Using Genetic Algorithms to Design Bayesian Reliability Experiments

Harry F. Martz
Group D-1, MS F600
Los Alamos National Laboratory
P.O. Box 1443
Los Alamos, NM USA 87545
hfm@lanl.gov

Michael S. Hamada
Group D-1, MS F600
Los Alamos National Laboratory
P.O. Box 1443
Los Alamos, NM USA 87545
hamada@lanl.gov

C. Shane Reese
Department of Statistics
Brigham Young University
Provo, UT USA 84602
reese@statmail.byu.edu

Alyson G. Wilson
Group D-1, MS F600
Los Alamos National Laboratory
P.O. Box 1443
Los Alamos, NM USA 87545
agw@lanl.gov

Abstract

The purpose of a lifetime reliability experiment is to quantify the effect of one or more factors on the lifetime of some device of interest. In such experiments, each tested device either fails or is still functioning when the experiment (under the given factor conditions) is terminated at a given time (right censored lifetimes/time-truncated testing). The purpose of this paper is to introduce genetic algorithms as a practical means for finding near-optimal Bayesian experimental designs for such experiments. The design criterion is the expected Shannon information gain provided by the experiment. The methodology is illustrated using a subset of the eleven-factor thermostat experiment considered by Bullington et al. (1993). A lognormal lifetime distribution is assumed.

1. Introduction

Bullington et al. (1993) used a 12-run Plackett-Burman design in an experiment designed to quantify the effect of 11 factors on the lifetime of industrial thermostats. Ten thermostats were manufactured and tested up to 7342 (x 1000) cycles for each of the 12 runs. Note that 22 of the 120 lifetimes were observed to be right-censored at 7342. Wu and Hamada (2000) used a Bayesian variable selection procedure to deduce the fact that the simple two-factor linear model $\mu_i = \alpha + \beta E_i + \gamma H_i + \delta E_i H_i$ is appropriate for these data. Here E_i is the beryllium copper grain size (in.) and H_i is the heat treatment. Further, suppose t_{ij} denotes the lifetime of the jth thermostat in the ith run. Following Wu and Hamada (2000), we henceforth assume that t_{ij} follows a lognormal distribution; thus, $y_{ij} \equiv \ln t_{ij} \sim N(\mu_i, \sigma)$, i.e., a lognormal regression model.

As a hypothetical extension to this experiment, we now consider a smaller Bayesian follow-up experiment for more precisely fitting μ_i. We will use the joint posterior distribution obtained from a Bayesian analysis of a subset of the original data as an informative prior distribution in the new experiment**.** The follow-up experiment will consist of testing 3 additional thermostats in each of 8 additional runs at (possibly new) factor levels $\{(E_i, H_i), i = 1, \ldots,8\}$. The objective is to determine these 8 pairs of factor levels using Bayesian experimental design methods.

Hamada et al. (2001) introduce the use of genetic algorithms (GAs) for finding near-optimal Bayesian experimental designs. As in Hamada et al. (2001), we seek a Bayesian design that maximizes the expected Shannon information gain, the so-called "optimal" Bayesian experimental design. Here the expectation must be taken over two classes of unknowns: the sample response data y_{ij}, and the four unknown parameters in μ_i as well as σ. This utility function is known to be appropriate when the purpose of the experiment is precise inference about $\theta = (\alpha, \beta, \gamma, \delta, \sigma)^T$. For this utility, the optimal Bayesian experimental design $\mathbf{X}^{\mathbf{OPT}}$ is the one that maximizes the *expected information gain (EIG)*

$$EIG(\boldsymbol{X}) = \iint \log[\pi(\theta \mid \boldsymbol{y}, \boldsymbol{X}) / \pi(\theta)] f(\boldsymbol{y} \mid \theta, \boldsymbol{X}) \pi(\theta) d\theta d\boldsymbol{y}, \quad (1)$$

where **y** denotes the vector of log lifetimes, both censored and uncensored, distributed according to the Gaussian sampling model $f(\mathbf{y} \mid \theta, \mathbf{X})$, $\pi(\theta)$ is the prior distribution, and $\pi(\theta \mid \mathbf{y}, \mathbf{X})$ is the posterior distribution of θ conditional on **y**. We denote designs that optimize (1) as optimal EIG designs. The main problem with (1) is that, for the lifetime experiment described above, the integration is intractable and simulation methods must be used. As in Hamada et al. (2001), we use a two-stage iterative process to find the design $\mathbf{X}^{\mathbf{OPT}}$ that nearly optimizes (1). In Stage 1 we use a GA to generate potentially high EIG designs, and in Stage 2 we use Monte Carlo simulation to approximate (1) for each of the candidate designs proposed in Stage 1.

Goldberg (1989), Michalewicz (1992), Holland (1992a) and Bäck (1996) are excellent textbooks on GAs, while Holland (1992b) provides a nice introductory tutorial. Also, the paper by Hamada et al. (2001) describes the use of a GA for obtaining near-optimal Bayesian experimental designs within the general context of regression models.

2. Brief Summary of Stage 1: Genetic Algorithms

A GA operates on a "population" of candidate "solutions" to the optimization problem. For each solution, a single chromosome completely defines an experimental design **X** where the length of each chromosome is simply the product of the number of runs and the number of factors. For convenience, suppose that the experiment of interest contains p factors, and n runs must be made (in our example, p = 2 and n = 8). If we consider each factor level as a gene, then each chromosome (or design **X**) has np genes whose values we seek.

A GA proceeds by defining an initial population of candidate solutions and subsequent populations of solutions obtained by use of the genetic operators of crossover and mutation within the context of an elitist GA as described by Hamada et al. (2001).

We assume that the design region is bounded, say, $L_i \le x_i \le U_i$ for the ith factor x_i. For our example, we have 0.008 in. $\le E_i \le$ 0.018 in. and 45 min. (at 600°) $\le H_i \le$ 240 min. (at 600°). To begin the GA process, we first generate an initial population of 10 random designs using independent uniform random draws for each factor in each of the 8 runs. We then evaluate the utility (the fitness) of each of these 10 random designs using the Stage 2 approach to be briefly described. These 10 designs are then ranked according to their utility, and this completes the first generation of the GA.

The second (and subsequent) GA generations are now populated using genetic crossover and mutation as described in Hamada et al. (2001). In the original GA, each new population completely replaces the previous one. It can then happen that the best (most fit) solution in population k + 1 is worse than the best solution in population k. Consequently, very good solutions can be lost forever. A solution to this problem is to use an "elitist" GA. At each generation we keep the best 10 designs (those with highest utilities) out of the 30 total designs (10 initial designs, 10 crossover designs and 10 mutated designs) which becomes the population of initial designs for the next generation.

We execute the above GA in batches of 100 generations in order to allow for, what is known in evolutionary biology, as "punctuated equilibrium." In simple terms, punctuated equilibrium is an observed genetic phenomenon in which mutations essentially decrease over time but with periodic upsets in this process (that is, periodic large-scale catastrophic mutations are occasionally permitted to occur). The best 10 solutions after a given batch has been completed become the initial set of designs for the next batch of 100 generations. After several batches of 100 generations of solutions have been obtained in this way, we finally report the design having the highest utility as our desired near-optimal Bayesian experimental design. In the next section we will illustrate the performance of this adaptation of a GA.

3. Brief Summary of Stage 2: Utility Estimation

The utility for each of the GA-produced candidate designs generated at Stage 1 is estimated in Stage 2. For a given candidate design **X**, we estimate the utility in (1) by Monte Carlo simulation. We sample the specified prior distribution $\pi(\theta)$ and assumed Gaussian sampling model $f(\mathbf{y} \mid \theta, \mathbf{X})$ conditional on θ and **X**. Because the posterior distribution is unavailable in closed form, we estimate (1) as

$$\hat{EIG}(\boldsymbol{X}) = \frac{1}{L}\sum_{l=1}^{L} \log\left[\frac{f(\boldsymbol{y}^{(l)} \mid \theta^{(l)}, \boldsymbol{X})}{f(\boldsymbol{y}^{(l)} \mid \boldsymbol{X})}\right] \tag{2}$$

where $\{(\theta^{(l)}, \mathbf{y}^{(l)}), l = 1, 2, \ldots, L\}$ denote L corresponding dependent pairs of randomly sampled values: $\theta^{(l)}$ from the prior distribution $\pi(\theta)$, and $\mathbf{y}^{(l)}$ conditionally from the $N(\mu_i, \sigma)$ sampling model. Any sampled y_{ij} value greater than ln(7342) is censored at this value. The corresponding censored data are used to calculate the combined (non-censored and censored) Gaussian likelihood in (2) given by

$$f(\mathbf{y}^{(l)} \mid \theta^{(l)}, \boldsymbol{X}) = \prod_{\substack{1 \le i \le 8 \\ j \in FAIL^{(l)}}} \frac{1}{\sigma}\phi\left(\frac{y_{ij} - \mu_i}{\sigma}\right) \prod_{\substack{1 \le i \le 8 \\ j \in CEN^{(l)}}} \left[1 - \Phi\left(\frac{\ln 7342 - \mu_i}{\sigma}\right)\right] \tag{3}$$

where $FAIL^{(l)}$ and $CEN^{(l)}$ each denote the set of failure and censored lifetimes, respectively, in the *l*th sample. In (3), $\varphi(\cdot)$ and $\Phi(\cdot)$ denote the usual standard normal pdf and cdf, respectively. We tentatively propose using a multivariate kernel density estimator to estimate the normalizing constant (the marginal density) in the denominator in (2).

4. Example

Again consider the example introduced in Section 1. Table 1 summarizes the marginal prior distributions based on an objective Bayesian analysis (using WinBUGS) of the original thermostat data using only a randomly selected subset of 3 lifetimes in each of the 12 runs. In this case, 5 out of the 36 lifetimes were found to be censored. The analysis of this subset of the data ensures that the joint prior distribution will be sufficiently diffuse so that the additional data from the follow-up experiment will have an impact.

Table 1: Summary of the marginal prior distributions of θ

Parameter	Mean	Std Dev	2.5%	5.0%	Median	95.0%	97.5%
α	6.293	0.148	6.012	6.054	6.290	6.543	6.597
β	-0.772	0.150	-1.076	-1.019	-0.770	-0.531	-0.482
γ	0.803	0.147	0.525	0.567	0.800	1.050	1.103
δ	-0.433	0.150	-0.735	-0.680	-0.431	-0.194	-0.151
σ	0.862	0.129	0.652	0.679	0.847	1.097	1.153

Unfortunately, the example results obtained by applying the above methodology were not yet available when this manuscript was submitted. Hopefully, however, the results will be available for presentation at the MMR 2002 conference.

References

Bäck, T. (1996). Evolutionary algorithms in theory and practice. New York: Oxford University Press.

Bullington, R. G., Lovin, S., Miller, D. M. and Woodall, W. H. (1993). Improvement of an industrial thermostat using designed experiments. *Journal of Quality Technology 25* (4), 262-270.

Goldberg, D. E. (1989). Genetic algorithms in search, optimization and machine learning. New York: Addison-Wesley.

Hamada, M., Martz, H. F., Reese, C. S. and Wilson, A. G. (2001). Finding near-optimal Bayesian experimental designs via genetic algorithms. *The American Statistician 55* (3), 175-181.

Holland, J. M. (1992a). Adaptation in natural and artificial systems. Cambridge, MA: MIT Press.

Holland, J. M. (1992b). Genetic algorithms. *Scientific American, July*, 66-72.

Michalewicz, Z. (1992). Genetic algorithms + data structures = evolution programs. New York: Springer-Verlag.

Wu, C. F. J. and Hamada, M. (2000). Experiments: Planning, analysis and parameter optimization. New York: Wiley.