Concentration: Introduction to Sequential Analysis

Master 2 Mathematics and Computer Science

Guillaume Aubrun, Aurélien Garivier, Rémi Gribonval
2020-2021
1. Sequential Decision Problems

2. The Simple Bandit Model

3. Classical strategies

4. Lower Bound

5. The Track-and-Stop Strategy
Sequential Decision Problems
Solving a Game

2 player game with finite number of actions

but too deep for exhaustive search of minimax action (by alpha-beta)
Example: Go ($\approx 10^{171}$ possible configurations)
Monte Carlo Tree Search

Heuristic search algorithm using random playouts / rollouts

Root Position

Random Playouts

MC Evaluation

9/10  3/10  4/10

src: https://www.remi-coulom.fr/
Dose Finding

![Graph showing the relationship between dose and toxicity percentage, with a target toxicity level indicated. The graph plots dose on the x-axis and toxicity (in %) on the y-axis. The data points show an increasing trend as the dose increases.]
Notez des films. Complétez votre profil et affinez vos recommandations personnalisées en notant plus de films ou en les ajoutant à votre wishlist!
Optimization

- Goal: maximize function $f: C \subset \mathbb{R}^d \rightarrow \mathbb{R}$ possibly observed with noise.

- Applications: computer experiment.

- Model: $f$ comes from a Gaussian Process, or when it has a small norm in the induced RKHS.
The Simple Bandit Model
Best-Arm Identification with Fixed Confidence

$K$ options = probability distributions $\nu = (\nu_a)_{1 \leq a \leq K}$

$\nu_a \in \mathcal{F}$ exponential family parameterized by its expectation $\mu_a$

At round $t$, you may:

- choose an option $A_t = \phi_t (A_1, X_1, \ldots, A_{t-1}, X_{t-1}) \in \{1, \ldots, K\}$
- observe a new independent sample $X_t \sim \nu_{A_t}$

so as to identify the best option $a^* = \arg\max_a \mu_a$ and $\mu^* = \max_a \mu_a$
as fast as possible: stopping time $\tau$. 
Best-Arm Identification with Fixed Confidence

\( K \) options = probability distributions \( \nu = (\nu_a)_{1 \leq a \leq K} \)
\( \nu_a \in \mathcal{F} \) exponential family parameterized by its expectation \( \mu_a \)

At round \( t \), you may:

- choose an option \( A_t = \phi_t (A_1, X_1, \ldots, A_{t-1}, X_{t-1}) \in \{1, \ldots, K\} \)
- observe a new independent sample \( X_t \sim \nu_{A_t} \)

so as to identify the best option \( a^* = \arg \max_a \mu_a \) and \( \mu^* = \max_a \mu_a \)

as fast as possible: stopping time \( \tau_\delta \).
Intuition: a Simple Example

Most simple setting: for all \( a \in \{1, \ldots, K\} \),

\[ \nu_a = \mathcal{N}(\mu_a, 1) \]

For example: \( \mu = [2, 1.75, 1.75, 1.6, 1.5] \).

At time \( t \):
- you have sampled \( n_a \) times the option \( a \)
- your empirical average is \( \bar{X}_{a,n_a} \).

\[ \text{if you stop at time } t, \text{ your probability of prefering arm } a \geq 2 \text{ to arm } a^* = 1 \text{ is:} \]

\[ \mathbb{P} \left( \bar{X}_{a,n_a} > \bar{X}_{1,n_1} \right) = \mathbb{P} \left( \frac{\bar{X}_{a,n_a} - \mu_a - (\bar{X}_{1,n_1} - \mu_1)}{\sqrt{1/n_1 + 1/n_a}} > \frac{\mu_1 - \mu_a}{\sqrt{1/n_1 + 1/n_a}} \right) \]

\[ = \Phi \left( \frac{\mu_1 - \mu_a}{\sqrt{1/n_1 + 1/n_a}} \right) \]

where \( \Phi(u) = \int_{u}^{\infty} \frac{e^{-u^2/2}}{\sqrt{2\pi}} \, du \).
Uniform Sampling

\[ P(\text{confusion}) \]

\[ \delta \]

\[ \frac{1}{2} \]

\begin{align*}
\text{arm 1} & : w = 0.20, n = 0 \\
\text{arm 2} & : w = 0.20, n = 0 \\
\text{arm 3} & : w = 0.20, n = 0 \\
\text{arm 4} & : w = 0.20, n = 0 \\
\text{arm 5} & : w = 0.20, n = 0
\end{align*}

\[ t = 0 \]
Uniform Sampling

P(confusion)

\[ t = 24 \]
Uniform Sampling

$P(\text{confusion})$

$t = 49$

\[
\begin{align*}
\delta & \quad \text{arm 1} \quad \text{arm 2} \quad \text{arm 3} \quad \text{arm 4} \quad \text{arm 5} \\
\text{w} & = 0.20 \quad \text{w} = 0.20 \quad \text{w} = 0.20 \quad \text{w} = 0.20 \quad \text{w} = 0.20 \\
\text{n} & = 10 \quad \text{n} = 10 \quad \text{n} = 10 \quad \text{n} = 10 \quad \text{n} = 9
\end{align*}
\]
Uniform Sampling

$P(\text{confusion})$

t = 74

<table>
<thead>
<tr>
<th>Arm</th>
<th>$w$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.20</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>0.20</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>0.20</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>0.20</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>0.20</td>
<td>14</td>
</tr>
</tbody>
</table>
Uniform Sampling

P(confusion)

$\delta$

$t = 99$

<table>
<thead>
<tr>
<th>Arm</th>
<th>$w$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.20</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>0.20</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>0.20</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>0.20</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>0.20</td>
<td>19</td>
</tr>
</tbody>
</table>
Uniform Sampling

P(confusion)

\[
t = 124
\]

\[
\begin{array}{cccccc}
\text{arm 1} & \text{arm 2} & \text{arm 3} & \text{arm 4} & \text{arm 5} \\
w & 0.20 & 0.20 & 0.20 & 0.20 & 0.20 \\
\end{array}
\]
Uniform Sampling

\[ P(\text{confusion}) \]

\[ t = 149 \]
Uniform Sampling

$P(\text{confusion})$

$t=174$

arms 1, 2, 3, 4, 5:
- arm 1: $w=0.20$, $n=35$
- arm 2: $w=0.20$, $n=35$
- arm 3: $w=0.20$, $n=35$
- arm 4: $w=0.20$, $n=35$
- arm 5: $w=0.20$, $n=34$
Uniform Sampling

\[ P(\text{confusion}) \]

t= 199
Uniform Sampling

\[ \Pr(\text{confusion}) \]

\[ t = 224 \]

\[
\begin{align*}
\delta &\quad \text{arm 1} & w = 0.20 & n = 45 \\
&\quad \text{arm 2} & w = 0.20 & n = 45 \\
&\quad \text{arm 3} & w = 0.20 & n = 45 \\
&\quad \text{arm 4} & w = 0.20 & n = 45 \\
&\quad \text{arm 5} & w = 0.20 & n = 44
\end{align*}
\]
Uniform Sampling

\[
P(\text{confusion})
\]

\[t = 249\]

\[
\begin{align*}
\text{arm 1} & : w = 0.20, n = 50 \\
\text{arm 2} & : w = 0.20, n = 50 \\
\text{arm 3} & : w = 0.20, n = 50 \\
\text{arm 4} & : w = 0.20, n = 50 \\
\text{arm 5} & : w = 0.20, n = 49
\end{align*}
\]
Uniform Sampling

\[ P(\text{confusion}) \]

\[ t = 274 \]

![Graph showing P(confusion) distribution for different arms with varying weights and sample sizes.](image-url)
Uniform Sampling

\[ P(\text{confusion}) \]

\[ t = 299 \]
Uniform Sampling

\[ P(\text{confusion}) \]

\[ t = 324 \]

\[ \delta \]

\begin{align*}
\text{arm 1} & : \ w = 0.20, \ n = 65 \\
\text{arm 2} & : \ w = 0.20, \ n = 65 \\
\text{arm 3} & : \ w = 0.20, \ n = 65 \\
\text{arm 4} & : \ w = 0.20, \ n = 65 \\
\text{arm 5} & : \ w = 0.20, \ n = 64
\end{align*}
Uniform Sampling

$P(\text{confusion})$

t = 349

- arm 1: $w = 0.20$, $n = 70$
- arm 2: $w = 0.20$, $n = 70$
- arm 3: $w = 0.20$, $n = 70$
- arm 4: $w = 0.20$, $n = 70$
- arm 5: $w = 0.20$, $n = 69$
Uniform Sampling

$P(\text{confusion})$

t = 374

\begin{align*}
\delta & \quad \text{arm 1} \quad \text{w= 0.20} \quad n= 75 \\
& \quad \text{arm 2} \quad \text{w= 0.20} \quad n= 75 \\
& \quad \text{arm 3} \quad \text{w= 0.20} \quad n= 75 \\
& \quad \text{arm 4} \quad \text{w= 0.20} \quad n= 75 \\
& \quad \text{arm 5} \quad \text{w= 0.20} \quad n= 74
\end{align*}
Uniform Sampling

$P(\text{confusion})$

$t = 399$

\begin{align*}
\text{arm 1} & : w = 0.20, n = 80 \\
\text{arm 2} & : w = 0.20, n = 80 \\
\text{arm 3} & : w = 0.20, n = 80 \\
\text{arm 4} & : w = 0.20, n = 80 \\
\text{arm 5} & : w = 0.20, n = 79
\end{align*}
Uniform Sampling

\[ P(\text{confusion}) \]

\[ t = 424 \]
Uniform Sampling

\[ P(\text{confusion}) \]

\[ t = 431 \]
Uniform Sampling

$P(\text{confusion})$

$t = 433$

$\delta$

arm 1: $w = 0.20$, $n = 87$

arm 2: $w = 0.20$, $n = 87$

arm 3: $w = 0.20$, $n = 87$

arm 4: $w = 0.20$, $n = 86$

arm 5: $w = 0.20$, $n = 86$
Intuition: Equalizing the Probabilities of Confusion

Most simple setting: for all \( a \in \{1, \ldots, K\} \),

\[ \nu_a = \mathcal{N}(\mu_a, 1) \]

For example: \( \mu = [2, 1.75, 1.75, 1.6, 1.5] \).

**Active Learning**

→ You allocate a relative budget \( w_a \) to option \( a \), with \( w_1 + \cdots + w_K = 1 \).

At time \( t \):

→ you have sampled \( n_a \approx w_a t \) times the option \( a \)

→ your empirical average is \( \bar{X}_{a,n_a} \).

→→ if you stop at time \( t \), your probability of preferring arm \( a \geq 2 \) to arm \( a^* = 1 \) is:

\[
\mathbb{P}(\bar{X}_{a,n_a} > \bar{X}_{1,n_1}) = \mathbb{P}\left( \frac{\bar{X}_{a,n_a} - \mu_a - (\bar{X}_{1,n_1} - \mu_1)}{\sqrt{1/n_1 + 1/n_a}} > \frac{\mu_1 - \mu_a}{\sqrt{1/n_1 + 1/n_a}} \right)
\]

\[= \Phi\left( \frac{\mu_1 - \mu_a}{\sqrt{1/n_1 + 1/n_a}} \right)\]
Improving: trial 1

\[ P(\text{confusion}) \]

\[ t = 0 \]

\[
\begin{array}{ccccc}
\text{arm 1} & \text{arm 2} & \text{arm 3} & \text{arm 4} & \text{arm 5} \\
\text{w} = 0.30 & \text{w} = 0.25 & \text{w} = 0.25 & \text{w} = 0.10 & \text{w} = 0.10 \\
\text{n} = 0 & \text{n} = 0 & \text{n} = 0 & \text{n} = 0 & \text{n} = 0
\end{array}
\]
Improving: trial 1

\[ P(\text{confusion}) \]

\[ t = 24 \]

\[ \begin{array}{cccccc}
\text{arm} & 1 & 2 & 3 & 4 & 5 \\
\text{w} & 0.30 & 0.25 & 0.25 & 0.10 & 0.10 \\
\text{n} & 7 & 6 & 6 & 3 & 2
\end{array} \]
Improving: trial 1

\[ P(\text{confusion}) \]

\[ t = 49 \]
Improving: trial 1

\[ P(\text{confusion}) \]

\[ t = 74 \]

\begin{align*}
\text{arm 1} & \quad w = 0.30 & \quad n = 22 \\
\text{arm 2} & \quad w = 0.25 & \quad n = 19 \\
\text{arm 3} & \quad w = 0.25 & \quad n = 18 \\
\text{arm 4} & \quad w = 0.10 & \quad n = 8 \\
\text{arm 5} & \quad w = 0.10 & \quad n = 7
\end{align*}
Improving: trial 1

P(confusion)

\[ t = 99 \]

- arm 1: \( w = 0.30 \), \( n = 29 \)
- arm 2: \( w = 0.25 \), \( n = 25 \)
- arm 3: \( w = 0.25 \), \( n = 25 \)
- arm 4: \( w = 0.10 \), \( n = 10 \)
- arm 5: \( w = 0.10 \), \( n = 10 \)
Improving: trial 1

P(confusion)

t= 124

<table>
<thead>
<tr>
<th>arm</th>
<th>w</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.30</td>
<td>37</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>31</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
<td>31</td>
</tr>
<tr>
<td>4</td>
<td>0.10</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>0.10</td>
<td>12</td>
</tr>
</tbody>
</table>
Improving: trial 1

\[ P(\text{confusion}) \]

t = 149

\[ \delta \]

\[ 1/2 \]

<table>
<thead>
<tr>
<th>Arm</th>
<th>w</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.30</td>
<td>45</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>37</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
<td>37</td>
</tr>
<tr>
<td>4</td>
<td>0.10</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>0.10</td>
<td>15</td>
</tr>
</tbody>
</table>
Improving: trial 1

P(confusion)

t = 174

\[
\begin{array}{c|c|c|c|c|c}
\text{arm} & \text{w} & \text{n} \\
\hline
1 & 0.30 & 52 \\
2 & 0.25 & 44 \\
3 & 0.25 & 43 \\
4 & 0.10 & 18 \\
5 & 0.10 & 17 \\
\end{array}
\]
Improving: trial 1

P(confusion)

t = 199

<table>
<thead>
<tr>
<th>Arm</th>
<th>w</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.30</td>
<td>59</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>0.10</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>0.10</td>
<td>20</td>
</tr>
</tbody>
</table>
Improving: trial 1

P(confusion)

$t = 224$

- arm 1: $w = 0.30$, $n = 67$
- arm 2: $w = 0.25$, $n = 56$
- arm 3: $w = 0.25$, $n = 56$
- arm 4: $w = 0.10$, $n = 23$
- arm 5: $w = 0.10$, $n = 22$
Improving: trial 1

\[ P(\text{confusion}) \]

\[ t = 249 \]

- **Arm 1**: \( w = 0.30 \), \( n = 75 \)
- **Arm 2**: \( w = 0.25 \), \( n = 62 \)
- **Arm 3**: \( w = 0.25 \), \( n = 62 \)
- **Arm 4**: \( w = 0.10 \), \( n = 25 \)
- **Arm 5**: \( w = 0.10 \), \( n = 25 \)
Improving: trial 1

\[ P(\text{confusion}) \]

\[ t = 274 \]
Improving: trial 1

$P(\text{confusion})$

$t = 299$

<table>
<thead>
<tr>
<th>Arm</th>
<th>$w$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.30</td>
<td>89</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>75</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
<td>75</td>
</tr>
<tr>
<td>4</td>
<td>0.10</td>
<td>30</td>
</tr>
<tr>
<td>5</td>
<td>0.10</td>
<td>30</td>
</tr>
</tbody>
</table>
Improving: trial 1

P(confusion)

\[ t = 317 \]

- arm 1: \( w = 0.30 \), \( n = 95 \)
- arm 2: \( w = 0.25 \), \( n = 79 \)
- arm 3: \( w = 0.25 \), \( n = 79 \)
- arm 4: \( w = 0.10 \), \( n = 32 \)
- arm 5: \( w = 0.10 \), \( n = 32 \)
Improving: trial 1

\[ P(\text{confusion}) \]

\[ t = 319 \]

\[ \delta \]

\[ \begin{array}{cccccc}
\text{arm 1} & \text{arm 2} & \text{arm 3} & \text{arm 4} & \text{arm 5} \\
w = 0.30 & w = 0.25 & w = 0.25 & w = 0.10 & w = 0.10 \\
n = 95 & n = 80 & n = 80 & n = 32 & n = 32 \\
\end{array} \]
Optimal Proportions

$P(\text{confusion})$

$\delta$

$1/2$

$\bullet \quad \bullet \quad \bullet \quad \bullet \quad t=0$

<table>
<thead>
<tr>
<th>arm 1</th>
<th>arm 2</th>
<th>arm 3</th>
<th>arm 4</th>
<th>arm 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>w= 0.37</td>
<td>w= 0.26</td>
<td>w= 0.26</td>
<td>w= 0.07</td>
<td>w= 0.04</td>
</tr>
<tr>
<td>n= 0</td>
<td>n= 0</td>
<td>n= 0</td>
<td>n= 0</td>
<td>n= 0</td>
</tr>
</tbody>
</table>
Optimal Proportions

P(confusion)

\[ t = 24 \]

\[ \begin{array}{cccccc}
\text{arm 1} & \text{arm 2} & \text{arm 3} & \text{arm 4} & \text{arm 5} \\
\text{w} = 0.37 & \text{w} = 0.26 & \text{w} = 0.26 & \text{w} = 0.07 & \text{w} = 0.04 \\
\text{n} = 9 & \text{n} = 6 & \text{n} = 6 & \text{n} = 2 & \text{n} = 1 \\
\end{array} \]
Optimal Proportions

P(confusion)

t = 49

\[
\begin{align*}
\text{arm 1: } w &= 0.37, \quad n = 18 \\
\text{arm 2: } w &= 0.26, \quad n = 13 \\
\text{arm 3: } w &= 0.26, \quad n = 12 \\
\text{arm 4: } w &= 0.07, \quad n = 4 \\
\text{arm 5: } w &= 0.04, \quad n = 2
\end{align*}
\]
Optimal Proportions

P(confusion)

t = 74

\[
\begin{array}{cccccc}
\text{arm 1} & \text{arm 2} & \text{arm 3} & \text{arm 4} & \text{arm 5} \\
w = 0.37 & w = 0.26 & w = 0.26 & w = 0.07 & w = 0.04 \\
n = 28 & n = 19 & n = 19 & n = 5 & n = 3
\end{array}
\]
Optimal Proportions

$P(\text{confusion})$

$t = 99$

- arm 1: $w = 0.37$, $n = 36$
- arm 2: $w = 0.26$, $n = 26$
- arm 3: $w = 0.26$, $n = 26$
- arm 4: $w = 0.07$, $n = 7$
- arm 5: $w = 0.04$, $n = 4$
Optimal Proportions

\[ P(\text{confusion}) \]

\[ t = 124 \]

- arm 1: \( w = 0.37 \), \( n = 46 \)
- arm 2: \( w = 0.26 \), \( n = 32 \)
- arm 3: \( w = 0.26 \), \( n = 32 \)
- arm 4: \( w = 0.07 \), \( n = 9 \)
- arm 5: \( w = 0.04 \), \( n = 5 \)
P(confusion)

$\delta$ - $1/2$

$t = 149$

<table>
<thead>
<tr>
<th>Arm</th>
<th>$w$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.37</td>
<td>55</td>
</tr>
<tr>
<td>2</td>
<td>0.26</td>
<td>39</td>
</tr>
<tr>
<td>3</td>
<td>0.26</td>
<td>38</td>
</tr>
<tr>
<td>4</td>
<td>0.07</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>0.04</td>
<td>6</td>
</tr>
</tbody>
</table>
Optimal Proportions

\[ P(\text{confusion}) \]

t = 174

\[ \delta \]

\[ \frac{1}{2} \]

<table>
<thead>
<tr>
<th>Arm</th>
<th>( w )</th>
<th>( n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.37</td>
<td>64</td>
</tr>
<tr>
<td>2</td>
<td>0.26</td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>0.26</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td>0.07</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>0.04</td>
<td>8</td>
</tr>
</tbody>
</table>
Optimal Proportions

P(confusion)

\[ t = 199 \]

<table>
<thead>
<tr>
<th>Arm</th>
<th>Value</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arm 1</td>
<td>0.37</td>
<td>73</td>
</tr>
<tr>
<td>Arm 2</td>
<td>0.26</td>
<td>52</td>
</tr>
<tr>
<td>Arm 3</td>
<td>0.26</td>
<td>51</td>
</tr>
<tr>
<td>Arm 4</td>
<td>0.07</td>
<td>14</td>
</tr>
<tr>
<td>Arm 5</td>
<td>0.04</td>
<td>9</td>
</tr>
</tbody>
</table>
Optimal Proportions

P(confusion)

t= 224
Optimal Proportions

\[ P(\text{confusion}) \]

\[ t = 249 \]

- Arm 1: \( w = 0.37 \), \( n = 92 \)
- Arm 2: \( w = 0.26 \), \( n = 64 \)
- Arm 3: \( w = 0.26 \), \( n = 64 \)
- Arm 4: \( w = 0.07 \), \( n = 18 \)
- Arm 5: \( w = 0.04 \), \( n = 11 \)
Optimal Proportions

\[ P(\text{confusion}) \]

\[ t = 274 \]
Optimal Proportions

$P(\text{confusion})$

$t = 283$

<table>
<thead>
<tr>
<th>Arm</th>
<th>Weight ($w$)</th>
<th>Sample Size ($n$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arm 1</td>
<td>0.37</td>
<td>105</td>
</tr>
<tr>
<td>Arm 2</td>
<td>0.26</td>
<td>73</td>
</tr>
<tr>
<td>Arm 3</td>
<td>0.26</td>
<td>73</td>
</tr>
<tr>
<td>Arm 4</td>
<td>0.07</td>
<td>20</td>
</tr>
<tr>
<td>Arm 5</td>
<td>0.04</td>
<td>12</td>
</tr>
</tbody>
</table>
P(confusion)

$\delta$ - 1/2

<table>
<thead>
<tr>
<th>Arm</th>
<th>w</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.37</td>
<td>108</td>
</tr>
<tr>
<td>2</td>
<td>0.26</td>
<td>76</td>
</tr>
<tr>
<td>3</td>
<td>0.26</td>
<td>76</td>
</tr>
<tr>
<td>4</td>
<td>0.07</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>0.04</td>
<td>13</td>
</tr>
</tbody>
</table>

$t = 294$
How to Turn this Intuition into a Theorem?

- The arms are not Gaussian (no formula for probability of confusion) → large deviations (Sanov, KL)
- You do not allocate a relative budget at first, but you use sequential sampling → no fixed-size samples: sequential experiment → tracking lemma
- How to compute the optimal proportions? → lower bound, game
- The parameters of the distribution are unknown → (sequential) estimation
- When should you stop? → Chernoff’s stopping rule
Exponential Families

\(\nu_1, \ldots, \nu_K\) belong to a one-dimensional exponential family

\[P_{\lambda, \Theta, b} = \{\nu_\theta, \theta \in \Theta : \nu_\theta \text{ has density } f_\theta(x) = \exp(\theta x - b(\theta)) \text{ w.r.t. } \lambda\}\]

Example: Gaussian, Bernoulli, Poisson distributions...

- \(\nu_\theta\) can be parametrized by its mean \(\mu = b(\theta) : \nu_\mu := \nu_{b^{-1}(\mu)}\)

Notation: Kullback-Leibler divergence

For a given exponential family,

\[d(\mu, \mu') := \text{KL}(\nu_\mu, \nu_{\mu'}) = \mathbb{E}_{X \sim \nu_\mu} \left[\log \frac{d\nu_\mu}{d\nu_{\mu'}}(X)\right]\]

is the KL-divergence between the distributions of mean \(\mu\) and \(\mu'\).

We identify \(\nu = (\nu^{\mu_1}, \ldots, \nu^{\mu_K})\) and \(\mu = (\mu_1, \ldots, \mu_K)\) and consider

\[S = \left\{\mu \in (b(\Theta))^K : \exists a \in \{1, \ldots, K\} : \mu_a > \max_{i \neq a} \mu_i\right\}\]
Classical strategies
LUCB: Lower-Upper Confidence Bounds

• Build confidence bounds $L_a(t)$ and $U_a(t)$ such that with probability at least $1 - \delta$, for all times $t \geq 1$ and all arms $a \in \{1, \ldots, K\}$:

$$\mu_a \in [L_a(t), U_a(t)] ,$$

• Sample alternately

$$\hat{a}(t) = \arg\max_{a \in \{1, \ldots, K\}} L_a(t) \quad \text{and} \quad \arg\max_{b \neq \hat{a}(t)} U_b(t)$$

• Stopping time $\tau_\delta = \text{the first time } t \text{ when}$

$$\exists \hat{a} \in \{1, \ldots, K\} : \forall a \neq \hat{a}, U_a(t) < L_{\hat{a}}(t)$$

Analysis: $\delta$-correct by nature, and with probability at least $1 - \delta$:

$$\tau_\delta \leq C \sum_{a \neq a^*} \frac{1}{(\mu^* - \mu_a)^2}$$

for some constant $C$. 
Racing: Successive Eliminations

- Proceed in rounds where, at each round, all active arms are sample once
- Keep a list of active arms = those which have not been eliminated
- At the end of each round, eliminate the arms which are provably suboptimal (with a global risk $\delta$)

Analysis: similarly, one finds a constant $C$ such that

$$\mathbb{E}[\tau_\delta] \leq C \sum_{a \neq a^*} \frac{1}{(\mu^* - \mu_a)^2}.$$
Lower Bound
Information in Bandit Models

**Theorem**  
[see Garivier, Ménard and Stoltz, M.O.R. to appear]

For all bandit problems $\mu$ and $\lambda$, all stopping time $\tau$ and $\sigma(\mathcal{F}_\tau)$–measurable random variables $Z$ with values in $[0, 1]$,

$$
\sum_{a=1}^{K} \mathbb{E}_\mu[N_a(\tau)] d(\mu_a, \lambda_a) \geq \text{kl}(\mathbb{E}_\mu[Z], \mathbb{E}_\lambda[Z]).
$$

Proof: if $I_\tau = (A_1, X_{A_1}, 1, \ldots, A_\tau, X_{A_\tau}, N_{A_\tau}(\tau))$,

$$
\sum_{a=1}^{K} \mathbb{E}_\mu[N_a(\tau)] d(\mu_a, \lambda_a) = \text{KL}(\mathbb{P}_{\mu}^{I_\tau}, \mathbb{P}_{\lambda}^{I_\tau}) \geq \text{KL}(\mathbb{P}_{\mu}^Z, \mathbb{P}_{\lambda}^Z) \geq \text{kl}(\mathbb{E}_\mu[Z], \mathbb{E}_\lambda[Z])
$$

by *tensorization* and *contraction* of entropy (and small lemma).
Let $\mu = (\mu_1, \ldots, \mu_K)$ and $\lambda = (\lambda_1, \ldots, \lambda_K)$ be two elements of $S$.

**Uniform $\delta$-correct Constraint** [Kaufmann, Cappé, Garivier ’15]

If $a^*(\mu) \neq a^*(\lambda)$, any $\delta$-correct algorithm satisfies

$$
\sum_{a=1}^{K} \mathbb{E}_\mu \left[ N_a(\tau_\delta) \right] d(\mu_a, \lambda_a) \geq kl(\delta, 1 - \delta)
$$

where $kl(p, q) = p \log \frac{p}{q} + (1 - p) \log \frac{1-p}{1-q}$.

Let $\text{Alt}(\mu) = \{ \lambda : a^*(\lambda) \neq a^*(\mu) \}$. Take: $\lambda_1 = m_2 - \epsilon$ $\lambda_2 = m_2 + \epsilon$

$$
\mathbb{E}_\mu \left[ N_1(\tau_\delta) \right] d(\mu_1, m_2 - \epsilon) + \mathbb{E}_\mu \left[ N_2(\tau_\delta) \right] d(\mu_2, m_2 + \epsilon) \geq kl(\delta, 1 - \delta)
$$
Let $\mu = (\mu_1, \ldots, \mu_K)$ and $\lambda = (\lambda_1, \ldots, \lambda_K)$ be two elements of $S$.

**Uniform $\delta$-correct Constraint** [Kaufmann, Cappé, Garivier ’15]

If $a^*(\mu) \neq a^*(\lambda)$, any $\delta$-correct algorithm satisfies

$$\sum_{a=1}^{K} \mathbb{E}_\mu [N_a(\tau_\delta)] \cdot d(\mu_a, \lambda_a) \geq \text{kl}(\delta, 1 - \delta)$$

where $\text{kl}(p, q) = p \log \frac{p}{q} + (1 - p) \log \frac{1-p}{1-q}$.

Let $\text{Alt}(\mu) = \{ \lambda : a^*(\lambda) \neq a^*(\mu) \}$. Take: $\lambda_1 = m_3 - \epsilon$ \quad $\lambda_3 = m_3 + \epsilon$

$$\mathbb{E}_\mu [N_1(\tau_\delta)] \cdot d(\mu_1, m_2 - \epsilon) + \mathbb{E}_\mu [N_2(\tau_\delta)] \cdot d(\mu_2, m_2 + \epsilon) \geq \text{kl}(\delta, 1 - \delta)$$

$$\mathbb{E}_\mu [N_1(\tau_\delta)] \cdot d(\mu_1, m_3 - \epsilon) + \mathbb{E}_\mu [N_3(\tau_\delta)] \cdot d(\mu_3, m_3 + \epsilon) \geq \text{kl}(\delta, 1 - \delta)$$
Let \( \mu = (\mu_1, \ldots, \mu_K) \) and \( \lambda = (\lambda_1, \ldots, \lambda_K) \) be two elements of \( S \).

**Uniform \( \delta \)-correct Constraint** [Kaufmann, Cappé, Garivier '15]

If \( a^*(\mu) \neq a^*(\lambda) \), any \( \delta \)-correct algorithm satisfies

\[
\sum_{a=1}^{K} \mathbb{E}_\mu \left[ N_a(\tau_\delta) \right] d(\mu_a, \lambda_a) \geq kl(\delta, 1-\delta)
\]

where \( kl(p, q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q} \).

Let \( \text{Alt}(\mu) = \{ \lambda : a^*(\lambda) \neq a^*(\mu) \} \). Take: \( \lambda_1 = m_4 - \epsilon \) \hfill \( \lambda_4 = m_4 + \epsilon \)

\[
\begin{align*}
\mathbb{E}_\mu \left[ N_1(\tau_\delta) \right] d(\mu_1, m_2 - \epsilon) + \mathbb{E}_\mu \left[ N_2(\tau_\delta) \right] d(\mu_2, m_2 + \epsilon) & \geq kl(\delta, 1-\delta) \\
\mathbb{E}_\mu \left[ N_1(\tau_\delta) \right] d(\mu_1, m_3 - \epsilon) + \mathbb{E}_\mu \left[ N_3(\tau_\delta) \right] d(\mu_3, m_3 + \epsilon) & \geq kl(\delta, 1-\delta) \\
\mathbb{E}_\mu \left[ N_1(\tau_\delta) \right] d(\mu_1, m_4 - \epsilon) + \mathbb{E}_\mu \left[ N_4(\tau_\delta) \right] d(\mu_4, m_4 + \epsilon) & \geq kl(\delta, 1-\delta)
\end{align*}
\]
Lower-Bounding the Sample Complexity

Let \( \mu = (\mu_1, \ldots, \mu_K) \) and \( \lambda = (\lambda_1, \ldots, \lambda_K) \) be two elements of \( S \).

**Uniform \( \delta \)-correct Constraint** [Kaufmann, Cappé, Garivier ’15]

If \( a^*(\mu) \neq a^*(\lambda) \), any \( \delta \)-correct algorithm satisfies

\[
\sum_{a=1}^{K} \mathbb{E}_\mu \left[ N_a(\tau_\delta) \right] d(\mu_a, \lambda_a) \geq \text{kl}(\delta, 1 - \delta)
\]

where \( \text{kl}(p, q) = p \log \frac{p}{q} + (1 - p) \log \frac{1-p}{1-q} \).

Let \( \text{Alt}(\mu) = \{ \lambda : a^*(\lambda) \neq a^*(\mu) \} \).

\[
\inf_{\lambda \in \text{Alt}(\mu)} \sum_{a=1}^{K} \mathbb{E}_\mu \left[ N_a(\tau_\delta) \right] d(\mu_a, \lambda_a) \geq \text{kl}(\delta, 1 - \delta)
\]

\[
\mathbb{E}_\mu [\tau_\delta] \times \inf_{\lambda \in \text{Alt}(\mu)} \sum_{a=1}^{K} \frac{\mathbb{E}_\mu \left[ N_a(\tau_\delta) \right]}{\mathbb{E}_\mu [\tau_\delta]} d(\mu_a, \lambda_a) \geq \text{kl}(\delta, 1 - \delta)
\]

\[
\mathbb{E}_\mu [\tau_\delta] \times \left( \sup_{w \in \Sigma_K} \inf_{\lambda \in \text{Alt}(\mu)} \sum_{a=1}^{K} w_a d(\mu_a, \lambda_a) \right) \geq \text{kl}(\delta, 1 - \delta)
\]
**Theorem** [Garivier and Kaufmann 2016]

For any $\delta$-correct algorithm,

$$\mathbb{E}_{\mu}[\tau_\delta] \geq T^*(\mu)\operatorname{kl}(\delta, 1 - \delta),$$

where

$$T^*(\mu)^{-1} = \sup_{w \in \Sigma_K} \inf_{\lambda \in \text{Alt}(\mu)} \left( \sum_{a=1}^{K} w_a d(\mu_a, \lambda_a) \right).$$

- $\operatorname{kl}(\delta, 1 - \delta) \sim \log(1/\delta)$ when $\delta \to 0$, $\operatorname{kl}(\delta, 1 - \delta) \geq \log \left( 1/(2.4\delta) \right)$

→ the optimal proportions of arm draws are

$$w^*(\mu) = \arg\max_{w \in \Sigma_K} \inf_{\lambda \in \text{Alt}(\mu)} \left( \sum_{a=1}^{K} w_a d(\mu_a, \lambda_a) \right)$$

→ they do not depend on $\delta$
Given a parameter $\mu = (\mu_1, \ldots, \mu_K)$:

- the statistician chooses proportions of arm draws $w = (w_a)_a$
- the opponent chooses an alternative model $\lambda$
- the payoff is the minimal number $T = T(w, \lambda)$ of draws necessary to ensure that he does not violate the $\delta$-PAC constraint

$$\sum_{a=1}^{K} T w_a d(\mu_a, \lambda_a) \geq \text{kl}(\delta, 1 - \delta)$$

- $T^*(\mu) \text{kl}(\delta, 1 - \delta) = \text{value of the game}$
- $w^* = \text{optimal action for the statistician}$
PAC-BAI as a Game

Given a parameter $\mu = (\mu_1, \ldots, \mu_K)$ such that $\mu_1 > \mu_2 \geq \cdots \geq \mu_K$:

- the statistician chooses proportions of arm draws $w = (w_a)_a$
- the opponent chooses an arm $a \in \{2, \ldots, K\}$ and $\lambda = \arg \min_{\lambda} w_1 d(\mu_1, \lambda) + w_a d(\mu_a, \lambda)$
- the payoff is the minimal number $T = T(w, a, \delta)$ of draws necessary to ensure that

$$Tw_1 d(\mu_1, \lambda_a - \epsilon) + Tw_a d(\mu_a, \lambda_a + \epsilon) \geq \text{kl}(\delta, 1 - \delta)$$

that is $T(w, a, \delta) = \frac{\text{kl}(\delta, 1 - \delta)}{w_1 d(\mu_1, \lambda_a - \epsilon) + w_a d(\mu_a, \lambda_a + \epsilon)}$

- $T^*(\mu) \text{kl}(\delta, 1 - \delta) = \text{value of the game}$
- $w^* = \text{optimal action for the statistician}$
Properties of $T^*(\mu)$ and $w^*(\mu)$

1. **Unique solution**, solution of scalar equations only
2. For all $\mu \in S$, for all $a$, $w^*_a(\mu) > 0$
3. $w^*$ is **continuous** in every $\mu \in S$
4. If $\mu_1 > \mu_2 \geq \ldots \geq \mu_K$, one has $w^*_2(\mu) \geq \ldots \geq w^*_K(\mu)$
   (one may have $w^*_1(\mu) < w^*_2(\mu)$)
5. **Case of two arms** [Kaufmann, Cappé, Garivier '14]

$$\mathbb{E}_\mu[T_{\delta}] \geq \frac{\text{kl}(\delta, 1 - \delta)}{d^*(\mu_1, \mu_2)} .$$

where $d^*$ is the 'reversed' Chernoff information

$$d^*(\mu_1, \mu_2) := d(\mu_1, \mu^*) = d(\mu_2, \mu^*) .$$

6. **Gaussian arms**: algebraic equation but no simple formula for $K \geq 3$.

$$\sum_{a=1}^{K} \frac{2\sigma^2}{\Delta^2_a} \leq T^*(\mu) \leq 2 \sum_{a=1}^{K} \frac{2\sigma^2}{\Delta^2_a} .$$
The Track-and-Stop Strategy
Outline

Sequential Decision Problems

The Simple Bandit Model

Classical strategies

Lower Bound

The Track-and-Stop Strategy

  Sampling Rule

  Stopping Rule

  Optimality
Sampling rule: Tracking the optimal proportions

\[ \hat{\mu}(t) = (\hat{\mu}_1(t), \ldots, \hat{\mu}_K(t)) \]: vector of empirical means

Introducing

\[ U_t = \left\{ a : N_a(t) < \sqrt{t} \right\}, \]

the arm sampled at round \( t + 1 \) is

\[
A_{t+1} \in \begin{cases} 
\argmin_{a \in U_t} N_a(t) & \text{if } U_t \neq \emptyset \quad (\text{forced exploration}) \\
\argmax_{1 \leq a \leq K} t w_a^*(\hat{\mu}(t)) - N_a(t) & \text{tracking}
\end{cases}
\]

**Lemma**

Under the Tracking sampling rule,

\[
P_\mu \left( \lim_{t \to \infty} \frac{N_a(t)}{t} = w_a^*(\mu) \right) = 1.
\]
Sequential Decision Problems

The Simple Bandit Model

Classical strategies

Lower Bound

The Track-and-Stop Strategy

Sampling Rule

Stopping Rule

Optimality
High values of the Generalized Likelihood Ratio

\[ Z_{a,b}(t) := \log \frac{\max\{\lambda: \lambda_a \geq \lambda_b\} \ dP_\lambda(X_1, \ldots, X_t)}{\max\{\lambda: \lambda_a \leq \lambda_b\} \ dP_\lambda(X_1, \ldots, X_t)} \]

\[ = N_a(t) \ d(\hat{\mu}_a(t), \hat{\mu}_{a,b}(t)) + N_b(t) \ d(\hat{\mu}_b(t), \hat{\mu}_{a,b}(t)) \quad \text{if } \hat{\mu}_a(t) > \hat{\mu}_b(t) \]

\[ -Z_{b,a}(t) \quad \text{otherwise} \]

reject the hypothesis that \((\mu_a \leq \mu_b)\).

We stop when one arm is assessed to be significantly larger than all other arms, according to a GLR test:

\[ \tau_\delta = \inf \{ t \in \mathbb{N} : \exists a \in \{1, \ldots, K\}, \forall b \neq a, Z_{a,b}(t) > \beta(t, \delta) \} \]

\[ = \inf \left\{ t \in \mathbb{N} : Z(t) := \max_{a \in \{1, \ldots, K\}} \min_{b \neq a} Z_{a,b}(t) > \beta(t, \delta) \right\} \]

Chernoff stopping rule [Chernoff '59]

Two other possible interpretations of the stopping rule:

\( \rightarrow \) MDL:

\[ Z_{a,b}(t) = (N_a(t) + N_b(t))H(\hat{\mu}_{a,b}(t)) - \left[ N_a(t)H(\hat{\mu}_a(t)) + N_b(t)H(\hat{\mu}_b(t)) \right] \]
High values of the Generalized Likelihood Ratio

\[ Z_{a,b}(t) := \log \frac{\max\{\lambda: \lambda_a \geq \lambda_b\} \, dP_{\lambda}(X_1, \ldots, X_t)}{\max\{\lambda: \lambda_a \leq \lambda_b\} \, dP_{\lambda}(X_1, \ldots, X_t)} \]

reject the hypothesis that \((\mu_a \leq \mu_b)\).

We stop when one arm is assessed to be significantly larger than all other arms, according to a GLR test:

\[
\tau_\delta = \inf \left\{ t \in \mathbb{N} : Z(t) := \max_{a \in \{1, \ldots, K\}} \min_{b \neq a} Z_{a,b}(t) > \beta(t, \delta) \right\}
\]

Chernoff stopping rule [Chernoff '59]

Two other possible interpretations of the stopping rule:

→ plug-in complexity estimate: if \( F(w, \mu) := \inf_{\lambda \in \text{Alt}(\mu)} K \sum_{a=1}^K w_a \, d\left(\mu_a, \lambda_a\right) \), stop when \( Z(t) = t \, F\left(\frac{N_a(t)}{t}, \hat{\mu}(t)\right) \geq \beta(t, \delta) \) instead of the lower bound \( \frac{t}{T^*(\mu)} = t \, F(w^*, \mu) \geq \text{kl}(\delta, 1 - \delta) \).
**Theorem**

The Chernoff rule is $\delta$-PAC for $\beta(t, \delta) = \log \left( \frac{2(K-1)t}{\delta} \right)$

<table>
<thead>
<tr>
<th>Lemma</th>
</tr>
</thead>
<tbody>
<tr>
<td>If $\mu_a &lt; \mu_b$, whatever the sampling rule,</td>
</tr>
<tr>
<td>$P_{\mu} \left( \exists t \in \mathbb{N} : Z_{a,b}(t) &gt; \log \left( \frac{2t}{\delta} \right) \right) \leq \delta$</td>
</tr>
</tbody>
</table>

The proof uses:

- Barron’s lemma (change of distribution)
- and Krichevsky-Trofimov’s universal distribution

(very information-theoretic ideas)
Outline

Sequential Decision Problems

The Simple Bandit Model

Classical strategies

Lower Bound

The Track-and-Stop Strategy

Sampling Rule

Stopping Rule

Optimality
Asymptotic Optimality of the T&S strategy

**Theorem** [Garivier and Kaufmann 2016]

The Track-and-Stop strategy, that uses

- the Tracking sampling rule
- the Chernoff stopping rule with $\beta(t, \delta) = \log \left( \frac{2(K-1)t}{\delta} \right)$
- and recommends $\hat{a}_{\tau_\delta} = \arg\max_{a=1, \ldots, K} \hat{\mu}_a(\tau_\delta)$

is $\delta$-PAC for every $\delta \in (0, 1)$ and satisfies

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_{\mu} [\tau_\delta]}{\log(1/\delta)} = T^*(\mu).$$
Why is the T&S Strategy asymptotically Optimal?

Chernoff's stopping rule

- $tF(w^*, \mu)$
- $z(t)$
- $\log(1/\delta)$
- $\beta(t, \delta)$

- $w^*$
- $\hat{w}(t)$
- $\mu$
- $\hat{\mu}(t)$

<table>
<thead>
<tr>
<th>Arm</th>
<th>WH</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.50</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.00</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0.00</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>0.50</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0.00</td>
<td>1</td>
</tr>
</tbody>
</table>
Why is the T&S Strategy asymptotically Optimal?

Chernoff's stopping rule

\[ t_F(w^2, \mu) \]

\[ z(t) \]

\[ \log(1/\delta) \]

\[ \beta(t, \delta) \]

\[ w^* \]

\[ \hat{w}(t) \]

\[ \mu \]

\[ \hat{\mu}(t) \]

<table>
<thead>
<tr>
<th>arm 1</th>
<th>arm 2</th>
<th>arm 3</th>
<th>arm 4</th>
<th>arm 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>wh= 0.39</td>
<td>wh= 0.21</td>
<td>wh= 0.32</td>
<td>wh= 0.05</td>
<td>wh= 0.03</td>
</tr>
</tbody>
</table>
Why is the T&S Strategy asymptotically Optimal?

Chernoff's stopping rule

\[ X_t = \text{tF}(w^X, \mu) \quad \log(1/\delta) \quad \beta(t, \delta) \]

\[ w^* \quad \hat{w}(t) \]

\[ \mu \quad \hat{\mu}(t) \]

<table>
<thead>
<tr>
<th>Arm</th>
<th>( \text{wh} )</th>
<th>( N )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.48</td>
<td>53</td>
</tr>
<tr>
<td>2</td>
<td>0.02</td>
<td>49</td>
</tr>
<tr>
<td>3</td>
<td>0.48</td>
<td>52</td>
</tr>
<tr>
<td>4</td>
<td>0.02</td>
<td>48</td>
</tr>
<tr>
<td>5</td>
<td>0.00</td>
<td>48</td>
</tr>
</tbody>
</table>
Why is the T&S Strategy asymptotically Optimal?

Chernoff's stopping rule

\[ tF(w^*, \mu) \]

\[ \log(1/\delta) \]

\[ \beta(t, \delta) \]

\[ w^* \]

\[ \hat{w}(t) \]

\[ \mu \]

\[ \hat{\mu}(t) \]

arm 1
wh = 0.41
N = 108

arm 2
wh = 0.07
N = 71

arm 3
wh = 0.38
N = 78

arm 4
wh = 0.13
N = 59

arm 5
wh = 0.01
N = 59
Why is the T&S Strategy asymptotically Optimal?

Chernoff's stopping rule

\[ \frac{tF(w^*, \mu)}{Z(t)} \]

\[ \log(1/\delta) \]

\[ \beta(t, \delta) \]

\[ w^* \]

\[ \hat{w}(t) \]

\[ \mu \]

\[ \hat{\mu}(t) \]

<table>
<thead>
<tr>
<th>Arm</th>
<th>( wh )</th>
<th>( N )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.36</td>
<td>156</td>
</tr>
<tr>
<td>2</td>
<td>0.16</td>
<td>93</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
<td>106</td>
</tr>
<tr>
<td>4</td>
<td>0.19</td>
<td>77</td>
</tr>
<tr>
<td>5</td>
<td>0.04</td>
<td>68</td>
</tr>
</tbody>
</table>
Why is the T&S Strategy asymptotically Optimal?

Chernoff's stopping rule

- $tF(w^*, \mu)$
- $\log(1/\delta)$
- $\beta(t, \delta)$
- $\hat{w}(t)$
- $\hat{\mu}(t)$
- $w^*$

- arm 1: $\text{wh}=0.36$, $N=193$
- arm 2: $\text{wh}=0.13$, $N=134$
- arm 3: $\text{wh}=0.27$, $N=141$
- arm 4: $\text{wh}=0.18$, $N=82$
- arm 5: $\text{wh}=0.05$, $N=75$
Why is the T&S Strategy asymptotically Optimal?

Chernoff's stopping rule

\[ t_F(w^*, \mu) \]
\[ \log(1/\delta) \]
\[ \beta(t, \delta) \]

\[ w^* \]
\[ \hat{w}(t) \]
\[ \mu \]
\[ \hat{\mu}(t) \]

arm 1: wh = 0.36, N = 231
arm 2: wh = 0.13, N = 164
arm 3: wh = 0.23, N = 170
arm 4: wh = 0.24, N = 102
arm 5: wh = 0.04, N = 83
Why is the T&S Strategy asymptotically Optimal?

Chernoff's stopping rule

\[ z(t) \quad t_{F}(\hat{w}^{*}, \mu) \quad \log(1/\delta) \quad \beta(t, \delta) \]

\[ \hat{w}(t) \quad w^{*} \quad \hat{\mu}(t) \quad \mu \]

<table>
<thead>
<tr>
<th>arm 1</th>
<th>arm 2</th>
<th>arm 3</th>
<th>arm 4</th>
<th>arm 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>wh = 0.36</td>
<td>wh = 0.13</td>
<td>wh = 0.27</td>
<td>wh = 0.20</td>
<td>wh = 0.03</td>
</tr>
<tr>
<td>N = 275</td>
<td>N = 182</td>
<td>N = 200</td>
<td>N = 129</td>
<td>N = 89</td>
</tr>
</tbody>
</table>
Why is the T&S Strategy asymptotically Optimal?

Chernoff's stopping rule

\[ t_F(w^*, \mu) \]

\[ \log(1/\delta) \]

\[ \beta(t, \delta) \]

\[ w^* \]

\[ \hat{w}(t) \]

\[ \mu \]

\[ \hat{\mu}(t) \]

arm 1: wh = 0.36, N = 337
arm 2: wh = 0.19, N = 199
arm 3: wh = 0.26, N = 225
arm 4: wh = 0.15, N = 144
arm 5: wh = 0.04, N = 95
Why is the T&S Strategy asymptotically Optimal?

Chernoff's stopping rule

$w^*$ $\hat{w}(t)$

$\mu$ $\hat{\mu}(t)$

$N = 382$ $N = 243$ $N = 250$ $N = 149$ $N = 101$
Why is the T&S Strategy asymptotically Optimal?

Chernoff's stopping rule

$F(w^*, \mu)$
$\log(1/\delta)$
$\beta(t, \delta)$

$w^*$
$\hat{w}(t)$

$\mu$
$\hat{\mu}(t)$

<table>
<thead>
<tr>
<th>Arm</th>
<th>$wh$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.36</td>
<td>424</td>
</tr>
<tr>
<td>2</td>
<td>0.17</td>
<td>288</td>
</tr>
<tr>
<td>3</td>
<td>0.27</td>
<td>281</td>
</tr>
<tr>
<td>4</td>
<td>0.14</td>
<td>150</td>
</tr>
<tr>
<td>5</td>
<td>0.06</td>
<td>107</td>
</tr>
</tbody>
</table>
Why is the T&S Strategy asymptotically Optimal?

Chernoff's stopping rule

- $\tilde{w}(t)$
- $\hat{\mu}(t)$
- $\mu$
- $w^*$
- $\beta(t, \delta)$
- $\log(1/\delta)$
- $\tilde{z}(t)$
- $F_w(\mu)$

Arm 1:
- $w_h = 0.36$
- $N = 462$

Arm 2:
- $w_h = 0.18$
- $N = 323$

Arm 3:
- $w_h = 0.28$
- $N = 320$

Arm 4:
- $w_h = 0.13$
- $N = 158$

Arm 5:
- $w_h = 0.06$
- $N = 112$
Why is the T&S Strategy asymptotically Optimal?

Chernoff's stopping rule

\[ X \sim \text{Beta}(\mu, \log(1/\delta)) \]

\[ w^* \succeq \hat{w}(t) \]

\[ \mu \succeq \hat{\mu}(t) \]

arm 1: wh = 0.35, N = 506
arm 2: wh = 0.20, N = 365
arm 3: wh = 0.26, N = 352
arm 4: wh = 0.12, N = 160
arm 5: wh = 0.06, N = 117
Why is the T&S Strategy asymptotically Optimal?

Chernoff's stopping rule

\[ X_{2(t)}^{tF(w^*, \mu)} \leq \beta(t, \delta) \]

\[ w^* \quad \hat{w}(t) \]

\[ \mu \quad \hat{\mu}(t) \]

- arm 1: wh = 0.35, N = 559
- arm 2: wh = 0.23, N = 398
- arm 3: wh = 0.23, N = 377
- arm 4: wh = 0.12, N = 170
- arm 5: wh = 0.07, N = 121
Why is the T&S Strategy asymptotically Optimal?

Chernoff's stopping rule

\[ \log(1/\delta) \]

\[ tF(w^*, t) \]

\[ X(t) \]

\[ X(\beta(t, \delta)) \]

\[ w^* \]

\[ \hat{w}(t) \]

\[ \mu \]

\[ \hat{\mu}(t) \]

\[ \text{arm 1: } wh = 0.36, N = 604 \]

\[ \text{arm 2: } wh = 0.22, N = 428 \]

\[ \text{arm 3: } wh = 0.24, N = 413 \]

\[ \text{arm 4: } wh = 0.12, N = 179 \]

\[ \text{arm 5: } wh = 0.06, N = 126 \]
Why is the T&S Strategy asymptotically Optimal?

Chernoff's stopping rule

\[ \log\left(\frac{1}{\delta}\right) \frac{tF(w^*, \mu)}{\beta(t, \delta)} \]

\[ w^* \quad \hat{w}(t) \]

\[ \mu \quad \hat{\mu}(t) \]

arm 1: wh = 0.36, N = 657
arm 2: wh = 0.22, N = 454
arm 3: wh = 0.25, N = 442
arm 4: wh = 0.11, N = 192
arm 5: wh = 0.06, N = 130
Why is the T&S Strategy asymptotically Optimal?

Chernoff's stopping rule

\[ \log(1/\delta) \]

\[ \frac{\log(1/\delta)}{\beta(t, \delta)} \]

\[ X_1 \]

\[ X_2 \]

\[ w^* \]

\[ \hat{w}(t) \]

\[ \mu \]

\[ \hat{\mu}(t) \]

\[ \text{arm 1: } \text{wh} = 0.36, \text{N} = 699 \]

\[ \text{arm 2: } \text{wh} = 0.21, \text{N} = 483 \]

\[ \text{arm 3: } \text{wh} = 0.26, \text{N} = 482 \]

\[ \text{arm 4: } \text{wh} = 0.12, \text{N} = 201 \]

\[ \text{arm 5: } \text{wh} = 0.06, \text{N} = 135 \]
Why is the T&S Strategy asymptotically Optimal?

Chernoff's stopping rule

\[ \log(1/\delta) \leq \frac{t}{\beta(t, \delta)} \]

\[ w^* \geq \hat{w}(t) \]

\[ \mu \geq \hat{\mu}(t) \]

<table>
<thead>
<tr>
<th>Arm</th>
<th>win</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.36</td>
<td>759</td>
</tr>
<tr>
<td>2</td>
<td>0.24</td>
<td>515</td>
</tr>
<tr>
<td>3</td>
<td>0.24</td>
<td>507</td>
</tr>
<tr>
<td>4</td>
<td>0.10</td>
<td>205</td>
</tr>
<tr>
<td>5</td>
<td>0.06</td>
<td>139</td>
</tr>
</tbody>
</table>
Why is the T&S Strategy asymptotically Optimal?

Chernoff's stopping rule

\[ \log(1/\delta) \]

\[ \chi^2_{(t, \delta)} \]

\[ w^* \]

\[ \hat{w}(t) \]

\[ \mu \]

\[ \hat{\mu}(t) \]

<table>
<thead>
<tr>
<th>Arm</th>
<th>(\text{wh} = )</th>
<th>(N = )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.36</td>
<td>803</td>
</tr>
<tr>
<td>2</td>
<td>0.24</td>
<td>547</td>
</tr>
<tr>
<td>3</td>
<td>0.24</td>
<td>544</td>
</tr>
<tr>
<td>4</td>
<td>0.09</td>
<td>213</td>
</tr>
<tr>
<td>5</td>
<td>0.06</td>
<td>143</td>
</tr>
</tbody>
</table>
Why is the T&S Strategy asymptotically Optimal?

Chernoff's stopping rule

\[ \log(1/\delta) \]

\[ \beta(t, \delta) \]

\[ w^* \]

\[ \hat{w}(t) \]

\[ \mu \]

\[ \hat{\mu}(t) \]

<table>
<thead>
<tr>
<th>Arm</th>
<th>wh</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.36</td>
<td>811</td>
</tr>
<tr>
<td>2</td>
<td>0.24</td>
<td>550</td>
</tr>
<tr>
<td>3</td>
<td>0.24</td>
<td>546</td>
</tr>
<tr>
<td>4</td>
<td>0.10</td>
<td>213</td>
</tr>
<tr>
<td>5</td>
<td>0.07</td>
<td>145</td>
</tr>
</tbody>
</table>
Sketch of proof (almost-sure convergence only)

- forced exploration $\implies N_a(t) \to \infty$ a.s. for all $a \in \{1, \ldots, K\}$
- $\hat{\mu}(t) \to \mu$ a.s.
- $\mathbf{w}^*(\hat{\mu}(t)) \to \mathbf{w}^*$ a.s.
- tracking rule: $\frac{N_a(t)}{t} \to \mathbf{w}^*_a$ a.s.

- but the mapping $F : (\mu', w) \mapsto \inf_{\lambda \in \text{Alt}(\mu')} \sum_{a=1}^{K} w_a d(\mu'_a, \lambda_a)$ is continuous at $(\mu, \mathbf{w}^*(\mu))$:

$Z(t) = t \times F\left(\hat{\mu}(t), (N_a(t)/t)_a^K\right) \sim t \times F(\mu, \mathbf{w}^*) = t \times T^*(\mu)^{-1}$

and for every $\epsilon > 0$ there exists $t_0$ such that

$t \geq t_0 \implies Z(t) \geq t \times (1 + \epsilon)^{-1} T^*(\mu)^{-1}$

$\implies$ Thus $\tau_\delta \leq t_0 \wedge \inf \left\{ t \in \mathbb{N} : (1 + \epsilon)^{-1} T^*(\mu)^{-1} t \geq \log(2(K - 1)t/\delta) \right\}$

and $\limsup_{\delta \to 0} \frac{\tau_\delta}{\log(1/\delta)} \leq (1 + \epsilon) T^*(\mu)$ a.s.
Numerical Experiments

- $\mu_1 = [0.5 \ 0.45 \ 0.43 \ 0.4] \Rightarrow w^*(\mu_1) = [0.42 \ 0.39 \ 0.14 \ 0.06]$
- $\mu_2 = [0.3 \ 0.21 \ 0.2 \ 0.19 \ 0.18] \Rightarrow w^*(\mu_2) = [0.34 \ 0.25 \ 0.18 \ 0.13 \ 0.10]$

In practice, set the threshold to $\beta(t, \delta) = \log \left( \frac{\log(t)+1}{\delta} \right)$ ($\delta$-PAC OK)

<table>
<thead>
<tr>
<th></th>
<th>Track-and-Stop</th>
<th>Chernoff-Racing</th>
<th>KL-LUCB</th>
<th>KL-Racing</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td>4052</td>
<td>4516</td>
<td>8437</td>
<td>9590</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>1406</td>
<td>3078</td>
<td>2716</td>
<td>3334</td>
</tr>
</tbody>
</table>

Table 1: Expected number of draws $E_{\mu}[\tau_\delta]$ for $\delta = 0.1$, averaged over $N = 3000$ experiments.

Empirically good even for ‘large’ values of the risk $\delta$

Racing is sub-optimal in general, because it plays $w_1 = w_2$

LUCB is sub-optimal in general, because it plays $w_1 = 1/2$
Perspectives

For best arm identification, we showed that

$$\limsup_{\delta \to 0} \inf_{\delta\text{-correct strategy}} \frac{E_{\mu}[T_\delta]}{\log(1/\delta)} = \left( \sup_{w \in \Sigma} \inf_{\lambda \in \text{Alt}(\mu)} \left( \sum_{a=1}^{K} w_a d(\mu_a, \lambda_a) \right) \right)^{-1}$$

and provided an efficient strategy asymptotically matching this bound.

Future work:

- * anytime stopping → gives a confidence level
- ** find an $\epsilon$-optimal arm (PAC-setting)
- * find the $m$-best arms
- *** design and analyze more stable algorithm (hint: optimism)
- *** give a simple algorithm with a finite-time analysis
  candidate: play action maximizing the expected increase of $Z(t)$
- *** extend to structured (dose, MCTS) and continuous settings