Bayesian Design for Random Walk Barriers

by Giovanni Parmigiani and Nicholas G. Polson
Duke University and University of Chicago

Abstract

We study the optimal design of absorption barriers and sample size for a simple random walk. High barriers yield more informative experiments, but may be more costly. We obtain optimal designs that are balancing the relative advantages of choosing a high barrier with few replications versus a lower barrier with more replications. We address the problem from a Bayesian, decision theoretic, viewpoint, by using a utility function based on a limiting form of Shannon information. After discussing properties of the optimal designs, we apply the results to a model describing the growth of cancer cells, and we illustrate how the selection of the prior distribution influences the solution.

Key words: Design; Shannon information; Stochastic growth model; Random walk.

1. Introduction and Overview

Consider a simple random walk N_t, $t = 0, 1, 2...$ with unknown probability Θ of an upward step, and assume that the walk starts at zero and has absorbing barriers at m_1 and $-m_2$, where m_1 and m_2 are positive integers. An experimenter is interested in learning about Θ from the observation of the barrier hitting time and of which barrier the walk hits. He is allowed to replicate the walk n times and to select the values of m_1 and m_2, and wishes to do so optimally. In this paper we consider the optimal choice of both the absorption barriers and the number of replicates n.

This problem constitutes an ideal setting to illustrate how one can address simultaneously two of the fundamental trade-offs associated with experimental design. The first, and most often encountered, is that between cost of collecting information and cost of uncertainty. The second, also of practical relevance in a variety of experimental situations, is that between few replications of an informative and expensive design and more replications of a less

informative but inexpensive design. Such trade-offs also arise, for example, in survival analysis and logistic regression.

In Section 2, we introduce a utility function accounting for the amount of information to be learnt about Θ and the cost of experimentation. The information measure used is based on a useful limiting form of Shannon information. In Section 3, we discuss properties of the optimal design. In particular, we show that if the cost is proportional to the number of steps of the walk, or if the 'exchange rate' between information and money is larger than 1/2, for then the simple Bernoulli trial ($m = 2$) with replication is the optimal design. When an additional penalty for replications is introduced, higher values of the barrier become attractive. Moreover, as long as the prior distribution is symmetric, it is optimal to select barriers equidistant from the starting point of the walk. Finally, in Section 4, we consider an application to the analysis of tumour growth, and illustrate it with various choices of prior distributions.

2. Utility and Information

2·1. *Choice of the Utility Function*

Let $\mathcal{E}$ denote the experiment consisting of n exchangeable replications of the walk, with absorption barriers at m_1 and m_2. Suppose that each replication provides the observation of the time to absorbtion $X = \min\{x \mid N_x = m_1 \text{ or } N_x = -m_2\}$, and of the indicator $Y = I_{\{N_X = m_1\}}$ of which barrier the process hits.

For a specified prior distribution on Θ, a natural measure of the information contained in $\mathcal{E}$, as pointed out, for example, in Lindley (1956), is the gain in Shannon information

$$I_n = E\left\{\log\left(\frac{p(\Theta|\mathcal{E})}{p(\Theta)}\right)\right\},$$

where the expectation is taken with respect to the joint distribution of the parameter Θ and the experiment $\mathcal{E}$. In many design problems, however, I_n is a difficult quantity to study. Polson (1988) proposed to replace it with a quantity based on the asymptotic expansion of I_n derived in the i.i.d. case by Ibragimov and H'asminsky (1973). Let Ω be the support of p and consider

the limit of I_n as the number of replications tends to infinity. Then, under suitable regularity conditions:

$$\lim_{n\to\infty}\left\{I_n - \frac{1}{2}\log\left(\frac{n}{2\pi e}\right)\right\} = \int_\Omega p(\theta)\log\left(\frac{i_{\mathcal{E}}(\theta)^{\frac{1}{2}}}{p(\theta)}\right)d\theta$$

where $i_{\mathcal{E}}(\theta)$ is Fisher information. Based on this limiting form we then adopt, as a criterion for the choice of the experiment $\mathcal{E}$, the expected normalized gain in information for large n, that is:

$$\frac{1}{2}\log\left(\frac{n}{2\pi e}\right) + \log\left(\frac{i_{\mathcal{E}}(\theta)^{\frac{1}{2}}}{p(\theta)}\right).$$

As of the experimentation cost, we assume that each replication entails a cost k, and that each of the Bernoulli trials underlying the random walk entails a cost c.

We can then write the utility associated with the experiment $\mathcal{E}$ for fixed θ as:

$$\mathcal{U}(\theta,\mathcal{E}) = \frac{1}{2}\log\left(\frac{n}{2\pi e}\right) + \log\left(\frac{i_{\mathcal{E}}(\theta)^{\frac{1}{2}}}{p(\theta)}\right) - \{cE(X|\theta)+k\}n \qquad (1)$$

This entails an expected utility function $U(m_1, m_2, n) = \int_\Omega \mathcal{U}(\theta,\mathcal{E})p(\theta)d\theta$ on the design parameters and the sample size. The choice of the optimal design and sample size consists then in the maximization of U with respect to m_1, m_2 and n.

The quantities k and c in the above expression are to be imagined as expressed in units of information. In most circumstances, the costs of replications and steps can be easily specified in terms of monetary amounts. If the information about Θ can also be translated into monetary terms, then k and c represent the ratio between the price of a replication (and a step) and the value of one unit of information. In general, however, the specification of the monetary value of one unit of information can be troublesome. In such cases, we find it attractive to think of the design problem as that of maximising the information under the constraint of a fixed expected cost for experimentation $[cE(X|\theta)+k]n$. Naturally, the first order conditions for the optimizations are the same, provided that one knows k/c and interprets the value of c in the unconstrained problem as the Lagrange multiplier of the

constrained problem. Of further help in the choice of the appropriate cost parameters are trade-off plots where the expected cost of experimentation is graphed against the maximum information obtainable at that expected cost.

2·2. *Fisher Information*

An important feature of the experimental setting considered here is that each replication is more informative the longer it takes for the walk to hit one of the barriers. Intuitively, a longer walk corresponds to a larger number of Bernoulli trials. In particular, the Fisher information, on which the utility function adopted here is based, incorporates this feature in a natural way. An application of Wald's identity (see, for example, Cox and Miller, 1965), leads to the expression:

$$i_{\mathcal{E}}(\theta) = \frac{E(X|\theta)}{\theta(1-\theta)} \tag{2}$$

The information in one replication is given by the information in a single Bernoulli trial, multiplied by the expected number of trials. In particular, in this case, it is useful to recall that:

$$E(X|\theta) = \frac{m_1\theta^{m_1}[\theta^{m_2} - (1-\theta)^{m_2}] - m_2(1-\theta)^{m_2}[\theta^{m_1} - (1-\theta)^{m_1}]}{(\theta^{m_1+m_2} - (1-\theta)^{m_1+m_2})\,(2\theta - 1)} \tag{3}$$

$$E(Y|\theta) = \frac{\theta^{m_1}[\theta^{m_2} - (1-\theta)^{m_2}]}{\theta^{m_1+m_2} - (1-\theta)^{m_1+m_2}}; \tag{4}$$

For ease of future reference, let us define $\varphi_\theta(m_1, m_2) = E(X|\theta)$ in (3). It can be verified that $\varphi_\theta(m_1, m_2)$ is increasing in both m_1 and m_2 for every θ. Figure 1 illustrates $i_{\mathcal{E}}(\theta)$ for various choices of m_1 and m_2. In passing, we note that the strong design dependence of Fisher information will affect the family of Jeffreys' priors for this experiment; as a consequence, the mechanical adoption of such priors will lead to choices of the design that conflict with standard expected utility theory and may even violate the dominance principle.

3. Properties of the Optimal Design

Based on Proposition 1, the problem of choosing the optimal design and sample size can be rewritten, neglecting irrelevant constants, as that of minimizing

$$\frac{1}{2}\log n + \frac{1}{2}\int_\Omega \log \varphi_\theta(m_1, m_2)p(\theta)d\theta - \left(c\int_\Omega \varphi_\theta(m_1, m_2)p(\theta)d\theta + k\right)n \quad (5)$$

with respect to (m_1, m_2, n). It can be verified that if the prior distribution is proper, the above integrals always exists, since φ is larger than one, and bounded above in θ for fixed design. The utility function depends on the design only through the function φ. For fixed θ and n, a higher φ increases the costs linearly and the information logarithmically. Note that if $c = 0$, increasing φ always increases expected utility so that the optimal design problem has no solution. Assume henceforth that $c > 0$.

To simplify the analysis we will from now on treat m_1, m_2 and n as real variables. For fixed m_1, m_2 it is possible to derive the optimal sample size

PROPERTY 1: The optimal sample size is:

$$n^* = \frac{1}{2}\left(c\int_\Omega \varphi_\theta(m_1, m_2)p(\theta)d\theta + k\right)^{-1} \quad (6)$$

PROOF: Differentiating (5) and setting the result to zero gives (6). Second order conditions are straightforward to verify. □

Substituting n^* in (5), and again neglecting irrelevant constants, reduces the optimal design problem to the maximization of:

$$U^* = \int_\Omega \log \varphi_\theta(m_1, m_2)p(\theta)d\theta - \log\left(\int_\Omega \varphi_\theta(m_1, m_2)p(\theta)d\theta + \frac{k}{c}\right). \quad (7)$$

A noticeable consequence of this on the analysis of the trade-offs is that the choice of the optimal design variables m_1 and m_2 only requires the specification of the ratio k/c, and therefore does not depend on the 'exchange rate' the experimenter decides to adopt between information and money. The exchange rate is only influencing the determination of n and the decision of whether or not to perform the experiment. This is a direct consequence of the fact that the cost component of the utility function is linear in n for fixed design parameters, and of the fact that a more expensive design affects the cost component of the utility function only through a term multiplying n. Finally, as expected, φ and n are inversely related. Heuristically, if c is

increased, it becomes comparatively more convenient to replicate the experiment many times with small barriers, rather than doing a few replications with relatively higher barriers.

We now give sufficient conditions for optimality and use them to derive some useful properties of the optimal design. Let

$$\varphi_\theta^{(i)}(m_1, m_2) = \frac{\partial \varphi}{\partial m_i}, \qquad i = 1, 2 \tag{8}$$

A system of sufficient conditions for optimality is given by the following.

PROPERTY 2: The optimum design variables m_1 and m_2 must satisfy:

$$\int_\Omega \varphi_\theta(m_1, m_2) p(\theta) d\theta + \frac{k}{c} = \frac{\int_\Omega \varphi_\theta^{(1)}(m_1, m_2) p(\theta) d\theta}{\int_\Omega \frac{\varphi_\theta^{(1)}(m_1, m_2)}{\varphi_\theta(m_1, m_2)} p(\theta) d\theta} = \frac{\int_\Omega \varphi_\theta^{(2)}(m_1, m_2) p(\theta) d\theta}{\int_\Omega \frac{\varphi_\theta^{(2)}(m_1, m_2)}{\varphi_\theta(m_1, m_2)} p(\theta) d\theta} \tag{9}$$

PROOF: From (7):

$$\frac{\partial U^*}{\partial m_i} = \int_\Omega \frac{\varphi_\theta^{(i)}(m_1, m_2)}{\varphi_\theta(m_1, m_2)} p(\theta) d\theta - \frac{\int_\Omega \varphi_\theta^{(i)}(m_1, m_2) p(\theta) d\theta}{\int_\Omega \varphi_\theta(m_1, m_2) p(\theta) d\theta + \frac{k}{c}}, \tag{10}$$

Setting the above to zero yields the desired conditions. □

The following two properties give sufficient conditions for the simple Bernoulli trial to be the optimal design.

PROPERTY 3: If $k = 0$ the optimal design is $m_1 = m_2 = 1$ for every choice of c and of the prior distribution. Consequently, the optimal sample size is $n^* = 1/(2c)$.

PROOF: From (3), we have: $\varphi_\theta(1, 1) = 1$ for every θ. Substituting into (9), all three terms are equal to unity. □

PROPERTY 4: If $c > 1/2$ then $m_1 = m_2 = 1$ for every k and prior distribution such that $n^* \geq 1$.

PROOF: Rewrite (10) as:

$$\frac{\partial U^*}{\partial m_i} = \int_\Omega \frac{\varphi_\theta^{(i)}(m_1, m_2)}{\varphi_\theta(m_1, m_2)} p(\theta) d\theta - 2cn^* \int_\Omega \varphi_\theta^{(i)}(m_1, m_2) p(\theta) d\theta$$

We know $\varphi \geq 1$, therefore:

$$\int_\Omega \frac{\varphi_\theta^{(i)}(m_1, m_2)}{\varphi_\theta(m_1, m_2)} p(\theta) d\theta < \int_\Omega \varphi_\theta^{(i)}(m_1, m_2) p(\theta) d\theta$$

and both $\partial U^*/\partial m_i$ are negative if $n^* > 1/2c$. This is guaranteed if $c > 1/2$. □

Finally, we give a condition for the optimal barriers to be symmetric.

PROPERTY 5: If the prior is symmetric about 1/2, that is if $p(\theta) = p(1-\theta)$, then $m_1 = m_2$ satisfies the optimality conditions.

PROOF: From (3), we have that $\varphi_\theta(m_1, m_2) = \varphi_{1-\theta}(m_2, m_1)$, and, therefore, using (8), that $\varphi^1_\theta(m_1, m_2) = \varphi^2_{1-\theta}(m_2, m_1)$. Also, from the symmetry of the prior distribution,

$$\begin{aligned}
\int_\Omega \varphi^{(1)}_\theta(m,m)p(\theta)d\theta &= \int_\Omega \varphi^{(2)}_{1-\theta}(m,m)p(\theta)d\theta = \int_\Omega \varphi^{(2)}_\theta(m,m)p(\theta)d\theta \\
\int_\Omega \frac{\varphi^{(1)}_\theta(m,m)}{\varphi_\theta(m,m)}p(\theta)d\theta &= \int_\Omega \frac{\varphi^{(2)}_{1-\theta}(m,m)}{\varphi_{1-\theta}(m,m)}p(\theta)d\theta = \int_\Omega \frac{\varphi^{(2)}_\theta(m,m)}{\varphi_\theta(m,m)}p(\theta)d\theta
\end{aligned}$$

so that the right hand equation of (9) is satisfied. □

4. An Application to Cancer Growth Models

A fruitful area of application for the problem considered in this paper is that of growth models. For example, Downham and Morgan (1973) and Downham and Green (1976) analyse the spread of cancer cells in a layer of competing normal cells. In this application, the position N_t of the walk represents the number of cancer cells after t cell divisions. The random walk is assumed to start at $N_0 = 1$, when the first abnormal cell appears, and to stop if no cancer cells are left, so that zero is the lower absorbtion barrier. Also, most empirical work is carried out based on experiments where there is an upper absorption barrier at a fixed number m of abnormal cells—typically related to the detectability threshold of the tumour. The parameter of interest is usually $\Gamma = \Theta/(1-\Theta)$, the relative division rate; Γ is believed to be larger than 1 and called carcinogenic advantage. Since the expected utility we adopt is invariant under one-to-one transformations of the parameter of interest, the conclusions reached hold for Γ as well. In this application the quantity k may be interpreted as the cost of one experimental unit, say a mice, whereas c represents a cost for waiting.

This particular growth model obtains as a special case of the model discussed in the previous sections by setting $m_2 = 1$ and $m_1 = m - 1$, $m \geq 2$. The expected number of steps and the probability of hitting the upper barrier

specialise to:

$$E(X|\theta) = \frac{m\theta^{m-1}}{\theta^m - (1-\theta)^m} - \frac{1}{2\theta - 1} \tag{11}$$

$$E(Y|\theta) = \frac{\theta - (1-\theta)}{\theta^m - (1-\theta)^m}\theta^{m-1}. \tag{12}$$

Figure 2 illustrates $E(X|\theta)$ for varying m.

Since $E(X|\theta) = 1$ at $m = 2$, Equation (2) provides further intuition to the results of Hinkley (1979), who found that $m = 2$ is the choice that guarantees the smallest discrepancy between observed and expected information.

More generally however,

$$\frac{\partial \varphi_\theta(m)}{\partial m} = \frac{\theta^{m-1}\left[\theta^m - (1-\theta)^m(1 + m\log\theta(1-\theta))\right]}{(\theta^m - (1-\theta)^m)^2} > 0. \tag{13}$$

Therefore, as expected, the conditional information contained in a single replication is an increasing function of m, for every θ, so that Bernoulli trials are not necessarily optimal.

Let us now briefly illustrate the choice of the design variable m for various prior distributions. First, we take Θ uniform on $(0, 1)$; then $m = 5$ is optimal at $\frac{k}{c} = 1$. Moreover, if $\Theta \sim Beta(4, 1)$, $m = 18$ is optimal for $\frac{k}{c} = 1$. Note that as the mass is placed on higher values of θ, a higher value of m becomes more convenient. This dependence becomes extreme if the prior is uniform on $(\frac{1}{2}, 1)$: for example $\frac{k}{c} = 1$ implies $m = 773$. This is not an implausible specification for a stochastic growth, where the experimenter usually believes that $\Theta > \frac{1}{2}$. The reason for such a drastic change is that the latter prior assigns a much smaller probability to less informative replications —such as replications where the lower barrier is hit at the first step— and therefore makes the few replications with a higher barrier a more attractive choice.

Acknowledgement

We thank Jay Kadane, Teddy Seidenfeld, Larry Wasserman and a referee for helpful comments.

References

Cox, D. R. & Miller, H. D. (1965). *The Theory of Stochastic Processes.* London, Methuen.

Downham, D. Y. & Morgan, K. (1973). A stochastic model for a two-dimensional growth on a square lattice. *Bull. Inst. Int. Statist.* **45**, 324–31.

Downham, D. Y. & Green, D. H. (1976). Inference for a two-dimensional stochastic growth model. *Biometrika* **63**, 551-4.

Hinkley, D. V. (1979) Likelihood inference for a simple growth model. *Biometrika*, **66**, 659-62.

Ibragimov, I. & H'asminsky. R. (1973). On the information contained in a sample about a parameter. *Second Intl. Symp. on Information Theory*, 295-309.

Kullback, S. & Leibler, R. A. (1951) On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79–86.

Lindley, D. V. (1956). On a Measure of the Information Provided by an Experiment. *Annals of Mathematical Statistics,* **27**, 986–1005.

Polson, N. G. (1988). *Bayesian perspectives on statistical modelling.* Unpublished Ph. D. thesis, University of Nottingham

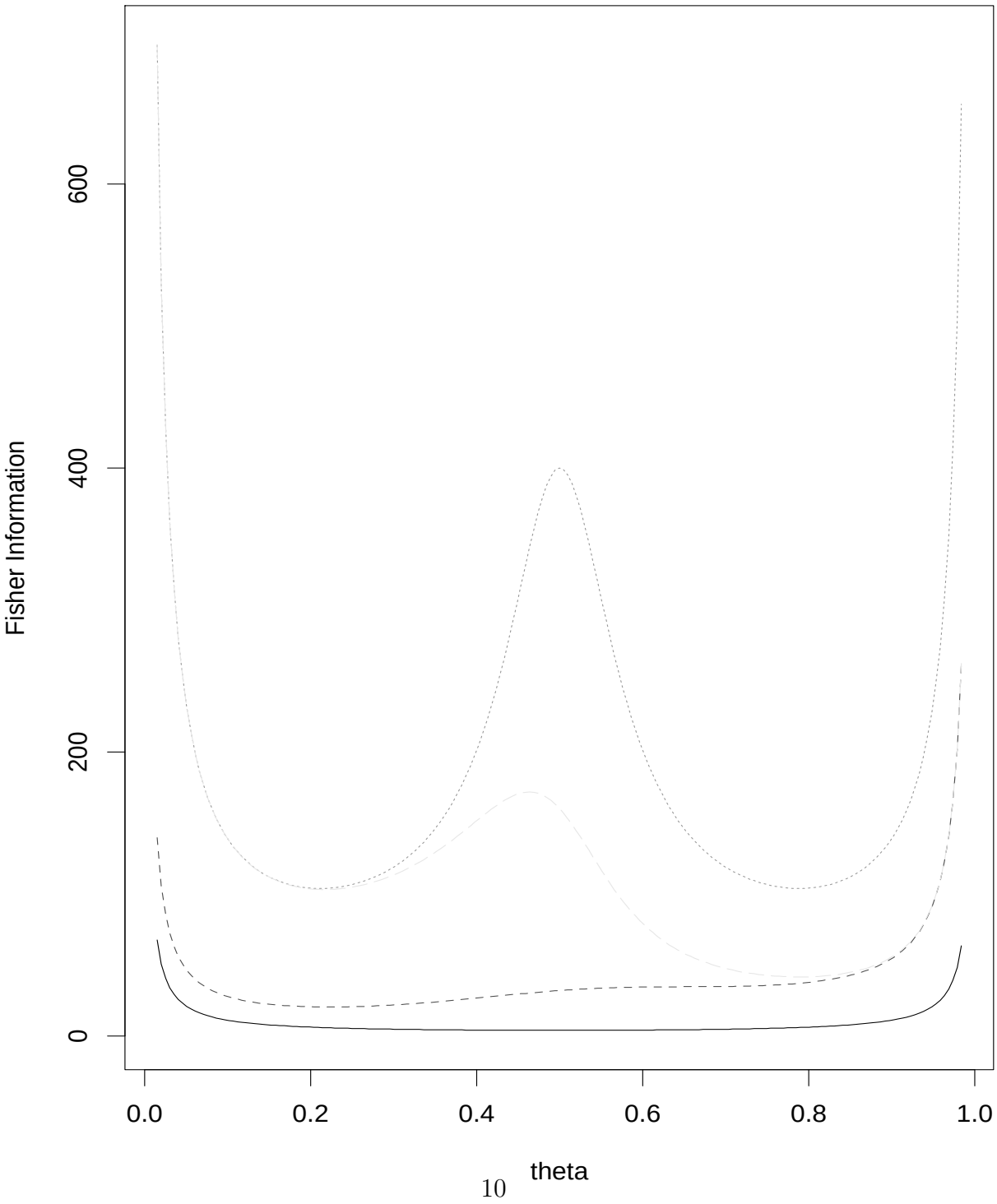

Figure 1: Fisher Information as a function of Theta, for varying values of the design parameters; from the bottom, $m_1 = m_2 = 1$, $m_1 = 4$ and $m_2 = 2$, $m_1 = 4$ and $m_2 = 10$, $m_1 = m_2 = 10$.