Bayesian design of pilot trials for complex interventions

August 22, 2017

1 Problem

We consider an external pilot trial of a complex intervention which aims to assess several outcomes (e.g. recruitment, adherence, data collection, efficacy) and to make a decision regarding a confirmatory trial based on these assessments. The decision is between three options: to abandon the evaluation of the intervention; to modify the intervention and/or the trial design in response to weaknesses identified in the pilot, and proceed to a confirmatory trial; and to proceed to confirmatory trial with no modifications. The same set of decisions has been considered for internal pilot trials assessing recruitment [1]. We assume that a prior distribution on all unknown parameters (including any nuisance parameters) has been specified. We also assume that the parameter space has been partitioned into three subsets corresponding to situations where the three decisions are considered to be optimal. We do not assume a utility function is available.

We aim to define one or more quantities which measure the quality of a proposed trial design and thus enable better design decisions. A trial design will specify both the sample size (which may include several variables, e.g. in a multilevel trial) and the decision rule.

2 Methods

We label the three decisions as:

- $d = -1$ - abandon evaluation;
- $d = 0$ - modify and proceed;
- $d = 1$ - proceed without modification.

The partition of the parameter space $\Phi = \Phi_{-1} \cup \Phi_0 \cup \Phi_1$ is such that decision d is optimal if and only if $\theta \in \Phi_d$. We propose to base our evaluations of a trial design on the nine probabilities of making each decision under each hypothesis

		Truth			
		-1	0	1	
Action	-1	$p_{-1,-1}$	$p_{0,-1}$	$p_{1,-1}$	A_{-1}
	0	$p_{-1,0}$	$p_{0,0}$	$p_{1,0}$	A_0
	1	$p_{-1,1}$	$p_{0,1}$	$p_{1,1}$	A_1
		Φ_{-1}	Φ_0	Φ_1	1

Table 1: Probabilities of decisions and hypotheses.

Φ_d. These are set out in Table 1. Specifically, we will focus on the unconditional probabilities of three events:

1. Running a futile confirmatory trial ($p_{-1,0} + p_{-1,1} + p_{0,1}$);
2. Making unnecessary adjustments ($p_{-1,0} + p_{1,0}$);
3. Discarding a promising intervention ($p_{0,-1} + p_{1,-1} + p_{0,1}$).

We propose to make the post-trial decision by following a pre-specified decision rule based on posterior probabilities $p_d = Pr[\phi \in \Phi_d \mid x]$, where x is the observed pilot data. A decision rule is a partition of the space $\mathcal{P} = \{(p_{-1}, p_1) \in [0,1]^2 \mid 0 \leq p_{-1} + p_1 \leq 1\}$ into three subsets Δ_d corresponding to the three decisions. An example decision rule is given in Figure 1. One-dimensional analogues of this, based on the posterior probabilities of only two hypotheses, are found throughout the Bayesian trial design literature e.g. [2].

The decision rule $a(x) : \mathcal{P} \to \{-1, 0, 1\}$ can be written as

$$a(x) = \max_{d \in \{-1,0,1\}} I[(p_{-1}, p_1) \in \Delta_d], \tag{1}$$

where the dependance on x come through the posterior probabilities p_{-1} and p_1. The probabilities in the cells of Table 1 are then given by

$$p_{i,j} = \mathbb{E}[I(\phi \in \Phi_i, a(x) = j)]. \tag{2}$$

A simple Monte Carlo estimate is then

$$p_{i,j} \approx \frac{1}{N} \sum_{k=1}^{N} I(\phi^{(k)} \in \Phi_i, a(x^{(k)}) = j), \tag{3}$$

where the $\phi^{(k)}, x^{(k)}$ are simulated from the joint distribution of ϕ, x. For each sample we need to compute $a(x^{(k)})$, which requires calculating the posterior probabilities p_{-1}, p_1. This typically requires sampling from the posterior distribution $\phi \mid x^{(k)}$, and such a nested Monte Carlo scheme (requiring a large number of samples in an inner loop for each iteration of an outer loop) will incur a significant computational burden. More importantly, when we are not

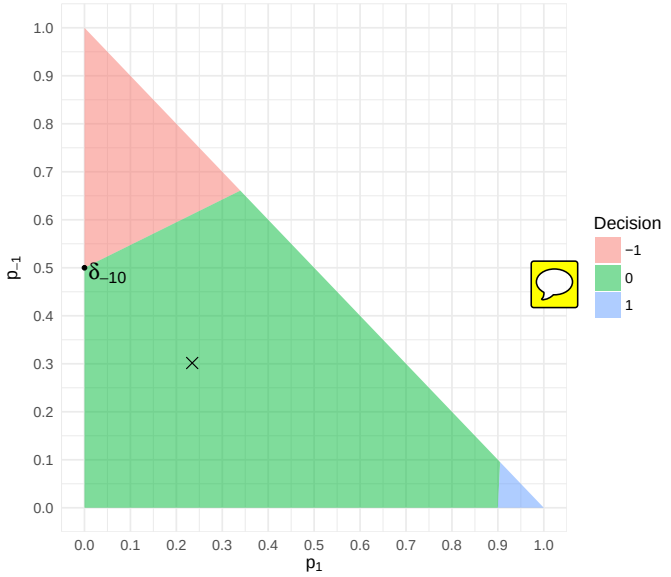

Figure 1: Example decision rule based on posterior probabilities p_{-1} and p_1. The cross marks the prior.

in a simple setting with conjugate priors, sampling from the posterior will require MCMC and will not be so reliable that the process can be automated N times as required (e.g. we can't automate checking for convergence of the Markov chains). To estimate the probabilities of Table 1 in a fast and reliable way, we adapt the ideas of [3] developed in the context of value of information calculation.

First, suppose that there exists a function $f(d, \phi)$ such that its expected value with respect to any posterior distribution $\phi \mid x$ is maximised at the same decision d that maximises expression (1), i.e.

$$\underset{d\in\{-1,0,1\}}{\arg\max}\ \mathbb{E}_{\phi|x}[f(d,\phi)] = \underset{d\in\{-1,0,1\}}{\arg\max}\ I[(p_{-1}, p_1) \in \Delta_d] = a(x). \tag{4}$$

Then, we note that the actual function $f(d, \phi)$ can be viewed as the sum of an expected value plus an error term,

$$f(d,\phi) = \mathbb{E}_{\phi|x}[f(d,\phi)] + \epsilon, \tag{5}$$

where the error term is a function of ϕ and x with zero mean [3]. For each action d we can think of $\mathbb{E}_{\phi|x}[f(d, \phi)]$ as just some unknown function of the data x, and we can construct a regression model approximating this function. To do so we generate a set of pairs of dependant variables $f(d, \phi)$ and independent variables $x \mid \phi$. To generate one pair we first sample ϕ from the prior distribution, then calculate $f(d, \phi)$, then sample x conditional on ϕ. We calculate appropriate

statistics $\bar{x}$ summarising the data x, and regress $f(d, \phi)$ against $\bar{x}$. We can then use the three regression models (one for each decision) to estimate the expectations in (4) to give a surrogate of the decision rule (1). We can then use this surrogate in the Monte Carlo estimate (3) to calculate the necessary probabilities whilst avoiding a nested MC sampling scheme.

We now consider how to generate a function $f(d, \phi)$ which satisfies (4). Let f take the piece-wise constant form

$$f(d, \phi) = \begin{cases} x_d \text{ if } \phi \in \Phi_{-1} \\ y_d \text{ if } \phi \in \Phi_0 \\ z_d \text{ if } \phi \in \Phi_1, \end{cases}$$

for $d \in \{-1, 0, 1\}$. Then $\mathbb{E}_{\phi|x}[f(d, \phi)] = p_{-1}x_d + (1 - p_{-1} - p_1)y_d + p_1 z_d$. Following the notation of Figure 1, we can impose a number of equality constraints on f's parameter values. For example, we require that when $p_{-1} = \delta_{-1,0}$ and $p_1 = 0$, the decisions -1 and 0 are equivalent and so

$$\delta_{-1,0}x_{-1} + (1 - \delta_{-1,0})y_{-1} = \delta_{-1,0}x_0 + (1 - \delta_{-1,0})y_0. \tag{6}$$

We may also ask for some inequalities to be satisfied, e.g. $x_{-1} \geq y_{-1}$ and $y_1 \leq z_1$. Finally, to avoid the trivial solution where all parameters are equal to 0, we ask that $x_{-1} = 1$. These constraints can be written as a linear system and a solution found using linear programming. For example, the rule of Figure 1 leads to the solution of

$$(x_{-1}, y_{-1}, \ldots, z_1) = (1, 0.028, -0.946, 0.514, 0.514, 0, 0, 0.028, 0.054). \tag{7}$$

Note that the function $f(d, \phi)$ depends on only the decision rule, and is not influenced by other characteristics of the problem such as the number of parameters, their prior distributions, or the specific form of the hypotheses Φ_d.

3 Illustration

We apply this method to a hypothetical cluster randomised pilot trial which aims to assess the average number of patients per cluster and the proportion of patients with complete data at follow-up. The trial will include $k = 12$ clusters, and we assume the number of patients in each cluster will be normally distributed with mean μ_r and known variance equal to 1. We denote by η_f the probability of obtaining complete follow-up data on a patient. We use a normal distribution as a prior for the mean cluster size, $\mu_r \sim N(m_r, s_r^2)$ with $m_r = 10$ and $s_r^2 = 1$. We use a beta distribution as a prior for the probability that a patient will have complete data $\eta_f \sim Beta(a_f, b_f)$, with $a_f = 22.4$ and $b = 9.6$ (corresponding to a mean of 0.7 and a prior sample size of 30). The joint prior distribution is illustrated, together with the partition of the parameter space into three hypotheses, in Figure 2.

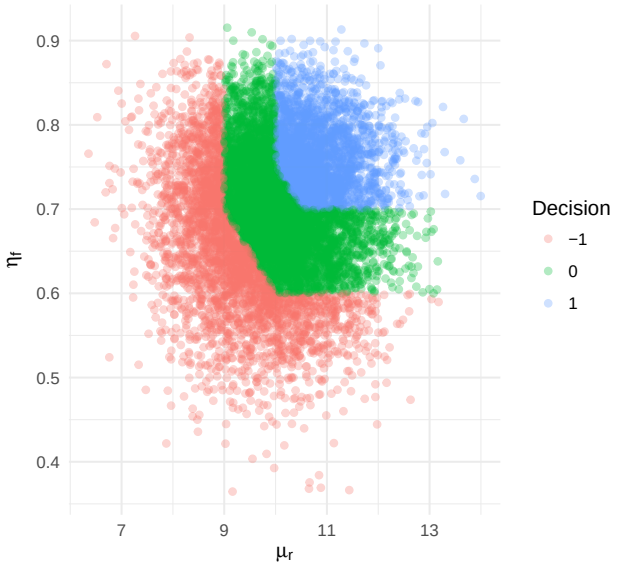

Figure 2: Samples from the prior distribution of mean cluster size and follow-up probability, classified according to the hypothesis they belong to.

We use the decision rule illustrated in Figure 1 with the aforementioned solution to the surrogate function $f(d, \phi)$. To build the regression models we generate 10^5 samples $\phi^{(i)} = (\mu_r^{(i)}, \eta_f^{(i)})$ from the joint prior distribution. For each sample we calculate $f(d, \phi^{(i)})$ for each action d and generate data $x^{(i)} \mid \phi^{(i)}$. Each data set is reduced to a vector $\bar{\mathbf{x}}^{(i)}$ of two statistics, the sample average of the cluster size and the sample proportion of patients with complete data. For each action d we regress $f(d, \phi^{(i)})$ against $\bar{\mathbf{x}}^{(i)}$ using the generalised additive model (GAM), a flexible non-parametric approach which assumes that the expectation of the dependant variable is a smooth function of the independent variables.

To evaluate the accuracy of the GAM approach, we simulate another 10^4 $\phi, \bar{\mathbf{x}}$ pairs and use the predictions of the models to estimate the expectations needed in (4) and thus determine the best decision d. We compare this with a procedure where MC estimates of the posterior probabilities are calculated using 10^4 samples from the posterior distribution $\phi \mid x$. This is feasible (although still not fast) in this example because we have conjugate priors. We can compare the decisions made using each approach with each other, and with the true underlying parameter value ϕ.

Table 2 illustrates the implications of the decision rule of Figure 1. The aforementioned unconditional error probabilities in this case are estimated to be:

1. Running a futile confirmatory trial; $p_{-1,0} + p_{-1,1} + p_{0,1} = 0.0938$.

		Truth			
		-1	0	1	
Action	-1	0.2093	0.0538	0.0006	0.2637
	0	0.0920	0.4101	0.1833	0.6854
	1	0.0000	0.0018	0.0491	0.0509
		0.3013	0.4657	0.2330	1

Table 2: Outcome probabilities using GAMs.

Figure 3: Sample estimates $(\hat{\mu}_r, \hat{\eta}_f)$ and the corresponding decision, according to the GAM model approximation.

2. Making an unnecessary adjustment; $p_{-1,0} + p_{1,0} = 0.2753$.

3. Discarding a promising intervention; $p_{0,-1} + p_{1,-1} + p_{0,1} = 0.0562$.

The decision rule is illustrated in the sample space in Figure 3, which plots the estimates $(\hat{\mu}_r, \hat{\eta}_f)$ for each sampled data set classified by the associated decision.

The accuracy of the GAM approach can be assessed by comparing the decisions with those made using the nested MC approach, for each of the sampled data. These differences are summarised in Table 3. We see that the decisions typically correspond well, disagreeing in only 0.0068% of the simulated instances.

If we want to reduce the unconditional error probabilities, we could consider adjusting the decision rule or the sample size of the trial. We consider the latter here, increasing the number of clusters from 12 to 50. Using the GAM approach,

		Nested MC			
		-1	0	1	
GAMs	-1	0.2637	0.0000	0.0000	0.2637
	0	0.0042	0.6788	0.0024	0.6854
	1	0.0000	0.0002	0.0507	0.0509
		0.2679	0.6790	0.0531	1

Table 3: Decisions made using GAMs and nested MC estimates.

this gives us error probabilities of 1) 0.042, 2) 0.1526 and 3) 0.0386. Again, a comparison with the nested MC approach shows broad agreement, with 0.0075% of cases not matching. For a fixed decision rule, we could choose an appropriate sample size by calculating these three quantities for a range of k values, plotting the results, and judging the trade-off between sample size and quality.

4 Discussion points

- We choose our action by maximising the expected value of $f(j, \phi)$, which is equivalent to following the subjective decision-theoretic approach where $f(j, \phi)$ is our utility function. Looking at the values of f which match the decision rule, they do not seem to match any intuitive utility description (e.g. the value of making decisions -1 and 0 are equal under hypothesis Φ_0).

- 
The GAM models assume that the expectations are smooth in terms of the independent variables, in our case sample statistics. When will this not be a realistic assumption? More generally, the models should be subject to regression diagnostics. How should they be assessed? What would indicate they have not approximated the unknown function sufficiently well?

- The form of the decision rule may be restricted according to some common sense / intuitive principles, and also mathematically, i.e. we require the existence of a non-trivial solution to the linear program. Explore each of these aspects to understand the limitations they impose.

- Both methods described, using GAM and using a nested MC scheme, will have some error in the expectations they estimate and therefore in the decisions they suggest and the probabilities of Table 1 they estimate. Find expressions for each of these.

- Explore to what extent the decision rule is fixed, or free to be adjusted in order to arrive at the desired unconditional error probabilities. Note that this will lead to a significantly more complex trial design optimisation problem.

- The surrogate function f should be evaluated to check it corresponds with the original decision rule. How best to do so?
- The surrogate function f is found as a feasible solution to a linear program. Can we go further and find an optimal solution, for example one which maximises the difference between the three decisions? 

References

[1] Lisa V Hampson, Paula R Williamson, Martin J Wilby, and Thomas Jaki. A framework for prospectively defining progression rules for internal pilot studies monitoring recruitment. *Statistical Methods in Medical Research*, 0(0):0962280217708906, 2017. PMID: 28589752.

[2] Joseph G. Ibrahim, Ming-Hui Chen, Mani Lakshminarayanan, Guanghan F. Liu, and Joseph F. Heyse. Bayesian probability of success for clinical trials using historical data. *Statistics in Medicine*, 34(2):249–264, oct 2014.

[3] Mark Strong, Jeremy E Oakley, Alan Brennan, and Penny Breeze. Estimating the expected value of sample information using the probabilistic sensitivity analysis sample: a fast, nonparametric regression-based method. *Medical Decision Making*, 35(5):570–583, 2015.