

Optimising external pilot trials of complex interventions using Bayesian statistical decision theory

Duncan T. Wilson

11 March, 2020

Introduction

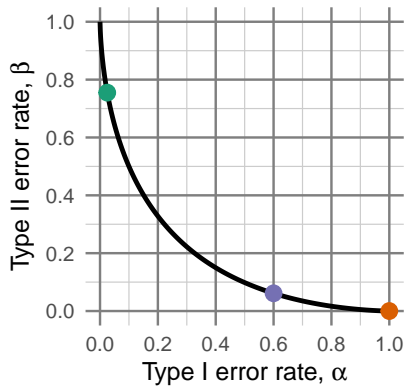
“What is the principal distinction between Bayesian and classical statistics? It is that Bayesian statistics is fundamentally **boring**. There is so little to do: *just specify the model and the prior, and turn the Bayesian handle*” - Philip Dawid

Introduction

“What is the principal distinction between Bayesian and classical statistics? It is that Bayesian statistics is fundamentally **boring**. There is so little to do: *just specify the model and the prior and the utility, and turn the Bayesian handle*” - Philip Dawid

We will consider how to apply the Bayesian decision-theoretic approach to problems of pilot trial design, considering first how we can define utility, and then how we do the handle turning.

Motivation



A massive over-simplification

Suppose

Bayesian Statistical Decision Theory

One way to define optimality is through Bayesian statistical decision theory. This assumes:

- i) We can represent our *beliefs* through a probability distribution, p ;

$$A \text{ more likely than } B \Leftrightarrow p(A) > p(B)$$

- ii) We can represent our *preferences* through a utility function, U ;

$$A \text{ is preferable to } B \Leftrightarrow E[u(A)] > E[u(B)]$$

It is axiomatic that this p and u exist (for an individual). But, how do we find them?

Part 1 - Utilities

Problem

Consider a two stage trial programme with an (external) pilot at stage $i = 1$ and a definitive trial at stage $i = 2$, with sample sizes n_i . We have a normally distributed primary outcome with known variance, and denote the sample mean difference by \bar{x}_i , $i = 1, 2$.

A frequentist analysis:

- ▶ At stage 1, proceed to stage 2 if and only if $\bar{x}_1 > d_1$;
- ▶ At stage 2, recommend the intervention over the control if and only if $\bar{x}_2 > d_2$.

The design problem is then: optimise (n_1, d_1, n_2, d_2) .

Attributes

To define our utility, we first need to choose a list of *attributes*: the things we have preferences about.

- i) The average change outcome in the population
- ii) The total sample size of the programme
- iii) Difference in treatment costs between intervention and control

Attributes [and assumptions]

To define our utility, we first need to choose a list of *attributes*: the things we have preferences about.

- i) The average change outcome in the population [everyone will get the recommended treatment]
- ii) The total sample size of the programme [no trial set-up costs]
- iii) Difference in treatment costs between intervention and control [known and deterministic]

Denote these attributes by d , n , and c .

Value

A useful intermediate step between attributes and utility is *value*: preferences under certainty.

$$(d, n, c) \text{ is preferred to } (d', n', c') \Leftrightarrow v(d, n, c) > v(d', n', c')$$

It is helpful to identify structure in our preferences. For example, we could say that

$$v(d, n, c) = k_d v_d(d) + k_n v_n(n) + k_c v_c(c),$$

i.e. the overall value is a weighted sum of three single-attribute value functions.

Q: When would this be reasonable to assume?

A: Under *pairwise preferential independance*

Single-attribute value functions

Under pairwise preferential independence, elicitation is now much simpler.

$$v_d(d) = d, \quad v_n(n) = n, \quad v_c(c) = c.$$

Note the implicit assumptions:

- ▶ A change in attribute level from d to $d + \Delta$ is independent of the starting level d ;
- ▶ Likewise for attribute n .

We are now left with determining the weights k_d , k_n and k_c .

Weights

We can scale everything to give $k_d + k_n + k_c = 1$. Then, we (at least) two questions:

- i) Specify a change in outcome, $\bar{b}ard$ that would justify increasing the total sample size from 0 to some value n^* :

$$(d = 0; n = 0; c = 1) \sim (d = \bar{d}; n = n^*; c = 1).$$

- ii) Specify a change in outcome \hat{d} that would justify switching from the current standard treatment to the intervention under study:

$$(d = 0; n; c = 1) \sim (d = \hat{d}; n; c = 0).$$

We can then solve for the weights.

Utility

Finally, we need to use our value function to obtain a utility function, which needs to represent **preferences under uncertainty**.

Remarkably, it turns out that if we have an additive value function, then our utility must* take the form:

$$u(d, n, c) = \begin{cases} 1 - e^{-\rho v(d, n, c)}, & \rho > 0 \\ v(d, n, c), & \rho = 0, \\ -1 + e^{-\rho v(d, n, c)}, & \rho < 0 \end{cases}$$

- ▶ We also require that one attribute is *utility independent* of the other two.

Note: - All utility functions are value functions, but not vice versa. - Any additive utility function implies an additive value function and risk-neutrality.

Attitude to risk

Utility elicitation is completed by determining $\rho \in \mathbb{R}$, which represents our **attitude to risk**.

We ask for a value d^* such that we would be indifferent between the following:

1. Obtaining d^* with certainty;
2. A gamble which will result in d_{min} with probability 0.5 and d_{max} with probability 0.5.

That is, we find d^* such that

$$u(d^*, n, c) = 0 : 5u(d_{min}, n, c) + 0 : 5u(d_{max}, n, c),$$

which we can solve for ρ . We will assume that we are risk-averse, giving $\rho > 0$.

Expected utility

Let $G_i = 1$ if there is a positive test result at stage i , and $G_i = 0$ otherwise. Then the expected utility of a programme z , conditional on the true effect μ , is

$$\begin{aligned} E_{G_1, G_2, | \mu, z}[u(\mu, G_1, G_2 | z)] = & Pr[G_1 = 1, G_2 = 1 | \mu] \left(1 - e^{\rho(k_d \mu + k_n(n_1 + n_2))}\right) \\ & Pr[G_1 = 1, G_2 = 0 | \mu] \left(1 - e^{-\rho(k_n(n_1 + n_2) + k_c)}\right) \\ & Pr[G_1 = 0 | \mu] \left(1 - e^{-\rho(k_n n_1 + k_c)}\right) \end{aligned}$$

We can then integrate over the prior distribution for μ :

$$E[u(\mu, G_1, G_2 | z)] = \int E_{G_1, G_2, | \mu, z}[u(\mu, G_1, G_2 | z)] p(\mu) d\mu.$$

Application: OK-Diabetes

- ▶ External pilot trial evaluating supported self-management for adults with learning disabilities and type II diabetes
- ▶ Sample size of $n_1 = 30$ based on a rule-of-thumb
- ▶ Asked by the funder to consider assessing the potential efficacy of the intervention, based on blood sugar levels
- ▶ Sample size increased to $n_1 = 56$ participants per arm, giving 0.82 power to detect a true mean reduction of 0.5% using a one-sided test with a type I error rate of 0.2

Application: OK-Diabetes

- ▶ Cost of new intervention:

$$(d = 0; n; c = 1) \sim (\hat{d} = 0.3; n; c = 0)$$

- ▶ Cost of sampling:

$$(d = 0; n = 0; c = 1) \sim (\bar{d} = 0.005; n = 50; c = 1)$$

- ▶ Attitude to risk:

$$u(d^* = 0.19, n, c) = 0.5u(d_{\min} = 0, n, c) + 0.5u(d_{\max} = 0.5, n, c)$$

- ▶ Prior on true treatment difference:

$$\mu \sim \text{Normal}, \quad \mathbb{E}[\mu] = 0, \quad \Pr[\mu > 0.5] = 0.2$$

Application: OK-Diabetes

Compare the optimal designs under two strategies:

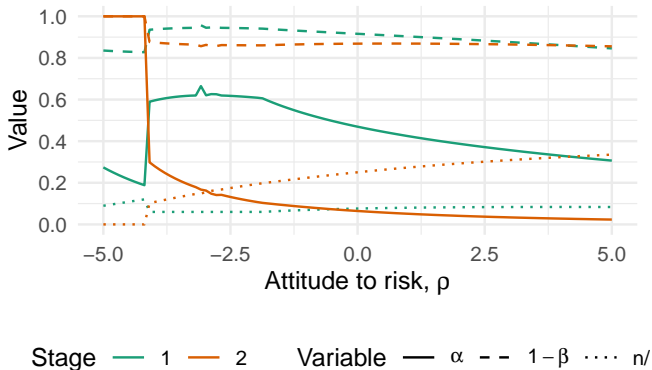
- ▶ Unrestricted: optimising over $\alpha_1, \beta_1, \alpha_2, \beta_2$
- ▶ No pilot test: optimising over α_2, β_2 subject to $\alpha_1 = 1, n_1 = 30$

Problem	n_1	n_2	α_1	β_1	α_2	β_2	Expected utility
Unrestricted	41	146	0.39	0.11	0.041	0.132	-0.42874
No pilot test	30	110	1.00	0.00	0.036	0.254	-0.42292

Note: the 0.00582 difference in expected utility approximately equals the cost of recruiting 66 participants.

Evaluation

How do optimal designs vary with utility function parameters?
Consider varying the attitude to risk, ρ :



Extensions

- ▶ Internal pilot
- ▶ More complex designs and models, e.g. unknown variance or binary outcome
- ▶ More complex utilities, e.g. more attributes

Part 2 - Computations

Re-cap

We have seen Bayesian solution to a frequentist trial design problem:

- ▶ External pilot trial followed by a definitive trial
- ▶ Both trials to be analysed using a hypothesis test
- ▶ Needed to choose type I and II error rates $\alpha_i, \beta_i, i = 1, 2$ at each stage
- ▶ Proposed to choose the rates which maximised expected utility, w.r.t. a prior distribution

We have focussed so far on how to choose an appropriate utility function.

A Bayesian approach

Rather than making a stop/go decision at the pilot stage through a frequentist hypothesis test, we could instead do a Bayesian decision-theoretic analysis:

- ▶ Given pilot data x_1 , get the posterior $\mu|x_1$
- ▶ Choose α_2, β_2 to maximise expected utility with respect to this posterior

With this analysis in mind, how large should the pilot trial be?

To simplify:

- ▶ Fix $\alpha_2 = 0.025$, so just need to choose β_2 or, equivalently, the definitive sample size n_2
- ▶ Assume the pilot data can be summarised in a single sufficient statistic (e.g. a sample mean)

Notation re-cap

μ : unknown parameter(s) of interest, with prior $p(\mu)$

n_1, n_2 : sample size for pilot (1) and definitive (2) trials

x_1, x_2 : data from the pilot (1) and definitive (2) trials

- Utility is a function of these:

$$u(\mu, n_1, n_2, x_1, x_2)$$

Maximising expected utility

Given pilot data x_1 from a pilot of sample size n_1 :

$$\max_{n_2} \mathbb{E}_{\mu, x_2 | x_1, n_1} [u(\mu, n_1, n_2, x_2)]$$

Maximising expected utility

Given a pilot of sample size n_1 , average over the pilot data x_1 :

$$\mathbb{E}_{x_1|n_1} \left(\max_{n_2} \mathbb{E}_{\mu, x_2|x_1, n_1} [u(\mu, n_1, n_2, x_2)] \right)$$

Maximising expected utility

Find optimal pilot sample size n_1 :

$$\max_{n_1} \left[\mathbb{E}_{x_1|n_1} \left(\max_{n_2} \mathbb{E}_{\mu, x_2|x_1, n_1} [u(\mu, n_1, n_2 \cdot x_2)] \right) \right]$$

Two-level nested Monte Carlo

For now, forget about choosing pilot sample size. Given some n_1 , how do we calculate

$$\mathbb{E}_{x_1|n_1} \left(\max_{n_2} \mathbb{E}_{\mu, x_2|x_1, n_1} [u(\mu, n_1, n_2, x_2)] \right)$$

We can:

- ▶ $\mathbb{E}_{x_1|n_1}$ - sample N pilot data $x_1^{(i)} \sim p(x_1, |n_1)$
- ▶ \max_{n_2} - exhaustive / bisection / other search
- ▶ $\mathbb{E}_{\mu, x_2|x_1, n_1}$ - sample M posterior parameters $\mu^{(j)} \sim p(\mu|x_1, n_1)$

Computational bottleneck: for each outer pilot data sample $x_1^{(i)}$, need to generate a set of posterior samples (through MCMC).

An alternative

Note that the inner optimisation problem maps pilot data x_1 to definitive sample size n_2 :

$$f(x_1) = \operatorname{argmax}_{n_2} \mathbb{E}_{\mu, x_2 | x_1, n_1} [u(\mu, n_1, n_2, x_2)]$$
$$f : \mathcal{X}_\infty \rightarrow \mathbb{N}$$

We can re-write our earlier expression using $f(x_1)$:

$$\mathbb{E}_{x_1} \left(\mathbb{E}_{\mu, x_2 | x_1} [u(\mu, n_1, n_2 = f(x_1), x_2)] \right)$$

Now, with the “max” removed, we can move around our expectations:

$$\begin{aligned} & \mathbb{E}_{x_1} \left(\mathbb{E}_{\mu, x_2 | x_1} [u(\mu, n_1, f(x_1), x_2)] \right) \\ &= \mathbb{E}_{\mu, x_1, x_2} \left(u(\mu, n_1, f(x_1), x_2) \right) \\ &= \mathbb{E}_{\mu, x_1} \left(\mathbb{E}_{x_2 | \mu, x_1} [u(\mu, n_1, f(x_1), x_2)] \right) \end{aligned}$$

Single-level Monte Carlo

We now have something we can estimate in a single-level Monte Carlo scheme:

$$\mathbb{E}_{\mu, x_1} \left(\mathbb{E}_{x_2 | \mu, x_1} [u(\mu, n_1, f(x_1), x_2)] \right)$$

- ▶ \mathbb{E}_{μ, x_1} - sample $N(\mu^{(i)}, x_1^{(i)}) \sim p(\mu, x_1 | n_1) = p(x_1 | \mu, n_1) p(\mu)$

$$\begin{aligned} \mathbb{E}_{x_2 | \mu, x_1} [u(\mu, n_1, f(x_1), x_2)] = & Pr[x_2 > c_2 | \mu, n_2] u(\mu, n_1, n_2, x_2 > c_2) + \\ & Pr[x_2 \leq c_2 | \mu, n_2] u(\mu, n_1, n_2, x_2 \leq c_2), \end{aligned}$$

where c_2 is the critical value in the definitive trial hypothesis test.

- ▶ $\mathbb{E}_{x_2 | \mu, x_1}$ - can be calculated exactly if the power function $Pr[x_2 > c_2 | \mu, n_2]$ is known

Optimisation

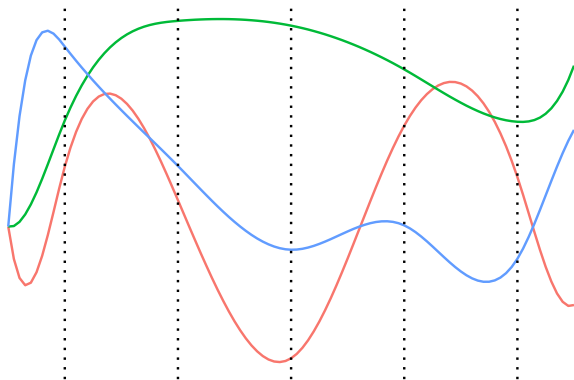
But, we do not know the optimal decision rule $f(x_1)$. Note that

$$\begin{aligned} & \mathbb{E}_{x_1} \left(\max_{n_2} \mathbb{E}_{\mu, x_2 | x_1} [u(\mu, n_1, n_2 \cdot x_2)] \right) \\ & \geq \\ & \mathbb{E}_{\mu, x_1} \left(\mathbb{E}_{x_2 | \mu, x_1, n_1} [u(\mu, n_1, f(x_1), x_2)] \right), \text{ for all } f. \end{aligned}$$

\Rightarrow if we can search over functions $f(x_1)$ and find that which maximises expected utility, this will coincide with the optimal decision rule.

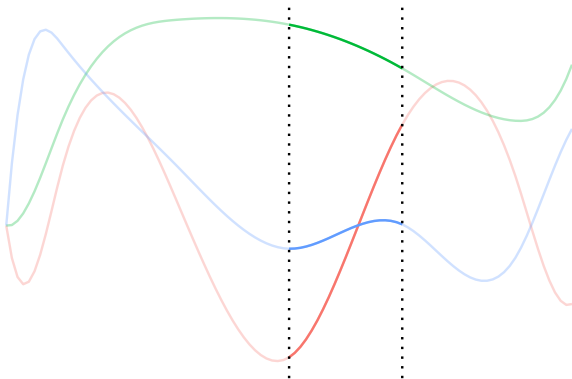
Splines

Spline: a piecewise polynomial defined by a set of *knots* and by the degree of the polynomial pieces.



Splines

Spline: a piecewise polynomial defined by a set of *knots* and by the degree of the polynomial pieces.



B-splines

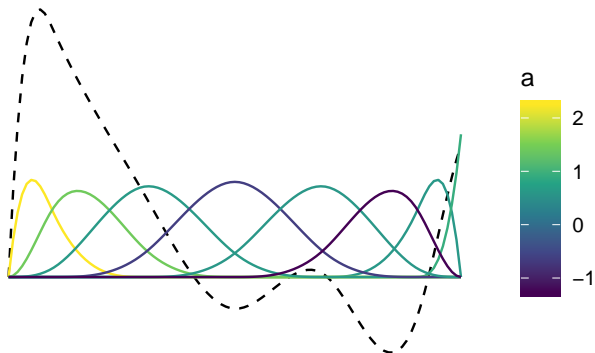
For a specific set of k knots, there is a set of cubic splines which serves as a basis for the space of all cubic splines defined over the same knots.

Denoting these so-called b-splines by $\phi_i(x)$, $k = 1, \dots, k + 4$, a function can then be written as

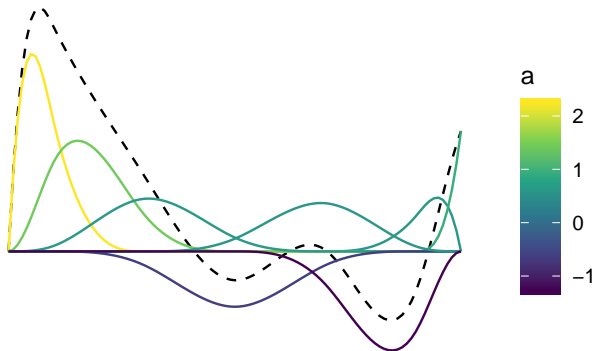
$$f_{\mathbf{a}}(x) = \sum_{i=1}^{k+4} a_i \phi_i(x),$$

So, functions are then defined by a vector of coefficients $\mathbf{a} \in \mathbb{R}^{k+4}$.

Example



Example



Optimisation

An algorithm for estimating the expected utility of a pilot trial with sample size n_1 :

- ▶ Choose a set of k knots
- ▶ Generate a sample $(\mu^{(i)}, x_1^{(i)}) \sim p(\mu, x_1 | n_1), i = 1, \dots, N$
- ▶ While not converged:
 - ▶ select candidate $\mathbf{a} \in \mathbb{R}^{k+4}$
 - ▶ calculate $\frac{1}{N} \sum_{i=1}^N \mathbb{E}_{x_2 | \mu^{(i)}, x_1^{(i)}} [u(\mu^{(i)}, n_1, f_{\mathbf{a}}(x_1^{(i)}), x_2)]$
 - ▶ check convergence
- ▶ Return maximum

Note that we use the same sample from $p(\mu, x_1 | n_1)$ at each iteration.

Application: OK-Diabetes

The function we are maximising is:

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{x_2 | \mu^{(i)}, x_1^{(i)}} [u(\mu^{(i)}, n_1, f_{\mathbf{a}}(x_1^{(i)}), x_2)] \\ &= \frac{1}{N} \sum_{i=1}^N \left[pow(n)(1 - e^{-\rho(k_d \mu + k_n(n_1 + n_2))}) + \right. \\ & \quad \left. (1 - pow(n))(1 - e^{-\rho(k_n(n_1 + n_2) + k_c)}) \right], \end{aligned}$$

where

$$pow(n) = \Phi \left(z_{1-\alpha/2} - \frac{\mu}{\sqrt{(2\sigma^2/n)}} \right).$$

We can get the derivative of this function with respect to the vector \mathbf{a} - helps dramatically with optimisation.

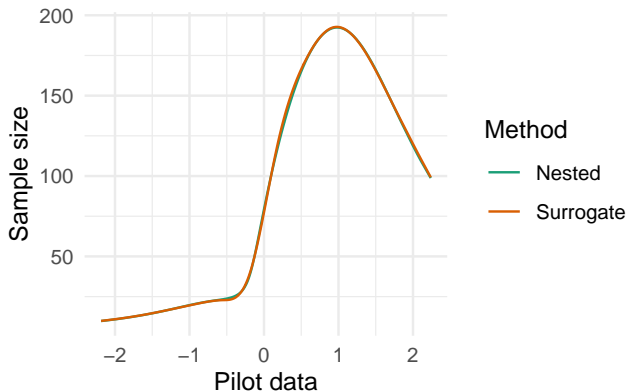
Application: OK-Diabetes

In this case (normal outcome, known variance, conjugate prior) we can get an exact expression for the expected utility of the definitive trial.

So, no need for MCMC sampling or numerical integration for the inner expectation \rightarrow a more accurate answer.

Application: OK-Diabetes

Computation time: 9s (surrogate), 66s (nested)



Extensions

- ▶ Optimising over n_1
- ▶ Adding a stop/go point
- ▶ Adding a decision rule mapping pilot data x_1 to optimal definitive type I error rate α_2
- ▶ Allowing for unknown variance

Unknown variance

Suppose we don't know the variance of the outcome, and have a prior on that as well as the mean.

- ▶ The pilot data can now be summarised by a 2D sufficient statistic, the sample mean x_1 and the sample variance y_1
- ▶ Now need to map this 2D space to choice of definitive trial sample size

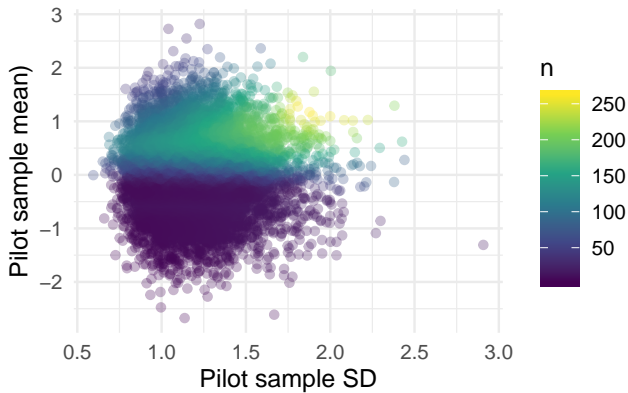
We can use tensor product splines:

$$\sum_{i=1}^{k+4} \sum_{j=1}^{k+4} a_{i,j} \phi_i(x_1) \psi_j(y_1)$$

- ▶ Still defined by vector \mathbf{a} , but now $\mathbf{a} \in \mathbb{R}^{(k+4)^2}$
- ▶ In our example with $k = 10$, solution space dimensions go from 14 to 196
- ▶ Impact is manageable *if* we have derivatives

Unknown variance

Computation time: 79s



Summary

We can now quickly calculate and maximise the expected utility of pilot trials, making a Bayesian decision-theoretic approach to pilot design and analysis feasible.

But:

- ▶ Need only 1 or 2 dimension sufficient statistic
- ▶ If 2 dimensions, need derivatives of objective function
- ▶ Need an analytic power function for the definitive trial
- ▶ How can we be sure of convergence to true global optimum?