

15.095: Machine Learning under a Modern Optimization Lens

Lecture 15: Randomization vs Optimization in Clinical Trials

Motivation

- Randomized trials is the golden standard for over a century.
- We want to test a cancer drug on mice. We build randomly 4 groups of ten mice each and give 3 drugs and a placebo. Is it sufficient?
- The cost of developing a new drug is approximately \$1 billion.
- A phase III trial involving 300 patients costs between \$10-30 million.
- Can we decrease it substantially?

Outline

- 1 Limitations of Randomization
- 2 MIO Approach
- 3 Optimization vs. Randomization in Reducing Discrepancies
- 4 Optimization, Randomization, and Bias
- 5 Optimization vs. Randomization in Making a Conclusion
- 6 Practical Significance

Outline

- 1 Limitations of Randomization
- 2 MIO Approach
- 3 Optimization vs. Randomization in Reducing Discrepancies
- 4 Optimization, Randomization, and Bias
- 5 Optimization vs. Randomization in Making a Conclusion
- 6 Practical Significance

Limitations of Randomization

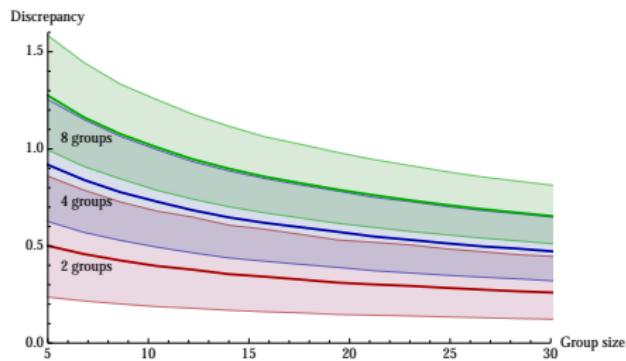


Figure: Average maximal pairwise discrepancy in means among randomly assigned groups of normal variates. The vertical axis is in units of standard deviation. The band denotes the average over- and under-shoot: $\mathbb{E}[X|X \geq \mathbb{E}X]$ and $\mathbb{E}[X|X \leq \mathbb{E}X]$ where X is maximal pairwise discrepancy.

Observations

- Discrepancy increases with the number of groups involved and decreases with increasing group size.
- Discrepancy can be **substantial**. Four groups of ten mice each has an average discrepancy of 0.66 standard deviations between some two of the groups.
- Doubling the group sizes to twenty leaves a discrepancy of 0.47 standard deviations.
- Discrepancy often larger than the effect of treatment.

Possible Solutions

- Increase group size until discrepancies decrease to acceptable levels.
- To reduce the expected discrepancy to below 0.1 standard deviations would require more than 400 subjects per group.
- For 0.01 standard deviations, more than 40,000 subjects per group would be necessary.

Outline

- 1 Limitations of Randomization
- 2 MIO Approach
- 3 Optimization vs. Randomization in Reducing Discrepancies
- 4 Optimization, Randomization, and Bias
- 5 Optimization vs. Randomization in Making a Conclusion
- 6 Practical Significance

MIO Approach

- Assign subjects so to minimize the discrepancies in centered first and second moments via integer optimization.
- After assignment, we randomize which group is given which treatment.
- Given w_i , $i = 1, \dots, n = mk$, create m groups each containing k subjects so that the discrepancy in means and ρ times the discrepancy in second moments is minimized between any two groups.
- Preprocess

$$w'_i = (w_i - \hat{\mu})/\hat{\sigma}, \quad \text{where} \quad \hat{\mu} = \sum_{i=1}^n w_i/n \quad \text{and} \quad \hat{\sigma}^2 = \sum_{i=1}^n (w_i - \hat{\mu})^2/n.$$

Decision Variables

$x_{ip} = 0$ or 1 to denote the assignment of subject i to group p . Auxiliary variables

$$\mu_p(x) = \frac{1}{k} \sum_{i=1}^n w'_i x_{ip} \quad \text{and} \quad \sigma_p^2(x) = \frac{1}{k} \sum_{i=1}^n (w'_i)^2 x_{ip},$$

Formulation

$$Z_m^{\text{opt}}(\rho) = \min_x \max_{p \neq q} (|\mu_p(x) - \mu_q(x)| + \rho |\sigma_p^2(x) - \sigma_q^2(x)|)$$

$$= \min_{x,d} d$$

s.t. $\forall p < q = 1, \dots, m :$

$$d \geq \mu_p(x) - \mu_q(x) + \rho \sigma_p^2(x) - \rho \sigma_q^2(x)$$

$$d \geq \mu_p(x) - \mu_q(x) + \rho \sigma_q^2(x) - \rho \sigma_p^2(x)$$

$$d \geq \mu_q(x) - \mu_p(x) + \rho \sigma_p^2(x) - \rho \sigma_q^2(x)$$

$$d \geq \mu_q(x) - \mu_p(x) + \rho \sigma_q^2(x) - \rho \sigma_p^2(x)$$

$$x_{ip} \in \{0, 1\}, \quad x_{ip} = 0 \quad \forall i < p.$$

$$\sum_{i=1}^n x_{ip} = k \quad \forall p = 1, \dots, m$$

$$\sum_{p=1}^m x_{ip} = 1 \quad \forall i = 1, \dots, n$$

Size and Progress

- MIO with $m(1 + 2n - m)/2$ binary variables and 1 continuous variable.
- Last constraint reduces redundancy due to permutation symmetry.
- Progress. $n = 40$ and $m = 4$ problem can be solved to full optimality in under twenty seconds

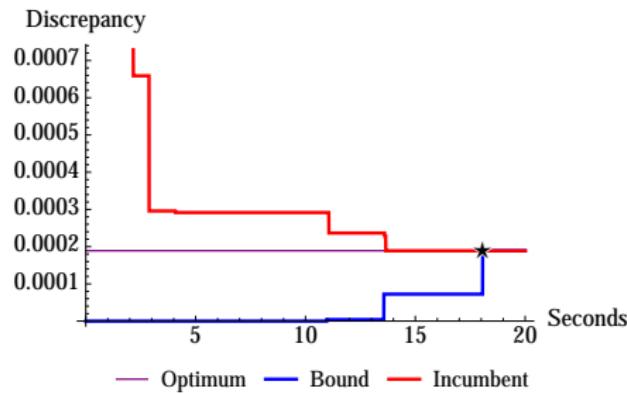


Figure: The progress of solving an instance of MIO with $n = 40$, $m = 4$.

Impact of Optimization

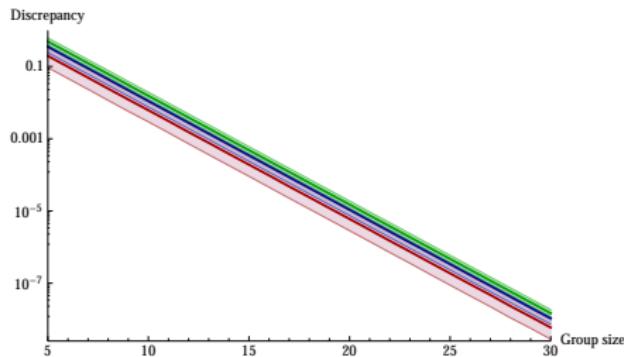


Figure: Discrepancy in means among optimally assigned groups of normal variates with $\rho = 0$. Note the vertical log scale.

Multiple Covariates

- We are interested in matching the first and second moments in a vector of r covariates where w_{is} denotes the s^{th} covariate of subject i .
- Given the tradeoff parameter ρ , we rewrite the optimization problem using $m(1 + 2n - m)/2$ binary and $1 + m(m - 1)r(r + 3)/4$ continuous variables

Multiple Covariates

$$\min d$$

$$\text{s.t. } x \in \{0, 1\}^{n \times m}, x_{ip} = 0 \forall i < p, d \geq 0$$

$$\sum_{i=1}^n x_{ip} = k \quad \forall p = 1, \dots, m$$

$$\sum_{p=1}^m x_{ip} = 1 \quad \forall i = 1, \dots, n$$

$$x_{ip} = 0 \quad \forall i < p$$

$$M \in \mathbb{R}^{\frac{m(m-1)}{2} \times r}, V \in \mathbb{R}^{\frac{m(m-1)}{2} \times \frac{r(r+1)}{2}}$$

Multiple Covariates

$\forall p = 1, \dots, m, q = p + 1, \dots, m :$

$$d \geq \sum_{s=1}^r M_{pq s} + \rho \sum_{s=1}^r V_{pq s s} + 2\rho \sum_{s=1}^r \sum_{s'=s+1}^r V_{pq s s'}$$

$\forall s = 1, \dots, r :$

$$M_{pq s} \geq \frac{1}{k} \sum_{i=1}^n w'_{is} (x_{ip} - x_{iq})$$

$$M_{pq s} \geq \frac{1}{k} \sum_{i=1}^n w'_{is} (x_{iq} - x_{ip})$$

$\forall s = 1, \dots, r, s' = s, \dots, r :$

$$V_{pq s s'} \geq \frac{1}{k} \sum_{i=1}^n w'_{is} w'_{is'} (x_{ip} - x_{iq})$$

$$V_{pq s s'} \geq \frac{1}{k} \sum_{i=1}^n w'_{is} w'_{is'} (x_{iq} - x_{ip}).$$

Outline

- 1 Limitations of Randomization
- 2 MIO Approach
- 3 Optimization vs. Randomization in Reducing Discrepancies**
- 4 Optimization, Randomization, and Bias
- 5 Optimization vs. Randomization in Making a Conclusion
- 6 Practical Significance

Optimization vs. Randomization in Reducing Discrepancies

The discrepancy of four groups of ten mice each under optimization with $\rho = 0.5$ is 0.0005 standard deviations in first moment compared with 0.66 standard deviations under randomization.

Optimization vs. Randomization in Reducing Discrepancies

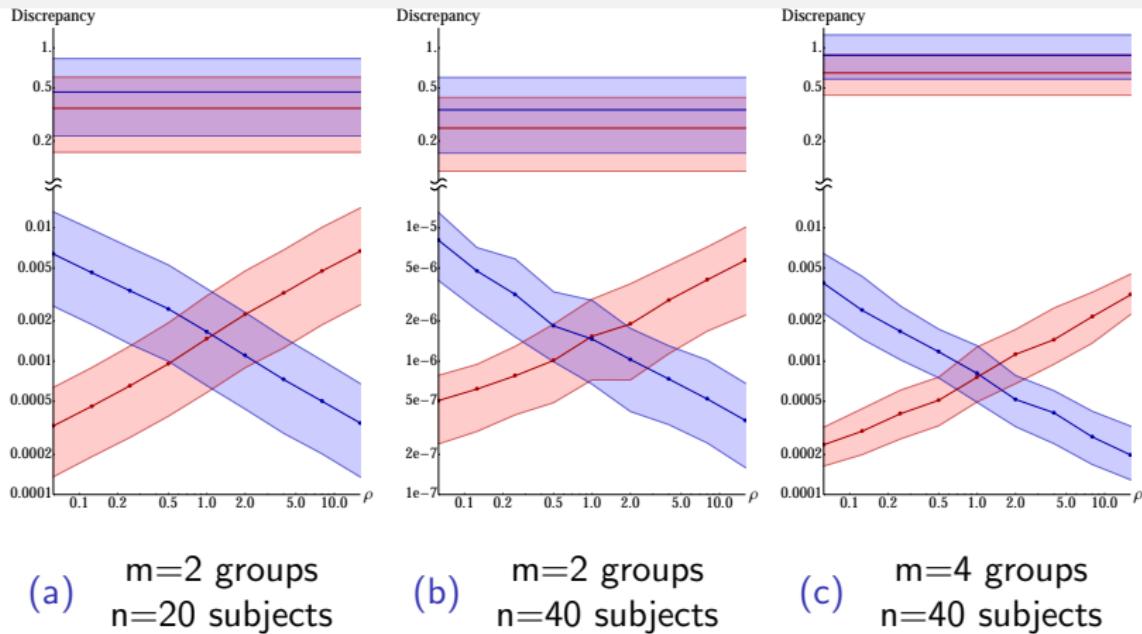


Figure: Upper halves correspond to randomization and lower ones to optimization. Red denotes discrepancy in mean and blue variance. The bands depict average under- and over-shoot. Note log scales and the break in the vertical axis.

Theoretical Backing

- Z_2^{rand} : discrepancy in means under randomization. When covariates are random with variance σ^2 ,

$$\mathbb{E}[Z_2^{\text{rand}}] \leq \sqrt{\frac{2}{k}}\sigma$$

and if normally distributed then

$$\mathbb{E}[Z_2^{\text{rand}}] = \frac{2}{\sqrt{\pi k}}\sigma.$$

- For $\rho = 0$, $m = 2$ for optimization, there is a $C > 0$ such that

$$\text{median}(Z_2^{\text{opt}}) \leq \frac{C}{2^{2k}}$$

and heuristic arguments from spin-glass theory provide the prediction

$$\mathbb{E}[Z_2^{\text{opt}}(0)] = \frac{2\pi\sigma}{2^k}.$$

- Comparing the asymptotic orders of Z_2^{rand} and Z_2^{opt} , we see an **exponential reduction** in discrepancies by optimization versus randomization.

Number of Samples needed

- For the normal distribution, if we want to limit discrepancy to less than $\epsilon\sigma$, we see a dramatic difference in the necessary number of subjects per group, k :

$$k^{\text{Opt}} = \left\lceil \log_2 \frac{2\pi}{\epsilon} \right\rceil, \quad k^{\text{Rand}} = \left\lceil \frac{4}{\pi\epsilon^2} \right\rceil.$$

- Using other modern methods for randomization, we see a dramatic improvement in optimization:

ϵ	k^{Opt}	k^{Rand} ($\approx k^{\text{FSM}}$)	k^{Pair}	k^{RR}
0.1	3	128	9	4
0.01	5	12833	65	83
0.001	7	1273240	514	8130
0.0001	8	127323955	4354	820143

Table: The number of subjects per group needed to guarantee an expected discrepancy no more than $\epsilon\sigma$ for $m = 2$ and $\rho = 0$.

Higher Moments

k	Method	1	2	3	4	5	\log
5	Opt	0.0513	0.286	1.43	2.67	9.75	0.498
	Rand	0.510	0.689	1.79	3.81	10.3	0.544
	Pair	0.184	0.498	1.27	3.29	8.93	0.345
	Re-rand	0.047	0.711	1.09	3.88	8.47	0.572
	FSM	0.508	0.553	1.76	3.33	10.2	0.440
10	Opt	0.00174	0.0145	0.906	1.47	6.87	0.338
	Rand	0.352	0.504	1.30	2.88	7.79	0.399
	Pair	0.0839	0.259	0.759	2.09	6.06	0.176
	Re-rand	0.0298	0.497	0.764	2.93	6.20	0.389
	FSM	0.374	0.334	1.33	2.26	7.90	0.264
20	Opt	1.23e-6	2.34e-6	0.600	1.04	5.23	0.221
	Rand	0.258	0.345	0.947	2.13	6.13	0.276
	Pair	0.0379	0.140	0.445	1.40	4.24	0.286
	Re-rand	0.0207	0.356	0.565	2.16	4.99	0.284
	FSM	0.249	0.190	0.896	1.50	5.89	0.146

Table: Col. ℓ corresponds to the average mismatch in the ℓ^{th} moments between the two groups and the last col. to the mismatch in $\log |w|$

Outline

- 1 Limitations of Randomization
- 2 MIO Approach
- 3 Optimization vs. Randomization in Reducing Discrepancies
- 4 Optimization, Randomization, and Bias**
- 5 Optimization vs. Randomization in Making a Conclusion
- 6 Practical Significance

Optimization, Randomization, and Bias

- Investigator bias, or the possibility that an investigator may construct groups in a manner that biases toward achieving a particular result. As a mechanical process, optimization guards against this possibility.
- Incidental disproportionate assignment of variables, measured or hidden, that directly affect the treatment.
- Randomization, given large enough samples, equalizes the apportionment of any one factor, but cannot eliminate discrepancies in hidden factors when samples are relatively small.
- As we randomize the identity of treatments, the observed difference in treatment effects will always be an *unbiased estimator* of the true population average difference.

Outline

- 1 Limitations of Randomization
- 2 MIO Approach
- 3 Optimization vs. Randomization in Reducing Discrepancies
- 4 Optimization, Randomization, and Bias
- 5 Optimization vs. Randomization in Making a Conclusion
- 6 Practical Significance

Optimization vs. Randomization in Making a Conclusion

- We want to test the null hypothesis that every subject $i = 1, \dots, n$ would have had the same response to treatment whether either of the two treatments were assigned.
- v_i denote the response measured for subject i after it was administered the treatment to which it was assigned.

Optimization vs. Randomization in Making a Conclusion

- ① Find an optimal assignment of these two groups (permuting randomly):

$$\{i_1, \dots, i_{n/2}\} \text{ and } \{i_{n/2+1}, \dots, i_n\}.$$

- ② Administer treatments and measure responses v_i .
 - ③ Compute $\delta = \frac{1}{k} (v_{i_1} + \dots + v_{i_{n/2}}) - \frac{1}{k} (v_{i_{n/2+1}} + \dots + v_{i_n})$.
 - ④ For $b = 1, \dots, B$:
 - ① Draw a random sample with replacement $w_{b,1}, \dots, w_{b,n}$ from w_1, \dots, w_n .
 - ② Find an optimal assignment of these to two groups (permuting randomly):
- $\{i_{b,1}, \dots, i_{b,n/2}\}$ and $\{i_{b,n/2+1}, \dots, i_{b,n}\}$.
- ③ Compute $\delta_b = \frac{1}{k} (v_{i_{b,1}} + \dots + v_{i_{b,n/2}}) - \frac{1}{k} (v_{i_{b,n/2+1}} + \dots + v_{i_{b,n}})$.
 - ⑤ Compute the p -value $p = \frac{1}{1+B} \left(1 + \sum_{b=1}^B \mathbb{I}[|\delta_b| \geq |\delta|] \right)$.
 - ⑥ To test our null hypothesis at a significance of α , we only reject it if $p \leq \alpha$.

An Example

- Two groups, each of k mice, with tumor weights initially normally distributed with mean 200mg and standard deviation 300mg (truncated to be nonnegative). Two treatments are considered: a placebo and a proposed treatment. Their effect on the tumor, allowed to grow for a period of a day, is of interest to the study.
- We consider a hidden reality where the growth of the tumors are dictated by the Gomp-ex model of tumor growth, that is

$$\frac{dw}{dt} = w(t) \left(a + \max \{0, b \log(w_c/w(t))\} \right),$$

where a and b are rate parameters and w_c is the critical weight that marks the change between exponential and logistic growth. We arbitrarily choose $a = 1 \frac{1}{\text{day}}$, $b = 5 \frac{1}{\text{day}}$, $w_c = 400\text{mg}$, and $t = 1\text{day}$. We pretend that tumors under either treatment grow according to this equation but subtract δ_0 from the final weights for the proposed treatment. We consider δ_0 being 0mg (no effect), 50mg (small effect), and 250mg (large effect).

An Example Continued ...

For various values of k and for several draws of initial weights, we consider assignments produced by randomization, our optimization approach ($\rho = 0.5$), pairwise matching, and re-randomization. We consider both the post-treatment estimate of the effect and the inference drawn on it at a significance of $\alpha = 0.05$, using our bootstrap test for our method and the standard randomization test for the others.

Cost-Saving Benefits

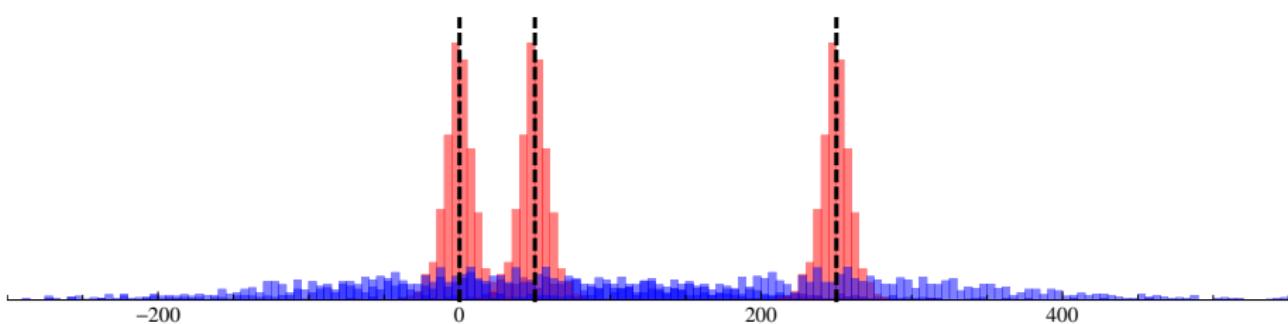


Figure: The distribution of estimates of effect size under optimization (red) and randomization (blue) for $k = 20$ and effect sizes 0mg, 50mg, and 250mg (dashed lines). The overlap of estimates under randomization of the nonzero effects and of the zero effect elucidate the low statistical power of randomization in detecting the nonzero effects.

Cost-Saving Benefits

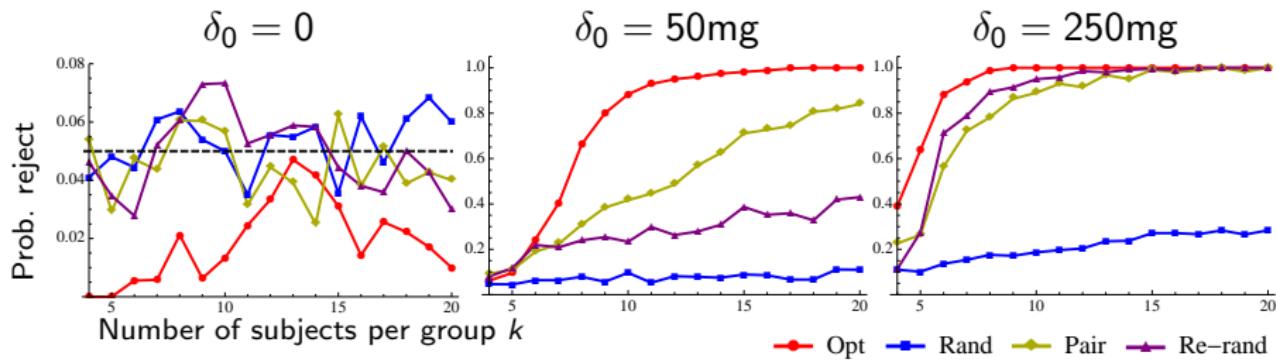


Figure: The probability of rejecting the null hypothesis of no effect for various effect sizes.

Outline

- 1 Limitations of Randomization
- 2 MIO Approach
- 3 Optimization vs. Randomization in Reducing Discrepancies
- 4 Optimization, Randomization, and Bias
- 5 Optimization vs. Randomization in Making a Conclusion
- 6 Practical Significance

Practical Significance

- Optimization produces groups that are exponentially more similar in mean and variance than those created by randomization, especially in situations in which group size is small, data variability is large, and numerous groups are needed for a single experiment.
- For each additional subject per group, optimization roughly halves the discrepancy in the covariate, whereas both randomization and subject-pair matchings offer quickly diminishing reductions.
- Making groups similar before treatment allows for statistical power beyond what can normally be hoped for with small samples.
- Optimization protects against experimental biases at least as well as randomization and that the advantage of optimized groups over randomized groups is substantial.
- Online application of this method is feasible