Sankhyā: The Indian Journal of Statistics
Special Issue on Bayesian Analysis
1998, Volume 60, Series A, Pt. 3, pp. 446-458

DESIGNING OBSERVATION TIMES FOR INTERVAL CENSORED DATA*

By GIOVANNI PARMIGIANI
Duke University, Durham

SUMMARY. This paper studies the optimal choice of observation times for duration data, presenting necessary conditions for the optimality of a sequence of observation times under two alternative criteria, and characterizing the form of the solution in special cases. A corollary of these results is a simple expression for the binary partition of a continuous random variable that is most informative for learning about an unknown parameter in its distribution.

1. **Introduction**

Observing a random process over time can sometime be difficult or costly so that efficiency in gathering information about the process can be substantially enhanced by carefully choosing observation times. In recent years, issues related to the choice of observation times have arisen in diverse application areas. Yet, relatively little methodology is available. Although recent progress in computing has provided a dramatic improvement in our ability to determine optimal or near optimal solutions, designing observation times is still challenging and further theoretical guidance is needed. From a statistical perspective, it is interesting to pose the problem: what are the observation strategies that can teach more about the process, for example in terms of providing information about process parameters?

This article considers this question in the context of duration data, that is data about the time until the occurrence of a given event of interest. A prototype example is the followup of patients after the administration of a medical treatment. Duration data are often collected by checking on the status of experimental units at regular intervals. In that case, actual durations are not

---

AMS (1991) *subject classification.* 62F15.
Key words and phrases. Survival data, Kullback-Leibler information criterion, Bayesian design of experiments, exponential distribution.
* Work supported by the NSF under grant DMS 9403818.

observed directly, but only known to belong to a certain interval of time, a circumstance termed interval-censoring. Sometimes intervals are small enough that little information is lost by censoring. However, when checking is expensive, intervals are wide, and inferences may be very sensitive to the choice of times. Equal spacing may be inefficient. Investigators need to best balance the cost of observations with the advantages of more accurate information.

Examples arise especially in reliability and biometry, and include investigations on the shelf life of chemicals, the recurrence time of diseases with an asymptomatic initial phase, bioassay experiments, the long term effects of medical treatments and others. Similar problems arise in determining a sequence of stress levels to apply to an item to induce a fracture, or the sequence of drug doses needed to provoke a certain response.

Bayesian inference and decision theory provide a general and successful theoretical framework for optimal decisions about observation times. Advantages are related the ease of modeling quantitatively both the investigator's expectations about the experimental results (including the relevant uncertainties), and the investigator's goals in performing the analysis. Chaloner and Verdinelli (1995) review issues and recent developments in Bayesian experimental design. We follow this approach here. Other examples of papers that use Bayesian methods to deal with designing sequences on the real line, albeit in different contexts, are Chaloner and Larntz (1989), Hatzis and Larntz (1992), Verdinelli, Polson and Singpurwalla (1992).

Our formulation of the design problem is general in that there is no constraint that the observations should be equally spaced, but it is limited to problems involving a single experimental unit. The focus is on two alternative objective functions: The first assumes that the experimenter is interested in learning about the distribution of unknown parameters in the probability model for the duration time. This is captured by considering a loss function that depends on the discrepancy between the posterior distribution of the unknown parameters after the experiment and the posterior distribution that would have been available if the event time could have been observed exactly. The second assumes that the experimenter is interested in producing a point estimate of a one-dimensional unknown parameter.

The paper discusses necessary conditions for optimality of a sequence of observation times, and a characterization of the form of the solution in special cases. Section 2 presents the problem formulation, Section 3 discusses general results for the information theoretic approach, Section 4 specializes those to exponential waiting times and Section 5 looks at squared error loss. Determining more general solutions in challenging, but our results may help tackle two kinds of important generalizations: multiple units observed in batch and multiple units observed sequentially. These issues are briefly addressed in Section 6.

2. Problem Definition

We will observe a single experimental unit until the occurrence of an event of interest. Our problem is to choose a sequence, or schedule, of observation times. For any positive integer n, denote a schedule by $\boldsymbol{\tau}^{(n)} = \{\tau_1^{(n)}, ..., \tau_n^{(n)}\}$, and let $\tau_0^{(n)} = 0$ and $\tau_{n+1}^{(n)} = \infty$. Here n is the number of planned observations and $\tau_i^{(n)}$ the time of the i-th observation. Because the data collection stops as soon as the event is detected, when n is finite a maximum of n observations are carried out. We will adopt the convention that infinity is allowed as a value for the observation time (meaning that no further observations are scheduled). Also $n = \infty$ will mean that the data collection will continue until the event is detected, no matter how many observations have already been taken. The superscript n will be dropped when no ambiguity arises.

The model for the event time distribution is as follows. We consider a parametric model indexed by the vector Λ in $\Omega \subseteq \Re^k$. Given $\Lambda = \lambda$, the time to event Y has density $g(y|\lambda)$ with support in $[0, \infty)$. In turn Λ has cumulative distribution function Π, with each marginal distribution either discrete or continuous. Π reflects the investigator's knowledge prior to performing the experiment. The resulting marginal density of Y is given by:

$$f(y) = \int_\Omega g(y|\lambda) d\Pi(\lambda), \qquad \ldots(1)$$

where the above is a k-dimensional integral.

If an event is observed at the $(i+1)$-st observation, the investigator's knowledge about Λ can be updated based on the information that the event must have occurred in the interval $(\tau_i, \tau_{i+1}]$. The posterior distribution for Λ is then, from Bayes' theorem:

$$\pi(\lambda|\tau_i < Y \leq \tau_{i+1}) = \frac{G(\tau_{i+1}|\lambda) - G(\tau_i|\lambda)}{F(\tau_{i+1}) - F(\tau_i)} \pi(\lambda), \qquad \ldots(2)$$

where F and G are the cumulative distribution functions corresponding to f and g. The narrower the interval containing the event, the more accurate the information and the more valuable the experiment. The relevant trade-off is therefore between the cost of observation and the accuracy of the posterior inference. The latter depends on the difference between the actual posterior distributions and the posterior distributions that would have been available had Y been observed exactly. Here we assume that each observation costs a fixed amount k and focus on two alternative strategies for quantifying the loss of information. The first is the Kullback-Leibler divergence between the "actual" and "ideal" posterior distributions. The second assumes that the posterior distribution is used in an estimation problem with squared error loss.

3. Information Theoretic Loss

Information theory offers general measures of the amount of information gained about a parameter from experimentation. Information theoretic criteria have been introduced by Lindley (1956) and are widely used in experimental design. Lindley's approach is the method of choice for measuring the information contained in an experiment, when the purpose of the investigation is not tied to a specific decision Verdinelli (1992), Parmigiani and Berry (1994).

Traditional information theoretic criteria focus on the divergence between the prior and the posterior distribution. In our context, we have a natural gold standard, that is the posterior distribution given the complete information $Y = y$, which is:

$$\pi(\lambda|Y = y) = \frac{g(y|\lambda)\pi(\lambda)}{f(y)}.$$

We can therefore introduce the following variant of traditional information theoretic criteria. We assume that, in addition to the cost incurred for observation, the experimenter loses the amount $cK_{\tau_{i+1},\tau_i}(y)$, where c is a known positive constant and $K_{\tau_{i+1},\tau_i}(y)$ is a Kullback-Leibler divergence between the posterior distributions given the interval-censored information $\tau_i < Y \leq \tau_{i+1}$ and under the posterior distribution given the complete information $Y = y$. In computing the divergence, the "ideal" posterior is used in the role of the true underlying distribution in computing the divergence.

In this formulation, $K_{\tau_{i+1},\tau_i}(y)$ can be thought of as the loss of information due to the fact that the knowledge of $Y = y$ has to be replaced by that of $\tau_i < Y \leq \tau_{i+1}$, and is given by:

$$\begin{aligned} K_{\tau_{i+1},\tau_i}(y) &= E\left\{\log \tfrac{\pi(\Lambda|Y=y)}{\pi(\Lambda|\tau_i<Y\leq\tau_{i+1})} \,|\, Y = y\right\} \\ &= \int_\Omega \log\left(\tfrac{\pi(\lambda|Y=y)}{\pi(\lambda|\tau_i<Y\leq\tau_{i+1})}\right)\pi(\lambda|Y = y)d\lambda. \end{aligned} \qquad \ldots(4)$$

If the cost of one observation is k, then the total loss conditional on $\tau_i < Y \leq \tau_{i+1}$ is:

$$L(\boldsymbol{\tau}^{(n)}, y) = k(i + 1) + cK_{\tau_{i+1},\tau_i}(y).$$

Calculating the expectation of L with respect to the marginal distribution of Y yields the risk function, or objective function:

$$R(\boldsymbol{\tau}^{(n)}) = \sum_{i=0}^{n-1}\int_{\tau_i}^{\tau_{i+1}}[k(i+1)+cK_{\tau_{i+1},\tau_i}(y)]f(y)dy+\int_{\tau_n}^{\infty}[kn+cK_{\infty,\tau_n}(y)]f(y)dy. \qquad \ldots(5)$$

Define $\tilde{Y}$ as the index of the observation at which the event is detected. Then, because $\log \pi(\Lambda|\tilde{Y})$ is constant as y varies within an interval, we can write:

$$\sum_{i=0}^{n} \int_{\tau_i}^{\tau_{i+1}} K_{\tau_{i+1},\tau_i}(y) f(y) dy = E\left\{\log \frac{\pi(\Lambda|Y)}{\pi(\Lambda|\tilde{Y})}\right\} = \phi(\pi, g) - E\{\log \pi(\Lambda|\tilde{Y})\}, \quad \ldots (6)$$

where expectations are taken with respect to the joint distribution of Λ and Y. The term $\phi(\pi, g)$ does not depend on the schedule $\boldsymbol{\tau}^{(n)}$. Hence minimizing the expectation of K is equivalent to minimizing the posterior entropy function $-E\{\log \pi(\Lambda|\tilde{Y})\}$.

In practical applications, the elicitation of the constants c and k for trading monetary units with information units can be difficult. However, because the risk function is a linear combination of expected loss of information and expected number of observations, the ratio of the costs can be thought of as the Lagrange multiplier in the constrained minimization of the expected number of observation for given expected information loss (Clyde and Chaloner 1996). Thus the optimality conditions derived using (4) can also be used to address constrained optimization problems, in which the elicitation is typically easier Verdinelli and Kadane (1992), Parmigiani and Polson (1992).

The following result summarizes the first order conditions for optimality.

Theorem 1. *Let $g(y|\lambda)$ be strictly positive, continuous and bounded in y for every $\lambda \in \Omega$. Then for $n > 1$ the optimal policy $\boldsymbol{\tau}^{(n)}$ must satisfy:*

$$\begin{aligned} \int_\Omega \log\left(\frac{\pi(\lambda|\tau_i<Y\le\tau_{i+1})}{\pi(\lambda|\tau_{i-1}<Y\le\tau_i)}\right)\pi(\lambda|Y=\tau_i)d\lambda &= \frac{k}{c} \qquad i = 1, 2, ..., n-1 \\ \int_\Omega \log\left(\frac{\pi(\lambda|Y>\tau_n)}{\pi(\lambda|\tau_{n-1}<Y\le\tau_n)}\right)\pi(\lambda|Y=\tau_i)d\lambda &= 0 \end{aligned} \quad \ldots (7)$$

(when $n = \infty$ the second line does not apply).

Proof. Using the definition of $K_{\tau_{i+1},\tau_i}(y)$, and equation (5), (4) becomes:

$$\begin{aligned} R(\boldsymbol{\tau}^{(n)}) &= k\sum_{i=0}^{n-1}(i+1)[F(\tau_{i+1}) - F(\tau_i)] + kn[1 - F(\tau_i)] \\ &\quad + c\phi - c\sum_{i=0}^{n}\int_\Omega \log \pi(\lambda|\tau_i < Y \le \tau_{i+1})[G(\tau_{i+1}|\lambda) - G(\tau_i|\lambda)]\pi(\lambda)d\lambda. \end{aligned}$$

Let $1 \le i < n$. Under the conditions of the theorem, integration and differentiation can be interchanged; so:

$$\begin{aligned}
&\tfrac{\partial}{\partial \tau_i} R(\boldsymbol{\tau}^{(n)}) \\
&= k\tfrac{\partial}{\partial \tau_i}\left[(i+1)[F(\tau_{i+1})-F(\tau_i)]+i[F(\tau_i)-F(\tau_{i-1})]\right] \\
&\quad -c\tfrac{\partial}{\partial \tau_i}\int_\Omega \log[\pi(\lambda|\tau_i<Y\le\tau_{i+1})]\,[G(\tau_{i+1}|\lambda)-G(\tau_i|\lambda)]\pi(\lambda)d\lambda \\
&\quad -c\tfrac{\partial}{\partial \tau_i}\int_\Omega [\log\pi(\lambda|\tau_{i-1}<Y\le\tau_i)]\,[G(\tau_i|\lambda)-G(\tau_{i-1}|\lambda)]\pi(\lambda)d\lambda \\
&= -kf(\tau_i)-c\int_\Omega \tfrac{\partial\log\pi(\lambda|\tau_i<Y\le\tau_{i+1})}{\partial\tau_i}[G(\tau_{i+1}|\lambda)-G(\tau_i|\lambda)]\pi(\lambda)d\lambda \\
&\quad +c\int_\Omega \log[\pi(\lambda|\tau_i<Y\le\tau_{i+1})]g(\tau_i|\lambda)\pi(\lambda)d\lambda \\
&\quad -c\int_\Omega \tfrac{\partial\log\pi(\lambda|\tau_{i-1}<Y\le\tau_i)}{\partial\tau_i}[G(\tau_i|\lambda)-G(\tau_{i-1}|\lambda)]\pi(\lambda)d\lambda \\
&\quad -c\int_\Omega \log[\pi(\lambda|\tau_{i-1}<Y\le\tau_i)]g(\tau_i|\lambda)\pi(\lambda)d\lambda.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\frac{\partial\log\pi(\lambda|\tau_i<Y\le\tau_{i+1})}{\partial\tau_i} &= \frac{f(\tau_{i+1})}{F(\tau_{i+1})-F(\tau_i)}-\frac{g(\tau_i|\lambda)}{G(\tau_{i+1}|\lambda)-G(\tau_i|\lambda)} \\
\frac{\partial\log\pi(\lambda|\tau_{i-1}<Y\le\tau_i)}{\partial\tau_i} &= \frac{g(\tau_i|\lambda)}{G(\tau_i|\lambda)-G(\tau_{i-1}|\lambda)}-\frac{f(\tau_i)}{F(\tau_i)-F(\tau_{i-1})}.
\end{aligned}$$

Integrating,

$$\begin{aligned}
\int_\Omega \frac{\partial\log\pi(\lambda|\tau_i<Y\le\tau_{i+1})}{\partial\tau_i}[G(\tau_{i+1}|\lambda)-G(\tau_i|\lambda)]\pi(\lambda)d\lambda &= 0 \\
\int_\Omega \frac{\partial\log\pi(\lambda|\tau_{i-1}<Y\le\tau_i)}{\partial\tau_i}[G(\tau_i|\lambda)-G(\tau_{i-1}|\lambda)]\pi(\lambda)d\lambda &= 0.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{\partial}{\partial\tau_i}R(\boldsymbol{\tau}^{(n)}) &= -kf(\tau_i)+c\int_\Omega g(\tau_i|\lambda)\log\pi(\lambda|\tau_{i-1}<Y\le\tau_i)d\lambda \\
&\quad -c\int_\Omega g(\tau_i|\lambda)\log\pi(\lambda|\tau_i<Y\le\tau_{i+1})d\lambda.
\end{aligned}$$

Setting this to zero gives the desired result. The case $i=n$ is similar.

Theorem 1 provides the starting point for developing efficient solution algorithms. For fixed n, the $i-th$ optimality condition depends only on τ_i and the two neighboring observation times. This determines a recursive relation between the optimal observation times. So long as $K_{\tau_{i+1},\tau_i}(y)$, decreases as a function of τ_{i+1} for given τ_i, which holds in many examples, the optimal continuation is unique, and one can start with a given τ_1, evaluate the optimal continuation, and then compute its risk. In this way we can reduce the problem to a two-dimensional optimization, with respect to τ_1 and n. Furthermore, if τ_n is an increasing function of τ_1, then the bottom equation in (7) provides grounds for iteratively adjusting the initial guess and implementing a successive substitution algorithms for a fast numerical solution. Of use in implementing numerical algorithms is the partial derivative of the left hand side of (6) with respect to τ_{i+1}, which is:

$$\int_\Omega \frac{g(\tau_{i+1}|\lambda)}{G(\tau_{i+1}|\lambda) - G(\tau_i|\lambda)} \pi(\lambda) d\lambda - \frac{f(\tau_{i+1})}{F(\tau_{i+1}) - F(\tau_i)}. \qquad \ldots (8)$$

The recursive nature of the solution also helps interpreting the conditions: for fixed τ_i, the optimality conditions prescribe to choose the value of next observation time τ_{i+1} such that the expected change in the logarithm of the posterior distribution, conditional on a event occurring exactly at τ_i, is equal to the ratio of the costs.

Equations (7) have a parallel in the work of Barlow, Hunter and Proschan (1963) on scheduling inspection times for randomly failing equipment. One important difference is that in their formulation the loss function, which represents the downtime of a system, is not bounded in the size of the interval between inspections. Therefore the optimal n is infinity and the optimization is only over τ_1. Similar situations arise in other reliability and maintenance problems (Parmigiani, 1993).

Interestingly, the loss of information depends only on the observed outcome of the experiment. Criteria that makes use of sampling properties of the design, such as Fisher information, or any expectation taken with respect to Jeffreys' prior, the optimality conditions would not enjoy the fundamental recursive property that makes the solution both interpretable and computationally accessible.

An interesting corollary problem is finding the most informative binary partition for an observation Y. This problem arises in applications when a continuous variable has to be turned into a categorical (in this case binary) variable: we may only observe whether or not it is greater than some point τ. What is the best τ? By setting $n = 1$ in Theorem 1, the optimal solution for τ must satisfy

$$\int_\Omega \log\left(\frac{G(\tau|\lambda)}{1 - G(\tau|\lambda)}\right) \pi(\lambda|Y = \tau) d\lambda = \log\left(\frac{F(\tau)}{1 - F(\tau)}\right). \qquad \ldots (9)$$

That is, the expected log-odds of the likelihood should equal the log-odds of the marginal distribution. Since exactly one observation will be taken regardless of the outcome, the costs k and c do not play a role.

4. Exponential Event Times with Conjugate Prior

An important special case is that of exponential events with unknown event rate λ and conjugate prior distribution, given by

$$\pi(\lambda) \propto \lambda^{a-1} e^{-b\lambda}.$$

In this case the solution is greatly simplified and takes an appealing form. Interestingly, equal spacing of observations is not necessarily optimal, and the solution takes a control-limit form in terms of the posterior predictive probability of an event. The following theorem summarizes the main result.

THEOREM 2. *For exponential events times with a conjugate prior distribution:*

$$\tau_{i+1}^{(n+1)} = \tau_1^{(n+1)} + \frac{b+\tau_1^{(n+1)}}{b}\tau_i^{(n)} \qquad i = 1, \ldots, n-1 \qquad \ldots(10)$$

$$\tau_{i+1}^{(n)} = \tau_1^{(n)} + \frac{b+\tau_1^{(n)}}{b}\tau_i^{(n)} \qquad i = 1, \ldots, n-2 \qquad \ldots(11)$$

PROOF. For fixed n, the optimal policy $\boldsymbol{\tau}^{(n)}$ must satisfy:

$$\int_\Omega \log\left(\frac{e^{-\lambda\tau_i} - e^{-\lambda\tau_{i+1}}}{e^{-\lambda\tau_{i-1}} - e^{-\lambda\tau_i}}\right)\frac{(b+\tau_i)^{a+1}}{\Gamma(a+1)}\lambda^a e^{-\lambda(b+\tau_i)}d\lambda$$

$$-\log\frac{\left(\frac{b}{b+\tau_i}\right)^a - \left(\frac{b}{b+\tau_{i+1}}\right)^a}{\left(\frac{b}{b+\tau_{i-1}}\right)^a - \left(\frac{b}{b+\tau_i}\right)^a} = \frac{k}{c}, \qquad \ldots(12)$$

for $i = 1, 2, ..., n-1$, and

$$\int_\Omega \log\left(\frac{e^{-\lambda\tau_n}}{e^{-\lambda\tau_{n-1}} - e^{-\lambda\tau_n}}\right)\frac{(b+\tau_n)^{a+1}}{\Gamma(a+1)}\lambda^a e^{-\lambda(b+\tau_n)}d\lambda$$

$$-\log\frac{\left(\frac{b}{b+\tau_n}\right)^a}{\left(\frac{b}{b+\tau_{n-1}}\right)^a - \left(\frac{b}{b+\tau_n}\right)^a} = 0. \qquad \ldots(13)$$

Suppose now that the prior parameter b is modified to b'; let the schedule $\boldsymbol{\sigma}^{(n)} = \{\sigma_1^{(n)}, ..., \sigma_n^{(n)}\}$ be the new optimum. Then:

$$\int_\Omega \log\left(\frac{e^{-\lambda\sigma_i} - e^{-\lambda\sigma_{i+1}}}{e^{-\lambda\sigma_{i-1}} - e^{-\lambda\sigma_i}}\right)\frac{(b'+\sigma_i)^{a+1}}{\Gamma(a+1)}\lambda^a e^{-\lambda(b'+\sigma_i)}d\lambda$$

$$-\log \frac{\left(\frac{b'}{b'+\sigma_i}\right)^a - \left(\frac{b'}{b'+\sigma_{i+1}}\right)^a}{\left(\frac{b'}{b'+\sigma_{i-1}}\right)^a - \left(\frac{b'}{b'+\sigma_i}\right)^a} = \frac{k}{c},$$

for $i = 1, 2, ..., n-1$, and

$$\int_\Omega \log\left(\frac{e^{-\lambda\sigma_n}}{e^{-\lambda\sigma_{n-1}} - e^{-\lambda\sigma_n}}\right)\frac{(b'+\sigma_n)^{a+1}}{\Gamma(a+1)}\lambda^a e^{-\lambda(b'+\sigma_n)}d\lambda$$

$$-\log \frac{\left(\frac{b'}{b'+\sigma_n}\right)^a}{\left(\frac{b'}{b'+\sigma_{n-1}}\right)^a - \left(\frac{b'}{b'+\sigma_n}\right)^a} = 0.$$

Transforming $\lambda = \theta\frac{b}{b'}$, the above becomes:

$$\int_\Omega \log\left(\frac{e^{-\theta\frac{b}{b'}\sigma_i} - e^{-\theta\frac{b}{b'}\sigma_{i+1}}}{e^{-\theta\frac{b}{b'}\sigma_{i-1}} - e^{-\theta\frac{b}{b'}\sigma_i}}\right)\frac{(b+\frac{b}{b'}\sigma_i)^{a+1}}{\Gamma(a+1)}e^{-\theta(b-\frac{b}{b'}\sigma_i)}d\theta$$

$$-\log \frac{\left(\frac{b}{b+\frac{b}{b'}\sigma_i}\right)^a - \left(\frac{b}{b+\frac{b}{b'}\sigma_{i+1}}\right)^a}{\left(\frac{b}{b+\frac{b}{b'}\sigma_{i-1}}\right)^a - \left(\frac{b}{b+\frac{b}{b'}\sigma_i}\right)^a} = \frac{k}{c}$$

For $i = 1, 2, ..., n-1$, and

$$\int_\Omega \log\left(\frac{e^{-\theta\frac{b}{b'}\sigma_n}}{e^{-\theta\frac{b}{b'}\sigma_{n-1}} - e^{-\theta\frac{b}{b'}\sigma_n}}\right)\frac{(b+\frac{b}{b'}\sigma_n)^{a+1}}{\Gamma(a+1)}e^{-\theta(b-\frac{b}{b'}\sigma_n)}d\theta$$

$$-\log \frac{\left(\frac{b}{b+\frac{b}{b'}\sigma_n}\right)^a}{\left(\frac{b}{b+\frac{b}{b'}\sigma_{n-1}}\right)^a - \left(\frac{b}{b+\frac{b}{b'}\sigma_n}\right)^a} = 0$$

Comparing these with (10) and (11) it is clear that the new optimal schedule satisfies: $\sigma_i = \frac{b'}{b}\tau_i$. In the conjugate prior distributions case, the posterior distribution for Λ after observing a survival until y is Gamma with parameters a and $b+y$. Hence, if $b' = b + \tau_1^{(n+1)}$, $\boldsymbol{\sigma}^{(n)}$ can be interpreted as the optimal continuation $\tau_i^{(n+1)}$, $i = 2, \ldots, n+1$, after an observation at $\tau_1^{(n+1)}$ revealed no event. Thus (10) follows. To prove (11) consider the optimality conditions for $\boldsymbol{\sigma}^{(n-1)}$ after changing b to b', and interpret $\boldsymbol{\sigma}^{(n-1)}$ as the optimal continuation $\tau_i^{(n)}$ $i = 2, \ldots, n$, after $\tau_1^{(n)}$ revealed no event. The same argument applies to all but the last optimality equation.

This theorem has relevant practical consequences. First, when $\boldsymbol{\tau}^{(n)}$ is known, $\boldsymbol{\tau}^{(n+1)}$ can be derived by the numerical solution of one equation instead of $n+1$. Second, the following holds.

Corollary 1. *For fixed n, the optimal policy $\boldsymbol{\tau}^{(n)}$ does not depend on k/c.*

Proof. From (10), the optimal $\tau_1^{(1)}$ is independent of k/c. Moreover, from (10):

$$\tau_2^{(2)} = \tau_1^{(2)} + \frac{b + \tau_1^{(2)}}{b}\tau_1^{(1)}.$$

Using the above, the optimality equation for $n = 2$, can be rewritten as a function of $\tau_1^{(2)}$ only. This function is then independent of k/c, so that $\boldsymbol{\tau}^{(2)}$ can be calculated independently of k/c. Iterating this argument gives the desired conclusion for an arbitrary n.

Thus, changes in the costs do not affect the observation schedule directly, but only through the choice of the optimal n. This makes the solution more robust and simpler to compute and interpret.

Further interesting features apply to the case $n = \infty$. Then the optimal policy is characterized by the property that it is best to inspect when the probability of event reaches a fixed threshold.

Corollary 2. *When $n = \infty$, the optimal policy satisfies:*

$$\frac{F(\tau_{i+1}^{(\infty)}) - F(\tau_i^{(\infty)})}{1 - F(\tau_i^{(\infty)})} = F(\tau_1^{(\infty)}) \qquad i \geq 1 \qquad \ldots (14)$$

Proof. If $n = \infty$, (11) becomes:

$$\tau_{i+1} = \tau_1 + \frac{b + \tau_1}{b}\tau_i \qquad i \geq 1. \qquad \ldots (15)$$

Now,

$$\frac{F(\tau_{i+1}) - F(\tau_i)}{1 - F(\tau_i)} = 1 - \left(\frac{b + \tau_i}{b + \tau_{i+1}}\right)^a.$$

Iteratively using (11),

$$\frac{F(\tau_{i+1}) - F(\tau_i)}{1 - F(\tau_i)} = 1 - \left(\frac{b + \tau_{i-1}}{b + \tau_i}\right)^a = 1 - \left(\frac{b}{b + \tau_1}\right)^a,$$

as required.

Finally, in the case $n = \infty$, the observation intervals increase geometrically for every i, as a direct consequence of (15).

Returning to the problem of finding the binary partition that is most informative about λ the optimal τ is the unique solution of:

$$b^{a+1}\zeta(a+2)\frac{e^{1+b/\tau}}{\tau^{a+1}} = \log\left[\left(\frac{b + \tau}{b}\right)^a - 1\right]$$

where ζ is the Riemann zeta function. Because the left hand side is always positive, the optimal τ is greater that the marginal median of Y.

5. Squared Error Loss

If the parameter space is one dimensional, and the main objective is the solution of a point estimation problem for $E(\Lambda)$ a popular objective function is based on the squared error estimation loss. This gives:

$$\begin{aligned} R(\boldsymbol{\tau}^{(n)}) &= k\sum_{i=0}^{n-1}(i+1)[F(\tau_{i+1})-F(\tau_i)]+kn[1-F(\tau_i)] \\ &\quad +c\sum_{i=0}^{n}\int_\Omega[\lambda-E\{\Lambda|\tau_i<Y\leq\tau_{i+1})]^2[G(\tau_{i+1}|\lambda)-G(\tau_i|\lambda)]\pi(\lambda)d\lambda. \end{aligned} \quad \ldots(16)$$

Similarly to Section 3, the following holds.

THEOREM 3. *Let $g(y|\lambda)$ be strictly positive, continuous and bounded in y for every $\lambda\in\Omega$. Then the optimal policy $\boldsymbol{\tau}^{(n)}$ must satisfy:*

$$\int_\Omega\{[\lambda-E\{\Lambda|\tau_{i-1}<Y\leq\tau_i)]^2-[\lambda-E\{\Lambda|\tau_i<Y\leq\tau_{i+1})]^2\}\pi(\lambda|Y=\tau_i)d\lambda=\frac{k}{c}$$

$$\int_\Omega\{[\lambda-E\{\Lambda|\tau_{n-1}<Y\leq\tau_n)]^2-[\lambda-E\{\Lambda|Y>\tau_n)]^2\}\pi(\lambda|Y=\tau_i)d\lambda=0.$$

for $i=1,2,....n-1$

PROOF. The proof is similar to that of Theorem (1). The only relevant modification is:

$$\begin{aligned} &\frac{\partial[\lambda-E\{\Lambda|\tau_i<Y\leq\tau_{i+1})]^2}{\partial\tau_i} \\ &=-2[\lambda-E\{\Lambda|\tau_i<Y\leq\tau_{i+1}\}]\int_\Omega\lambda\frac{\partial\pi(\Lambda|\tau_i<Y\leq\tau_{i+1})}{\partial\tau_i}d\lambda \\ &=-2[\lambda-E\{\Lambda|\tau_i<Y\leq\tau_{i+1}\}]\phi(\tau_i,\tau_{i+1}) \end{aligned}$$

where $\phi(\tau_i,\tau_{i+1})$ is constant in λ. Hence, integrating, the above term vanishes.

6. Discussion

In this paper we focus on censored duration data in which knowledge of each duration is limited to the interval containing it. We considered the problem of choosing the observation times to minimize the loss of information, and we developed optimal solutions for key pilot problems based on recursive dynamic techniques.

The results provide building blocks for solving more complex and interesting problems. First, one can consider observing a batch of exchangeable units rather than a single one. At each observation time, the number of failed units is observed, and the next interval is chosen on the basis of the result. Formula (9) can be easily obtained in the $n = 1$ case, providing the basis to derive sufficient conditions for stopping observation of the batch. The problem is related the monotone sequential design of rodent bioassay experiments considered by Louis (1989) from the point of view of asymptotic efficiency, and also to the work of Pötzelberger and Felsenstein (1993) addressing the asymptotic loss of information from grouping observations.

Another interesting case arises when several exchangeable units are observed consecutively. To illustrate, consider two units with event times Y and U; in analogy with the results of this paper, schedule $\boldsymbol{\tau}^{(n)}$ is followed for the first unit, and schedule $\boldsymbol{\zeta}^{(m)}$ is followed for the second unit. After both units are followed and events are observed at the $i+1$ and $j+1$ observation, the posterior distribution for Λ is proportional to $[G(\tau_{i+1}^{(n)}|\lambda) - G(\tau_i^{(n)}|\lambda)][G(\zeta_{j+1}^{(m)}|\lambda) - G(\zeta_j^{(m)}|\lambda)]\pi(\lambda)$. A solution can be derived by backwards induction on units: after the first unit is observed to have failed, the problem is identical to that of this paper, with the prior distribution replaced by the posterior distribution given the outcome of the first testing.

Acknowledgement. Helpful discussions with Don Berry, Jay Kadane and Nick Polson are gratefully acknowledged.

References

Barlow, R. E., Hunter, L. C., and Proschan, F. (1963), Optimum checking procedures. *Jour. Society for Industrial and Applied Mathematics*, **4**, 1078–1095.

Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors (1992).,*Bayesian Statistics 4*. Oxford University Press, Oxford.

Chaloner, K. and Larntz, K. (1989), Optimal Bayesian design applied to logistic regression experiments. *Jour. Statistical Planning and Inference*, **21**, 191–208.

Chaloner, K. and Verdinelli, I. (1995), Bayesian experimental design: A review. *Statistical Science*, **10**, 273–304.

Clyde, M. A. and Chaloner, K. (1996). The equivalence of constrained and weighted designs in multiple objective design problems. *Jour. Amer. Statist. Assoc.*, **91**, 1236–1244.

Hatzis, C. and Larntz, K. (1992), Optimal design in nonlinear multiresponse estimation: Poisson model for filter feeding. *Biometrics*, **48**, 1235–1248.

Lindley, D. V. (1956), On a measure of the information provided by an experiment. *Ann. Math. Statist.*, **27**, 986–1005.

Louis, T. A. (1989), Efficient monotone sequential design. *Report* MS-R8705, Department of Mathematical Statistics, C.W.I., Amsterdam.

Parmigiani, G. (1993), Scheduling inspections in reliability. In Basu, A. P., editor, *Advances in Reliability*, pages 303–319. Elsevier/North-Holland.

Parmigiani, G. and Berry, D. A. (1994), Applications of Lindley information measure to the design of clinical experiments. In *Aspects of Uncertainty. A Tribute to D. V. Lindley*, pages 351–362. John Wiley & Sons.

Parmigiani, G. and Polson, N. G. (1992), Bayesian design for random walk barriers. In (Bernardo *et al.*, 1992), pages 715–721.

Verdinelli, I. (1992), Advances in Bayesian experimental designs. In (Bernardo *et al.*, 1992), pages 467–482.

Verdinelli, I. and Kadane, J. B. (1992), Bayesian designs for maximizing information and outcome. *Jour. Amer. Statist. Assoc.*, **87**, 510–515.

Verdinelli, I., Polson, N. G., and Singpurwalla, N. (1992), Shannon information and Bayesian design for prediction in accelerated life testing. In Barlow, R. E. and C, C., editors, *Reliability and Decision Making.* Elsevier.

Giovanni Parmigiani
Institute of Statistics and Decision Sciences
Duke University
Box 90251
Durham, North Carolina, USA
e-mail : gp@stat.duke.edu