A flexible non-parametric dose-finding design for Phase II clinical trials

Pavel Mozgunov, Thomas Jaki

Medical and Pharmaceutical Statistics Research Unit,
Department of Mathematics and Statistics, Lancaster University, UK
p.mozgunov@lancaster.ac.uk

Acknowledgement: This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 633567.

Motivation (I)

Consider a **small population** sequential Phase II trial with **two arms** and **binary outcomes** which aims to find the **superior arm**.

Motivation (I)

Consider a **small population** sequential Phase II trial with **two arms** and **binary outcomes** which aims to find the **superior arm**. Assume that

- 10 outcomes observed for each arm

Motivation (I)

Consider a **small population** sequential Phase II trial with **two arms** and **binary outcomes** which aims to find the **superior arm**. Assume that

- 10 outcomes observed for each arm
- 4 successes on 1st arm
- 6 successes on 2st arm

Motivation (I)

Consider a **small population** sequential Phase II trial with **two arms** and **binary outcomes** which aims to find the **superior arm**. Assume that

- 10 outcomes observed for each arm
- 4 successes on 1st arm
- 6 successes on 2st arm

Q: To which arm a next patient should be assigned?

Motivation (I)

Consider a **small population** sequential Phase II trial with **two arms** and **binary outcomes** which aims to find the **superior arm**. Assume that

- 10 outcomes observed for each arm
- 4 successes on 1st arm
- 6 successes on 2st arm

Q: To which arm a next patient should be assigned?

Keeping in mind that, we would like to

- make a reliable recommendation (high statistical power)

Motivation (I)

Consider a **small population** sequential Phase II trial with **two arms** and **binary outcomes** which aims to find the **superior arm**. Assume that

- 10 outcomes observed for each arm
- 4 successes on 1st arm
- 6 successes on 2st arm

Q: To which arm a next patient should be assigned?

Keeping in mind that, we would like to

- make a reliable recommendation (high statistical power)
- maximize the proportion of the population on the superior arm

Motivation (I)

Consider a **small population** sequential Phase II trial with **two arms** and **binary outcomes** which aims to find the **superior arm**. Assume that

- 10 outcomes observed for each arm
- 4 successes on 1st arm
- 6 successes on 2st arm

Q: To which arm a next patient should be assigned?

Keeping in mind that, we would like to

- make a reliable recommendation (high statistical power)
- maximize the proportion of the population on the superior arm

"Earn vs Learn" trade-off

Motivation (II)

1. **Option 1. Earn**

   Assign a next patients to 2nd arm

Motivation (II)

1. **Option 1. Earn**

   Assign a next patients to 2nd arm

   Challenges:

   - Selection can lock in the suboptimal arm
   - Low statistical power by the end of the trial

Motivation (II)

1. **Option 1. Earn**

   Assign a next patients to 2nd arm

   Challenges:

   - Selection can lock in the suboptimal arm
   - Low statistical power by the end of the trial

2. **Option 2. Learn**

   Assign a next patient to an arm about which we have the least information

Motivation (II)

1. **Option 1. Earn**

   Assign a next patients to 2nd arm

   Challenges:

   - Selection can lock in the suboptimal arm
   - Low statistical power by the end of the trial

2. **Option 2. Learn**

   Assign a next patient to an arm about which we have the least information (e.g. the Shannon information)

Motivation (II)

1. **Option 1. Earn**

   Assign a next patients to 2nd arm

   Challenges:

   - Selection can lock in the suboptimal arm
   - Low statistical power by the end of the trial

2. **Option 2. Learn**

   Assign a next patient to an arm about which we have the least information (e.g. the Shannon information)

   Challenges:

   - Unethical (low number of treated patients)

Current approaches

- Fixed randomization
- Thompson Sampling (proportional to a probability being the best)
  Low expected number of successes

Current approaches

- Fixed randomization
- Thompson Sampling (proportional to a probability being the best)
  Low expected number of successes
- Current belief (maximum point estimate)
  Low statistical power, **high variance** of the expected number of success

Current approaches

- Fixed randomization
- Thompson Sampling (proportional to a probability being the best)
  Low expected number of successes
- Current belief (maximum point estimate)
  Low statistical power, **high variance** of the expected number of success
- Optimal multi-arm bandit (MAB) and the dynamic programming
  Low statistical power

Back to information measures

The Shannon information (entropy)

$$h(f) = -\int_{\mathbb{R}} f(z) \log f(z) \mathrm{d}z.$$

Back to information measures

The Shannon information (entropy)

$$h(f) = -\int_{\mathbb{R}} f(z) \log f(z) \mathrm{d}z.$$

In the example above,

$$h(\text{arm } 1) = h(\text{arm } 2).$$

This information **does not reflect** our specific interest in the **superior arm**

Back to information measures

The Shannon information (entropy)

$$h(f) = -\int_{\mathbb{R}} f(z) \log f(z) \mathrm{d}z.$$

In the example above,

$$h(\text{arm } 1) = h(\text{arm } 2).$$

This information **does not reflect** our specific interest in the **superior arm**

It shows the amount of information needed to answer the question

<u>What is the success probability?</u>

Back to information measures

The Shannon information (entropy)

$$h(f) = -\int_{\mathbb{R}} f(z)\log f(z)\mathrm{d}z.$$

In the example above,

$$h(\text{arm } 1) = h(\text{arm } 2).$$

This information **does not reflect** our specific interest in the **superior arm**

It shows the amount of information needed to answer the question

What is the success probability?

Q: Can we quantify this interest in the information measure?

Weighted information

Consider a twofold experiment:

Weighted information

Consider a twofold experiment:

(i) what is the probability of success (large n)

Weighted information

Consider a twofold experiment:

(i) what is the probability of success (large n)

(ii) is the probability of success close to the target (small and moderate n)

Weighted information

Consider a twofold experiment:

(i) what is the probability of success (large n)

(ii) is the probability of success close to the target (small and moderate n)

A: The *weighted Shannon information*

$$h_{\phi}(f) = -\int_{\mathbb{R}} \phi(z) f(z) \log f(z) \mathrm{d}z.$$

Weighted information

Consider a twofold experiment:

(i) what is the probability of success (large n)

(ii) is the probability of success close to the target (small and moderate n)

A: The *weighted Shannon information*

$$h_{\phi}(f) = -\int_{\mathbb{R}} \phi(z) f(z) \log f(z) \mathrm{d}z.$$

Due to ethical constraints we concentrate on the question (ii) alone and on the corresponding measure of the information

$$h_{\phi}(f) - h(f)$$

Methods

Consider the probability of success as a RV with Beta prior $\mathcal{B}(\nu+1, \beta-\nu+1)$.

Methods

Consider the probability of success as a RV with Beta prior $\mathcal{B}(\nu+1, \beta-\nu+1)$.
After x successes in n trial $\rightarrow$ Beta posterior $\mathcal{B}(x+\nu+1, n-x+\beta-\nu+1)$.

Methods

Consider the probability of success as a RV with Beta prior $\mathcal{B}(\nu+1, \beta-\nu+1)$.
After x successes in n trial $\rightarrow$ Beta posterior $\mathcal{B}(x+\nu+1, n-x+\beta-\nu+1)$.

- α is the true probability of success
- γ is the target probability (for instance, $\gamma = 0.999$)

Methods

Consider the probability of success as a RV with Beta prior $\mathcal{B}(\nu+1, \beta-\nu+1)$.
After x successes in n trial $\rightarrow$ Beta posterior $\mathcal{B}(x+\nu+1, n-x+\beta-\nu+1)$.

- α is the true probability of success
- γ is the target probability (for instance, $\gamma = 0.999$)

Beta-form of the *weight function*

$$\phi_n(\mathbf{p}) = C(x, \gamma, n) p^{\gamma n^{\kappa}} (1-p)^{(1-\gamma) n^{\kappa}}$$

Methods

Consider the probability of success as a RV with Beta prior $\mathcal{B}(\nu+1, \beta-\nu+1)$. After x successes in n trial $\rightarrow$ Beta posterior $\mathcal{B}(x+\nu+1, n-x+\beta-\nu+1)$.

- α is the true probability of success
- γ is the target probability (for instance, $\gamma = 0.999$)

Beta-form of the *weight function*

$$\phi_n(\mathbf{p}) = C(x,\gamma,n)p^{\gamma n^{\kappa}}(1-p)^{(1-\gamma)n^{\kappa}}$$

Theorem

Let $h(f_n)$ and $h^{\phi_n}(f_n)$ be the standard and weighted differential entropies. Then,

$$\lim_{n\to\infty}\left(\left[h^{\phi_n}(f_n)-h(f_n)\right]-\frac{1}{2}\left(\frac{(\alpha-\gamma)^2}{\alpha(1-\alpha)}\right)n^{2\kappa-1}+\omega\right)=0$$

Design

'Plug-in' modal estimator of a success probability of the arm j

$$\hat{p}_{n_j} = \frac{x_j + \nu_j}{n_j + \beta_j},\ j = 1, \ldots, m$$

Design

'Plug-in' modal estimator of a success probability of the arm j

$$\hat{p}_{n_j} = \frac{x_j + \nu_j}{n_j + \beta_j},\ j = 1, \ldots, m$$

for

$$\hat{\delta}_{n_j}^{(\kappa)} = \frac{(\hat{p}_{n_j} - \gamma)^2}{\hat{p}_{n_j}(1 - \hat{p}_{n_j})} n_j^{2\kappa - 1}.$$

Design

'Plug-in' modal estimator of a success probability of the arm j

$$\hat{p}_{n_j} = \frac{x_j + \nu_j}{n_j + \beta_j},\ j = 1, \ldots, m$$

for

$$\hat{\delta}_{n_j}^{(\kappa)} = \frac{(\hat{p}_{n_j} - \gamma)^2}{\hat{p}_{n_j}(1 - \hat{p}_{n_j})} n_j^{2\kappa - 1}.$$

Arm selection algorithm:

1. Start from $\hat{\delta}_{\beta_i}^{(\kappa)}$, $i = 1, \ldots, m$
2. Observed n_i and x_i outcomes for the arm A_i, $i = 1, \ldots, m$
3. Arm A_j is selected if it satisfies

$$\hat{\delta}_{n_j}^{(\kappa)} = \inf_{i=1,\ldots,m} \hat{\delta}_{n_i}^{(\kappa)}.$$

4. Repeat 2-3 until the total number of patients is reached.

Illustration. Two arms trial

Consider the trial with $m = 2$ arms ($\alpha_1 = 0.5$ and $\alpha_2 = 0.7$), $n = 75$ patients

Illustration. Two arms trial

Consider the trial with $m = 2$ arms ($\alpha_1 = 0.5$ and $\alpha_2 = 0.7$), $n = 75$ patients

$$\text{Prior :} \qquad \hat{p} = (0.99, 0.99); \quad \beta = (2, 2)$$

Illustration. Two arms trial

Consider the trial with $m = 2$ arms ($\alpha_1 = 0.5$ and $\alpha_2 = 0.7$), $n = 75$ patients

$$\text{Prior :} \qquad \hat{p} = (0.99, 0.99); \quad \beta = (2, 2)$$

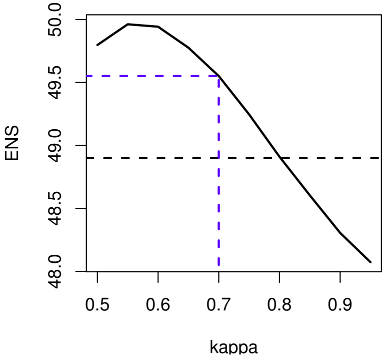
Alternative: Constrained rand. dynamic programming (Williamson et.al, 2016)

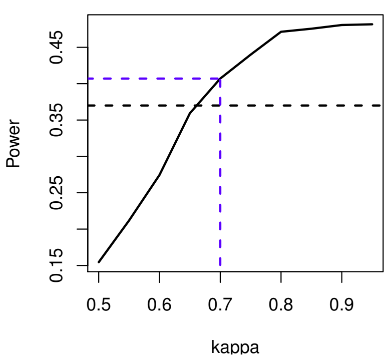
Illustration. Two arms trial

Consider the trial with $m = 2$ arms ($\alpha_1 = 0.5$ and $\alpha_2 = 0.7$), $n = 75$ patients

$$\text{Prior :} \qquad \hat{p} = (0.99, 0.99); \quad \beta = (2, 2)$$

Alternative: Constrained rand. dynamic programming (Williamson et.al, 2016)





Numerical study

We consider two trials with $m = 4$ treatments (Villar et.al, 2015)

Numerical study

We consider two trials with $m = 4$ treatments (Villar et.al, 2015)

Trial 1: $N_1 = 423$, $p = [0.3, 0.3, 0.3, 0.5]^{\mathrm{T}}$

Numerical study

We consider two trials with $m = 4$ treatments (Villar et.al, 2015)

Trial 1: $N_1 = 423$, $p = [0.3, 0.3, 0.3, 0.5]^{\mathrm{T}}$

Trial 2: $N_2 = 80$, $p = [0.3, 0.4, 0.5, 0.6]^{\mathrm{T}}$.

Numerical study

We consider two trials with $m = 4$ treatments (Villar et.al, 2015)

Trial 1: $N_1 = 423$, $p = [0.3, 0.3, 0.3, 0.5]^{\mathrm{T}}$

Trial 2: $N_2 = 80$, $p = [0.3, 0.4, 0.5, 0.6]^{\mathrm{T}}$.

$$\text{Hypothesis} \quad H_0: \ p_0 \geq p_i \text{ for } i = 1, 2, 3$$

with the family-wise error rate calculated at $p_0 = \ldots = p_3 = 0.3$

Numerical study

We consider two trials with $m = 4$ treatments (Villar et.al, 2015)

Trial 1: $N_1 = 423$, $p = [0.3, 0.3, 0.3, 0.5]^{\mathrm{T}}$

Trial 2: $N_2 = 80$, $p = [0.3, 0.4, 0.5, 0.6]^{\mathrm{T}}$.

$$\text{Hypothesis} \quad H_0 : \; p_0 \geq p_i \text{ for } i = 1, 2, 3$$

with the family-wise error rate calculated at $p_0 = \ldots = p_3 = 0.3$

$$\text{Prior :} \qquad \hat{p} = (0.99, 0.99, 0.99, 0.99); \quad \beta = (5, 2, 2, 2)$$

Numerical study

We consider two trials with $m = 4$ treatments (Villar et.al, 2015)
Trial 1: $N_1 = 423$, $p = [0.3, 0.3, 0.3, 0.5]^{\mathrm{T}}$
Trial 2: $N_2 = 80$, $p = [0.3, 0.4, 0.5, 0.6]^{\mathrm{T}}$.

$$\text{Hypothesis} \quad H_0 : \ p_0 \geq p_i \text{ for } i = 1, 2, 3$$

with the family-wise error rate calculated at $p_0 = \ldots = p_3 = 0.3$

$$\text{Prior :} \qquad \hat{p} = (0.99, 0.99, 0.99, 0.99); \quad \beta = (5, 2, 2, 2)$$

We study:

- the type-I error rate (α)
- statistical power ($1 - \eta$)
- expected number of successes (ENS)

Numerical study

We consider two trials with $m = 4$ treatments (Villar et.al, 2015)
Trial 1: $N_1 = 423$, $p = [0.3, 0.3, 0.3, 0.5]^{\mathrm{T}}$
Trial 2: $N_2 = 80$, $p = [0.3, 0.4, 0.5, 0.6]^{\mathrm{T}}$.

$$\text{Hypothesis} \quad H_0: \; p_0 \geq p_i \text{ for } i = 1, 2, 3$$

with the family-wise error rate calculated at $p_0 = \ldots = p_3 = 0.3$

$$\text{Prior}: \qquad \hat{p} = (0.99, 0.99, 0.99, 0.99); \quad \beta = (5, 2, 2, 2)$$

We study:

- the type-I error rate (α)
- statistical power ($1 - \eta$)
- expected number of successes (ENS)

Comparators:

- MAB approach based on the Gittins index
- Fixed randomization

Numerical study. Results

Trial 1

Method	$H_0 : p_0 = p_1 = p_2 = p_3 = 0.3$			$H_1 : p_0 = p_1 = p_2 = 0.3, p_3 = 0.5$		
	α	$p^*(s.e)$	ENS(s.e.)	$(1-\eta)$	$p^*(s.e.)$	ENS (s.e.)
MAB	0.05	0.25 (0.18)	126.7 (9.4)	0.43	0.83 (0.10)	198.3 (13.7)
WE ($\kappa = 0.55$)	0.05	0.22 (0.20)	126.9 (9.4)	0.55	0.83 (0.18)	197.1 (17.8)

Numerical study. Results

Trial 1

Method	$H_0 : p_0 = p_1 = p_2 = p_3 = 0.3$			$H_1 : p_0 = p_1 = p_2 = 0.3, p_3 = 0.5$		
	α	$p^*(s.e)$	ENS(s.e.)	$(1-\eta)$	$p^*(s.e.)$	ENS (s.e.)
MAB	0.05	0.25 (0.18)	126.7 (9.4)	0.43	0.83 (0.10)	198.3 (13.7)
WE ($\kappa = 0.55$)	0.05	0.22 (0.20)	126.9 (9.4)	0.55	0.83 (0.18)	197.1 (17.8)
FR	0.05	0.25 (0.02)	126.9 (9.4)	0.82	0.25 (0.02)	147.9 (9.6)
WE ($\kappa = 0.65$)	0.05	0.23 (0.13)	126.9 (9.4)	0.87	0.74 (0.10)	189.3 (13.7)

Numerical study. Results

Trial 1

Method	$H_0 : p_0 = p_1 = p_2 = p_3 = 0.3$			$H_1 : p_0 = p_1 = p_2 = 0.3, p_3 = 0.5$		
	α	$p^*(s.e)$	ENS(s.e.)	$(1-\eta)$	$p^*(s.e.)$	ENS (s.e.)
MAB	0.05	0.25 (0.18)	126.7 (9.4)	0.43	0.83 (0.10)	198.3 (13.7)
WE ($\kappa = 0.55$)	0.05	0.22 (0.20)	126.9 (9.4)	0.55	0.83 (0.18)	197.1 (17.8)
FR	0.05	0.25 (0.02)	126.9 (9.4)	0.82	0.25 (0.02)	147.9 (9.6)
WE ($\kappa = 0.65$)	0.05	0.23 (0.13)	126.9 (9.4)	0.87	0.74 (0.10)	189.3 (13.7)

Trial 2

Method	$H_0 : p_0 = p_1 = p_2 = p_3 = 0.3$			$H_1 : p_0 = 0.3, p_1 = 0.4, p_2 = 0.5, p_3 = 0.6$		
	α	$p^*(s.e)$	ENS(s.e.)	$(1-\eta)$	$p^*(s.e.)$	ENS (s.e.)
MAB	0.00	0.25 (0.13)	24.0 (4.10)	0.00	0.49 (0.21)	41.6 (5.4)
WE ($\kappa = 0.55$)	0.01	0.20 (0.15)	24.0 (4.10)	0.11	0.50 (0.27)	40.7 (5.9)

Numerical study. Results

Trial 1

Method	$H_0: p_0 = p_1 = p_2 = p_3 = 0.3$			$H_1: p_0 = p_1 = p_2 = 0.3, p_3 = 0.5$		
	α	$p^*(s.e)$	ENS(s.e.)	$(1-\eta)$	$p^*(s.e.)$	ENS (s.e.)
MAB	0.05	0.25 (0.18)	126.7 (9.4)	0.43	0.83 (0.10)	198.3 (13.7)
WE ($\kappa = 0.55$)	0.05	0.22 (0.20)	126.9 (9.4)	0.55	0.83 (0.18)	197.1 (17.8)
FR	0.05	0.25 (0.02)	126.9 (9.4)	0.82	0.25 (0.02)	147.9 (9.6)
WE ($\kappa = 0.65$)	0.05	0.23 (0.13)	126.9 (9.4)	0.87	0.74 (0.10)	189.3 (13.7)

Trial 2

Method	$H_0: p_0 = p_1 = p_2 = p_3 = 0.3$			$H_1: p_0 = 0.3, p_1 = 0.4, p_2 = 0.5, p_3 = 0.6$		
	α	$p^*(s.e)$	ENS(s.e.)	$(1-\eta)$	$p^*(s.e.)$	ENS (s.e.)
MAB	0.00	0.25 (0.13)	24.0 (4.10)	0.00	0.49 (0.21)	41.6 (5.4)
WE ($\kappa = 0.55$)	0.01	0.20 (0.15)	24.0 (4.10)	0.11	0.50 (0.27)	40.7 (5.9)
FR	0.05	0.25 (0.04)	24.0 (4.10)	0.50	0.25 (0.04)	36.0 (4.3)
WE ($\kappa = 0.65$)	0.05	0.24 (0.07)	24.0 (4.05)	0.52	0.47 (0.21)	40.2 (4.8)

Conclusion

- Simple, intuitevely clear, can be computed by non-statisticians

Conclusion

- Simple, intuitevely clear, can be computed by non-statisticians
- Penalty parameter κ reflects the trade-off between ENS and Power

Conclusion

- Simple, intuitevely clear, can be computed by non-statisticians
- Penalty parameter κ reflects the trade-off between ENS and Power
- Performs better than currently used approaches

	MAB	FR
Power	higher	same
ENS	same	higher

Conclusion

- Simple, intuitevely clear, can be computed by non-statisticians
- Penalty parameter κ reflects the trade-off between ENS and Power
- Performs better than currently used approaches

	MAB	FR
Power	higher	same
ENS	same	higher

- Can be applied to any trial with the target $\gamma \in (0, 1)$

Conclusion

- Simple, intuitevely clear, can be computed by non-statisticians
- Penalty parameter κ reflects the trade-off between ENS and Power
- Performs better than currently used approaches

	MAB	FR
Power	higher	same
ENS	same	higher

- Can be applied to any trial with the target $\gamma \in (0, 1)$
- Theoretical result: the design is consistent
- The criterion can be generalized for multinomial outcomes