From Multi-armed Bandit Problems to Response-Adaptive Randomisation: Implementing Optimality Criteria in Clinical Trials”

Sofía S. Villar

The Multi-armed Bandit Problem (MABP)

The Gittins Index and the solution to a MABP

Introducing Randomisation to the Gittins index (FLGI)
  Increasing Power of Bandit rules

Introducing Covariates to the Gittins index (CARA FLGI)

Discussion
The *(Classic)* Multi-armed Bandit Problem

Maximise total expected gains over time:

\[ \text{Learning about the success rates of the slot machines just enough to maximise average total profit} \]

\[ \sum_{t} \mathbb{E}[Y_{1,t}] \quad \sum_{t} \mathbb{E}[Y_{2,t}] \quad \ldots \quad \sum_{t} \mathbb{E}[Y_{K,t}] \]
Although their scope is much more general, the most common scenario chosen to motivate the MABP in the literature is that of a *clinical trial* which has the aim of balancing two separate goals:

**G1** To correctly identify the best treatment (learning).

**G2** To best (most effectively) treat as many patients as possible (earning).

Traditional *clinical trials* are designed to meet requirements on error probabilities, which relates to the *learning* element of this dilemma.

The ethical conflict around these goals becomes more acute (*suboptimality gap grows*) when: the population with a disease is small, the disease is life-threatening and/or there are multiple potential treatments to study.
Trial design as a (classic) Multi-armed Bandit Problem

Maximise total expected *patient benefit* over time:

\[ Y_{1,t} \quad Y_{2,t} \quad \ldots \quad Y_{K,t} \]

*learn about the treatments’ efficacy just enough to maximise patients’ outcomes over the population*
Theoretical work (a sample of it)
Optimality in Clinical Trials I

Optimality in terms of patient benefit (70’s 80’s)

• “A procedure which maximizes the expected number of successes in a clinical trial involving two treatments can usually be found only by backward induction.” Berry, D.A. (1978)

• “Multi-armed bandit problems are similar, but with more than two arms. Their chief practical motivation comes from clinical trials” Gittins and Jones (1979)

• “The number of observations needed to obtain a given level of precision can be minimized by using a fixed-sample rule, but this involves too many applications of the inferior treatment. ... sequential allocation rules can achieve a similar pattern of error probabilities for a small fraction of the expected cost to the volunteers” Bather, J. (1985)
Optimality in terms of patient benefit (from 2000’s and from the 30’s!)

- The optimal sample size of a 2-armed RCT (optimal in terms of patient benefit) is $\propto \sqrt{N}$. Cheng et al (2003)

- Suggested to randomise patients with a probability $\propto$ the posterior probability of an arm being superior than the other. “This would be important in cases where either the rate of accumulation of data is slow or the individuals treated are valuable, or both.” Thompson, W. (1933)
Armitage (1985):

“Yet most of the theoretical work done in this tradition, over the last 20 years or so, has found no application whatsoever in the actual conduct of trials. This lack of contact between theory and practice seems to me quite deplorable. Either the theoreticians have got hold of the wrong problem, or the practising triallists have shown a culpable lack of awareness of relevant theoretical developments, or both. In any case, the situation does not reflect particularly well on the statistical community”
The Multi-armed Bandit Problem (MABP)

The Gittins Index and the solution to a MABP

Introducing Randomisation to the Gittins index (FLGI)
  Increasing Power of Bandit rules

Introducing Covariates to the Gittins index (CARA FLGI)

Discussion
The Curse of Dimensionality
MABP and Computational Feasibility

• Solution to the **MABP** according to Bellman’s principle of optimality exists but is **computationally expensive**. Prohibitively so for most realistic scenarios.

• The curse of dimensionality was (till early 80’s) **the single most important limitation** to its applicability in practice (in any context).

Armitage (1985):

“The problem can now be seen as essentially the ‘two-armed bandit’ problem for a finite horizon. The solution to this can in principle be obtained by dynamic programming methods, but in practice the computation involved is prohibitive except for trivially small horizons.”
The Gittins Index: Divide and Conquer!

Classic MABP & Infinite Horizon Case

Theorem ('74, '79, '89): The Expected Total (discounted) Reward is maximised by playing at each time $t$ the machine having the largest “dynamic allocation index”: $G_k(s_k,t, f_k,t)$ (some values in Table below).

<table>
<thead>
<tr>
<th>$f_0/s_0$</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>
</tbody>
</table>

The Gittins indices (GI) for different $(s_0, f_0)$ pairs (Gittins and Jones, 1979; Gittins et al., 2011) → Huge computational gains!
The Gittins index for a clinical trial
An example of the index rule in practice

<table>
<thead>
<tr>
<th>Time</th>
<th>Gittins Index</th>
<th>Allocation Prob.</th>
<th>Action</th>
<th>Outcome</th>
<th>Patient</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>t</strong></td>
<td><strong>G₀</strong></td>
<td><strong>G₁</strong></td>
<td><strong>P(a₀,t = 1)</strong></td>
<td><strong>P(a₁,t = 1)</strong></td>
<td><strong>a₀,t</strong></td>
</tr>
<tr>
<td>0</td>
<td>0.8699</td>
<td>0.8699</td>
<td>1/2</td>
<td>1/2</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0.8699</td>
<td>0.9102</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.8699</td>
<td>0.7844</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0.9102</td>
<td>0.7844</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>0.9285</td>
<td>0.7844</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0.8268</td>
<td>0.7844</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Treatment decisions using the Gittins indices in a 2-arm trial example

<table>
<thead>
<tr>
<th></th>
<th>H₀ : p₀ = p₁ = 0.3</th>
<th>H₁ : p₀ = 0.3 , p₁ = 0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>α</strong></td>
<td><em><em>p</em> (s.e.)</em>*</td>
<td><strong>ENS (s.e.)</strong></td>
</tr>
<tr>
<td><strong>FR</strong></td>
<td>0.052 0.500 (0.04)</td>
<td>44.3 (5.6)</td>
</tr>
<tr>
<td><strong>GI</strong></td>
<td>0.053 0.501 (0.26)</td>
<td>44.4 (5.6)</td>
</tr>
<tr>
<td><strong>UB</strong></td>
<td>44.4 (0.0)</td>
<td>1 74.0 (0.0)</td>
</tr>
</tbody>
</table>

Comparison of the OCs of different two-arm trial designs of size T = 148. α: type I error; 1 − β: power; p*: expected number of patients assigned to best treatment; ENS: expected number of patient successes; UB: upper bound.
The Gittins index for a clinical trial

An example of the index rule in practice

<table>
<thead>
<tr>
<th>Time</th>
<th>Gittins Index</th>
<th>Allocation Prob.</th>
<th>Action</th>
<th>Outcome</th>
<th>Patient</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>$G_0$</td>
<td>$G_1$</td>
<td>$P(a_0,t = 1)$</td>
<td>$P(a_1,t = 1)$</td>
<td>$a_{0,t}$</td>
</tr>
<tr>
<td>0</td>
<td>0.8699</td>
<td>0.8699</td>
<td>1/2</td>
<td>1/2</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0.8699</td>
<td>0.9102</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.8699</td>
<td>0.7844</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0.9102</td>
<td>0.7844</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>0.9285</td>
<td>0.7844</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0.8268</td>
<td>0.7844</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Treatment decisions using the Gittins indices in a 2-arm trial example

<table>
<thead>
<tr>
<th></th>
<th>$H_0 : p_0 = p_1 = 0.3$</th>
<th>$H_1 : p_0 = 0.3, p_1 = 0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha$</td>
<td>$p^*$ (s.e.)</td>
</tr>
<tr>
<td>FR</td>
<td>0.052</td>
<td>0.500 (0.04)</td>
</tr>
<tr>
<td>GI</td>
<td>0.053</td>
<td>0.501 (0.26)</td>
</tr>
<tr>
<td>UB</td>
<td></td>
<td>44.4 (0.0)</td>
</tr>
</tbody>
</table>

Comparison of the OCs of different two-arm trial designs of size $T = 148$. $\alpha$: type I error; $1 - \beta$: power; $p^*$: expected number of patients assigned to best treatment; ENS: expected number of patient successes; UB: upper bound.
Gittins (1979) “Their chief practical motivation comes from clinical trials...”

Despite being computationally feasible for multi-armed trials (and simpler than DP to summarise), index rules have not been applied to a trial yet.

Important barriers to its use in practice include (Villar et al, 2015a):

1. Its fully sequential nature: outcomes must be immediately available.
2. Decisions are not randomized: treatment allocation bias, covariate imbalance. Basis for inference.
3. Given an objective degree of discrimination between two treatments, it lacks a sufficient/comparable level of statistical power.
4. It does not incorporate potentially important prognostic covariates.
5. Others: bias in estimation of treatment effect (overestimation of treatment effect), the effect of patient drift, etc.
Gittins (1979)

“Their chief practical motivation comes from clinical trials...”

Despite being computationally feasible for multi-armed trials (and simpler than DP to summarise), index rules have not been applied to a trial yet.

Important barriers to its use in practice include (Villar et al, 2015a):

1. Its fully sequential nature: outcomes must be immediately available.

2. Decisions are not randomized: treatment allocation bias, covariate imbalance. Basis for inference.

3. Given an objective degree of discrimination between two treatments, it lacks a sufficient/comparable level of statistical power.

4. It does not incorporate potentially important prognostic covariates.

5. Others: bias in estimation of treatment effect (overestimation of treatment effect), the effect of patient drift, etc.
The Multi-armed Bandit Problem (MABP)

The Gittins Index and the solution to a MABP

Introducing Randomisation to the Gittins index (FLGI)
  Increasing Power of Bandit rules

Introducing Covariates to the Gittins index (CARA FLGI)

Discussion
Assume that \( T \) patients are enrolled sequentially in groups of size \( b \) over \( J \) stages, so that \( J \times b = T \). In Villar et al (2015b) we defined group allocation probabilities based on the GI as follows:

Simplest example: \( b = 2 \). Priors: control \((s_{(0,0)}, f_{(0,0)}) = (1, 2)\) and experimental \((s_{(1,0)}, f_{(1,0)}) = (1, 1)\)

What is the (patient-average) probability of each arm being allocated in the next block using the GI (and given the priors)?

\[
\pi_{1,0} = \frac{(0 \times 1) + (0 \times 1/2 + 1/2 \times 1/2)}{2} = 1/8, \quad \pi_{1,1} = \frac{(1 \times 1) + (1 \times 1/2 + 1/2 \times 1/2)}{2} = 7/8.
\]
Assume that $T$ patients are enrolled sequentially in groups of size $b$ over $J$ stages, so that $J \times b = T$. In Villar et al (2015b) we defined group allocation probabilities based on the GI as follows:

Simplest example: $b = 2$. Priors: control $(s_{0,0}, f_{0,0}) = (1, 2)$ and experimental $(s_{1,0}, f_{1,0}) = (1, 1)$

What is the (patient-average) probability of each arm being allocated in the next block using the GI (and given the priors)?

\[
\pi_{1,0} = \frac{(0 \times 1) + (0 \times 1/2 + 1/2 \times 1/2)}{2} = 1/8, \quad \pi_{1,1} = \frac{(1 \times 1) + (1 \times 1/2 + 1/2 \times 1/2)}{2} = 7/8.
\]
Just as for the MABP, the computational cost of the exact FLGI probabilities grows with the number of arms ($K$) and $b$ (block size).

Computation in practice can be done via Monte Carlo simulation. Example: $P = [1 \ 1; 2 \ 1; 1 \ 2; 2 \ 2]$ ($K = 4$) and block $b = 9$ then $\pi \approx [0.2646; 0.5901; 0.0246; 0.1208]$ after $5 \times 10^2$ replicas.

P1 For equal priors the algorithm defines equal allocation probabilities.

P2 As the block size tends to grow (in the limit it equals the trial size), the design tends to a balanced design (given initial equipoise).

P3 If the block is of only 1 patient (i.e. there is an interim after every patient), the FLGI rule recovers the GI rule.
The Multi-armed Bandit Problem (MABP)

The Gittins Index and the solution to a MABP

Introducing Randomisation to the Gittins index (FLGI)
   Increasing Power of Bandit rules

Introducing Covariates to the Gittins index (CARA FLGI)

Discussion
NeoSphere is a 4-arm ER trial in breast cancer with 417 patients. The response rates reported were 29.0%, 45.8%, 16.8% and 24.0%.

\[ H_1 : \mathbf{p}_1 = [0.29 \ 0.458 \ 0.168 \ 0.24] \]

<table>
<thead>
<tr>
<th>Power ((1 - \beta))</th>
<th>Patient Benefit (p^*) (s.e.)</th>
<th>ENS (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ER (block=417)</strong></td>
<td>0.653</td>
<td>0.250 (0.02)</td>
</tr>
<tr>
<td><strong>FLGI (block=9)</strong></td>
<td>0.177</td>
<td>0.804 (0.09)</td>
</tr>
<tr>
<td><strong>GI (block=1)</strong></td>
<td>0.140</td>
<td>0.840 (0.10)</td>
</tr>
<tr>
<td><strong>UB</strong></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

with the \(\pi_{k,j}\) probabilities computed via Monte Carlo simulation.

- **Effects of randomisation:** (slight) increase in power/ (slight) reduction in ENS (patient benefit)

- **Increase power levels:** apply FLGI to experimental arms only.
  Allocation to control arm fixed at FR level (25%) (Trippa et al, 2012)
NeoSphere is a 4-arm ER trial in breast cancer with 417 patients. The response rates reported were 29.0%, 45.8%, 16.8% and 24.0%.

\[ H_1 : p_1 = [0.29 \ 0.458 \ 0.168 \ 0.24] \]

<table>
<thead>
<tr>
<th></th>
<th>Power ( (1 - \beta) )</th>
<th>Patient Benefit ( p^* ) (s.e.)</th>
<th>ENS (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER (block=417)</td>
<td>0.653</td>
<td>0.250 (0.02)</td>
<td>120.88 (9.34)</td>
</tr>
<tr>
<td>C FLGI (block=9)</td>
<td>0.816</td>
<td>0.665 (0.06)</td>
<td>166.40 (11.9)</td>
</tr>
<tr>
<td>FLGI (block=9)</td>
<td>0.177</td>
<td>0.804 (0.09)</td>
<td>174.11 (13.3)</td>
</tr>
<tr>
<td>GI (block=1)</td>
<td>0.140</td>
<td>0.840 (0.10)</td>
<td>177.97 (13.0)</td>
</tr>
<tr>
<td>UB</td>
<td></td>
<td>1</td>
<td>190.99 (0.00)</td>
</tr>
</tbody>
</table>

with the \( \pi_{k,j} \) probabilities computed via Monte Carlo simulation.

- **Effects of randomisation:** (slight) increase in power/ (slight) reduction in ENS (patient benefit)

- **Increase power levels:** apply FLGI to experimental arms only. Allocation to control arm fixed at FR level (25%) (Trippa et al, 2012)
Outline

The Multi-armed Bandit Problem (MABP)

The Gittins Index and the solution to a MABP

Introducing Randomisation to the Gittins index (FLGI)
  Increasing Power of Bandit rules

Introducing Covariates to the Gittins index (CARA FLGI)

Discussion
FDA has recently approved several cancer drugs for use in patients whose tumours have specific genetic characteristics.

This has strengthened the promise of "personalised medicine" - the tailoring of treatment to the individual characteristics of each patient.

How can trials find treatments that work for a subgroup of patients?

The challenge is on how to do so in contexts in which there are several promising treatments and relatively few patients to test them - even fewer if a treatment works only within a subgroup.

Some recent trials have used covariate-adjusted response-adaptive (CARA) randomisation (Rosenberger et al, 2001) to more quickly identify superior treatments among several, mainly treatments that work better within subgroups. E.g., I-SPY II or BATTLE trial.
Incorporating Covariate Information to the Gittins Index
Increasing Patient Benefit by Personalising Treatment

MABP with covariates: let patient outcome $Y_{k,t} \sim Bernoulli(p_k(z_t))$ where $Z_t \sim Bernoulli(q)$ (with $q$ known).

E.g., $p_k(z_n) = \text{Expit} (\alpha_k + \beta_k z_t) \forall t$, where $\text{Expit}(u) = \frac{\exp(u)}{1+\exp(u)}$.

For patient $t$, we observe their covariate value $z_t$ then we treat them.

- Associated MABP with Dynamic Programming: computational complexity even larger than in the classic case. (Deterministic)

Q: Can we define a simple index rule in this case? Little work in the literature: Clayton ’89; Woodroofe ’79

- Villar and Rosenberger (2017) proposed a heuristic (extended) Gittins index rule for a binary endpoint with a discrete covariate with $C$ levels.
(1) We consider a MABP with \( K \) experimental arms, a control arm and \( T \) patients. Before arm \( k \) is allocated to patient \( t \), a binary covariate \( Z_t \) is observed. Immediately after, a binary response \( Y_{t,n} \) is observed.

(2) Reformulate the above MABP: for every treatment-covariate combination there exists a combination arm \( k_z \). E.g., the arm “00” corresponds to the control arm and covariate negative patients.

New reformulated MABP has \( 2(K+1) \) combinations arms (with rate \( p_{kt} \)) and patients are optimally allocated to arms with the constraint that they are only allowed arms feasible given their biomarker profile.

(3) We defined a modified GI rule: each patient gets the treatment with the highest GI among the arms available for their biomarker profile.

(4) From this modified GI, a randomised group allocation procedure is defined as in Villar et al (2015b) but for every covariate value (and block) we have a different vector of allocation probabilities \( \pi_{k,j}(Z) \).
The CARA FLGI in Practice
Simulation Results

3-arm trial 300 patients $p_{k0} = (0.22; 0.34; 0.49)$, $p_{k,1} = (0.47; 0.71; 0.37)$. Treatment-covariate interaction: best arm for covariate negative patients is arm 2 while for covariate positive patients is arm 1.

<table>
<thead>
<tr>
<th></th>
<th>Power $1 - \beta_0$</th>
<th>Power $1 - \beta_1$</th>
<th>$p_0^*$ (s.d)</th>
<th>$p_1^*$ (s.d)</th>
<th>ENS (s.d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER ($b=300$)</td>
<td>0.82</td>
<td>0.63</td>
<td>0.33 (0.04)</td>
<td>0.33 (0.04)</td>
<td>130.71 (9.3)</td>
</tr>
<tr>
<td>CARA CFLGI ($b=10$)</td>
<td><strong>0.85</strong></td>
<td><strong>0.79</strong></td>
<td><strong>0.55 (0.16)</strong></td>
<td><strong>0.62 (0.06)</strong></td>
<td><strong>148.36 (9.6)</strong></td>
</tr>
<tr>
<td>CARA FLGI ($b=10$)</td>
<td>0.13</td>
<td>0.03</td>
<td>0.75 (0.22)</td>
<td>0.86 (0.16)</td>
<td>166.73 (11.2)</td>
</tr>
<tr>
<td>CARA GI ($b=1$)</td>
<td>0.11</td>
<td>0.03</td>
<td>0.78 (0.24)</td>
<td>0.88 (0.18)</td>
<td>169.39 (11.4)</td>
</tr>
</tbody>
</table>

CARA FLGI probabilities (Monte Carlo simulation), $T = 300$, $p_z = 0.5$ and 5000 runs.

- Treatment-covariate interactions are detected by the CARA (Covariate-Adjusted Response Adaptive) FLGI procedure but its statistical power is very low.
- In a multi-armed case the CARA CFLGI addresses the power limitation (though in a two-arm setting power may be insufficient).
Outline

The Multi-armed Bandit Problem (MABP)

The Gittins Index and the solution to a MABP

Introducing Randomisation to the Gittins index (FLGI)
   Increasing Power of Bandit rules

Introducing Covariates to the Gittins index (CARA FLGI)

Discussion
Armitage (1985)

“I close with two specific suggestions: first, that statisticians concerned with the development of optimization models and those concerned directly in clinical trials should meet to discuss the feasibility of these methods for various sorts of trials; secondly, that members of the two groups should work in collaboration on specific trials so as to foster closer understanding and to explore the possibilities in a realistic setting.”

- Designing implementable optimal designs still requires dialogue between theory and practice. Such a dialogue can potentially result in sound solutions for the current challenges in clinical trials.

- Explicitly including patient benefit as an optimisation goal can greatly improve trials. Reporting on patient benefit properties of designs should become as standard as reporting expected error rates.


Thompson, W. R. *On the likelihood that one unknown probability exceeds another in view of the evidence of two samples*. Biometrika 25(3/4) 285–294 (1933)


What do we mean by computational infeasibility

State space & Dynamic Programming for $T=7$
Earn-learn dilemma and block size
How to select block size? Should we ramp up accrual?

![Graph showing Earn-learn dilemma and block size selection.](image-url)