Proceedings of the 2015 Winter Simulation Conference
L. Yilmaz, W. K V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds.

OPTIMAL SEQUENTIAL SAMPLING WITH DELAYED OBSERVATIONS AND UNKNOWN VARIANCE

Stephen E. Chick

Technology and Operations Management
INSEAD
Boulevard de Constance
77300 Fontainebleau, FRANCE

Martin Forster

Department of Economics and Related Studies
University of York
Heslington
York YO42 2GA, UK

Paolo Pertile

Department of Economics
University of Verona
Via dell'Artigliere 19
37129 Verona, ITALY

ABSTRACT

Sequential stochastic optimization has been used in many contexts, from simulation, to e-commerce, to clinical trials. Much of this analysis assumes that observations are made soon after a sampling decision is made, so that the next sampling decision can benefit from the most recent data. This assumption is not true in a number of contexts, including clinical trials. In this paper we extend sequential sampling tools from simulation optimization to be useful when there exists a delay in observing the data from sampling, with a specific focus on the situation in which the sampling variance is unknown. We demonstrate the benefits of doing so by benchmarking the optimization algorithms with data from a published clinical trial.

1 INTRODUCTION

Recent advances in simulation optimization are enabling effective use of sequential sampling to achieve optimization goals. Some of these techniques are based on Bayesian principles (Chick and Inoue 2001, Frazier, Powell, and Dayanik 2008), and others on heuristics (Chen 1996, Branke and Schmidt 2004, Chen and Kelton 2005, Chen, Yücesan, Dai, and Chen 2010) or frequentist techniques (Kim and Nelson 2006, Xu, Nelson, and Hong 2010). Significant advances are also found in others areas of operations (Gittins and Jones 1974, Bertsimas and Mersereau 2007, Caro and Gallien 2007, Ahuja and Birge 2015), biostatistics and health economics (Berry and Eick 1995, Pertile, Forster, and La Torre 2014).

One key decision in sequential sampling is whether to continue sampling, or whether to stop and make a so-called implementation decision. In a simulation context, sequential sampling decisions might represent allocations of alternative system designs to CPUs in an attempt to infer the mean performance of the various alternatives in a way which helps the analyst identify the system design which has the largest mean performance. In a health or medical context, sequential sampling decisions might represent the allocation of patient pairs in a clinical trial, in an effort to infer whether a new treatment or clinical process is better than an existing one. The vast majority of this literature assumes that outcomes are observed, alone or in groups, before the next decision to take samples or to stop sampling arises. That is, this type of analysis often assumes that delays in observing outcomes are not important in the decision process.

But delays arise very naturally in many contexts, and can have practical importance. For example, clinical trials and health technology assessments often involve delays between the time that a sampling decision is made and the time that the outcome from that sample is ultimately observed. For example, Marple et al. (2010) measured recovery two weeks after administering a comparator antibiotic for treating acute sinusitis and Connor et al. (2015) measured the primary end point at 90 days in an adaptive trial carried out by the stroke hyperglycemia insulin network effort.

This paper focuses on the design of a sequential experiment in which the observation on the outcome of interest arrives with a fixed delay and the sampling variance is unknown. We assume that a number $\tau \geq 0$ sampling allocation decisions are made between making an allocation and observing the data from that allocation. This means that, when at least τ samples have been allocated and sampling continues, there will be τ samples 'in the pipeline', whose outcomes are yet to be observed. We are interested in knowing for how long sampling should continue before stopping, observing the remaining pipeline data, and making an implementation decision.

Chick, Forster, and Pertile (2015) (hereafter CFP) derive a Bayesian decision-theoretic model of optimal sequential sampling which allows for samples to arrive with a fixed delay when the sampling variance is known. The authors apply the model to clinical trials and health technology assessments which seek to compare two alternatives, one a 'new' health technology, the other an existing one. They allow for on-line ('earn while you learn') and off-line learning as well as the absence or presence of two types of sampling cost (marginal cost per sample, and discounting costs). Their model extends work in simulation optimization (Chick and Gans 2009, Chick and Frazier 2012), biostatistics and health economics (Pertile, Forster, and La Torre 2014).

This paper extends the model of CFP, summarized in section 2 to the case of an unknown sampling variance. The extension proceeds in two steps: the first considers the use of predictive distributions in one stage sampling plans; the second involves the specification of a diffusion model when sampling variances are unknown. For the first step, we follow the standard technique of using a conjugate normal-inverse-gamma distribution for the unknown mean and variance of a normal distribution (DeGroot 1970, Chick and Inoue 2001), and extend the treatment in a straightforward way to allow for delayed samples. For the second step, the extension of an optimal stopping time for unknown variances is more subtle. When sampling variances are known and the unknown mean is inferred through sampling, the continuation region $\mathscr{C}$ is the set of sufficient statistics for the unknown mean as a function of time t, such as $(\hat{\mu}_t, t)$, for which it is optimal to continue sampling. This can be handled by solving a free boundary problem for a heat equation in one dimension, say $\hat{\mu}_t$. When the sampling variance is also unknown, the sufficient statistic has an extra dimension, say $(\hat{\mu}_t, \hat{\sigma}_t^2, t)$, which adds an extra dimension, and associated numerical computation, to the free boundary problem. At the same time, observing a diffusion over an interval of positive measure gives full information about the sampling variance, but in applications the diffusion is really used as a surrogate for observations at discrete time points (e.g., integer number of samples).

We seek an approximation to the stopping boundary of the continuation set which is not too challenging computationally and which is appropriate to use when sampling and observing outcomes takes place at discrete time points. One approach is to compute the solution to the diffusion under the assumption that the sampling variance is known, and then plug in the sample variance estimator to rescale the optimal stopping boundaries in an appropriate way. This approach has been tried in the past (Chick and Frazier 2012, for example) but does not account for uncertainty in the sampling variance when modeling the progression of the posterior mean on the sample path of the diffusion. We propose two alternatives below to handle the approximation of the optimal stopping boundaries. Our goal is to model better the effect of not knowing the sampling variance on the evolution of the sample path of the posterior mean. This, in turn, influences the diffusion approximation which determines the optimal stopping boundaries. We hope to do so in a way which does not require solving a higher dimensional free boundary problem for a diffusion than is required for the case of a known sampling variance.

These two alternative approaches are presented in section 3. The first endogenizes the increased variance inside the trinomial grid which is used to solve the free boundary problem. The second adapts the KG_* approach (Chick and Frazier 2009, Frazier and Powell 2010), which has typically assumed a known sampling variance, to the context here of an unknown sampling variance and potentially delayed observations. Section 4 presents preliminary output.

There are several other papers which also model delays in samples. Hardwick, Oehmke, and Stout (2006) account for Poisson arrivals and exponential delays, and develop heuristics to minimize losses. Caro and Yoo (2010) show that certain bandit problems with stationary random delays satisfy an indexability criterion as long as the order in which samples are observed is the same as the order in which they are allocated. Hampson and Jennison (2013) incorporate delay in a group sequential trial within a Neyman-Pearson hypothesis testing framework, minimizing the expected sample size of the trial subject to satisfying prespecified Type I and Type II error rates.

2 OPTIMAL SEQUENTIAL SAMPLING WITH DELAY: KNOWN SAMPLING VARIANCE

This section recalls the model of CFP. A risk neutral decision-maker faces a choice between using a new technology or an existing one to earn a monetary reward from treating a defined number of patients. The difference between the expected reward of the new technology and the existing one (the sampling mean) is unknown to the decision maker; the sampling variance is known. The decision maker has the option to sample at a constant marginal cost before deciding which technology to use. The outcome of each sampling allocation is observed with delay. If sampling commences, the solution to the optimal stopping problem indicates whether the sampling statistics so far justify continuing to sample after each outcome is observed, or whether the analyst should stop sampling so as to wait for the data in the 'pipeline' in order to inform the adoption decision.

This framework is directly applicable to simulation optimization by allowing the difference in expected rewards to be the difference in unknown means of two alternatives which are simulated in pairs (with either CRN or independent random numbers), or comparing an alternative with unknown mean against a known standard. At present, the framework does not allow for more than two alternatives, although we are working to overcome that limitation. Another limitation which might be more significant in the simulation context is the assumption that the delays are of fixed duration. This might not be a good assumption if the run times of different alternatives vary widely.

2.1 Basic Framework

CFP consider a two-armed, sequential experiment in which patients are allocated at random, and in a pairwise manner, to either a control (standard) technology or a new technology. There is a cost $c \in \mathbb{R}_{\geq 0} \equiv [0, \infty)$ per pairwise allocation made. The experiment evaluates which of the two technologies should be used to treat $P \in \mathbb{R}_{\geq 0}$ patients upon stopping the trial. A switching cost $I \in \mathbb{R}_{\geq 0}$ is incurred if the decision is made to move to the new technology. No cost is incurred if the decision is made to continue with the standard technology.

The data collected for each patient is the measure of effectiveness, denoted by the random variable $E_{\mathrm{n}} \in \mathbb{R}$ if the patient is assigned to the new technology and $E_{\mathrm{c}} \in \mathbb{R}$ if the patient is assigned to the control. The patient-level costs of using each technology are the random variables $C_{\mathrm{n}} \in \mathbb{R}_{\geq 0}$ and $C_{\mathrm{c}} \in \mathbb{R}_{\geq 0}$. CFP assume that all patients complete their assigned course of treatment, there is no loss to follow up, and E_{n}, E_{c}, C_{n} and C_{c} are observed without measurement error.

Following standard approaches in Bayesian decision-theoretic models (e.g., Pertile, Forster, and La Torre 2014, Lewis, Lipsky, and Berry 2007 and Berry and Ho 1988), a common unit of measurement is used to value benefits and costs. The realisation of the individual level incremental net monetary benefit (INMB) of the new technology versus the existing one for pairwise allocation i is:

$$X_i = \lambda(E_{\mathrm{n},i} - E_{\mathrm{c},i}) - \delta_{\mathrm{CE}}(C_{\mathrm{n},i} - C_{\mathrm{c},i}), \tag{1}$$

where $\lambda \in \mathbb{R}_{>0} \equiv (0,\infty)$ is the monetary value of one unit of effectiveness and $\delta_{\text{CE}} = 1$ if the experiment assesses cost-effectiveness and $\delta_{\text{CE}} = 0$ if it assesses effectiveness only. The authors assume that $X_i \sim \text{Normal}(W, \sigma_X^2)$, $i = 1,2,\ldots,T_{\max}$, where $T_{\max}$ is the maximum number of pairwise allocations which can be made. W is assumed to be unknown and σ_X^2 is assumed known. The prior distribution on W is assumed to be $\text{Normal}(\mu_0, \sigma_0^2)$. Then $n_0 = \sigma_X^2/\sigma_0^2$ is the so-called effective number of samples in the prior distribution, to be consistent with Bayesian nomenclature.

The annual rate of accrual is constant and equal to $R \in \mathbb{Z}_{>0}$. Realisations of X arrive with a delay of $\tau \in \mathbb{Z}_{\geq 0}$, $\tau < T_{\max}$, where $T_{\max}$ is the maximum number of patient pairs which may be observed. Realizations are used to update the prior/posterior distribution of W in a sequential manner. The model permits future benefits and costs to be down-weighted using a discount factor. The annual continuous discount rate is $\rho_{\text{year}} \geq 0$, so that $\rho = \rho_{\text{year}}/R$ is the continuous discount rate when time units are per patient pair allocation, and $\tilde{\rho} = \exp[\rho_{\text{year}}/R] - 1$ is the discrete time discount rate per patient pair. CFP let $\delta_{\text{on}} = 1$ if the rewards for patients in the trial are to be included in the reward function (known as 'online learning') and $\delta_{\text{on}} = 0$ if they are not ('offline learning').

Define $\mathbb{T} \equiv \{0,1,\ldots,T_{\max}\}$ and define $T \in \mathbb{T}$ as the time at which pairwise allocations cease to be made. It is assumed that sampling cannot be restarted once pairwise allocations cease to be made. At each $t \in \mathbb{T}\backslash\{T_{\max}\}$, an action a_t must be chosen from the set of available actions, $\mathscr{A} \equiv \{0,1\}$, such that $a_t = 0$ denotes ceasing to make pairwise allocations (so $T = t$) and $a_t = 1$ denotes making another pairwise allocation ($T > t$).

For $t \leq \tau$, a_t is chosen only on the basis of prior information. For $\tau < t < T_{\max}$, the action can be a function of the realisations $\{X_i\}_{1 \leq i \leq t-\tau}$. For $t = \tau,\ldots,T_{\max}-1$, the ordering of events is as follows: action a_t is chosen; realisation $X_{t+1-\tau}$ is observed; prior on W is updated. If sampling continues as far as $t = T_{\max}$, $T = T_{\max}$ and sampling stops at this terminal stage.

Once sampling is stopped, the decision maker waits to observe all outcomes for the pipeline subjects – those who have been treated but whose outcomes are yet to be observed – before making the technology adoption decision. Define $\mathscr{D} \in \{\text{n},\text{c}\}$ as the decision concerning whether to choose the new technology ($\mathscr{D} = \text{n}$) or the existing one ($\mathscr{D} = \text{c}$). This adoption decision is made at time 0 if $a_0 = 0$, because no samples will be observed. It is made at time $T + \tau$ if $a_0 = 1$, because of the delay in the arrival of observations.

More compactly, the adoption decision is made at time $\mathbf{1}_{T>0}(T+\tau)$, where $\mathbf{1}_F$ is the indicator function, equal to 1 if the event F is realized and 0 otherwise. The reward from selecting technology $\mathscr{D}$ is $\mathbf{1}_{\mathscr{D}=\text{n}}(PW - I)$, ignoring the cost of sampling and discounting.

Define $\mathscr{F} = (\mathscr{F}_t)_{t\in\bar{\mathbb{T}}}$ as the natural filtration generated by the realisations $(\{X_i\}_{i\leq(t-\tau)})$ for $t \in \bar{\mathbb{T}} \equiv \{0,1,\ldots,T_{\max}+\tau\}$. Note that $\mathscr{F}_t = \mathscr{F}_0$ for $t \in \{0,1,\ldots,\tau\}$ due to the delay in the arrival of observations. Define a variable tracking the *effective sample size* in the posterior distribution for W given information available to time $t \in \bar{\mathbb{T}}$,

$$n_t = n_0 + (t-\tau)^+, \tag{2}$$

where $(m)^+ = \max(0,m)$. Define:

$$Y_t = \mu_0 n_0 + \sum_{i=1}^{(t-\tau)^+} X_i, \tag{3}$$

where the sum is equal to 0 if the upper bound for the summation is less than 1.

Posterior beliefs about W at time t have a normal distribution

$$W \,|\, \mathscr{F}_t \sim \text{Normal}(\mu_t, \sigma_X^2/n_t), \text{ where:} \tag{4a}$$

$$\mu_t = Y_t/n_t. \tag{4b}$$

(y_t, n_t) may be used as a sufficient statistic for W conditional on $\mathscr{F}_t$ and (y_t, t) may be used as a state because it also provides information about the number of pipeline subjects.

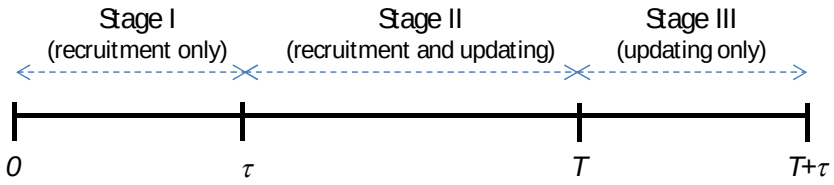


Figure 1: Stages of the optimal sequential sampling with delay problem with stopping time T and delay τ.

A policy π defines a mapping $f(y_t,t) : \mathbb{R} \times \mathbb{T} \backslash \{T_{\max}\} \to \mathscr{A}$ from states to a decision to stop or to continue sampling, which in turn determines T. A policy π also specifies the choice of the new technology or control, $\mathscr{D} \in \{\mathrm{n}, \mathrm{c}\}$, as above. By construction, $T \in \mathbb{T}$ is a stopping time of the filtration $\mathscr{F}$; $\mathscr{D} \in \{\mathrm{n}, \mathrm{c}\}$ is $\mathscr{F}_{\mathbf{1}_{T>0}(T+\tau)}$-measurable and π is measurable with respect to $\mathscr{F}$. Let Π be the set of all policies which are measurable in this way with respect to $\mathscr{F}$.

The expected reward from a policy $\pi \in \Pi$ depends on the parameters of the prior distribution (μ_0, n_0), and is determined by the cost of samples, benefits to patients during the trial (in the case of online learning), and benefits from the technology adoption decision:

$$V^{\pi}(\mu_0, n_0) = \mathbb{E}_{\pi}\left[\left\{\sum_{t=0}^{T-1} \frac{-c + \delta_{\mathrm{on}} X_{t+1}}{(1+\tilde{\rho})^t}\right\} + \frac{\mathbf{1}_{\mathscr{D}=\mathrm{n}}(PW - I)}{(1+\tilde{\rho})^{\mathbf{1}_{T>0}(T+\tau)}} \middle| \mu_0, n_0\right]. \tag{5}$$

The denominators account for discounting. The term $\mathbf{1}_{T>0}(T+\tau)$ indicates that a penalty for discounting is only relevant for the terminal reward if at least one pairwise allocation is made. The *optimal sequential sampling with delay problem* of CFP is defined formally to be that of finding a policy $\pi^* \in \Pi$ such that

$$V^{\pi^*}(\mu_0, n_0) = \sup_{\pi \in \Pi} V^{\pi}(\mu_0, n_0). \tag{6}$$

CFP define three distinct stages of the trial in order to characterise the optimal policy. These stages are illustrated in Figure 1. In *Stage I* (recruitment only, $t \in \{0, 1, \ldots, \tau - 1\}$) pairwise allocations are made sequentially, no outcomes are observed owing to the delay. In *Stage II* (updating and recruitment, $t \in \{\tau, \tau+1, \ldots, T-1\}$) pairwise allocations are made, realisations $X_{t+1-\tau}$ for pipeline subjects arrive sequentially and are used to carry out Bayesian updating. In *Stage III* (updating only, $t \in \{T, T+1, \ldots, T+\tau\}$) no pairwise allocations are made, observations on pipeline subjects arrive sequentially and are used to carry out Bayesian updating.

2.2 Solution

CFP prove a number of structural results about the solution to this problem. First, they write the Bellman equation for this problem and show that the following policy is optimal: at each step, maximize the expected reward from choosing to continue to sample or to stop sampling in the Bellman equation and, when all data are observed, pick the new alternative if it has a posterior mean of at least I/P (the threshold determined by spreading adoption costs over all patients to benefit from the new technology); to continue with the existing technology otherwise. CFP also observe that the set of (μ_t, t) for which it is optimal to continue sampling is symmetric above and below the line $\mu = I/P$ when there is off-line learning and $\tilde{\rho} = 0$.

In order to describe the posterior expected reward after all observations on pipeline subjects are observed, given that stage III has just been entered, let $Z_{t,s}$ be the posterior expected INMB at the patient level, given that t pairwise allocations have been made and that s outcomes are yet to be observed. Given the known

sampling variance assumptions above, $Z_{t,s}$ and its predictive distribution (DeGroot 1970) are:

$$Z_{t,s} = \mathbb{E}[\mu_{t+s} \mid \mathscr{F}_t, X_{t+1}, X_{t+2}, \ldots, X_{t+s}]; \tag{7}$$

$$Z_{t,s} \sim \text{Normal}\left(\mu_t, \frac{\sigma_X^2 s}{n_t(n_t+s)}\right). \tag{8}$$

As is customary, $Z_{t,0}$ is taken to be a point mass at μ_t.

The optimal reward upon entering stage III at time T is shown to be:

$$G(\mu_t, t) = (1+\tilde{\rho})^{-\tau \mathbf{1}_{t>0}} \mathbb{E}[(PZ_{T,\min(T,\tau)} - I)^+ \mid (\mu_t, t), T = t]. \tag{9}$$

where (μ_t, t) are the sufficient statistics given that t samples have already been allocated, and $s = \min\{T, \tau\}$ is the number of pipeline samples to be observed.

CFP describe the optimal reward during stage II is conceptually found by writing a continuous time approximation to the Bellman equation which gives the expected reward of stopping or of continuing to sample. This involves writing and solving a free boundary problem for a diffusion model / heat equation. They solve this free boundary problem numerically by using a trinomial grid to approximate the resulting diffusion model.

CFP solve for the optimal reward during stage I by using the predictive distribution for the expected reward from the optimal one stage policy of length $\beta = \tau$ or less, and compare it with the expected reward of continuing to stage II. The 'Optimal Bayes One Stage' policy chooses a sample size $s^*(\mu_0; \beta)$ so as to maximise the net benefit of sampling and selecting a technology, in expectation, over fixed-length sampling policies of length β or less:

$$s^*(\mu_0; \beta) = \arg \max_{s=0,1,\ldots,\beta} \left\{ \sum_{t=0}^{s-1} \frac{-c + \delta_{\text{on}} \mu_0}{(1+\tilde{\rho})^t} \right\} + \mathbb{E}\left[\frac{(PZ_{0,s} - I)^+}{(1+\tilde{\rho})^{\mathbf{1}_{s>0}(s+\tau)}} \middle| (\mu_0, 0) \right]. \tag{10}$$

These techniques all require modification in order to handle the case of unknown sampling variances. The solutions for stages I and III are relatively straightforward, but the solution for stage II is more subtle.

3 OPTIMAL SEQUENTIAL SAMPLING WITH DELAY: UNKNOWN SAMPLING VARIANCE

This section describes a standard Bayesian inference model to handle the case of unknown variances. It then summarizes how the model of CFP can be extended to handle the case of unknown sampling variances.

We will refer to several properties of the Student t distribution. Let $\phi_\nu(x)$ and $\Phi_\nu(x)$ be the density function and cumulative distribution function, respectively, of a standard Student t distribution with ν degrees of freedom. If T_ν is a standard Student t random variable with ν degrees of freedom, we say that $\mu + T_\nu/\sqrt{\kappa}$ is a three parameter Student t random variable, denoted $\texttt{St}(\mu, \kappa, \nu)$, with precision κ. When $\nu > 2$, the variance is $\kappa^{-1}\nu/(\nu-2)$. Let $\Psi_\nu[s] = \mathbb{E}[(T_n u - s)^+] = \int_s^\infty (x-s)\phi_\nu(x)dx$ denote the standard Student t linear loss function. Note that $\Psi_\nu[s] = \frac{\nu+s^2}{\nu-1}\phi_\nu(s) - s\Phi_\nu(-s)$ for $\nu > 1$ (Chick and Inoue 2001).

3.1 Model

We now suppose that the samples X_i are normally distributed and conditionally independent, given the unknown mean and *unknown* sampling variance. Because they are unknown, we assume they are random variables whose values are to be inferred. Let ς be the random variable whose realization is the sampling variance ${\sigma_X}^2$. Then

$$X_i \mid W, \varsigma \overset{iid}{\sim} \text{Normal}(W, \varsigma).$$

We presume that the prior distribution for each unknown mean and variance is in the family of conjugate priors for normally distributed samples with unknown means and variances (DeGroot 1970, § 9.6),

$$\begin{aligned} \varsigma &\sim \texttt{InvGamma}(\xi_0, \chi_0), \\ W \mid \varsigma &\sim \text{Normal}(\mu_0, \varsigma/\eta_0), \end{aligned} \tag{11}$$

where $\xi_0 > 1$ and χ_0 are shape and scale parameters of an inverse-gamma distribution with mode χ_0/ξ_0 (one's best *a priori* guess of σ_X^2), a finite mean $\mathbb{E}[\varsigma] = \chi_0/(\xi_0 - 1)$, $\mathbb{E}[1/\varsigma] = \xi_0/\chi_0$ and $\mathrm{Var}[1/\varsigma] = \xi_0/\chi_0^2$, and where μ_0 and η_0 describe the a priori mean and variance of the unknown sampling mean. It follows that W is a $\mathtt{St}(\mu_0, \xi_0\eta_0/\chi_0, 2\xi_0)$ random variable and that $\mathrm{Var}[W] = \chi_0/[(\xi_0 - 1)\eta_0]$. Thus, if ξ_0 exceeds 1, then the *a priori* variance of W exists. For further discussion on choosing these parameters of the prior distribution, see DeGroot (1970) or Bernardo and Smith (1994) or Chick and Frazier (2009).

With the prior distribution in Eq. (11), the posterior distribution has the same form. Given information to time t, the data from the $t+1$st sample can be used to update the posterior distribution as follows:

$$\begin{aligned} \varsigma \,|\, x_{t+1}, \mathscr{F}_t &\sim \mathtt{InvGamma}(\xi_{t+1}, \chi_{t+1}), \\ W \,|\, \varsigma, x_{t+1}, \mathscr{F}_t &\sim \mathrm{Normal}(\mu_{t+1}, \varsigma/\eta_{t+1}), \\ W \,|\, x_{t+1}, \mathscr{F}_t &\sim \mathtt{St}(\mu_{t+1}, \eta_{t+1}\xi_{t+1}/\chi_{t+1}, 2\xi_{t+1}), \end{aligned} \tag{12}$$

where, for $t = \tau, \tau+1, \ldots$, we have $\xi_{t+1} = \xi_t + 1/2$, $\chi_{t+1} = \chi_t + \frac{\eta_t}{2(\eta_t+1)}(\mu_t - x_{t+1-\tau})^2$, $\eta_{t+1} = \eta_t + 1$, and $\mu_{t+1} = (\eta_t\mu_t + x_{t+1-\tau})/\eta_t$ (given a simple adaptation of DeGroot 1970 to account for delays). For $t = 0, 1, \ldots, \tau - 1$, no data arrives, so $\xi_{t+1} = \xi_t$, $\chi_{t+1} = \chi_t$, $\eta_{t+1} = \eta_t$, and $\mu_{t+1} = \mu_t$ for those values of t.

With these modifications, the state vector for the case of an unknown sampling variance is (μ_t, χ_t, t), in comparison with the state (μ_t, t) for the case of a known sampling variance (ξ_t and η_t are functions of t, given ξ_0 and η_0). This state vector is sufficient to summarize $\mathscr{F}_t$ for purposes of inference about X.

This leads us to a key requirement for generalizing our model to handle an unknown sampling variance in section 3.2. Specifically, with this model we can obtain the distribution of the posterior distribution to be realized after pipeline samples come in, given that sampling stops at time $T = t$, with state (μ_t, χ_t, t), and with $\min(T, \tau)$ pipeline samples to arrive. Thus, we substitute Eq. (8) with (Chick and Inoue 2001)

$$Z_{t,s} \sim \mathtt{St}\left(\mu_t, \frac{\xi_t\eta_t(\eta_t + s)}{\chi_t s}, 2\xi_t\right). \tag{13}$$

3.2 Solution

The optimal solution to the sequential sampling with delay problem depends on solving stages I, II and III as illustrated in Figure 1 and as discussed in section 2.2 for the case of known sampling variances.

Stage I and stage III are straightforward to modify to account for unknown variances. In particular, the terminal reward function in Eq. (9) for stage III is modified to handle the case of unknown variances by taking the expectation in its right hand side with respect to the Student-t distribution for $Z_{t,s}$ in Eq. (13) rather than with respect to the normal distribution in Eq. (8). A similar change is sufficient to modify the expectation in the right hand size of Eq. (10) for Stage I.

A more interesting challenge arises when attempting to adapt the solution for stage II, and there are several potential plans of attack to compute (at least approximately) the optimal solution. The solution for stage II in CFP is to use an optimal stopping problem for a continuous time diffusion, in the spirit of Chernoff (1961) and many after, adapted to handle delayed observations. That technique computes stopping boundaries which define a continuation set. That technique is appropriate for the case of known sampling variances, but might introduce suboptimality if used with plug-in estimators for the unknown sampling variance.

3.2.1 Scaling for Optimal Sequential Sampling with Known Sampling Variance and No Delay

We use results for the case of zero delay ($\tau = 0$) and known σ_X^2 to motivate new results to handle the case of $\tau \geq 0$ and unknown σ_X^2, at least when one of c or ρ are 0.

When the discount rate is 0 and $c > 0$, Chick and Frazier (2012) give results for a special case ($\tau = 0$, known σ_X^2, $\rho = 0$, large $T_{\max}$ and offline learning) which justify bounding the continuation set $\mathscr{C}$ above

and below by stopping boundaries:

$$b(n_t) = (I/P) \pm (c/P)^{1/3} \sigma_X^{2/3} \tilde{b}_2(\sigma_X^{2/3}/((c/P)^{2/3} n_t)) \tag{14}$$

where $c^{1/3}\sigma_X^{2/3}$ is the cube root of the inverse of the sampling efficiency (Hammersley and Hanscomb 1964, p. 22) and $\tilde{b}_2(s)$ is a specific increasing function s with $\tilde{b}_2(0) = 0$ (Chick and Frazier 2012).

Importantly, both the mean and time are scaled in this result. The appropriate scaling factor for μ_t is $(c/P)^{1/3}\sigma_X^{2/3}$ above and below I/P if there is no discounting and a positive marginal cost of sampling. Time is scaled for $\tilde{b}_2$ via $\sigma_X^{2/3}/((c/P)^{2/3} n_t)$.

When the discount rate is positive and c=0, (Chick and Gans 2009, Online Companion) give an approximation to the optimal stopping boundary for a special case ($\tau = 0$, known σ_X^2, $\rho > 0$, large $T_{\max}$ and offline learning) which justify bounding the continuation set $\mathscr{C}$ above by

$$b(n_t) = (I/P) + \sigma_X \sqrt{\rho/P}\, \tilde{b}_1(1/\rho n_t) \tag{15}$$

where $\tilde{b}_1(s)$ in Eq. (16) makes a slight improvement upon the approximation of **?**Online Companion]chick:09a

$$\tilde{b}_1(s) \quad = \quad \begin{cases} s/\sqrt{2} & \text{if } s \leq 1/7 \\ \exp\left[-.0275(\log s)^2 + .8797 \log s - .5024\right] & \text{if } 1/7 < s \leq 100 \\ \sqrt{s}\left[2\log s - \log\log s - \log 16\pi\right]^{1/2} & \text{if } 100 < s. \end{cases} \tag{16}$$

Thus, the appropriate scaling factor for μ_t is $\sigma_X\sqrt{\rho/P}$ about I/P if there is positive discounting. Time is scaled for $\tilde{b}_1$ via $1/(\rho n_t)$. The associated stopping boundary is influenced by both of these scalings. These results do not characterize the functional form of the lower boundary in our context when the discount rate is positive – that boundary is computed here using the trinomial grid – but suggests that the same scaling factor can be used to scale upper and lower boundaries about I/P.

We compute the case of $c > 0$, $\rho > 0$ numerically, and note a similar baseline of I/P when there is offline learning. In the implementation reported below, we forced the stopping boundaries to be monotone in order to manage to numerical stability issues. These issues arose for certain parameter combinations especially when the variance is unknown.

3.2.2 Approximation to Optimal Stopping Boundaries with Unknown Sampling Variance, Fixed Delay

We now turn to several potential mechanisms to approximate the optimal stopping boundary for the case of delayed samples and an unknown sampling variance. Each approximation to the optimal stopping boundary results in a different stopping rule: stopping continues as long as the statistics of the process are inside of the stopping boundaries. The numerical results in section 4 illustrate the shape of those stopping boundaries.

The first mechanism is a plug-in estimator for the sampling variance in the formula for the stopping boundaries. Chick and Frazier (2012) suggested this technique for the case of $\tau = 0$, $\rho = 0$. This plug-in approach is adapted to our context by substituting $\mathbb{E}[\varsigma \mid \mathscr{F}_t] = \chi_t/(\xi_t - 1)$ for σ_X^2 in Eq. (14) for the case of $c > 0, \rho = 0$ and in Eq. (15) for the case of $c = 0, \rho > 0$. Note that this substitution may change both the space and time scaling of the boundaries, but is implementable by solving the PDE only once by assuming the variance is known to be $\chi_0/(\xi_0 - 1)$ and making the appropriate substitutions.

This first mechanism does not endogenize the variance of the sampling mean as it evolves on sample paths of inference. That is, the free boundary problem (PDE) which is used in the diffusion to compute the optimal stopping time does not fully account for the variance in the mean.

The second mechanism we propose now is to endogenize the variance of the posterior means in the PDE by more fully modeling the variance in posterior mean over the interval $[t, t+h]$, where h is the time step in the trinomial grid for solving the PDE. This is done using the variance of $Z_{t,h}$ implied by Eq. (13) rather than by Eq. (8). This technique increases the volatility of the diffusion, as desired, especially

when few samples have been observed, because the variance of a Student-t random variable is a factor of $\nu/(\nu-2)$ times bigger than the variance of a normal distribution with the same precision parameter. This is computed for a fixed value of σ_X^2 and the boundaries are then rescaled with a plug-in estimator as above.

The third mechanism we propose is to adapt the KG$_*$ framework to our context of unknown sampling variances and delayed observations, with online or offline learning, and with zero or positive discounting. Prior work on the KG$_*$ focused largely on its application to case of a known sampling variance. The KG$_*$ idea was originally proposed and used in numerical experiments by Chick and Frazier (2009) and was then later analyzed more thoroughly by Frazier and Powell (2010).

To recall, the KG$_*$ idea is roughly based on the idea of one-stage sampling as follows. A one-stage sampling policy is one for which exactly s samples are observed, and then a selection decision must be made. In this context, this corresponds to having s pipeline samples together with information $\mathscr{F}_t$ to time t, and using that information to select whether to adopt the new technology or to retain the existing technology. Informally, if there is a one-stage sampling policy of length $s \geq 1$ such that the expected value of information from sampling exceeds its cost, then KG$_*$ suggests to continue sampling. If not, then KG$_*$ suggests stopping.

We now apply that idea to our context, and allow for both unknown sampling variances and for delays in observations of samples. When in stage II of sampling, the expected value of sampling s more samples, waiting to observe the outcomes, and selecting an alternative optimally is:

$$\hat{V}_s(t) = \mathbb{E}\left[\left\{\sum_{t=0}^{s-1} \frac{-c+\delta_{\text{on}} X_{t+1}}{(1+\tilde{\rho})^t}\right\} + \frac{\mathbf{1}_{\mathscr{D}=\text{n}}(PZ_{t,s}-I)}{(1+\tilde{\rho})^{(s+\tau)}}\Big|\mathscr{F}_t\right]. \tag{17}$$

Delay and the unknown variance are modeled via the distribution of $Z_{t,s}$ in Eq. (13). For $t \geq \tau$ let

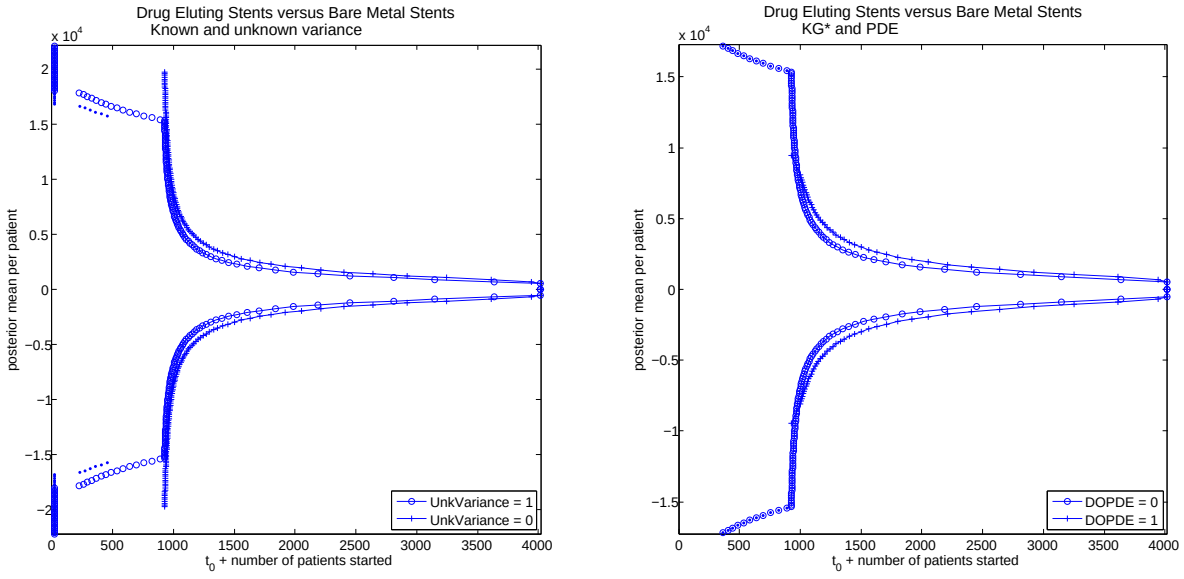
$$\hat{V}^*(t) = \max_{s=0,1,\ldots,T_{\max}-t}\left\{\sum_{i=0}^{s-1} \frac{-c+\delta_{\text{on}}\mu_t}{(1+\tilde{\rho})^i}\right\} + \mathbb{E}\left[\frac{(PZ_{t,s}-I)^+}{(1+\tilde{\rho})^{(s+\tau)}}\Big|\mathscr{F}_t\right]. \tag{18}$$

Formally, if $\hat{s}^*(t)=0$ is the arg max of $\hat{V}^*(t)$ in Eq. (18), then the KG$_*$ stopping rule, adapted to this context, says to stop sampling. If $\hat{s}^*(t)>0$ is the arg max of $\hat{V}^*(t)$ in Eq. (18), then the KG$_*$ stopping role says to continue sampling. Ties for the arg max can be broken by taking the larger value of s. In summary, sampling stops unless there is a feasible non-zero length one-stage policy whose expected reward is at least as great as the expected reward of stopping.

The maximum in Eq. (18) may be computationally tedious to compute when $T_{\max}$ is even a moderate amount larger than t. It is therefore interesting to get approximations to values of s which are close to the maximizer of $\hat{V}_s(t)$ for a given t. When the discount rate is zero and the sampling variance is known, analytic approximations have been given by Frazier and Powell (2010) for off-line learning and by Ryzhov, Frazier, and Powell (2010) for online learning. An interesting area of future work would be to identify good approximations for the general case of unknown variances. In computational results below, we test the maximum over $s \in \{1, 2^{1/2}, 2, 2^{3/2}, \ldots, \min(128, T_{\max}-t)\}$ in order to approximate the continuation set for KG$_*$ in a computationally reasonable way.

4 NUMERICAL RESULTS

Figure 2 displays results motivated by an application from a published study. Moses et al. (2003) and Cohen et al. (2004) report on a clinical trial and health technology assessment to compare drug eluting stents (DES) with bare metal stents (BMS). We use data from those papers and some additional assumptions. We assume the same rate of patient enrolment and a one year time delay, so $\tau = 904$ (based on the study's enrolment of 529 patient pairs in 7 months); and $\sigma_X = 17538$. Additional assumptions include $I=0$; $P = 5\times 10^6$; $c = 800$; $T_{\max} = 4000$; $\mu_0 = 0$; zero discounting; offline learning; $n_0 = 20$ for the known variance case, and $\eta_0 = 20, \xi_0 = 40$ for the unknown variance case. CFP study this problem under the assumption of a known sampling variance. Here, we explore the case of unknown variance.



(a) Known ('+') and unknown ('o') variance with PDE (with t_0 in graph set to n_0 from known variance model).

(b) PDE ('+') and KG$_*$ ('o') with unknown variance (with t_0 in graph set to η_0 from unknown variance model).

Figure 2: Advantage of the Optimal Bayes Sequential policy over two alternative policies using the illustrative simulations (averages $\pm$ 1.96 standard errors are also shown)

Figure 2(a) illustrates the first two mechanisms for approximating the optimal stopping boundaries described in section 3.2.2. The first mechanism, which assumes a known sampling variance in the PDE, is labeled with 'UnkVariance = 0'. The second mechanism is labeled 'UnkVariance = 1'. The x-axis gives n_0 + the number of patients started, the vertical axis gives the posterior mean INMB per patient. The delay is seen at $n_0 + \tau \approx 924$, after which stage II starts. From 924 to 4024 on the x-axis, we see the stopping boundary (with 'o') when the PDE endogenizes the uncertainty due to an unknown sampling variance is inside of the boundary (with '+') when the sampling variance is assumed known.

We further interpret Figure 2(a). Stage 1: If the prior mean μ_0 exceeds 17500 then it is optimal to not run a trial and to immediately adopt. If μ_0 is between 16000 and 17500, then the dots (for the first mechanism) and circles (for the second mechanism) in the range of 200-900 on the x-axis indicate it is optimal to run a one-stage trial of the given length. If μ_0 is between -16000 and 16000, then it is optimal with both mechanisms to take at least τ samples and to continue into stage II (sequential sampling). Sampling continues until the posterior mean crosses the line with the '+' symbols (for the first mechanism) or circles (for the second mechanism), after the appropriate corrections to those curves are made with the plug-in estimator for σ_X^2.

Figure 2(b) serves to explain the second mechanism (as above; labeled now with 'DOPDE = 1') and third mechanism (adaptation of KG$_*$; labeled with 'DOPDE = 0') for approximating the optimal stopping boundaries described in section 3.2.2. As KG$_*$ uses a subset of potential future stopping plans (namely, the one-stage sampling plans), it is not surprising that the continuation set for KG$_*$ is inside of the continuation set for the PDE-based stopping rule. The KG$_*$ continuation set has, however, the advantage of not requiring a free boundary problem for a diffusion to be solved.

5 CONCLUSIONS

This paper proposed new estimators for the continuation sets of stochastic optimization tools for the special case of comparing two alternatives. Contributions are new proposals to handle unknown sampling variances

in a setting which can handle delayed samples, in addition to other (more studied) features such as online and offline learning, discounting and marginal costs per samples.

This is preliminary work for a more detailed study of the properties of the proposed approximations to the continuation sets, and additional numerical analysis. Initial results suggest that endogenizing the sampling variance in the PDE which determines optimal stopping boundaries seems to lead to somewhat smaller continuation sets as compared to similar continuation sets when the sampling variance is known. The KG_* approach was also extended to this context, and provides an alternative which does not require the solution of a free boundary PDE. Further work will also explore practical implications for costs and benefits for clinical trials.

REFERENCES

Ahuja, V., and J. Birge. 2015. "Response-adaptive designs for clinical trials: simultaneous learning from multiple patients". *European Journal of Operations Research*. DOI:10.1016/j.ejor.2015.06.077, in press.

Bernardo, J. M., and A. F. M. Smith. 1994. *Bayesian Theory*. Chichester, UK: Wiley.

Berry, D., and S. Eick. 1995. "Adaptive assignment versus balanced randomization in clinical trials: a decision analysis". *Statistics in Medicine* 14 (3): 231–246.

Berry, D. A., and C. Ho. 1988. "One-sided sequential stopping boundaries for clinical trials: a decision-theoretic approach". *Biometrics* 44:219–227.

Bertsimas, D., and A. J. Mersereau. 2007. "A learning approach for interactive marketing to a customer segment". *Operations Research* 55 (6): 1120–1135.

Branke, J., and C. Schmidt. 2004. "Sequential Sampling in Noisy Environments". In *Parallel Problem Solving from Nature*, edited by X. Yao et al., Volume 3242 of *LNCS*, 202–211: Springer.

Caro, F., and J. Gallien. 2007. "Dyamic assortment with demand learning for seasonal consumer goods". *Management Science* 53 (2): 276–292.

Caro, F., and O. S. Yoo. 2010. "Indexability of Bandit Problems with Response Delays". *Probability in the Engineering and Informational Sciences* 24:349–374.

Chen, C.-H. 1996. "A Lower Bound for the Correct Subset-Selection Probability and Its Application to Discrete Event Simulations". *IEEE Transactions on Automatic Control* 41 (8): 1227–1231.

Chen, C.-H., E. Yücesan, L. Dai, and H. Chen. 2010. "Efficient Computation of Optimal Budget Allocation for Discrete Event Simulation Experiment". *IIE Transactions* 42 (1): 60–70.

Chen, E. J., and W. D. Kelton. 2005. "Sequential Selection Procedures: Using Sample Means to Improve Efficiency". *European Journal of Operational Research* 166:133–153.

Chernoff, H. 1961. "Sequential tests for the mean of a normal distribution". In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 79–91.

Chick, S.E. and M. Forster and P. Pertile 2015. "A Bayesian Decision-Theoretic Model of Sequential Experimentation with Delayed Response". INSEAD Technical Report.

Chick, S. E., and P. Frazier. 2009. "The conjunction of the knowledge gradient and the economic approach to simulation selection". In *Proceedings of the 2009 Winter Simulation Conference*, 528–539. Piscataway, New Jersey: IEEE, Inc.

Chick, S. E., and P. I. Frazier. 2012. "Sequential Sampling for Selection with Economics of Selection Procedures". *Management Science* 58 (3): 550–569.

Chick, S. E., and N. Gans. 2009. "Economic Analysis of Simulation Selection Problems". *Management Science* 55 (3): 421–437.

Chick, S. E., and K. Inoue. 2001. "New Two-Stage and Sequential Procedures for Selecting the Best Simulated System". *Operations Research* 49 (5): 732–743.

Cohen, D. J., A. Bakhai, C. Shi, L. Githiora, T. Lavelle, R. Berezin, and others. 2004. "Cost-effectiveness of sirolimus-eluting stents for treatment of complex coronary stenoses". *Circulation* 110:508–514.

Connor, J. T., K. R. Broglio, V. Durkalski, W. J. Meurer, and K. C. Johnston. 2015. "The Stroke Hyperglycemia Insulin Network Effort (SHINE) trial: an adaptive trial design case study". *Trials* 16 (72).

DeGroot, M. 1970. *Optimal Statistical Decisions*. First ed. New York: McGraw-Hill.
Frazier, P., W. B. Powell, and S. Dayanik. 2008. "A Knowledge-Gradient Policy for Sequential Information Collection". *SIAM Journal on Control and Optimization* 47:2410–2439.
Frazier, P. I., and W. B. Powell. 2010. "Paradoxes in Learning and the Marginal Value of Information". *Decision Analysis* 7 (4): 378–403.
Gittins, J. C., and D. M. Jones. 1974. "A dynamic allocation index for the sequential design of experiments". In *Progress in Statistics*, edited by J. Gani, 241–266. Amsterdam: North-Holland.
Hammersley, J. M., and D. C. Hanscomb. 1964. *Monte Carlo Methods*. London: Methuen.
Hampson, L., and C. Jennison. 2013. "Group sequential tests for delayed responses". *Journal of the Royal Statistical Society, Series B* 75:3–54.
Hardwick, J., R. Oehmke, and Q. F. Stout. 2006. "New adaptive designs for delayed response models". *Journal of Statistical Planning and Inference* 136:1940–1955.
Kim, S.-H., and B. L. Nelson. 2006. "On the Asymptotic Validity of Fully Sequential Selection Procedures for Steady-State Simulation". *Operations Research* 54 (3): 475–488.
Lewis, R. J., A. M. Lipsky, and D. A. Berry. 2007. "Bayesian decision-theoretic group sequential clinical trial design based on a quadratic loss function: a frequentist evaluation". *Clinical Trials* 4:5–14.
Marple, B. F., C. S. Roberts, J. R. Frytak et al. 2010. "Azithromycin extended release vs. amoxicillin/clavulanate: symptom resolution in acute sinusitis". *American Journal of Otolaryngology - Head and Neck Medicine and Surgery* 31:1–8.
Moses, J. W., M. B. Leon, J. J. Popma et al. 2003. "Sirolimus-eluting stents versus standard stents in patients with stenosis in a native coronary artery". *The New England Journal of Medicine* 349 (14): 1315–1323.
Pertile, P., M. Forster, and D. La Torre. 2014. "Optimal Bayesian sequential sampling rules for the economic evaluation of health technologies". *Journal of the Royal Statistical Society, Series A* 177 (2): 419–438.
Ryzhov, I. O., P. I. Frazier, and W. B. Powell. 2010. "On the robustness of a one-period look-ahead policy in multi-armed bandit problems". *Procedia Computer Science* 1 (1): 1635–1644.
Xu, J., B. Nelson, and L. Hong. 2010. "Industrial Strength COMPASS: A comprehensive algorithm and software for optimization via simulation". *ACM TOMACS* 20 (1): 3.

AUTHOR BIOGRAPHIES

STEPHEN E. CHICK is Professor of Technology and Operations Management and the Novartis Chair of Healthcare Management at INSEAD. He has a BS degree from Stanford and an MS and PhD from UC Berkeley. He works in the areas of simulation analysis, sequential optimization, health care management, and Bayesian inference. He is Department Editor of the Simulation area in the journal Operations Research. His email address is stephen.chick@insead.edu.

MARTIN FORSTER is a Lecturer in the Department of Economics and Related Studies at the University of York. He holds MSc and DPhil degrees from the University of York. His research interests include dynamic economic models, with applications in health care and inequality and Bayesian statistics, with applications to health technology assessment. He is Programme Co-Director of York's Distance Learning programmes in Health Economics. His email address is martin.forster@york.ac.uk.

PAOLO PERTILE is an Associate Professor in the Department of Economics at the University of Verona. He holds a PhD from the Catholic University in Milan. He research interests include real options approaches in health technologies, public finance and policy, and pharmaceutical reimbursement. His email address is paolo.pertile@univr.it.