The multi-armed bandit model and the optimal design of clinical trials: benefits and drawbacks

Sofía S. Villar

Research Associate
MRC Biostatistics Unit, Cambridge
Department of Mathematics and Statistics, Lancaster University

Bilbao
July 17th, 2013
Outline

Multi-armed Bandit Problems

Clinical Trials

Bandits and Clinical Trials Design
What is a multi-armed bandit problem?
What is a multi-armed bandit problem?
What is a multi-armed bandit problem?
A three-armed Bandit

A greedy gambler’s problem!

Red →

Green →

Blue →

Rewards = 7 + β8 + β²6 + β³5 + β⁴4 + ... 

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Gittins’ *Classic* Bandits

Rewards = $7 + \beta 8 + \beta^2 6 + \beta^3 5 + \beta^4 4 + \ldots$

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Gittins’ Classic Bandits

Rewards = \[ 7 + \beta 8 + \beta^2 6 + \beta^3 5 + \beta^4 4 + \ldots \]

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Gittins’ Classic Bandits

Rewards = 7 + \beta 8 + \beta^2 6 + \beta^3 5 + \beta^4 4 + \ldots

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Gittins’ Classic Bandits

Rewards = \( 7 + \beta 8 + \beta^2 6 + \beta^3 5 + \beta^4 4 + \ldots \)

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Red → 7, 8, 1, 0, …
Green →
Blue →

Gittins’ *Classic* Bandits

Rewards = 7 + β8 + β²6 + β³5 + β⁴4 + …

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit  
A greedy gambler’s problem!

Red → 7, 8, 1, 0, …

Green → 4, 2, 3, 6, …

Blue →

Gittins’ *Classic* Bandits

\[ \text{Rewards} = 7 + \beta^8 + \beta^{26} + \beta^{35} + \beta^{44} + \ldots \]

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Red → 7, 8, 1, 0, ...
Green → 4, 2, 3, 6, ...
Blue →

Gittins’ Classic Bandits

Rewards = 7 + β⁸ + β²⁶ + β³⁵ + β⁴⁴ + ...

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Red $\rightarrow$ 7, 8, 1, 0, …
Green $\rightarrow$ 4, 2, 3, 6, …
Blue $\rightarrow$

Gittins’ Classic Bandits

Rewards$= 7 + \beta 8 + \beta^2 6 + \beta^3 5 + \beta^4 4 + \ldots$

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Red $\rightarrow$ 7, 8, 1, 0, ...

Green $\rightarrow$ 4, 2, 3, 6, ...

Blue $\rightarrow$

Gittins’ *Classic* Bandits

Rewards$= 7 + \beta 8 + \beta^2 6 + \beta^3 5 + \beta^4 4 + \ldots$

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Red $\rightarrow 7, 8, 1, 0, \ldots$

Green $\rightarrow 4, 2, 3, 6, \ldots$

Blue $\rightarrow 6, 5, 1, 6, \ldots$

Gittins’ Classic Bandits

Rewards $= 7 + \beta 8 + \beta^2 6 + \beta^3 5 + \beta^4 4 + \ldots$

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Red  →  7, 8, 1, 0, ...  
Green → 4, 2, 3, 6, ...  
Blue  → 6, 5, 1, 6, ...  

Gittins’ Classic Bandits

Rewards = 7 + β 8 + β 2 6 + β 3 5 + β 4 4 + ... 

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Red → 7, 8, 1, 0, ...
Green → 4, 2, 3, 6, ...
Blue → 6, 5, 1, 6, ...

Gittins’ Classic Bandits

Rewards = 7 + β8 + β²6 + β³5 + β⁴4 + ...

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Red → 7, 8, 1, 0, …
Green → 4, 2, 3, 6, …
Blue → 6, 5, 1, 6, …

Gittins’ Classic Bandits

Rewards = 7 + β8 + β26 + β35 + β44 + …

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit

A greedy gambler’s problem!

Red $\rightarrow$ 7, 8, 1, 0, …

Green $\rightarrow$ 4, 2, 3, 6, …

Blue $\rightarrow$ 6, 5, 1, 6, …

Gittins’ Classic Bandits

Rewards$= 7 + \beta 8 + \beta^2 6 + \beta^3 5 + \beta^4 4 + \ldots$

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Gittins’ Classic Bandits

Rewards = $7 + \beta 8 + \beta^2 6 + \beta^3 5 + \beta^4 4 + \ldots$

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Red $\rightarrow 7, 8, 1, 0, \ldots$
Green $\rightarrow 4, 2, 3, 6, \ldots$
Blue $\rightarrow 6, 5, 1, 6, \ldots$

Gittins’ Classic Bandits

Rewards $= 7 + \beta 8 + \beta^2 6 + \beta^3 5 + \beta^4 4 + \ldots$

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Red $\rightarrow$ 7, 8, 1, 0, …
Green $\rightarrow$ 4, 2, 3, 6, …
Blue $\rightarrow$ 6, 5, 1, 6, …

Gittins’ Classic Bandits

Rewards $= 7 + \beta 8 + \beta^2 6 + \beta^3 5 + \beta^4 4 + \ldots$

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Rewards = $7 + \beta^8 + \beta^2 6 + \beta^3 5 + \beta^4 4 + \ldots$

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Red $\rightarrow$ 7, 8, 1, 0, ...  
Green $\rightarrow$ 4, 2, 3, 6, ...  
Blue $\rightarrow$ 6, 5, 1, 6, ... 

Gittins’ Classic Bandits

Rewards = $7 + \beta 8 + \beta^2 6 + \beta^3 5 + \beta^4 4 + ...$

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit

A greedy gambler’s problem!

Red → 7, 8, 1, 0, ...  
Green → 4, 2, 3, 6, ...  
Blue → 6, 5, 1, 6, ...  

Rewards = 7 + β^8 + β^2 6 + β^3 5 + β^4 4 + ...  

How to play to make the maximum total (expected discounted) profit?  

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Red → 8, 1, 0, ...
Green → 4, 2, 3, 6, ... → 7,
Blue → 6, 5, 1, 6, ...

Gittins’ Classic Bandits

Rewards = 7 + β8 + β²6 + β³5 + β⁴4 + ... 

How to play to make the maximum total (expected discounted) profit? 

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Gittins’ Classic Bandits

Red → , 1, 0, ...

Green → 4, 2, 3, 6, ... → 7, 8,

Blue → 6, 5, 1, 6, ...

Rewards = 7 + β8 + β26 + β35 + β44 + ... 

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Red $\rightarrow$, $1, 0, \ldots$
Green $\rightarrow 4, 2, 3, 6, \ldots \rightarrow 7, 8, 6, 5,$
Blue $\rightarrow$, $1, 6, \ldots$

Rewards$= 7 + \beta 8 + \beta^2 6 + \beta^3 5 + \beta^4 4 + \ldots$

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Gittins’ Classic Bandits

Red $\rightarrow$, 1, 0, ...
Green $\rightarrow$ 2, 3, 6, ... $\rightarrow$ 7, 8, 6, 5, 4, ...
Blue $\rightarrow$, 1, 6, ...

Rewards$= 7 + \beta 8 + \beta^2 6 + \beta^3 5 + \beta^4 4 + \ldots$

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Red $\rightarrow \ldots, 1, 0, \ldots$

Green $\rightarrow \ldots, 2, 3, 6, \ldots \rightarrow 7, 8, 6, 5, 4, \ldots$

Blue $\rightarrow \ldots, 1, 6, \ldots$

Rewards $= 7 + \beta 8 + \beta^2 6 + \beta^3 5 + \beta^4 4 + \ldots$

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Red $\rightarrow$ , , 1, 0, …
Green $\rightarrow$ , 2, 3, 6, … $\rightarrow$ 7, 8, 6, 5, 4 …
Blue $\rightarrow$ , , 1, 6, …

Gittins’ *Classic* Bandits

Rewards$=7 + \beta 8 + \beta^2 6 + \beta^3 5 + \beta^4 4 + \ldots$ ($0 \leq \beta < 1$)

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Red → , 1, 0, …
Green → 2, 3, 6, … → 7, 8, 6, 5, 4, …
Blue → , 1, 6, …

Gittins’ Classic Bandits

Rewards = 7 + β8 + β6 + β5 + β4 + … (0 ≤ β < 1)

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A three-armed Bandit
A greedy gambler’s problem!

Red \rightarrow, 1, 0, \ldots

Green \rightarrow, 2, 3, 6, \ldots \rightarrow 7, 8, 6, 5, 4, \ldots

Blue \rightarrow, 1, 6, \ldots

Gittins’ Classic Bandits

Rewards\begin{align*} &= 7 + \beta 8 + \beta^2 6 + \beta^3 5 + \beta^4 4 + \ldots \quad (0 \leq \beta < 1) \\
\end{align*}

How to play to make the maximum total (expected discounted) profit?

Optimal dynamic resource allocation model: job scheduling, research planning, etc.
A multi-armed Bandit
A collection of MCPs

$K$ arms, each one defining a Markov Control Process (MCPs) consisting of:

- **State variable:** $x_k(t) \in \mathbb{X}_k$
- **Action space:** $A_k = \{0, 1\}$
- **Immediate rewards:** one-period reward function
  $R_k(x_k(t), a_k(t)) \triangleq x_k(t)a_k(t)$
- **One-period Dynamics:** Markovian transition probabilities
  $p_k(x, y) \triangleq \mathbb{P}_k(x \rightarrow y)$ for every $x, y \in \mathbb{X}_k$ (when $a_k(t) = 1$)
A multi-armed Bandit
A collection of MCPs

$K$ arms, each one defining a Markov Control Process (MCPs) consisting of:

- **State variable:** $x_k(t) \in X_k$

- **Action space:** $A_k = \{0, 1\}$

- **Immediate rewards:** one-period reward function
  $R_k(x_k(t), a_k(t)) \triangleq x_k(t)a_k(t)$

- **One-period Dynamics:** Markovian transition probabilities
  $p_k(x, y) \triangleq P_k(x \rightarrow y)$ for every $x, y \in X_k$ (when $a_k(t) = 1$)
A multi-armed Bandit
A collection of MCPs

$K$ arms, each one defining a Markov Control Process (MCPs) consisting of:

- **State variable:** $x_k(t) \in \mathbb{X}_k$
- **Action space:** $A_k = \{0, 1\}$
- **Immediate rewards:** one-period reward function $R_k(x_k(t), a_k(t)) \triangleq x_k(t)a_k(t)$
- **One-period Dynamics:** Markovian transition probabilities $p_k(x, y) \triangleq \mathbb{P}_k(x \rightarrow y)$ for every $x, y \in \mathbb{X}_k$ (when $a_k(t) = 1$)
A multi-armed Bandit
A collection of MCPs

\( K \) arms, each one defining a Markov Control Process (MCPs) consisting of:

- **State variable:** \( x_k(t) \in \mathbb{X}_k \)
- **Action space:** \( A_k = \{0, 1\} \)
- **Immediate rewards:** one-period reward function \( R_k(x_k(t), a_k(t)) \triangleq x_k(t)a_k(t) \)
- **One-period Dynamics:** Markovian transition probabilities \( p_k(x, y) \triangleq \mathbb{P}_k(x \rightarrow y) \) for every \( x, y \in \mathbb{X}_k \) (when \( a_k(t) = 1 \))
A multi-armed Bandit
A collection of MCPs

$K$ arms, each one defining a Markov Control Process (MCPs) consisting of:

- **State variable:** $x_k(t) \in \mathbb{X}_k$
- **Action space:** $A_k = \{0, 1\}$
- **Immediate rewards:** one-period reward function $R_k(x_k(t), a_k(t)) \triangleq x_k(t)a_k(t)$
- **One-period Dynamics:** Markovian transition probabilities $p_k(x, y) \triangleq \mathbb{P}_k(x \rightarrow y)$ for every $x, y \in \mathbb{X}_k$ (when $a_k(t) = 1$)
A multi-armed Bandit
A collection of MCPs

\( K \) arms, each one defining a Markov Control Process (MCPs) consisting of:

- **State variable:** \( x_k(t) \in X_k \)
- **Action space:** \( A_k = \{0, 1\} \)
- **Immediate rewards:** one-period reward function \( R_k(x_k(t), a_k(t)) \triangleq x_k(t)a_k(t) \)
- **One-period Dynamics:** Markovian transition probabilities \( p_k(x, y) \triangleq \mathbb{P}_k(x \rightarrow y) \) for every \( x, y \in X_k \) (when \( a_k(t) = 1 \))

**Objective:** maximize the expected total \( \beta \)-discounted reward:

\[
\mathbb{E} \left[ \sum_{t=0}^{\infty} R_k(x_k(t), a_k(t)) \beta^t \right] \quad (1)
\]
The Dynamic Programming (DP) equation is:

- Solving (2) by the traditional techniques is hindered by the size of the state space.
- If each bandit has $E$ possible states, then the problem's state space is $E^K$.
- Closed-form solutions are rare, so the backward-induction algorithm is also affected by the (truncated) final horizon $T$. 
The Dynamic Programming (DP) equation is:

\[ V^*(x_1, \ldots, x_K) = \max_k \left\{ x_k + \beta \sum_y P_k(x_k, y) V^*(\ldots, x_{k-1}, y, x_{k+1}, \ldots) \right\} \]

(2)

- Solving (2) by the traditional techniques is hindered by the size of the state space.
- If each bandit has \( E \) possible states, then the problem’s state space is \( E^K \).
- Closed-form solutions are rare, so the backward-induction algorithm is also affected by the (truncated) final horizon \( T \).
The optimal solution via Dynamic programming

The curse of dimensionality

The Dynamic Programming (DP) equation is:

\[ V^*(x_1, \ldots, x_K) = \max_k \left\{ x_k + \beta \sum_y P_k(x_k, y) V^*(\ldots, x_{k-1}, y, x_{k+1}, \ldots) \right\} \]

(2)

- Solving (2) by the traditional techniques is hindered by the size of the state space.
- If each bandit has \( E \) possible states, then the problem’s state space is \( E^K \).
- Closed-form solutions are rare, so the backward-induction algorithm is also affected by the (truncated) final horizon \( T \).
The Dynamic Programming (DP) equation is:

\[
V^*(x_1, \ldots, x_K) = \max_k \{ x_k + \beta \sum_y P_k(x_k, y) V^*(\ldots, x_{k-1}, y, x_{k+1}, \ldots) \} 
\]

(2)

- Solving (2) by the traditional techniques is hindered by the size of the state space.
- If each bandit has \( E \) possible states, then the problem’s state space is \( E^K \).
- Closed-form solutions are rare, so the backward-induction algorithm is also affected by the (truncated) final horizon \( T \).
The Dynamic Programming (DP) equation is:

\[ V^*(x_1, \ldots, x_K) = \max_k \left\{ x_k + \beta \sum_y P_k(x_k, y) V^*(\ldots, x_{k-1}, y, x_{k+1}, \ldots) \right\} \]  

(2)

- Solving (2) by the traditional techniques is hindered by the size of the state space.
- If each bandit has \( E \) possible states, then the problem’s state space is \( E^K \).
- Closed-form solutions are rare, so the backward-induction algorithm is also affected by the (truncated) final horizon \( T \).
The Dynamic Programming (DP) equation is:

\[ V^*(x_1, \ldots, x_K) = \max_k \{ x_k + \beta \sum_y P_k(x_k, y) V^*(\ldots, x_{k-1}, y, x_{k+1}, \ldots) \} \]

(2)

- Solving (2) by the traditional techniques is hindered by the size of the state space.
- If each bandit has \( E \) possible states, then the problem’s state space is \( E^K \).
- Closed-form solutions are rare, so the backward-induction algorithm is also affected by the (truncated) final horizon \( T \).
A one-armed Bernoulli Bandit

Computational Complexity

Simplest possible scenario:

Only two possible outcomes and 1 arm.

Blue →

Time Horizon: $T$ (say, e.g., 100 plays).

Assume a Bernoulli distribution (with unknown parameter $p$).
Simplest possible scenario:

Only two possible outcomes and 1 arm.

Blue →

Time Horizon: $T$ (say, e.g., 100 plays).

Assume a Bernoulli distribution (with unknown parameter $p$).
A one-armed Bernoulli Bandit

Computational Complexity

Simplest possible scenario:

Only two possible outcomes and 1 arm.

Blue →

Time Horizon: \( T \) (say, e.g., 100 plays).

Assume a Bernoulli distribution (with unknown parameter \( p \)).
Simplest possible scenario:

Only two possible outcomes and 1 arm.

Blue →

Time Horizon: \( T \) (say, e.g., 100 plays).

Assume a Bernoulli distribution (with unknown parameter \( p \)).
A one-armed Bernoulli Bandit

Computational Complexity

Simplest possible scenario:

Only two possible outcomes and 1 arm.

Blue $\rightarrow$

Time Horizon: $T$ (say, e.g., 100 plays).

Assume a Bernoulli distribution (with unknown parameter $p$).
A one-armed Bernoulli Bandit
Computational Complexity

Simplest possible scenario:
Only two possible outcomes and 1 arm.
Blue →

Time Horizon: $T$ (say, e.g., 100 plays).
Assume a Bernoulli distribution (with unknown parameter $p$).
Simplest possible scenario:

Only two possible outcomes and 1 arm.

Blue $\rightarrow$ 1, 0, 1, 1, 1, 0, 0, ... 

Time Horizon: $T$ (say, e.g., 100 plays).

Assume a Bernoulli distribution (with unknown parameter $p$).
Simplest possible scenario:
Only two possible outcomes and 1 arm.
Blue $\rightarrow 1, 0, 1, 1, 1, 0, 0, \ldots$

Time Horizon: $T$ (say, e.g., 100 plays).

Assume a Bernoulli distribution (with unknown parameter $p$).
Simplest possible scenario:

Only two possible outcomes and 1 arm.

Blue $\rightarrow$ 1, 0, 1, 1, 1, 0, 0, \ldots

Time Horizon: $T$ (say, e.g., 100 plays).

Assume a Bernoulli distribution (with unknown parameter $p$).
A one-armed bernoulli Bandit

Computational Complexity

\[ t=0 \]
A one-armed bernoulli Bandit
Computational Complexity
A one-armed bernoulli Bandit
Computational Complexity

\[ t = 0 \]

\[ 1 \]

\[ t = 0 \]

\[ 0 \]
A one-armed bernoulli Bandit

Computational Complexity

$t=0$

1

0

0,1

1

0,1,1

1,1,0

1,0

0,1,0

0,0,1

0,0,0

1,1,1
A one-armed Bernoulli Bandit

Computational Complexity

\[ t=0 \]

\[ 1 \]

\[ 0 \]

\[ 0,0 \]

\[ 0,1 \]

\[ 1 \]

\[ 1,1 \]

\[ 1,1,1 \]

\[ 1,1,0 \]

\[ 1,0 \]

\[ 1,0,1 \]

\[ 1,0,0 \]
A one-armed bernoulli Bandit
Computational Complexity

\[
\begin{array}{c}
\text{t=0} \\
1 \\
0
\end{array}
\quad
\begin{array}{c}
1,1 \\
0,0 \\
0,1
\end{array}
\]
A one-armed Bernoulli Bandit

Computational Complexity

\[ t = 0 \]

0

0,0

0,1

1

1,0

1,1

1,0

1,1
A one-armed Bernoulli Bandit

Computational Complexity

- $t=0$
- $0,1,0$
- $0,0$
- $1,0$
- $0,0,1$
- $0,0,0$
- $1,1$
- $1,1,1$
- $1,1,0$
- $1,0$
- $1,0,1$
- $1,0,0$
- $0,1,1$
- $0,1,0$
A one-armed bernoulli Bandit
Computational Complexity
A one-armed Bernoulli Bandit

Computational Complexity

\[
\begin{array}{c}
t=0 \\
0 \quad 1 \quad 1,0 \quad 1,0,0 \\
0,0 \quad 1,0,1 \\
0,0,0 \quad 1,1,0 \\
0,0,1 \quad 1,1,1 \\
0,1 \quad 1,1,0 \\
0,1,1 \quad 0,1,0 \\
1,0,0
\end{array}
\]
A one-armed bernoulli Bandit

Computational Complexity

```
t=0
  0
  1
  1,0
    1,0,0
    1,0,1
  1,1
    1,1,0
    1,1,1
  0
    0,0
      0,0,0
      0,0,1
    0,1
      0,1,1
      0,1,0
```
A one-armed Bernoulli Bandit

Computational Complexity

t = 0

\[ \begin{array}{c}
0,0 \\
0,1 \\
1,0 \\
1,1 \\
1,0,0 \\
1,0,1 \\
1,1,0 \\
1,1,1 \\
0,0,0 \\
0,0,1 \\
0,1,0 \\
0,1,1 \\
1,0,0 \\
1,0,1 \\
1,1,0 \\
1,1,1 \\
\end{array} \]
A one-armed Bernoulli Bandit

Computational Complexity

Simplest possible scenario:

Only \textbf{two} possible outcomes and \textbf{1 arm}.

\textbf{Blue} $\rightarrow$ \textbf{1, 0, 1, 1, 1, 0, 0, \ldots}$

Time Horizon: $T$ (say, e.g., 100 plays).

Assume a Bernoulli distribution (with unknown parameter $p$).

The total possible sequence of results: $2^T$.

If $K$-armed, then $2^{TK}$

If multinomial (say $R$ possible outcomes), then $R^{TK}$.

Even with Dynamic Programming it is a computational unfeasible problem for $T = 100$, $K = 3$ and $R = 2$. 
A one-armed Bernoulli Bandit

Computational Complexity

Simplest possible scenario:
Only two possible outcomes and 1 arm.
Blue → 1, 0, 1, 1, 1, 0, 0, ...

Time Horizon: $T$ (say, e.g., 100 plays).
Assume a Bernoulli distribution (with unknown parameter $p$).

The total possible sequence of results: $2^T$.

If $K$-armed, then $2^{TK}$

If multinomial (say $R$ possible outcomes), then $R^{TK}$.

Even with Dynamic Programming it is a computational unfeasible problem for $T = 100$, $K = 3$ and $R = 2$. 
A one-armed Bernoulli Bandit

Computational Complexity

Simplest possible scenario:
Only two possible outcomes and 1 arm.
Blue → 1, 0, 1, 1, 1, 0, 0, . . .

Time Horizon: $T$ (say, e.g., 100 plays).
Assume a Bernoulli distribution (with unknown parameter $p$).

The total possible sequence of results: $2^T$.

If $K$-armed, then $2^{TK}$

If multinomial (say $R$ possible outcomes), then $R^{TK}$.

Even with Dynamic Programming it is a computational unfeasible problem for $T = 100$, $K = 3$ and $R = 2$. 
Simplest possible scenario:
Only two possible outcomes and 1 arm.
Blue → 1, 0, 1, 1, 0, 0, …

Time Horizon: $T$ (say, e.g., 100 plays).
Assume a Bernoulli distribution (with unknown parameter $p$).

The total possible sequence of results: $2^T$.

If $K$-armed, then $2^{TK}$

If multinomial (say $R$ possible outcomes), then $R^{TK}$.

Even with Dynamic Programming it is a computational unfeasible problem for $T = 100$, $K = 3$ and $R = 2$. 
Simplest possible scenario:

Only two possible outcomes and 1 arm.

Blue $\rightarrow$ 1, 0, 1, 1, 1, 0, 0, ... 

Time Horizon: $T$ (say, e.g., 100 plays).

Assume a Bernoulli distribution (with unknown parameter $p$).

The total possible sequence of results: $2^T$.

If $K$-armed, then $2^{TK}$

If multinomial (say $R$ possible outcomes), then $R^{TK}$.

Even with Dynamic Programming it is a computational unfeasible problem for $T = 100$, $K = 3$ and $R = 2$. 
The multi-armed Bandit and the Gittins Index
Divide and conquer strategy!

Theorem ('74, '79, '89): Reward is maximized by always continuing to pull the arm having the greatest value of a dynamic allocation index:

Computation time: it grows quadratically on $T$ for a given $x_k$.

Instances such as $T = 100$, $K = 3$ and $R = 2$ (and more) become computationally feasible!
The multi-armed Bandit and the Gittins Index
Divide and conquer strategy!

**Theorem (’74, ’79, ’89):** Reward is maximized by always continuing to pull the arm having the greatest value of a dynamic allocation index:

\[
G_k(x_k) = \sup_{\tau \geq 1} \frac{\mathbb{E}_{x_k(0) = x_k} \sum_{t=0}^{\tau-1} x_k(t) \beta^t}{\mathbb{E}_{x_k(0) = x_k} \sum_{t=0}^{\tau-1} \beta^t}
\]  \quad (3)

Computation time: it grows quadratically on \( T \) for a given \( x_k \).

Instances such as \( T = 100, K = 3 \) and \( R = 2 \) (and more) become computationally feasible!
The multi-armed Bandit and the Gittins Index
Divide and conquer strategy!

Theorem ('74, '79, '89): Reward is maximized by always continuing to pull the arm having the greatest value of a dynamic allocation index:

\[ G_k(x_k) = \sup_{\tau \geq 1} \frac{\text{Expected Discounted Reward up to } \tau}{\text{Expected Discounted time up to } \tau} \]  \hspace{1cm} (3)

Computation time: it grows quadratically on \( T \) for a given \( x_k \).

Instances such as \( T = 100, K = 3 \) and \( R = 2 \) (and more) become computationally feasible!
The multi-armed Bandit and the Gittins Index

Divide and conquer strategy!

Theorem ('74, '79, '89): Reward is maximized by always continuing to pull the arm having the greatest value of a dynamic allocation index:

\[ G_k(x_k) = \sup_{\tau \geq 1} \frac{\text{Expected Discounted Reward up to } \tau}{\text{Expected Discounted time up to } \tau} \]  \hspace{1cm} (3)

Computation time: it grows \textit{quadratically} on \( T \) for a given \( x_k \).

Instances such as \( T = 100, K = 3 \) and \( R = 2 \) (and more) become \textit{computationally feasible!}
The multi-armed Bandit and the Gittins Index

Divide and conquer strategy!

Theorem (’74, ’79, ’89): Reward is maximized by always continuing to pull the arm having the greatest value of a dynamic allocation index:

$$G_k(x_k) = \sup_{\tau \geq 1} \frac{\text{Expected Discounted Reward up to } \tau}{\text{Expected Discounted time up to } \tau}$$ (3)

Computation time: it grows quadratically on \( T \) for a given \( x_k \).

Instances such as \( T = 100, \ K = 3 \) and \( R = 2 \) (and more) become computationally feasible!
The Gittins Index Breakthrough
Comparing Computational Cost: $K = 3$ and $R = 2$
The Gittins Index Breakthrough

Comparing Computational Cost: $K = 3$ and $R = 2$
The Gittins Index Breakthrough
Comparing Computational Cost: $K = 3$ and $R = 2$
The Gittins Index Breakthrough
Comparing Computational Cost: $K = 3$ and $R = 2$
Outline

Multi-armed Bandit Problems

Clinical Trials

Bandits and Clinical Trials Design
• The current gold standard design is known as **Randomised Controlled Trials (RCT)**.

• Basically, after the assessment of eligibility and patient recruitment, patients are allocated **randomly** to the treatments (one of them being a control treatment).

• The first published RCT appeared in the 1948 in Tuberculosis. (Sir Austin Bradford Hill - MRC funded research)

• Randomization minimizes allocation bias, balancing both known and unknown prognostic factors, in the assignment of treatments.
The current gold standard design is known as Randomised Controlled Trials (RCT).

Basically, after the assessment of eligibility and patient recruitment, patients are allocated randomly to the treatments (one of them being a control treatment).

The first published RCT appeared in the 1948 in Tuberculosis. (Sir Austin Bradford Hill - MRC funded research)

Randomization minimizes allocation bias, balancing both known and unknown prognostic factors, in the assignment of treatments.
The current gold standard design is known as **Randomised Controlled Trials (RCT)**.

Basically, after the assessment of eligibility and patient recruitment, patients are allocated *randomly* to the treatments (one of them being a control treatment).

The first published RCT appeared in the 1948 in Tuberculosis. *(Sir Austin Bradford Hill - MRC funded research)*

Randomization minimizes allocation bias, balancing both known and unknown prognostic factors, in the assignment of treatments.
• The current gold standard design is known as **Randomised Controlled Trials (RCT)**.

• Basically, after the assessment of eligibility and patient recruitment, patients are allocated **randomly** to the treatments (one of them being a control treatment).

• The first published RCT appeared in the 1948 in Tuberculosis. (Sir Austin Bradford Hill - MRC funded research)

• Randomization minimizes allocation bias, balancing both known and unknown prognostic factors, in the assignment of treatments.
• The use of RCTs lead to the use of the scientific method in medical research and had a significant impact in the knowledge of the incidence and causes of diseases.

• Nowadays the pharmaceutical industry is in a plateau: lengthy and expensive trials, high rates of drug failures and increased safety concerns.

• Average cost of developing a new drug in US in 2009 was estimated 1.3 billion U.S. dollars.

• Pressure to shorten trials, reduce the number of required patients and produce conclusive results.
The use of RCTs lead to the use of the **scientific method** in medical research and had a significant impact in the knowledge of the incidence and causes of diseases.

Nowadays the pharmaceutical industry is in a **plateau**: lengthy and expensive trials, high rates of drug failures and increased safety concerns.

Average cost of developing a new drug in US in 2009 was estimated **1.3 billion U.S. dollars**.

Pressure to shorten trials, reduce the number of required patients and produce conclusive results.
• The use of RCTs lead to the use of the **scientific method** in medical research and had a significant impact in the knowledge of the incidence and causes of diseases.

• Nowadays the pharmaceutical industry is in a **plateau**: lengthy and expensive trials, high rates of drug failures and increased safety concerns.

• Average cost of developing a new drug in US in 2009 was estimated **1.3 billion U.S. dollars**.

• Pressure to shorten trials, reduce the number of required patients and produce conclusive results.
The use of RCTs lead to the use of the scientific method in medical research and had a significant impact in the knowledge of the incidence and causes of diseases.

Nowadays the pharmaceutical industry is in a plateau: lengthy and expensive trials, high rates of drug failures and increased safety concerns.

Average cost of developing a new drug in US in 2009 was estimated 1.3 billion U.S. dollars.

Pressure to shorten trials, reduce the number of required patients and produce conclusive results.
• In RCT patient allocation is randomised but **fixed** during the trial. (Main goal is the *learning* by the end of the trial)

• Adaptive designs allow to **change** patient allocation rules as data is gathered (*Learning* during the trial)

• If used properly, it provides **efficiency gains** (smaller sample size, an increased chance of correctly answering the clinical question of interest, etc.).

• Determining the **statistical properties** of a particular adaptive design requires careful consideration to avoid biased studies.
• In RCT patient allocation is randomised but fixed during the trial. (Main goal is the *learning* by the end of the trial)

• Adaptive designs allow to change patient allocation rules as data is gathered (*Learning* during the trial)

• If used properly, it provides efficiency gains (smaller sample size, an increased chance of correctly answering the clinical question of interest, etc.).

• Determining the statistical properties of a particular adaptive design requires careful consideration to avoid biased studies.
• In RCT patient allocation is randomised but fixed during the trial. (Main goal is the learning by the end of the trial)

• Adaptive designs allow to change patient allocation rules as data is gathered (Learning during the trial)

• If used properly, it provides efficiency gains (smaller sample size, an increased chance of correctly answering the clinical question of interest, etc.).

• Determining the statistical properties of a particular adaptive design requires careful consideration to avoid biased studies.
Traditional Clinical Trial Design III
Adaptive Designs

- In RCT patient allocation is randomised but fixed during the trial. (Main goal is the learning by the end of the trial)
- Adaptive designs allow to change patient allocation rules as data is gathered (Learning during the trial)
- If used properly, it provides efficiency gains (smaller sample size, an increased chance of correctly answering the clinical question of interest, etc.).
- Determining the statistical properties of a particular adaptive design requires careful consideration to avoid biased studies.
Outline

Multi-armed Bandit Problems

Clinical Trials

Bandits and Clinical Trials Design
Bayesian Bernoulli MABP

Elements

- **Treatment/bandit k**: with unknown $0 \leq p_k \leq 1$.
- **Patient/Time (decision) periods**: $t = 0, 1, \ldots, M_h \to \infty$.
- **Information State space**: $I_{k,t} \triangleq (s_{k,t}, f_{k,t}) \in \mathbb{R}^2$, for $t \leq \infty$.
- **Action Set**: $a_{k,t} \triangleq \{0, 1\}$.
- **Patient-to Patient (one-period) Information State Dynamics**:

  $$I_{k,t+1} = \begin{cases} (s_{k,t} + 1, f_{k,t}) , & \text{if } a_{k,t} = 1 \quad \text{w.p} \quad \frac{s_{k,t}}{s_{k,t} + f_{k,t}} , \\ (s_{k,t}, f_{k,t} + 1) , & \text{if } a_{k,t} = 1 \quad \text{w.p} \quad \frac{f_{k,t}}{s_{k,t} + f_{k,t}} , \\ I_{k,t} = (s_{k,t}, f_{k,t}) , & \text{if } a_{k,t} = 0 \quad \text{w.p} \quad 1 , \end{cases}$$

- **Patient Specific (one-period) Expected Rewards**:

  $$r(I_{k,t-}, a_{k,t}) \triangleq d^t \frac{s_{k,t-1}}{s_{k,t-1} + f_{k,t-1}} \quad \text{with} \quad 0 \leq d \leq 1 \quad t = 1, 2, \ldots$$
$K$ arms/treatments with a binary outcome: success/failure.

Q: How to allocate patients to treatments to maximise the total mean number of successes?

Let $K = 2$, $t = 12$ and $(2, 2), (4, 4)$.

Posterior means = 0.5. Gittins indexes are:
$K$ arms/treatments with a binary outcome: success/failure.

Q: How to allocate patients to treatments to maximise the total mean number of successes?

Let $K = 2$, $t = 12$ and $(2, 2), (4, 4)$.

Posterior means = 0.5. Gittins indexes are:
$K$ arms/treatments with a binary outcome: success/failure.

Q: How to allocate patients to treatments to maximise the total mean number of successes?

Let $K = 2$, $t = 12$ and $(2, 2), (4, 4)$.

Posterior means $= 0.5$. Gittins indexes are:
The Gittins index for a clinical trial
An influential example of Adaptive sampling

$K$ arms/treatments with a binary outcome: success/failure.

**Q:** How to allocate patients to treatments to maximise the total mean number of successes?

Let $K = 2$, $t = 12$ and $(2, 2)$, $(4, 4)$.

Posterior means $= 0.5$. Gittins indexes are:
The Gittins index for a clinical trial
An influential example of Adaptive sampling

\( K \) arms/treatments with a binary outcome: success/failure.

**Q:** How to allocate patients to treatments to maximise the total mean number of successes?

Let \( K = 2, \ t = 12 \) and \((2, 2), (4, 4)\).

Posterior means = 0.5. Gittins indexes are:

<table>
<thead>
<tr>
<th>f/s</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.8699</td>
<td>.9102</td>
<td>.9285</td>
<td>.9395</td>
<td>.9470</td>
<td>.9525</td>
</tr>
<tr>
<td>2</td>
<td>.7005</td>
<td>.7844</td>
<td>.8268</td>
<td>.8533</td>
<td>.8719</td>
<td>.8857</td>
</tr>
<tr>
<td>3</td>
<td>.5671</td>
<td>.6726</td>
<td>.7308</td>
<td>.7696</td>
<td>.7973</td>
<td>.8184</td>
</tr>
<tr>
<td>4</td>
<td>.4701</td>
<td>.5806</td>
<td>.6490</td>
<td>.6952</td>
<td>.7295</td>
<td>.7561</td>
</tr>
<tr>
<td>5</td>
<td>.3969</td>
<td>.5093</td>
<td>.5798</td>
<td>.6311</td>
<td>.6697</td>
<td>.6998</td>
</tr>
<tr>
<td>6</td>
<td>.3415</td>
<td>.4509</td>
<td>.5225</td>
<td>.5756</td>
<td>.6172</td>
<td>.6504</td>
</tr>
<tr>
<td>7</td>
<td>.2979</td>
<td>.4029</td>
<td>.4747</td>
<td>.5277</td>
<td>.5710</td>
<td>.6061</td>
</tr>
</tbody>
</table>
The Gittins index for a clinical trial
An influential example of Adaptive sampling

$K$ arms/treatments with a binary outcome: sucess/failure.

Q: How to allocate patients to treatments to maximise the total mean number of successes?

Let $K = 2$, $t = 12$ and $(2, 2)$, $(4, 4)$.

Posterior means = 0.5. Gittins indexes are:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.8699</td>
<td>.9102</td>
<td>.9285</td>
<td>.9395</td>
<td>.9470</td>
<td>.9525</td>
</tr>
<tr>
<td>2</td>
<td>.7005</td>
<td>.7844</td>
<td>.8268</td>
<td>.8533</td>
<td>.8719</td>
<td>.8857</td>
</tr>
<tr>
<td>3</td>
<td>.5671</td>
<td>.6726</td>
<td>.7308</td>
<td>.7696</td>
<td>.7973</td>
<td>.8184</td>
</tr>
<tr>
<td>4</td>
<td>.4701</td>
<td>.5806</td>
<td>.6490</td>
<td>.6952</td>
<td>.7295</td>
<td>.7561</td>
</tr>
<tr>
<td>5</td>
<td>.3969</td>
<td>.5093</td>
<td>.5798</td>
<td>.6311</td>
<td>.6697</td>
<td>.6998</td>
</tr>
<tr>
<td>6</td>
<td>.3415</td>
<td>.4509</td>
<td>.5225</td>
<td>.5756</td>
<td>.6172</td>
<td>.6504</td>
</tr>
<tr>
<td>7</td>
<td>.2979</td>
<td>.4029</td>
<td>.4747</td>
<td>.5277</td>
<td>.5710</td>
<td>.6061</td>
</tr>
</tbody>
</table>
The Gittins index for a clinical trial
An influential example of Adaptive sampling

\( K \) arms/treatments with a binary outcome: success/failure.

**Q:** How to allocate patients to treatments to maximise the total mean number of successes?

Let \( K = 2 \), \( t = 12 \) and \((2, 2), (4, 4)\).

Posterior means = 0.5. Gittins indexes are:

<table>
<thead>
<tr>
<th>f/s</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8699</td>
<td>0.9102</td>
<td>0.9285</td>
<td>0.9395</td>
<td>0.9470</td>
<td>0.9525</td>
</tr>
<tr>
<td>2</td>
<td>0.7005</td>
<td>0.7844</td>
<td>0.8268</td>
<td>0.8533</td>
<td>0.8719</td>
<td>0.8857</td>
</tr>
<tr>
<td>3</td>
<td>0.5671</td>
<td>0.6726</td>
<td>0.7308</td>
<td>0.7696</td>
<td>0.7973</td>
<td>0.8184</td>
</tr>
<tr>
<td>4</td>
<td>0.4701</td>
<td>0.5806</td>
<td>0.6490</td>
<td>0.6952</td>
<td>0.7295</td>
<td>0.7561</td>
</tr>
<tr>
<td>5</td>
<td>0.3969</td>
<td>0.5093</td>
<td>0.5798</td>
<td>0.6311</td>
<td>0.6697</td>
<td>0.6998</td>
</tr>
<tr>
<td>6</td>
<td>0.3415</td>
<td>0.4509</td>
<td>0.5225</td>
<td>0.5756</td>
<td>0.6172</td>
<td>0.6504</td>
</tr>
<tr>
<td>7</td>
<td>0.2979</td>
<td>0.4029</td>
<td>0.4747</td>
<td>0.5277</td>
<td>0.5710</td>
<td>0.6061</td>
</tr>
</tbody>
</table>
The Gittins index for a clinical trial
An influential example of Adaptive sampling

K arms/treatments with a binary outcome: success/failure.

Q: How to allocate patients to treatments to maximise the total mean number of successes?

Let K = 2, t = 12 and (2, 2), (4, 4).

Posterior means = 0.5. Gittins indexes are:

<table>
<thead>
<tr>
<th>f/s</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.8699</td>
<td>.9102</td>
<td>.9285</td>
<td>.9395</td>
<td>.9470</td>
<td>.9525</td>
</tr>
<tr>
<td>2</td>
<td>.7005</td>
<td>.7844</td>
<td>.8268</td>
<td>.8533</td>
<td>.8719</td>
<td>.8857</td>
</tr>
<tr>
<td>3</td>
<td>.5671</td>
<td>.6726</td>
<td>.7308</td>
<td>.7696</td>
<td>.7973</td>
<td>.8184</td>
</tr>
<tr>
<td>4</td>
<td>.4701</td>
<td>.5806</td>
<td>.6490</td>
<td>.6952</td>
<td>.7295</td>
<td>.7561</td>
</tr>
<tr>
<td>5</td>
<td>.3969</td>
<td>.5093</td>
<td>.5798</td>
<td>.6311</td>
<td>.6697</td>
<td>.6998</td>
</tr>
<tr>
<td>6</td>
<td>.3415</td>
<td>.4509</td>
<td>.5225</td>
<td>.5756</td>
<td>.6172</td>
<td>.6504</td>
</tr>
<tr>
<td>7</td>
<td>.2979</td>
<td>.4029</td>
<td>.4747</td>
<td>.5277</td>
<td>.5710</td>
<td>.6061</td>
</tr>
</tbody>
</table>
Let $K = 2$ and Simulations $= 10^3$.

OC  Mean number of successes (ES),
    Mean Proportion of patients allocated to $p_2 (p^*)$,
    Mean p-value $\alpha$ and mean power $(1 - \beta)$.
Let $K = 2$ and Simulations $= 10^3$.

**OC** Mean number of successes (ES),
Mean Proportion of patients allocated to $p_2 \ (p^*)$,
Mean p-value $\alpha$ and mean power $(1 - \beta)$. 
The Gittins Index vs. RCT
Comparing operating criteria: a natural trade-off

Let $K = 2$ and Simulations $= 10^3$.

**OC** Mean number of successes (ES),
Mean Proportion of patients allocated to $p_2$ ($p^*$),
Mean p-value $\alpha$ and mean power $(1 - \beta)$.

<table>
<thead>
<tr>
<th>$T = 138$</th>
<th>ES</th>
<th>$p^*$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCT</td>
<td>41.3577</td>
<td>.5</td>
<td>0.0495</td>
</tr>
<tr>
<td>Gittins</td>
<td>41.3260</td>
<td>0.5080</td>
<td>0.0520</td>
</tr>
<tr>
<td>Best</td>
<td>41.4</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
Let $K = 2$ and Simulations $= 10^3$.

**OC** Mean number of successes (ES), Mean Proportion of patients allocated to $p_2$ ($p^*$), Mean p-value $\alpha$ and mean power $(1 - \beta)$.

<table>
<thead>
<tr>
<th></th>
<th>$p_1 = 0.3$</th>
<th>$p_2 = 0.3$</th>
<th>$p^*$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 138$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>RCT</strong></td>
<td>41.3577</td>
<td>0.5</td>
<td>0.0495</td>
<td></td>
</tr>
<tr>
<td><strong>Gittins</strong></td>
<td>41.3260</td>
<td>0.5080</td>
<td>0.0520</td>
<td></td>
</tr>
<tr>
<td><strong>Best</strong></td>
<td>41.4</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>$p_1 = 0.3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_2 = 0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T = 138$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>RCT</strong></td>
<td>55.2054</td>
<td>0.5</td>
<td>0.7785</td>
<td></td>
</tr>
<tr>
<td><strong>Gittins</strong></td>
<td>65.5930</td>
<td>0.8651</td>
<td>0.5400</td>
<td></td>
</tr>
<tr>
<td><strong>Best</strong></td>
<td>69</td>
<td>1</td>
<td>–</td>
<td></td>
</tr>
</tbody>
</table>
• Gittins’ work inspired many extensions, yet practically none was applied to the actual conduct of clinical trials. Reasons: Index computation, immediate response, lack of study of its statistical properties, controversial idea ... other reasons?

• There is a trade-off between being Bayes optimal and achieving power constraints. How to strike a balance?

• RCTs have as main goal maximising the learning (prioritising future patients), ADs try to maximise health of the patients in the trial, given the information available. (learn and exploit!)

• The multi-arm bandit approach “learns and exploits", sensible approach for clinical trials on rare diseases.
Discussion

- Gittins’ work inspired many extensions, yet practically none was applied to the actual conduct of clinical trials. **Reasons:** Index computation, immediate response, lack of study of its statistical properties, controversial idea ... **other reasons?**

- There is a trade-off between being Bayes optimal and achieving power constraints. **How to strike a balance?**

- RCTs have as main goal maximising the learning (prioritising future patients), ADs try to maximise health of the patients in the trial, given the information available. (**learn and exploit!**)

- The multi-arm bandit approach “*learns and exploits*”, sensible approach for clinical trials on **rare diseases**.
• Gittins’ work inspired many extensions, yet practically none was applied to the actual conduct of clinical trials. **Reasons:** Index computation, immediate response, lack of study of its statistical properties, controversial idea ... other reasons?

• There is a trade-off between being Bayes optimal and achieving power constraints. How to strike a balance?

• RCTs have as main goal maximising the learning (prioritising future patients), ADs try to maximise health of the patients in the trial, given the information available. *(learn and exploit!)*

• The multi-arm bandit approach "*learns and exploits*", sensible approach for clinical trials on rare diseases.
• Gittins’ work inspired many extensions, yet practically none was applied to the actual conduct of clinical trials. Reasons: Index computation, immediate response, lack of study of its statistical properties, controversial idea ... other reasons?

• There is a trade-off between being Bayes optimal and achieving power constraints. How to strike a balance?

• RCTs have as main goal maximising the learning (prioritising future patients), ADs try to maximise health of the patients in the trial, given the information available. (learn and exploit!)

• The multi-arm bandit approach “learns and exploits”, sensible approach for clinical trials on rare diseases.
Discussion

• Gittins’ work inspired many extensions, yet practically none was applied to the actual conduct of clinical trials. Reasons: Index computation, immediate response, lack of study of its statistical properties, controversial idea ... other reasons?

• There is a trade-off between being Bayes optimal and achieving power constraints. How to strike a balance?

• RCTs have as main goal maximising the learning (prioritising future patients), ADs try to maximise health of the patients in the trial, given the information available. (learn and exploit!)

• The multi-arm bandit approach “learns and exploits”, sensible approach for clinical trials on rare diseases.
• Gittins’ work inspired many extensions, yet practically none was applied to the actual conduct of clinical trials. **Reasons:** Index computation, immediate response, lack of study of its statistical properties, controversial idea ... **other reasons?**

• There is a trade-off between being Bayes optimal and achieving power constraints. **How to strike a balance?**

• RCTs have as main goal maximising the learning (prioritising future patients), ADs try to maximise health of the patients in the trial, given the information available. (learn and exploit!)

• The multi-arm bandit approach “learns and exploits”, sensible approach for clinical trials on rare diseases.
Discussion

- Gittins’ work inspired many extensions, yet practically none was applied to the actual conduct of clinical trials. **Reasons:** Index computation, immediate response, lack of study of its statistical properties, controversial idea ... **other reasons?**

- There is a trade-off between being Bayes optimal and achieving power constraints. **How to strike a balance?**

- RCTs have as main goal maximising the learning (prioritising future patients), ADs try to maximise health of the patients in the trial, given the information available. **(learn and exploit!)**

- The multi-arm bandit approach “learns and exploits”, sensible approach for clinical trials on rare diseases.
Discussion

• Gittins’ work inspired many extensions, yet practically none was applied to the actual conduct of clinical trials. Reasons: Index computation, immediate response, lack of study of its statistical properties, controversial idea ... other reasons?

• There is a trade-off between being Bayes optimal and achieving power constraints. How to strike a balance?

• RCTs have as main goal maximising the learning (prioritising future patients), ADs try to maximise health of the patients in the trial, given the information available. (learn and exploit!)

• The multi-arm bandit approach “learns and exploits”, sensible approach for clinical trials on rare diseases.
• Gittins’ work inspired many extensions, yet practically none was applied to the actual conduct of clinical trials. **Reasons:** Index computation, immediate response, lack of study of its statistical properties, controversial idea ... other reasons?

• There is a trade-off between being Bayes optimal and achieving power constraints. **How to strike a balance?**

• RCTs have as main goal maximising the learning (prioritising future patients), ADs try to maximise health of the patients in the trial, given the information available. *(learn and exploit!)*

• The multi-arm bandit approach “learns and exploits”, sensible approach for clinical trials on rare diseases.
Discussion

- Gittins’ work inspired many extensions, yet practically none was applied to the actual conduct of clinical trials. **Reasons:** Index computation, immediate response, lack of study of its statistical properties, controversial idea ... other reasons?

- There is a trade-off between being Bayes optimal and achieving power constraints. **How to strike a balance?**

- RCTs have as main goal maximising the learning (prioritising future patients), ADs try to maximise health of the patients in the trial, given the information available. (learn and exploit!)

- The multi-arm bandit approach “learns and exploits”, sensible approach for clinical trials on rare diseases.
Thanks for the attention! 😊