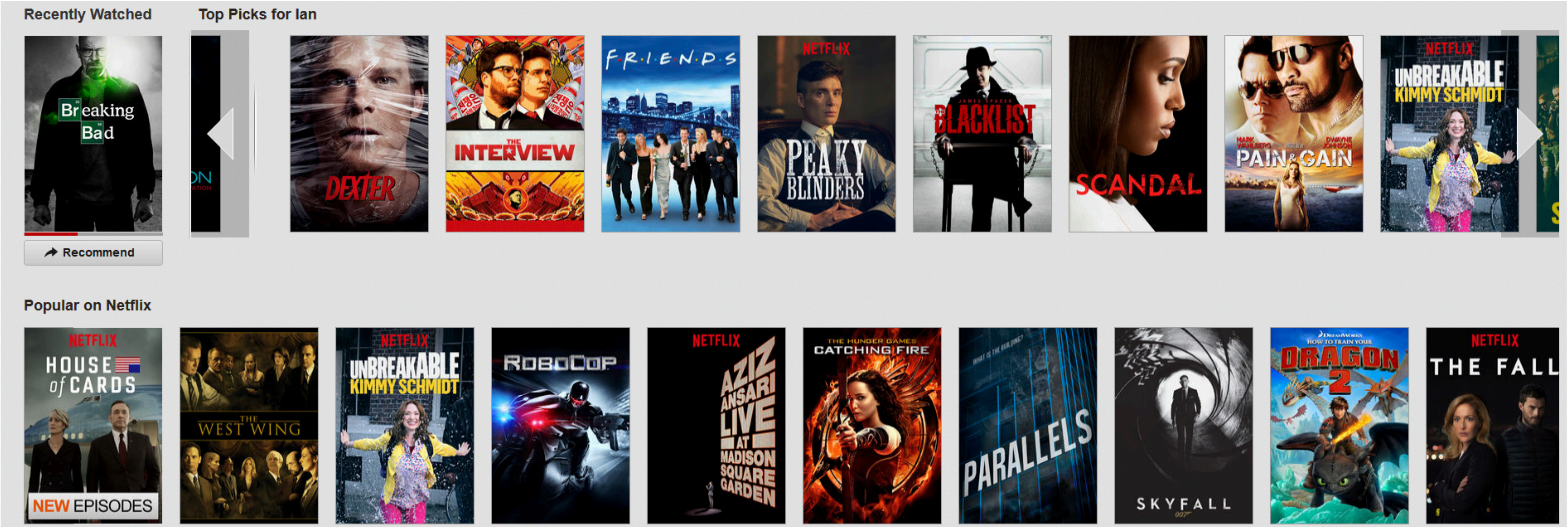
Applications: Clinical trials



- Four vaccines (or experimental drugs). Which ones to give to patients
- 'it seems apparent that a considerable saving of individuals otherwise sacrificed to the inferior (drug) treatment might be effected' Thompson, 1933
-

Applications

- Placing advertisements on a Google search
- Web construction amongst many options
- Recommendation systems
  - Movies to recommend
  - Facebook posts to show
  - News paper articles to bring to your attention
  - Price to offer for a digital good
- Travel route to recommend amongst many



Selecting the best player

-  

To separate prob. 0.6 from 0.4 with 95% certainty
need around 150 samples

Popular algorithm

Our friend: Hoeffding

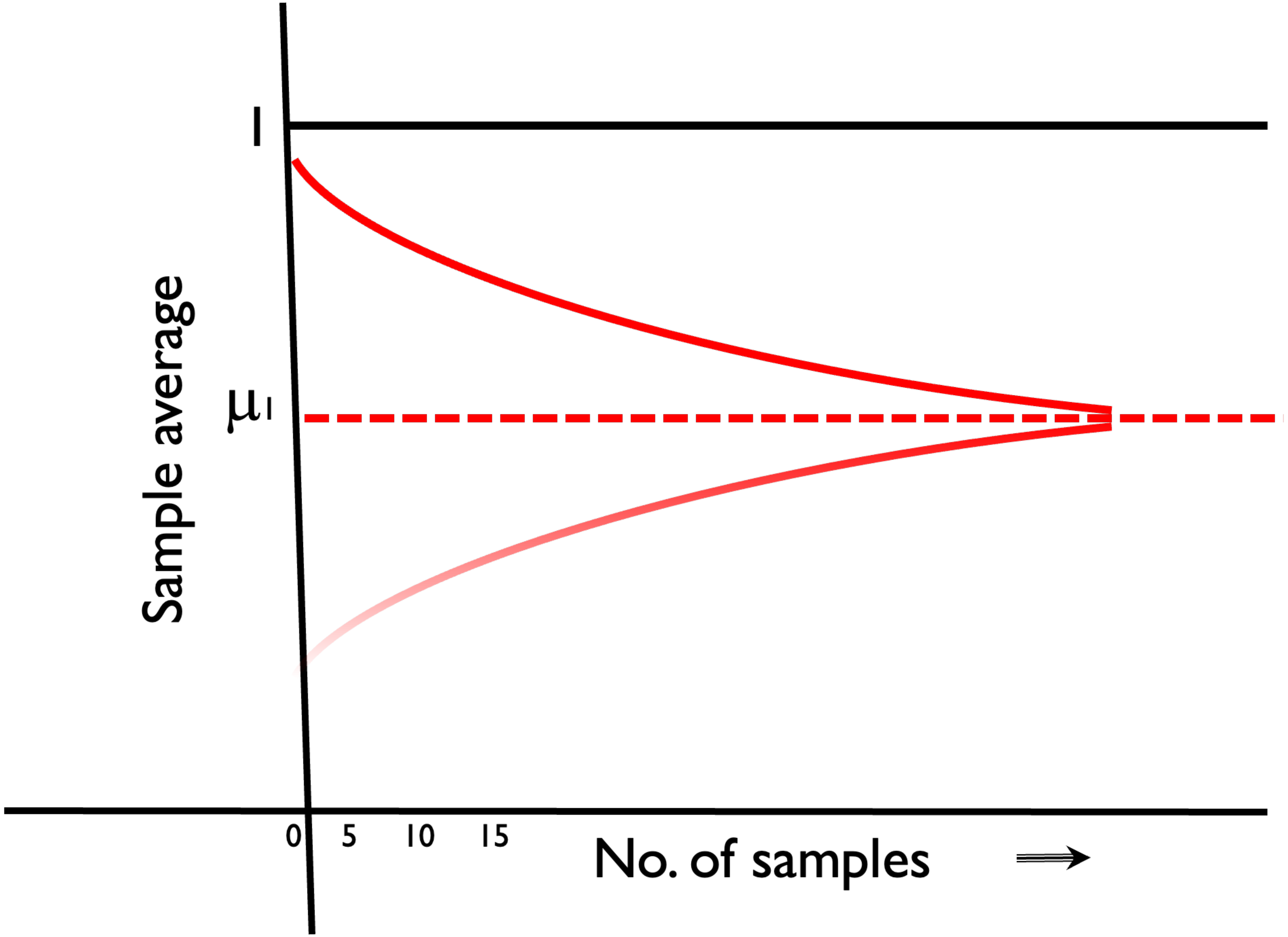
Each $X_i \in [-1,1]$ are iid with zero mean.

Hoeffding's - $$P\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq \epsilon\right) \leq \exp(-n\epsilon^2/2)\,.$$

$$P\left(\text{there exists t:}\frac{1}{t}\sum_{i=1}^{t} X_i \geq \alpha_t\right) \leq \sum_t P\left(\frac{1}{t}\sum_{i=1}^{t} X_i \geq \alpha_t\right) \leq \delta\,.$$

Where $\alpha_t = \sqrt{\frac{4\log t/\delta}{t}}$

$\bar{X}_t \in \mu \pm \alpha_t$ **for all t with probability 1-δ**



The successive rejection algorithm for arm rewards in [0,1]

Dar, Mannor, Mansour 2006

1. Sample each arm once

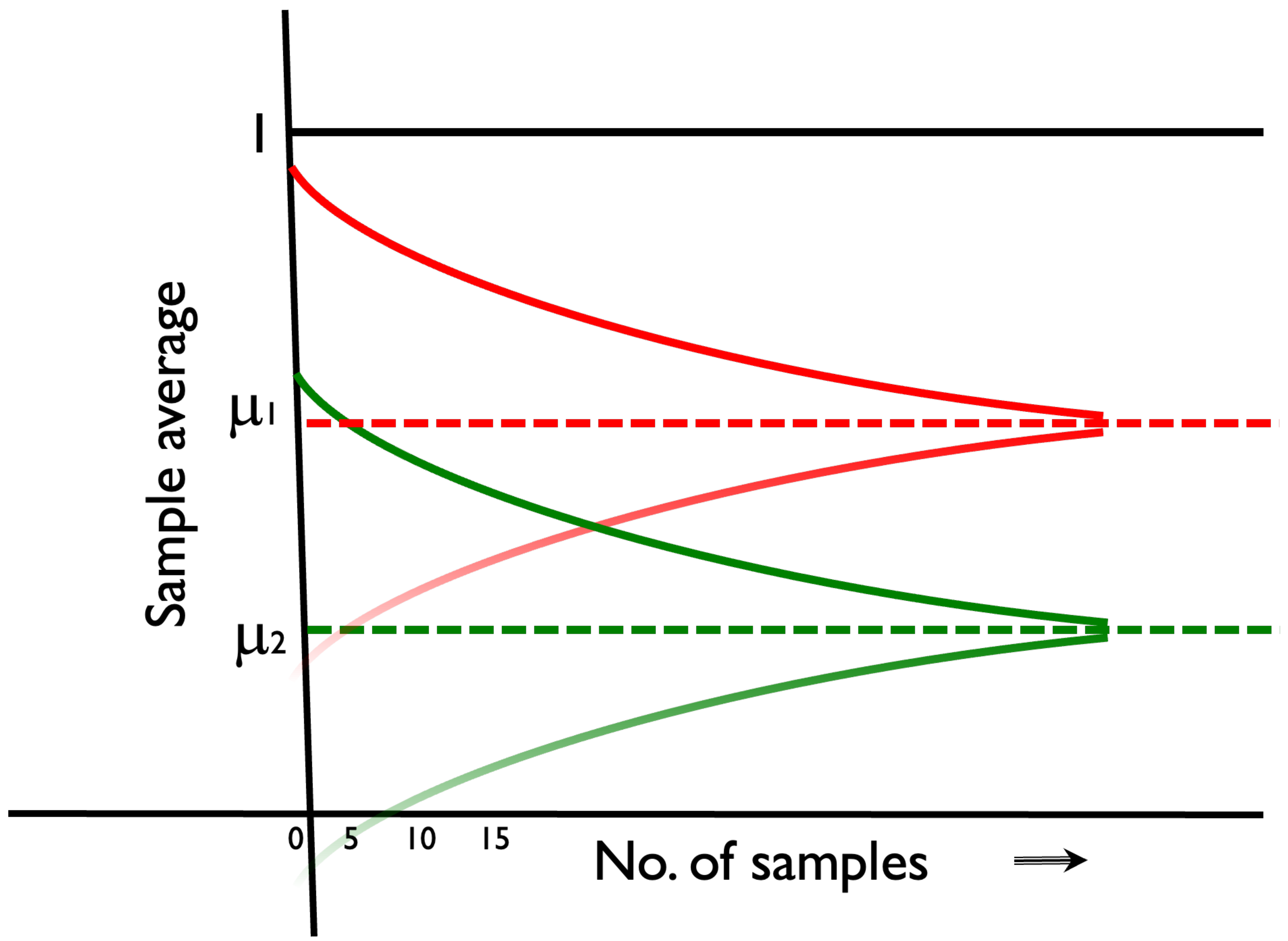
2. If at sample t,

$$\bar{X}_{\max}(t) - \bar{X}_j(t) \geq 2\alpha_t$$

then remove arm j from consideration. $\alpha_t = \sqrt{\dfrac{4\log(Kt/\delta)}{t}}$

Repeat till one arm left

Isolating the high probability tubes that contain sample averages



$\alpha_t = \sqrt{\log(Kt^2/\delta)/t}$

Best arm never rejected

$\bar{X}_1(t) \geq \mu_1 - \alpha_t$

$\bar{X}_a(t) \leq \mu_a + \alpha_t$

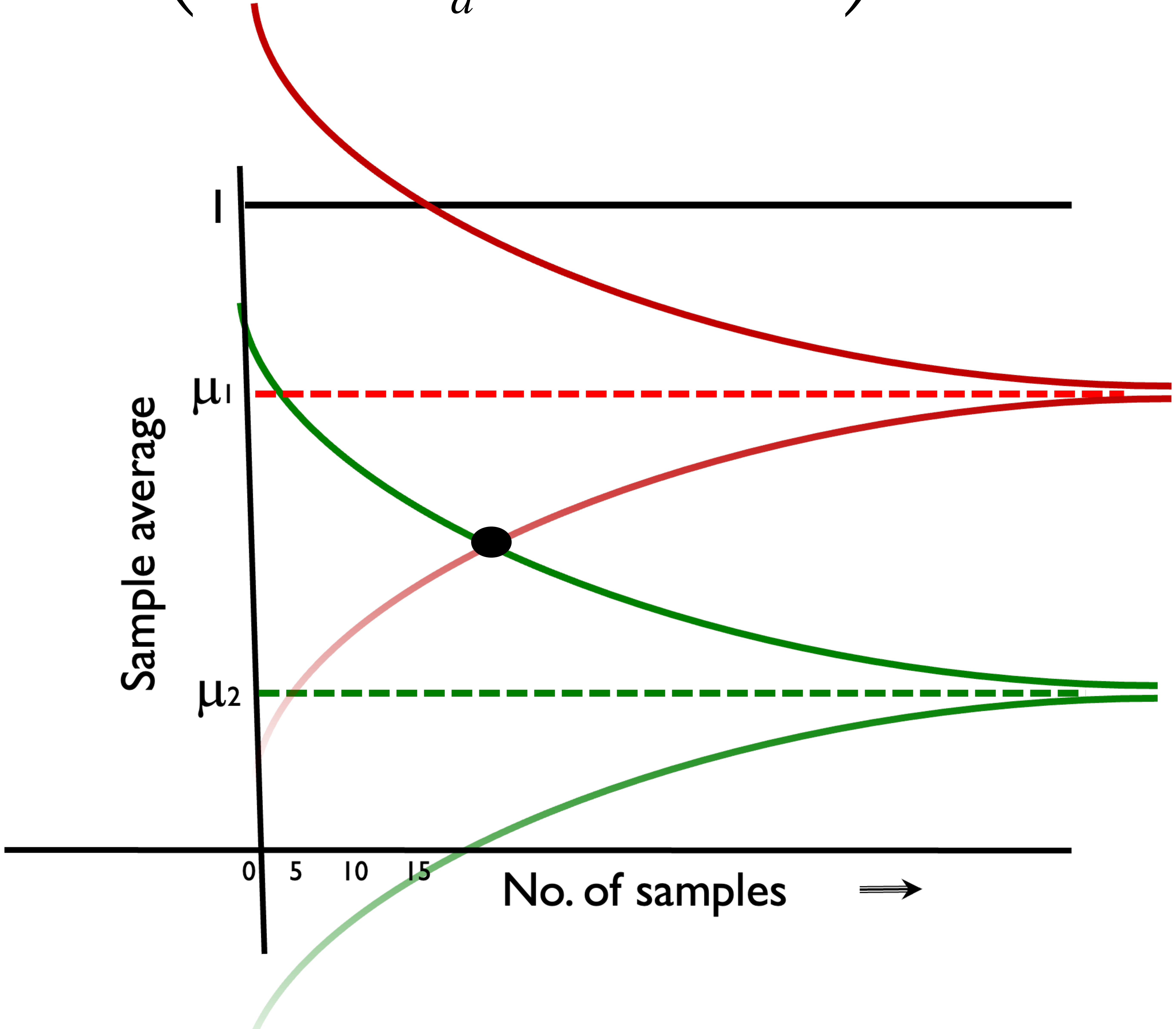
So

$\bar{X}_a(t) - \bar{X}_1(t) \leq 2\alpha_t - (\mu_1 - \mu_a)$

Samples needed $K\,O\left(\log(1/\delta)\sum_a \frac{1}{(\mu_{\max}-\mu_a)^2}\right)$

Consider tubes

$\bar{X}_t \in \mu \pm 2\alpha_t$



Lower bounds and algorithms that match even the constant in the lower bounds

A trivial lower bound

- Suppose each arm receives $\log(1/\delta)^{\alpha}$ samples for $\alpha \in (0,1)$.

- Consider large deviations approximation for sample average

$$P(\bar{X}_n \approx a) \approx \exp(-nI(a)) \text{ where } I(a) > 0 \text{ for } a \neq EX$$

- If $n = \log(1/\delta)^{\alpha}$, then $P(\bar{X}_n \approx a) \approx \delta^{\frac{I(a)}{\log(1/\delta)^{1-\alpha}}} > \delta$ for small δ

- Thus order $\log(1/\delta)$ samples necessary

Large deviations result (Sanov's Thm.)

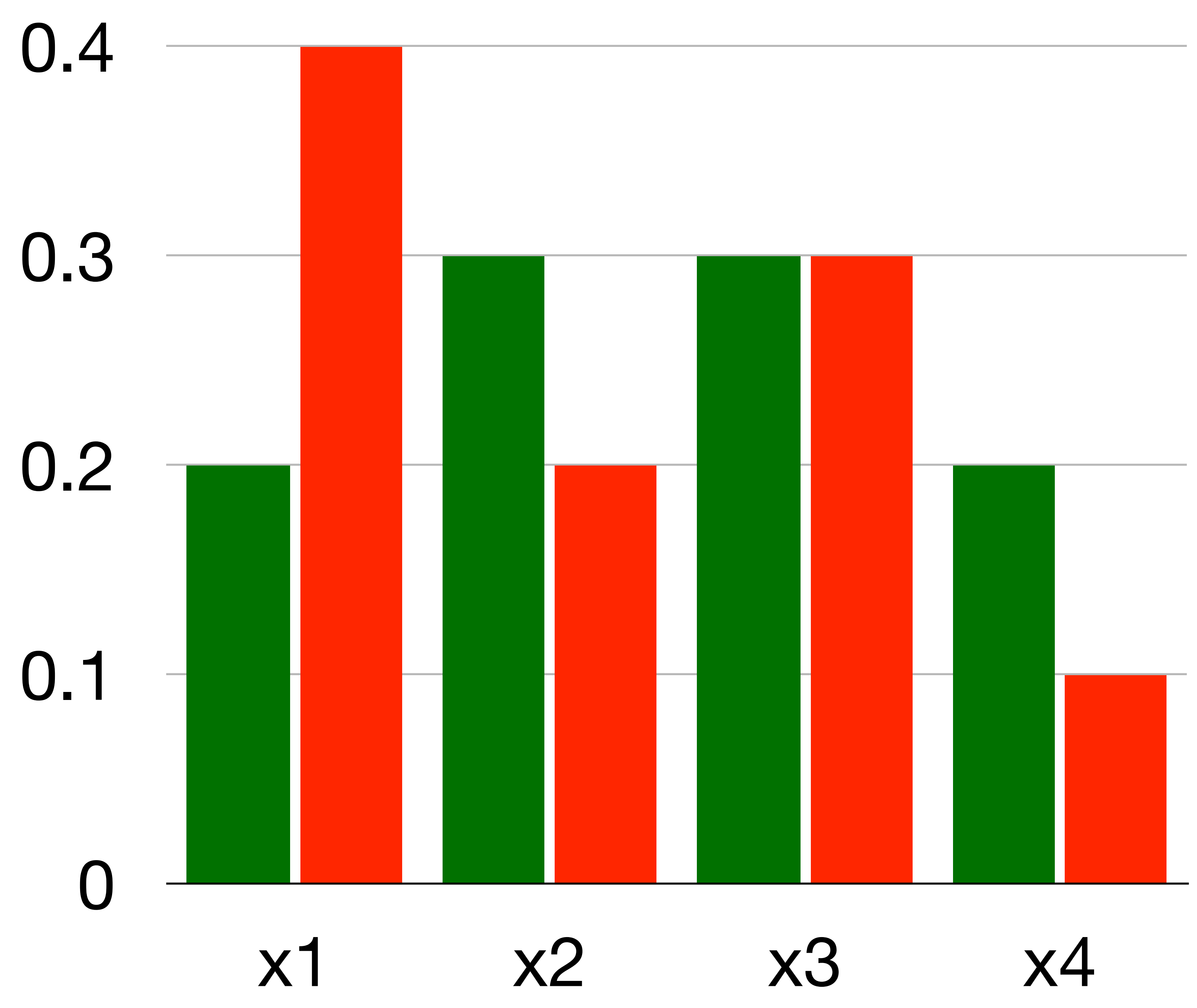
Green is the true distribution μ. Red is the empirical distribution ν (based on generated samples

$$(X_1, X_2, \ldots, X_n))$$

Probability of seeing empirical distribution ν when the true distribution is μ is

$$\approx \exp(-nKL(\nu \mid \mu))$$

where $KL(\nu \mid \mu) = \sum_{i=1}^{4} \nu_i \log\left(\frac{\nu_i}{\mu_i}\right)$



Lower bound: A heuristic argument

Two arms: Observed distributions from N_1 and N_2 samples are $\hat{\mu}_1(N_1)$ and $\hat{\mu}_2(N_2)$

With high probability, each $\hat{\mu}_i(N_i) \approx \mu_i$, so that if

$m(\hat{\mu}_1) > m(\hat{\mu}_2)$ suggests that $m(\mu_1) > m(\mu_2)$

As a skeptical scientist, you wonder its likelihood if true distributions are

$$(\nu_1, \nu_2) : m(\nu_1) < m(\nu_2)$$

Likelihood under alternate hypothesis

Let $A^c = \{\nu_1, \nu_2 : m(\nu_1) < m(\nu_2)\}$

Worst-case likelihood of incorrect assessment

$$\max_{(\nu_1,\nu_2)\in A^c} \; e^{-N_1 KL(\hat{\mu}_1|\nu_1)} \; e^{-N_2 KL(\hat{\mu}_2|\nu_2)}$$

$$\approx \max_{(\nu_1,\nu_2)\in A^c} \; e^{-N_1 KL(\mu_1|\nu_1)+N_2 KL(\mu_2|\nu_2)}$$

This needs to be kept less than δ

In many arms setting

Let $A^c = \{\nu = (\nu_1, \ldots, \nu_K) : m(\nu_1) < \max_{i \geq 2} m(\nu_i)\}$

Worst-case likelihood of incorrect assessment.

$$\max_{\nu \in A^c} \; e^{-\sum_{i \leq K} N_i KL(\mu_i | \nu_i))}$$

This needs to be less than δ.

The lower bound optimisation problem

$$\text{Minimize} \sum_a N_a$$

s.t.

$$\inf_{\{\nu: m(\nu_1) < \max_{i \geq 2} m(\nu_i)\}} \sum_{a \leq K} N_a KL(\mu_a | \nu_a) \geq \log(1/\delta)$$

The Data Processing Inequality

$$KL(P_X | Q_X) \geq KL(P_{g(X)} | Q_{g(X)}) .$$

Consider bandit algorithm output $X = (a_1, R_{a_1, N_{a_1}(1)}, \ldots, a_T, R_{a_T, N_{a_T}(T)})$ till time T.

Let P_μ correspond to the measure associated with X when arms dist. is $\mu = (\mu_a : a \leq K)$.

P_ν when the arms distribution is $\nu = (\nu_a : a \leq K)$.

Then, from Data PI

$$KL(P_\mu(X) \,|\, P_\nu(X)) \geq KL(P_\mu(I_E) \,|\, P_\nu(I_E))$$

$$KL(P_\mu(X) \,|\, P_\nu(X)) = \sum_{a=1}^{K} E_{P_\mu} N_a(T) KL(\mu_a \,|\, \nu_a) \,.$$

$$KL(P_\mu(X)\,|\,P_\nu(X)) = E_{P_\mu}\left(\sum_{t=1}^{T}\log\frac{d\mu_{a_t}}{d\nu_{a_t}}(R_{a_t,N_{a_t}(t)})\right) = E_{P_\mu}\left(\sum_{t=1}^{T}\sum_{a=1}^{K}\log\frac{d\mu_a}{d\nu_a}(R_{a,N_a(t)})I(a_t=a)\right).$$

Now, the conditional expectation

$$E\left[\log\left(\frac{d\mu_a}{d\nu_a}(R_{a,N_a(t)})\right)I(a_t=a)\,|\,\mathcal{F}_t\right] = KL(\mu_a\,|\,\nu_a)I(a_t=a)$$

and thus through iterative conditioning over time periods,

$$KL(P_\mu(X)\,|\,P_\nu(X)) = \sum_{a=1}^{K} E_{P_\mu}N_a(T)KL(\mu_a\,|\,\nu_a).$$

The lower bound optimisation problem

$$\text{Minimize} \sum_a N_a$$

s.t.

$$\inf_{\{\nu : m(\nu_1) < \max_{i \geq 2} m(\nu_i)\}} \sum_{a \leq K} N_a KL(\mu_a | \nu_a) \geq \log(1/\delta)$$

Good time to Summarise!

Simplifying the lower bound optimisation problem

Minimize $\sum_a N_a$

$$\inf_{\nu : m(\nu_1) < m(\nu_a)} N_1 KL(\mu_1 | \nu_1) + N_a KL(\mu_a | \nu_a) \geq \log(1/\delta), \quad \text{for } a \geq 2$$

$$N_1 KL(\mu_1 | x_{1,a}) + N_a KL(\mu_a | x_{1,a}) \geq \log(1/\delta), \quad \text{for } a \geq 2 \quad x_{1,a} = \frac{N_1 \mu_1 + N_a \mu_a}{N_1 + N_a}$$

Since $\{\nu : m(\nu_1) < \max_{a \geq 2} m(\nu_a)\} = \cup_{a \geq 2} \{\nu : m(\nu_1) < m(\nu_a)\}$

And distributions restricted to **single parameter exponential family**

When to stop: Generalized likelihood ratio based

Compute logarithm of

$$\frac{\text{Maximum likelihood of data}}{\text{Maximum likelihood of data under alternate hypthesis}}$$

This equals $\min_{\nu \in \hat{A}^c} \left(\sum_{a \le K} N_a KL(\hat{\mu}_a | \nu_a) \right)$

Stop when the statistic exceeds $\log(1/\delta)$ + smaller order terms

Top-2 algorithms

Top two β optimal algorithms are gaining interest (DR 16, JDBHK 22)

Index $\mathscr{I}_a$ empirical version of $N_1 KL(\mu_1 | x_{1,a}) + N_a KL(\mu_1 | x_{1,a})$

1. Please don't starve any arm

2. Select arm with largest sample mean with prob β.

3. Select challenger arm with smallest index with prob $1-\beta$

4. Stop (generalised likelihood ratio test) when

$$\min_a \mathscr{I}_a \geq \log(1/\delta) + \text{smaller order terms}$$

Recall the lower bound problem minimise $\sum_{a \leq K} N_a$

S. t. $N_1 KL(\mu_1 | x_{1,a}) + N_a KL(\mu_a | x_{1,a}) \geq \log(1/\delta) \quad \forall a$

Has a **unique** strictly positive solution that satisfies

$$N_1 KL(\mu_1 | x_{1,a}) + N_a KL(\mu_a | x_{1,a}) = \log(1/\delta) \quad \forall a$$

$$\sum_a \frac{KL(\mu_1, x_{1,a})}{KL(\mu_a, x_{1,a})} = 1.$$

Optimal top 2 algorithm

g denotes empirical $\sum_a \frac{KL(\mu_1, x_{1,a})}{KL(\mu_a, x_{1,a})}$. $\mathscr{I}_a$ empirical $N_1 KL(\mu_1 | x_{1,a}) + N_a KL(\mu_1 | x_{1,a})$

1. Please don't starve any arm
2. If g >1, sample arm 1
3. If g< 1 sample arm with the empirical smallest index $\mathscr{I}_a$
4. Stop when $\min_a \mathscr{I}_a \geq \log(1/\delta)$+smaller order terms

After enough samples, the algorithm closely tracks a **fluid path**

Under fluid path

1) Empirical dist. $\hat{\mu}$ is equal to μ

2) Once g=1 it stays one

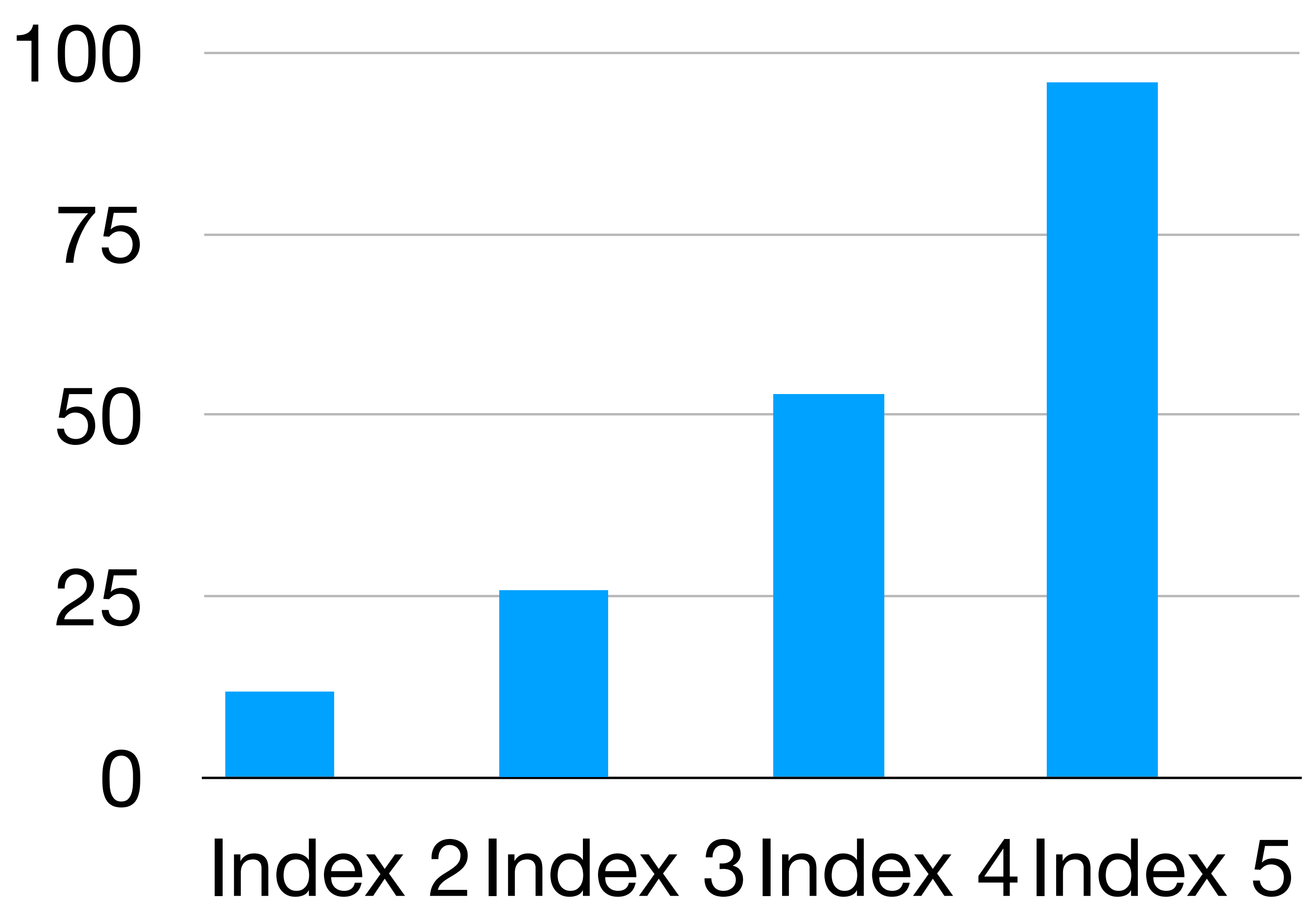
3) Once two indexes become equal they stay equal

Fluid view

Suppose after initial exploration, $g = \sum_a \frac{KL(\mu_1, x_{1,a})}{KL(\mu_a, x_{1,a})} > 1$

Then samples given to arm 1 till $\sum_a \frac{KL(\mu_1, x_{1,a})}{KL(\mu_a, x_{1,a})} = 1$

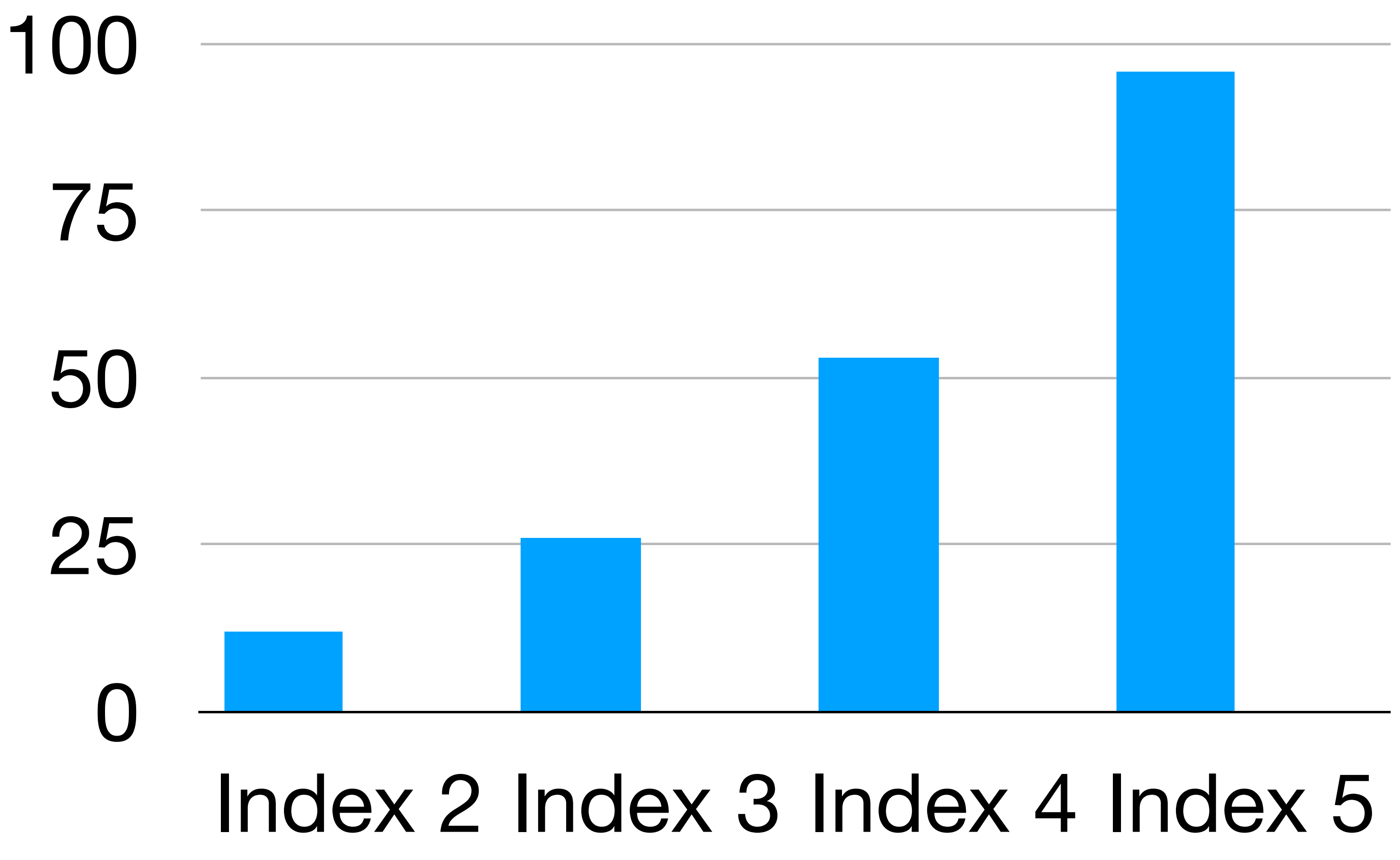
Indexes have order



Feed arm 1 and minimum index(s) while maintaining $\sum_a \frac{KL(\mu_1, x_{1,a})}{KL(\mu_a, x_{1,a})} = 1$

(Recall index $\mathcal{I}_a = N_1(n)KL(\mu_1 \,|\, x_{1,a}) + N_a(n)KL(\mu_a \,|\, x_{1,a})$ for total samples n)

- Smallest indexes increase at least linearly
- Larger ones increase sub linearly
- Once they meet they move together
- System becomes stationary once all indexes are equal



Key analysis step: **Implicit function theorem**

It shows that the fluid solution satisfies ODEs concatenated together

The algorithm after sufficiently large samples closely tracks the fluid path

Implicit function thm works because Jacabian of constraints

$$\sum_a \frac{KL(\mu_1, x_{1,a})}{KL(\mu_a, x_{1,a})} = 1 \text{ And}$$

$$N_1 KL(\mu_1 | x_{1,a}) + N_a KL(\mu_a | x_{1,a}) = I \text{ for } a \in B$$

$$N_1 + \sum_{a \in B} N_a + \sum_{a \in B^c} N_a = N$$

with respect to $(N_1, N_a, a \in B, I)$ after transformations is invertible.

Thus $(N_1, N_a, a \in B)$ of perturbed system are close by.

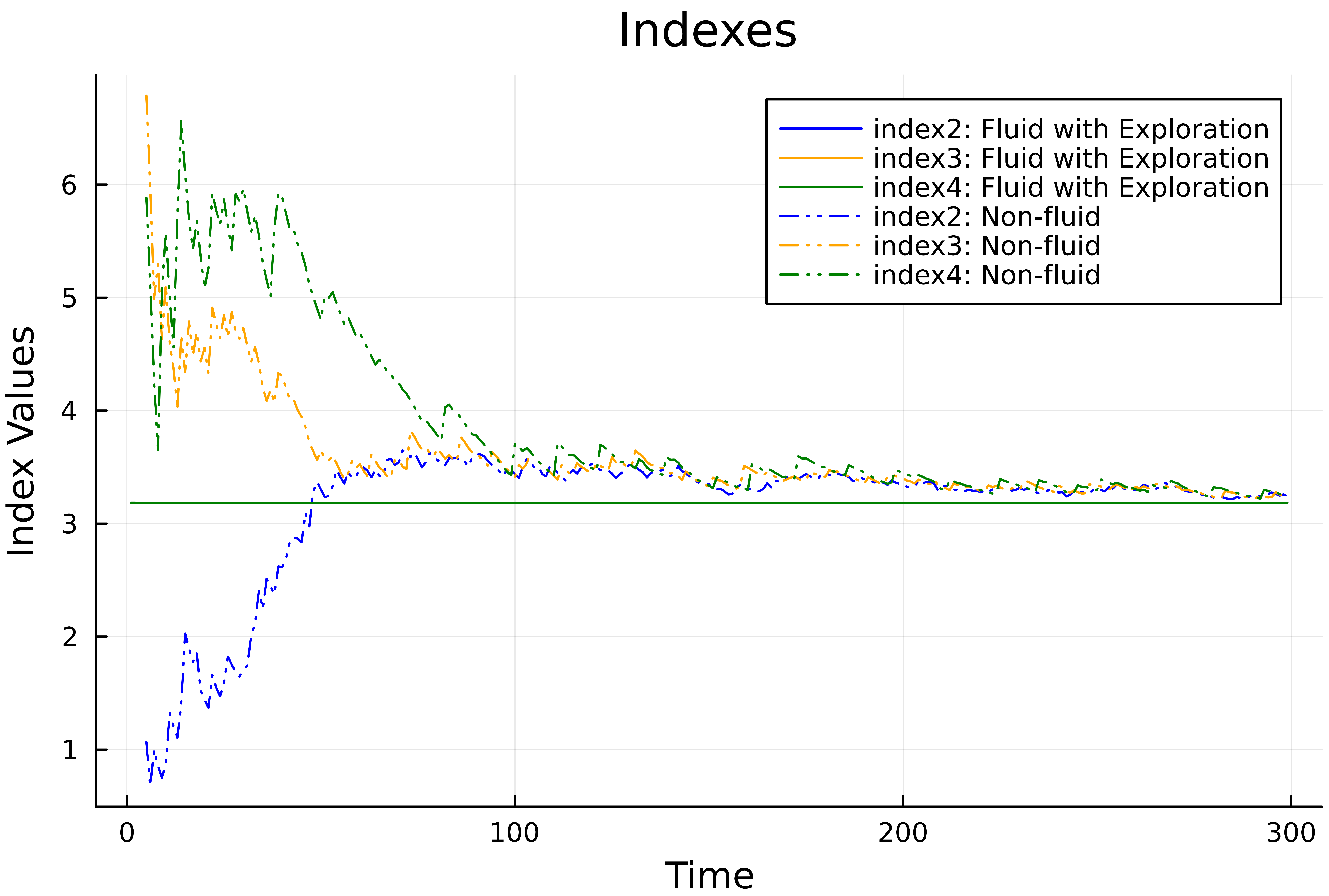
Due Implicit Function Theorem, we have $\left(\frac{dN_b}{dN} = N_b'\right)$

$$N_1' = \frac{N_1 h_B}{(N_1 + \sum_{a\in B} N_a)h_B + d_B^{-1}h(N)}$$

$$N_b' = \frac{N_b h_B + d_{b,b}^{-1}h(N)}{(N_1 + \sum_{a\in B} N_a)h_B + d_B^{-1}h(N)} \text{ for all } b \in B$$

Let $h_a = \frac{\partial g}{\partial N_a}$, $d_{1,a} = d(\mu_1, x_{1,a})$ and $d_{a,a} = d(\mu_a, x_{1,a})$ $h_B = \sum_{a\in B} h_a d_{a,a}^{-1}$, $h(N) = \sum_{a\in B^c/1} h_a N_a$ and $d_B = \left(\sum_{a\in B} d_{a,a}^{-1}\right)^{-1}$.

Simulation



Indexes
index2: Fluid with Exploration
index3: Fluid with Exploration
index4: Fluid with Exploration
index2: Non-fluid
index3: Non-fluid
index4: Non-fluid
Index Values
Time
0
100
200
300
1
2
3
4
5
6

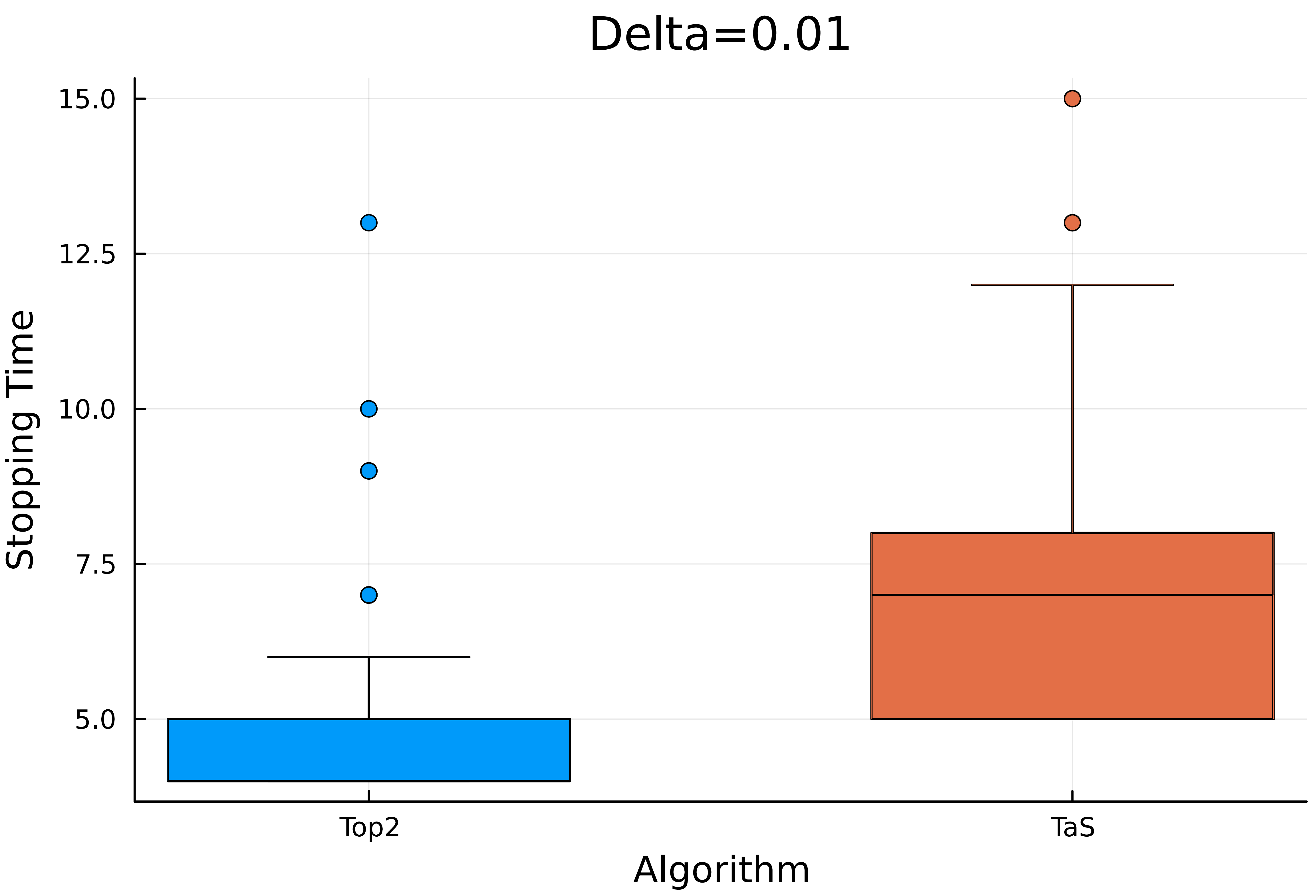
Due Implicit Function Theorem, we have $\left(\frac{dN_b}{dN} = N_b'\right)$

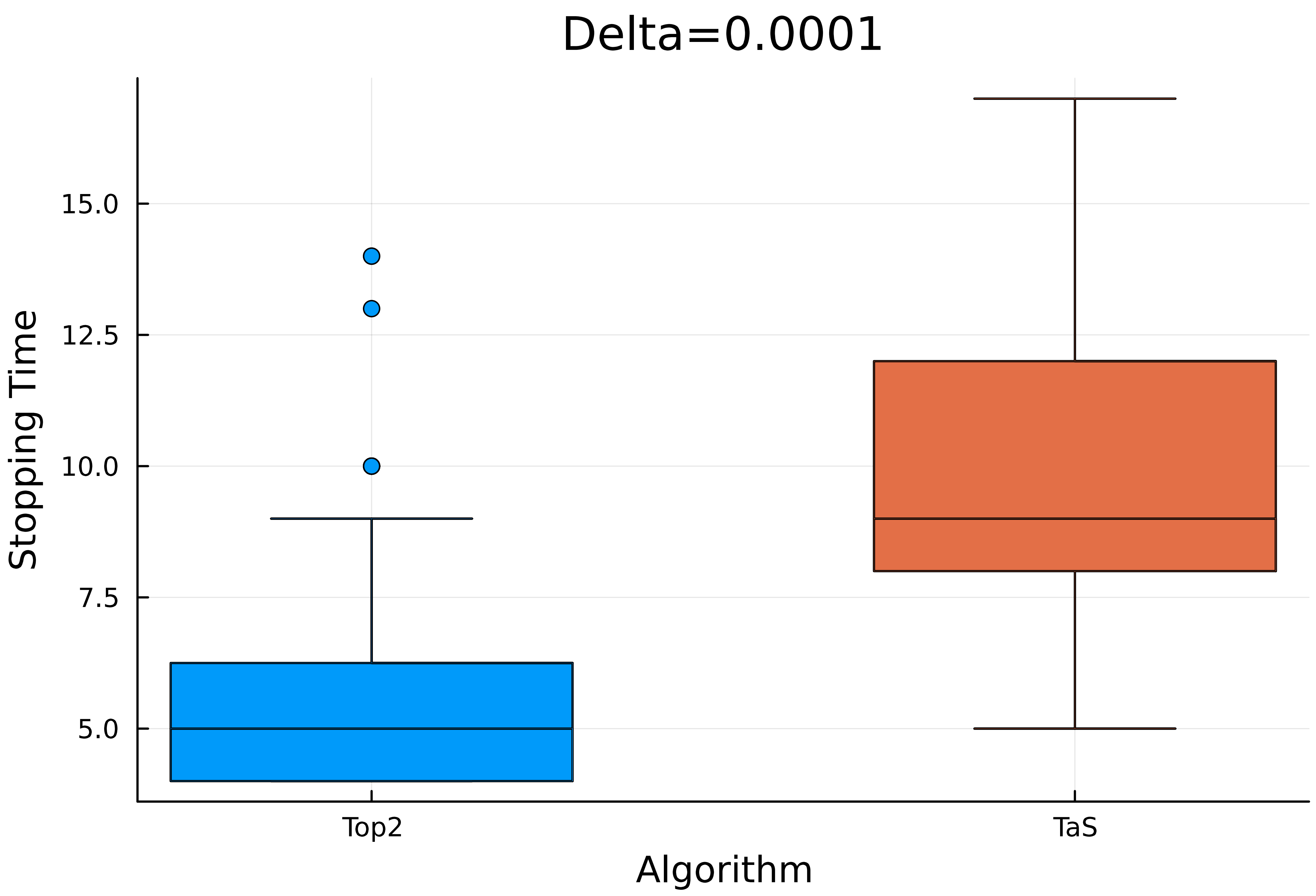
$$N_1' = \frac{N_1 h_B}{(N_1 + \sum_{a \in B} N_a)h_B + d_B^{-1} h(N)}$$

$$N_b' = \frac{N_b h_B + d_{b,b}^{-1} h(N)}{(N_1 + \sum_{a \in B} N_a)h_B + d_B^{-1} h(N)} \text{ for all } b \in B$$

Let $h_a = \frac{\partial g}{\partial N_a}$, $d_{1,a} = d(\mu_1, x_{1,a})$ and $d_{a,a} = d(\mu_a, x_{1,a})$ $h_B = \sum_{a \in B} h_a d_{a,a}^{-1}$, $h(N) = \sum_{a \in B^c/1} h_a N_a$ and $d_B = \left(\sum_{a \in B} d_{a,a}^{-1}\right)^{-1}$.

Simulation


Delta=0.01
Stopping Time
15.0
12.5
10.0
7.5
5.0
Top2
TaS
Algorithm


Delta=0.0001
Stopping Time
15.0
12.5
10.0
7.5
5.0
Top2
TaS
Algorithm

Controlling the probability of error

Recall we stop when $\min_{\nu \in \hat{A}^c} \left(\sum_{a \le K} N_a KL(\hat{\mu}_a | \nu_a) \right) \ge \log(1/\delta) +$ small

If you stop wrong, $\hat{A}^c$ contains the true probability vector μ.

Need to bound $P(\sum_{a \le K} N_a KL(\hat{\mu}_a | \mu_a)) \ge \log(1/\delta) +$ small, for any n) by δ.

Dual representations, exponential concave inequalities and mixture martingales cleverly used for this.

Extending to general distributions

- Consider

$$\mathscr{L} := \{\eta \in \mathscr{P}(\mathfrak{R}) : \mathbb{E}_{X\sim\eta}(|X|^{1+\epsilon} \leq B\}$$

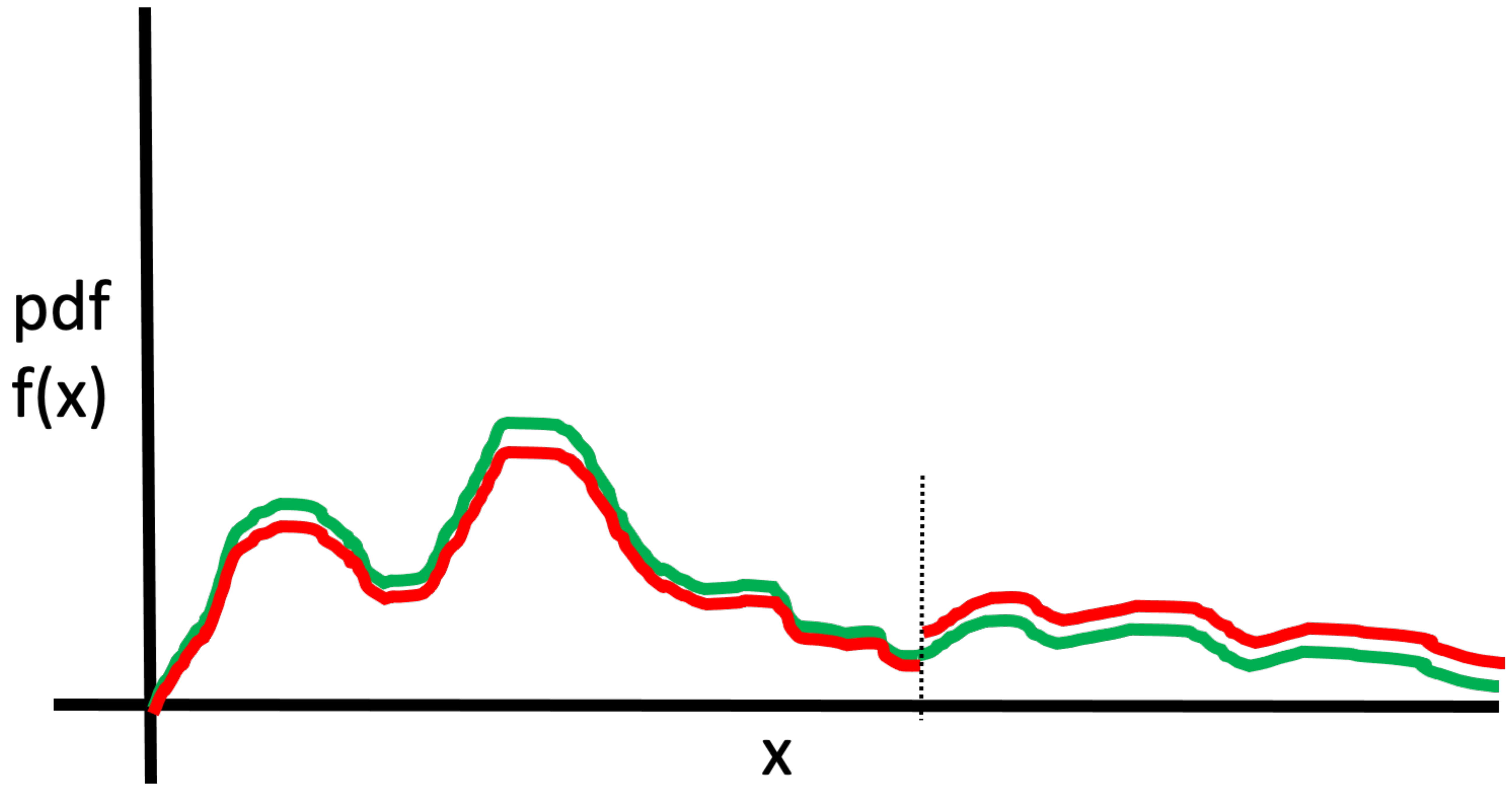
- Define $KL_{inf}(\mu_a, x) = \inf_{\nu \in \mathscr{L} : m(\nu) > x} KL(\mu_a, \nu)$

Some conditions on the underlying distributions are necessary

Easy to find two distributions whose KL distance is arbitrarily close, but means are arbitrarily far.

These are difficult to separate

Some restrictions on distributions necessary - bounded, have sub-Gaussian tails, variance or other moments bounded



Understanding $KL_{\inf}(\eta, x)$

It equals $\inf_{\kappa} \sum_i \log\left(\frac{\eta_i}{\kappa_i}\right)\eta_i$ such that

$\sum_i |y_i|^{1+\epsilon}\kappa_i \leq B$, $\sum_i y_i\kappa_i \geq x$ and $\sum_i \kappa_i = 1$.

This is a convex program and is solved through Lagrangian duality.

Using duality, $KL_{\inf}(\eta, x)$ can be seen to equal

$$\max_{(\lambda_1,\lambda_2)\in\mathcal{R}_2} E_\eta \log(1-(X-x)\lambda_1-(B-|X|^{1+\epsilon}\lambda_2),$$ where

For empirical distribution $\hat{\mu}_a(n)$ we have $KL_{\inf}(\hat{\mu}_a(t), m(\mu_a))$ equals

$$\max_{(\lambda_1,\lambda_2)\in\mathcal{R}_2} \frac{1}{N_a(n)} \sum_{i=1}^{N_a(n)} \log(1-(X_i-m(\mu_a))\lambda_1-(B-|X_i|^{1+\epsilon})\lambda_2)).$$

In developing concentration inequality for this, the maximum function poses difficulties. We observe that inside the maximum we have a sum of exp-concave functions.

Sum of exp concave functions: a useful inequality

Let $\Lambda \subseteq \Re^d$ be a compact and convex subset and q be the uniform distribution on Λ. Let $g_t : \Lambda \to \Re$ be any series of exp-concave functions. Then

$$\max_{\lambda \in \Lambda} \sum_{t=1}^{T} g_t(\lambda) \leq \log E_{\lambda \sim q} e^{\sum_{t=1}^{T} g_t(\lambda)} + d \log(T+1) + 1.$$

Thus $\max_{\lambda \in \Lambda} \exp\left(\sum_{t=1}^{T} g_t(\lambda)\right)$ is close to the expectation $E_{\lambda \sim q} e^{\sum_{t=1}^{T} g_t(\lambda)}$.

The latter is a mixture of super-martingales and hence is a super martingale.

Ville's inequality

Ville's inequality: For a non-negative super martingale $(M_n : n \geq 0)$,

$$P(\exists n : M_n \geq x) \leq \frac{EM_0}{x}.$$

Conclusion

Discussed the best arm identification problem, applications and a popular algorithm

Introduced the optimal top-2 approach and discussed its fluid behaviour

Argued that our algorithm closely tracks the fluid behaviour when generated samples are large

Outlined how δ guarantees are shown