Statistica Sinica **11**(2001), 1-21

BOOTSTRAP METHODS FOR THE UP AND DOWN TEST ON PYROTECHNICS SENSITIVITY ANALYSIS

M. T. Chao and C. D. Fuh

Academia Sinica

Abstract: We start with a data set recently obtained from a Bruceton test. The data come from the study of CS-M-3 ignitor in a military experiment and are analyzed by the up-and-down method of Dixon and Mood (1948). We reexamine the method and develop a more appropriate inference that takes account of the special dependent data structure. Two bootstrap confidence interval procedures, percentile and bootstrap-t, are introduced to find approximate confidence intervals for the parameters of interest. A simulation study shows that the bootstrap-t, with proper bias corrections, gives better coverage probability, but is considerably more computer-intensive than non-bias-corrected versions. This leads to the development of an importance resampling technique which can reduce the CPU time by a factor of 10 or more. Finally, we apply the proposed procedure to analyze our data set.

Key words and phrases: Bootstrap, importance resampling, Markov chains, maximum likelihood estimate, probit model, sequential design, up and down method.

1. Introduction

Table 1.1 summarizes the testing of the CS-M-3 ignitor in an experiment carried out by the Chung-Shan Institute of Science and Technology in Taiwan. One wishes to estimate the probability of reliable and safe functioning of components. Here, the problem is to determine the reliability and safety of fuse explosive trains. The test is a destructive one, 1 denotes explosion and 0 denotes non-explosion. The x-axis is time and the y-axis denotes the stimulus level in the log domain of the dosage, with volt as unit. There are 43 items in the test. The experiment was performed at initial stimulus level .874, with step size $\Delta = .01$ through a total of six levels. This is called a Bruceton test when the data is analyzed by the classical up-and-down method of Dixon and Mood (1948). (Bruceton is the name of a military installation in Pennsylvania, see Mood (1998).) For parametric inference on binary data with various stimulus levels, this method is still actively used in Pyrotechnics Sensitivity Analysis (PSA), and is documented in certain military standards (e.g., MIL-STD-331B, MIL-STD-322B).

Table 1.1. Experiment of CS-M-3 ignitor by the Bruceton test.

.884	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.874	1	-	1	-	-	-	-	-	-	-	-	-	-	-	0	-	1	-	0	-	1	-	1	-	-	-	-	-	1	-	-	-	1	-	1	-	-	-	1	-	-	-	1
.864	-	0	-	1	-	-	-	-	-	-	-	-	-	0	-	-	-	0	-	-	-	0	-	1	-	-	-	0	-	1	-	0	-	0	-	1	-	0	-	1	-	0	-
.854	-	-	-	-	1	-	1	-	1	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	1	-	0	-	-	-	0	-	-	-	-	-	0	-	-	-	0	-	-
.844	-	-	-	-	-	0	-	0	-	1	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.834	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

A Bruceton run can be described as follows. Start from an initial stimulus level and assume that increasing the stimulus level will increase the explosion probability. When the response is 0 (non-explosive), we increase the stimulus level by one unit at the next stage; if the response is 1 (explosive), we decrease the stimulus level by one unit at the next stage. The process continues until a certain fixed number of observations is obtained.

An implicit assumption underlying this quantal response model is that there is an unobservable random variable X with distribution F which represents the critical value of this item, in the sense that the response at level x, $Y(x)$, takes the value 1 if and only if $X \leq x$. Here $Y(x) = 1$ or 0 according as the test item is exploded or not. Although F, the distribution function of the critical stimulus levels, could be any distribution function, normally a parametric family is assumed. In particular, two distribution forms have been commonly used, the probit and the logit, perhaps after a transformation of x. For the probit model, $F(x)$ is modeled as an integrated normal distribution function

$$F(x) = P\{Y(x) = 1\} = \int_{-\infty}^{(x-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt, \tag{1.1}$$

where μ and σ are two unknown parameters. For the logit model, $F(x)$ is the logistic distribution

$$F(x) = P\{Y(x) = 1\} = \frac{1}{1 + \exp\{-(x-\mu)/\sigma\}}, \tag{1.2}$$

where μ and σ also denote two unknown parameters. Given the data set, the main objective is to estimate x_p, the pth quantile of F, for p close to unity.

A salient feature of the Bruceton test is that observations are concentrated around stimulus levels that produce the median value of F, and this holds true irrespective of the choice of the initial stimulus level and without the prior knowledge of the value of the median. When the data in Table 1.1 are observed from the y-axis, one can see more data points around the median level from a histogram-like picture. This is an important feature when the value of the location parameter is of primary interest. However, for the data set in Table 1.1, the objective is to estimate the high quantiles of F. This requires a parametric model. We assume the probit model and estimate the unknown parameters μ and σ^2, and the quantile x_p via extrapolation, i.e., by using the corresponding parametric form of the underlying distribution.

Under the probit, logit and double exponential distributions, approximate D-, A- and c-optimal designs can be found in Sitter and Wu (1993) and Chao and Fuh (1999). The simulation results in Chao and Fuh (1999) show that in most cases, the up-and-down test exhibits an efficiency level of 70% or more compared to the best D- or A-optimal design, and this is achieved *without* prior knowledge of μ and σ^2. With respect to estimation, the up-and-down design is an efficient means of data collection.

Here we make no attempt to improve the data collection process, since the data have already been collected. Instead, we take the data as given and proceed to make the best use of it. A major drawback of the classical analysis is that the normal approximation intervals used in the Bruceton test may not have coverage probability close to the nominal values in case of a small to moderate number of failures. This is easily seen in a small sample simulation study, see Tables 3.1 and 3.3. We investigate the possibility of using computer-intensive data analysis methods, such as bootstrap methods, to analyze the data set. It is found that the bias corrected bootstrap-t, the most computer-intensive of all, gives the most reliable nominal values. This motivated our search for a better simulation procedure. Since the event corresponding to the nominal coverage is a moderate tail event, it is natural to look into large deviation based techniques for the importance resampling. That operates extra difficulties here, the data obtained from an up-and-down method follow a Markov chain. To cope with this, a tilting formula based on the Poisson equation is developed for the actual simulation. In Section 2, we explore the specific Markovian structure of the data, and then propose a bootstrap algorithm for the Bruceton test. Simulation studies for the comparison of the various bootstrap confidence intervals are reported in Section 3. Our main contribution is reported in Section 4, where we propose an importance resampling method to facilitate variance reduction of the Monte-Carlo simulation. In Section 5, we apply the proposed statistical procedure to analyze the data set in Table 1.1. Conclusion and further remarks are in Section 6. Theoretical issues related to the bootstrap method and importance resampling are included in the Appendix.

2. Markov Chain Representation and Bootstrap Confidence Intervals

To estimate μ, σ and x_p in the probit model (1.1) using the up-and-down method, we first represent the data described in Table 1.1 as a Markov chain.

The idea of Markov chain representation to study non-parametric estimation in the up-and-down method originated with Derman (1957). For further work along this line, see Wetherill (1963) and Wetherill and Glazebrook (1986) and the references therein. Recently, Thomas (1994) used a Markov chain representation with a slightly different model to investigate the ignition sensitivity of thermal-battery heat pellets. The utility of the Markov chain representation can be described as follows. For the up-and-down method, by the intrinsic relation between the stimulus level X and the response variable Y, we can reproduce the data set $\{(X_t, Y_t), t = 1, \ldots, n-1, X_n\}$ from the ordered set of stimulus levels $\{X_t, t = 1, \ldots, n\}$, and the latter can be formulated as a Markov chain with the specific transition probability matrix P described in (2.1) below. Therefore, in the analysis that follows, we can ignore the values of the Y's and concentrate on the X's; the classical method of maximum likelihood estimate in parametric Markov chains can be applied to make point estimation. The statistical inference for ergodic Markov chains is well documented in Anderson and Goodman (1957), Billingsley (1961), and Basawa and Prakasa Rao (1980).

Now consider the data set $(x, y) = \{(x_0, y_0), \ldots, (x_n, y_n)\}$ produced by the up-and-down method, where x_t is the stimulus level for the tth component and y_t is the response value for the tth component. Recall that y_t is 0 or 1, representing "non-explosive" or "explosive" respectively. For ease of exposition, suppose there are five stimulus levels $a_0, \ldots, a_4$ with $a_i < a_j$ for $i < j$. Let m_j be the number of explosive units, and n_j be the number of non-explosive units at stimulus level j. The likelihood function is

$$L_n((\mu, \sigma^2)) = L_n((\mu, \sigma^2)|(x, y)) = K \prod_{j=0}^{4} p_j^{n_j} (1 - p_j)^{m_j},$$

where $p_j = \Phi((a_j - \mu)/\sigma)$ for the probit model (1.1), $p_j = (1 + \exp\{(a_j - \mu)/\sigma\})^{-1}$ for the logit model (1.2), and where K is some constant.

For given x values, the conditional likelihood is simply a product but the sampling distribution of the maximum likelihood estimates of μ and σ^2 depends on x (see (5.2), (5.5) and (A.7) below), and the $x's$ are not independent due to the up-and-down design.

Now, it is easy to see that $\{X_t, t = 1, \ldots, n\}$ forms a Markov chain on a state space $\{a_0, \ldots, a_4\}$ with transition probability matrix

$$P = \begin{pmatrix} p_0 & 1-p_0 & 0 & 0 & 0 \\ p_1 & 0 & 1-p_1 & 0 & 0 \\ 0 & p_2 & 0 & 1-p_2 & 0 \\ 0 & 0 & p_3 & 0 & 1-p_3 \\ 0 & 0 & 0 & p_4 & 1-p_4 \end{pmatrix}. \tag{2.1}$$

In practice, only 4 to 7 different levels are used in testing, see Section 5 for details.

Based on an approximation of the likelihood function, Dixon and Mood ((1948), Equations (1) and (2)) derived approximate maximum likelihood estimates of μ, σ and x_p, for the probit model (1.1). Hampton, Blum and Ayres ((1973), Equations (23) and (24)) derived approximate maximum likelihood estimates for the logit model (1.2). Consistency and asymptotic normality of the estimates follows from standard theory (see Billingsley (1961)).

Although the asymptotic normality of MLE's can be used to find approximate confidence intervals of μ, σ and x_p, coverage probabilities are often far from their nominal values. In addition, difficulty in computing the asymptotic variance-covariance matrix makes this approach less suitable for application. The bootstrap method is an alternative, and perhaps better, technique for this type of problem. It is known that in the i.i.d. setting, bootstrap confidence intervals are not only asymptotically more *accurate* than the classical normal approximation intervals, they are also more *correct* (see Efron and Tibshirani (1993)). In the case of ergodic finite state parametric Markov chains, these properties are expected to be valid as well (see Fuh and Lai (1998)).

The bootstrap (Efron (1979)) was designed for i.i.d. settings to evaluate the sampling distribution of an estimate. In particular, it can be used to estimate the bias and variance of an estimate, and to produce confidence intervals. The application of bootstrap methods to Markov chains is discussed by Kulperger and Prakasa Rao (1990), Athreya and Fuh (1992), Datta and McCormick (1993) and Fuh (1993). In these papers, asymptotic properties have been verified for various bootstrap methods in the nonparametric case. Little is known, however, about the parametric case except for the i.i.d. setup. Here we propose a bootstrap algorithm for parametric Markov chains, and introduce two procedures (percentile and bootstrap-t) to give approximate confidence intervals. The parametric bootstrap-t is second-order accurate. The percentile procedure is very easy to implement but is only first-order accurate. Theoretical studies of the bootstrap methods is given in Appendix 1.

Bootstrap algorithms:

Let $\mathbf{x} = \{x_0, \ldots, x_n\}$ be a realization of the Markov chain $\{X_t;\ t \geq 0\}$ with transition probability $P = (p_{ij}(\theta))$, where $\theta = (\theta_1, \ldots, \theta_p)$ is a vector of unknown parameters. Let $\hat{\theta}$ be the maximum likelihood estimate (MLE) of θ. The bootstrap algorithm to approximate the sampling distribution H_n of $R(\mathbf{x}, \theta) := \sqrt{n}(\hat{\theta} - \theta)$ is as follows.

(1) With $\hat{P}_n =: P(\hat{\theta})$ as its transition probability, generate a Markov chain realization of n steps $\mathbf{x}^* = \{x_0^*, \ldots, x_n^*\}$. Call this a bootstrap sample, and let $\hat{\theta}^*$ be the MLE of θ based on $\mathbf{x}^*$.

(2) Approximate the sampling distribution H_n of $R(\mathbf{x}, \theta)$ by the conditional distribution H_n^* of $R(\mathbf{x}^*, \hat{\theta}_n) = \sqrt{n}(\hat{\theta}^* - \hat{\theta})$ given $\mathbf{x}$, which can be done by Monte-Carlo simulation .

Later, we compare two procedures (percentile and bootstrap-t, with and without bias correction) to give approximate confidence intervals for the parameters θ ($= \mu, \sigma^2$, or x_p in our case) of interest. For the percentile bootstrap confidence interval, we repeatedly generate bootstrap samples $\mathbf{x}^*$ according to the above bootstrap algorithm, and replications $\hat{\theta}^* = \theta(\mathbf{x}^*)$ are computed. Let $\hat{G}$ be the cumulative distribution function of $\hat{\theta}^*$. The $1 - 2\alpha$ *percentile interval* is defined by the α and $1 - \alpha$ percentiles of $\hat{G}$:

$$[\hat{\theta}_l, \hat{\theta}_u] = [\hat{G}^{-1}(\alpha), \hat{G}^{-1}(1-\alpha)] = [\hat{\theta}^{*(\alpha)}, \hat{\theta}^{*(1-\alpha)}].$$

The bootstrap-t estimates the percentiles of a studentized statistic $T = \sqrt{n}(\hat{\theta} - \theta)/\hat{\sigma}$ by bootstrapping, where $\hat{\sigma}^2$ is the variance estimator. Each $\mathbf{x}^*$ gives a pair $(\hat{\theta}^*, \hat{\sigma}^*)$, yielding $T^* = \sqrt{n}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^*$, a bootstrap replication.

In the Bruceton test, data were originally designed to estimate the location accurately, and the approximate MLE is a complicated function of the sample averages (see (5.5)). That makes the variance estimator $\hat{\sigma}^2$ systematically biased downward (see Hampton, Blum and Aryes (1973)), and hence a modified technique is required for more accurate interval estimation. A general approach in the bootstrap literature is to apply nested levels of bootstrap sampling. In principle, a nested bootstrap might involve more than two levels, but in practice the computational burden would ordinarily be too large for more than two levels to be worthwhile, and we use with two. In the first level, a resample size B_1 is used to compute percentiles; resample size B_2 is needed to estimate the bias of standard error. Hence the overall number of bootstrap samples is $B = B_1 \times B_2$, a formidable number needed for interval estimation. This idea leads us to study a bias corrected bootstrap-t algorithm in Section 3 and an importance resampling technique in Section 4 for efficient Monte-Carlo simulation. General treatment for the nested bootstrap algorithm can be found in Chapters 5 and 12 of Efron and Tibshirani (1993) and Sections 3.9, 4.5 and 5.6 of Davison and Hinkley (1997).

3. A Simulation Study

3.1. Design of the simulations

A small sample comparison for bootstrapping confidence intervals of the up-and-down test is presented in this section. We compare all the approximate confidence intervals (normal approximation, percentile and bootstrap-t, with and without bias-correction) for the parameters μ, σ and x_p, in the probit model

(1.1). The performance of this comparison is based on coverage probability and average length of the corresponding confidence intervals. The nominal coverage probability for the confidence intervals is .95 in all cases. The results are shown in Tables 3.1 to 3.4, in which three factors are used:

(1) Sample sizes of $n = 50$ and 100.

(2) Step sizes of $\Delta = .8\sigma,\ 1.0\sigma,\ 1.2\sigma$ and 1.5σ.

(3) With and without bias correction (to be discussed below) for the standard deviation estimator.

For this simulation study, we consider (1.1) with $\mu = 0,\ \sigma = .02$ and $x_{.999} = .0168$. The original sample is simulated from an ergodic Markov chain with transition probability matrix

$$P = \begin{pmatrix} .00003 & .99997 & .00000 & .00000 & .00000 & .00000 \\ .00621 & .00000 & .99379 & .00000 & .00000 & .00000 \\ .00000 & .15866 & .00000 & .84134 & .00000 & .00000 \\ .00000 & .00000 & .69146 & .00000 & .30854 & .00000 \\ .00000 & .00000 & .00000 & .97725 & .00000 & .02275 \\ .00000 & .00000 & .00000 & .00000 & .99977 & .00023 \end{pmatrix},$$

and stationary distribution $\pi = (.211,\ .241,\ .217,\ .186,\ .146)^t$, where t denotes transpose. The bootstrap replication size for the ordinal bootstrap confidence intervals is $B = 2000$. The true 95% interquantile range $(t_{.025}, t_{.975})$ is given for reference, endpoints were obtained (with step size $\Delta = 1.2\sigma$) from the appropriate quantiles of the empirical distributions based on a large simulation (180,000 replications for each case). Computations were performed using FORTRAN-77 programs on the IBM workstation 397 of the Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, ROC. The pseudo-random numbers were generated by using IMSL routines. All tests were compared on the basis of the same random numbers, samples of different size were nested.

In the bootstrap-t confidence interval, we need a variance estimator for each parameter. An approximate variance estimator formula for $\hat{\mu}$ and $\hat{\sigma}$ is discussed on page 121 and shown in Figure 2 of Dixon and Mood (1948). Here we use the approximate formula in the Fortran program provided by McMains (1984).

From a pilot Monte-Carlo investigation, it has been shown that the MLE's in the Bruceton test, based upon the assumption of either a probit model or a logit model, gives a biased estimate of the standard deviation. For the logit model, a simulation study by Hampton, Blum and Aryes (1973) showed that the bias decreases as the sample size increases and the bias is greater as the product of the standard deviation and the step size $\sigma\Delta$ increases, that is, as the step becomes small with respect to the value of $1/\sigma$. A bias correction used in a standard Bruceton test (see MIL-STD-1512 and MIL-STD-1576) is to multiply the variance estimator $\hat{\sigma}$ by 1.059. In this study, based on the idea of a nested bootstrap, we propose a data driven bias-corrected interval estimate for $\hat{\sigma}$ in (5.5) as follows.

Bias-corrected bootstrap confidence interval for $\hat{\sigma}$

Let $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$, where $\hat{\mu}$ is given in (5.3).

(1) With $\hat{P} =: P(\hat{\theta})$ as its transition probability, generate the b_1th original bootstrap $\mathbf{x}^* = \{x_0^*, \ldots, x_n^*\}$, and let $\hat{\theta}^*$ be the MLE of θ based on $\mathbf{x}^*$, $b_1 = 1, \ldots, B_1$.

(2) With $\tilde{P} =: P(\hat{\theta}^*)$ as its transition probability, generate the B_2 second-level bootstrap $\mathbf{x}^{**} = \{x_0^{**}, \ldots, x_n^{**}\}$. Let $\widehat{bias}_{B_2} = \hat{\sigma}^*/\hat{\sigma}^{**}$ be the bootstrap bias estimator of the bias of $\hat{\sigma}$; define $\bar{\sigma}^* = \hat{\sigma}^* \widehat{bias}_{B_2}$.

(3) The bias-corrected bootstrap is obtained from B_1 bootstrap replications of $\sqrt{n}(\hat{\sigma}^* - \hat{\theta})/\bar{\sigma}^*$.

The following notation is used in Tables 3.1-3.4 below: CI_a – confidence interval of a; CP – coverage probability; n – sample size; AL – average length; NA – normal approximation; Δ – step size; P – percentile method; Bt – Bootstrap-t method; T–True.

3.2. Simulation results

Note from Tables 3.1-3.4 that when the sample size is 50, the coverage probabilities of the bootstrap-t confidence interval for the location parameter μ is pretty convincing for all step sizes from $.8\sigma$ to 1.5σ, although the average length is a bit wider than that of normal approximation. The coverage probabilities for bootstrap-t confidence interval of σ (with bias correction) are around .930 for $\Delta = 1.2\sigma$, better than the classical normal approximation. The same holds for the bootstrap-t confidence interval for $x_{.999}$. We also observe that the performance of bootstrapping confidence interval (without bias correction) for σ and $x_{.999}$ is worse than that of the bias-corrected (multiply by 1.059) normal approximation.

When the sample size n is 100, the Bruceton test provides more reliable confidence intervals, especially for estimating σ and the .999th quantiles. In the case $\Delta = 1.2\sigma$, the coverage probability of normal approximation goes to .919 for σ and .920 for $x_{.999}$, while the coverage probability of bootstrap-t (with bias correction) goes to .939 for σ and .934 for $x_{.999}$. Thus, the bootstrap-t again provides a better interval estimation for the .999th quantile. This is consistent with the general belief that the bootstrap-t confidence intervals is second-order accurate.

In general, coverage probabilities for the bootstrap methods with bias correction are better than those of the normal approximation. Also, the bootstrap-t slightly outperforms the percentile method. For estimating the high percentile x_p, we need to have n larger than 100 and, at the same time, to use the bias corrected bootstrap-t method.

Table 3.1. Comparison of approximate confidence intervals for $n = 50$. $\mathrm{CI}_\mu = (-.0072, .0080)$, $\mathrm{CI}_\sigma = (.0107, .0308)$ and $\mathrm{CI}_{x_{.999}} = (.032, .0972)$.

	μ		σ		$x_{.999}$	
	CP	AL	CP	AL	CP	AL
T ($\Delta = .8\sigma$)		.0152		.0201		.0650
NA	.9126	.0145	.8705	.0219	.8748	.0691
P	.9153	.0252	.8726	.0160	.8762	.0491
Bt	.9518	.0365	.8388	.0293	.8003	.0817
T ($\Delta = 1.0\sigma$)		.0152		.0201		.0650
NA	.9167	.0151	.8810	.0207	.8900	.0658
P	.9411	.0261	.7619	.0159	.8225	.0496
Bt	.9610	.0359	.8471	.0258	.8190	.0726
T ($\Delta = 1.2\sigma$)		.0152		.0201		.0650
NA	.9220	.0156	.8958	.0200	.8860	.0637
P	.9400	.0232	.8215	.0157	.8388	.0496
Bt	.9757	.0303	.8796	.0246	.8281	.0713
T ($\Delta = 1.5\sigma$)		.0152		.0201		.0650
NA	.9238	.0164	.8825	.0193	.9033	.0618
P	.9312	.0248	.8878	.0165	.8870	.0531
Bt	.9674	.0326	.8739	.0261	.8822	.0791

Table 3.2. Comparison of approximate confidence intervals for $n = 50$ with bias correction.

	μ		σ		$x_{.999}$	
	CP	AL	CP	AL	CP	AL
P ($\Delta = .8\sigma$)	.9216	.0253	.8752	.0175	.8267	.0533
Bt	.9548	.0250	.9158	.0169	.9142	.0566
P ($\Delta = 1.0\sigma$)	.9516	.0264	.8235	.0170	.8522	.0526
Bt	.9516	.0245	.9117	.0135	.9124	.0377
P ($\Delta = 1.2\sigma$)	.9536	.0234	.8693	.0166	.8851	.0522
Bt	.9684	.0290	.9293	.0221	.9249	.0456
P ($\Delta = 1.5\sigma$)	.9334	.0244	.8993	.0171	.9241	.0547
Bt	.9665	.0297	.9130	.0227	.9172	.0707

Table 3.3. Comparison of approximate confidence intervals for $n = 100$. $\mathrm{CI}_\mu = (-.0054, .0056)$, $\mathrm{CI}_\sigma = (.0129, .0278)$ and $\mathrm{CI}_{x_{.999}} = (.0399, .0862)$.

	μ		σ		$x_{.999}$	
	CP	AL	CP	AL	CP	AL
T ($\Delta = .8\sigma$)		.0110		.0149		.0463
NA	.9279	.0106	.8982	.0159	.9079	.0503
P	.9436	.0177	.7738	.0126	.7950	.0387
Bt	.9788	.0226	.8556	.0189	.8416	.0546
T ($\Delta = 1.0\sigma$)		.0110		.0149		.0463
NA	.9368	.0109	.9111	.0150	.9223	.0476
P	.9434	.0248	.8421	.0125	.8919	.0401
Bt	.9602	.0315	.8705	.0167	.8501	.0506
T ($\Delta = 1.2\sigma$)		.0110		.0149		.0463
NA	.9367	.0112	.9190	.0144	.9205	.0460
P	.9588	.0232	.8471	.0120	.8751	.0403
Bt	.9448	.0280	.8937	.0156	.8885	.0491
T ($\Delta = 1.5\sigma$)		.0110		.0149		.0463
NA	.9396	.0117	.9295	.0138	.9251	.0444
P	.9387	.0152	.9058	.0122	.8719	.0401
Bt	.9438	.0175	.8906	.0159	.8928	.0501

Table 3.4. Comparison of approximate confidence intervals for $n = 100$ with bias correction.

	μ		σ		$x_{.999}$	
	CP	AL	CP	AL	CP	AL
P ($\Delta = .8\sigma$)	.9521	.0170	.8564	.0135	.8603	.0413
Bt	.9509	.0205	.9302	.0176	.9388	.0515
P ($\Delta = 1.0\sigma$)	.9430	.0244	.8851	.0132	.9176	.0416
Bt	.9583	.0295	.9444	.0157	.9334	.0481
P ($\Delta = 1.2\sigma$)	.9668	.0235	.8973	.0126	.9180	.0415
Bt	.9450	.0276	.9391	.0145	.9337	.0464
P ($\Delta = 1.5\sigma$)	.9516	.0156	.9430	.0126	.9147	.0411
Bt	.9576	.0171	.9419	.0144	.9329	.0457

4. Importance Resampling

In Section 2, we proposed a bootstrap algorithm for parametric Markov chains. There the bootstrap estimate of the expectation of a statistic of interest $s = s(\mathbf{x})$ is

$$\hat{e} = E_{\hat{P}_n} s(\mathbf{x}^*). \tag{4.1}$$

Typically, it is not easy to compute $\hat{e}$ analytically. Instead, we approximate it by a Monte-Carlo approximation

$$\hat{e}_B = \frac{1}{B}\sum_{b=1}^{B} s(\mathbf{x}^{*b}), \tag{4.2}$$

where each $\mathbf{x}^{*b}$ is a bootstrap sample of size n produced by $\hat{P}_n$. Note that $\hat{e}_B \to \hat{e}$ as $B \to \infty$ according to the Law of Large Numbers; furthermore $E(\hat{e}_B) = \hat{e}$ and $Var(\hat{e}_B) = c/B$ so that the error (standard deviation of $\hat{e}_B - \hat{e}$) goes to zero at the rate $1/\sqrt{B}$.

Since highly accurate interval estimation is required in the Bruceton test, and the bootstrap-t (with bias correction) brings much computation, we look to develop more efficient Monte-Carlo simulation methods. A natural candidate is importance resampling. Importance resampling is a scheme to simulate a rare event under the bootstrap setup. Good references for importance sampling are Glynn and Iglehart (1989), and Asmussen and Rubinstein (1995).

Suppose $\mathbf{x} = \{x_0, \dots, x_n\}$ is a realization of an ergodic Markov chain $\{X_t; t \geq 0\}$ on a finite state space $S = \{0, \dots, k\}$, with transition probability $P = (p_{ij})$ and invariant stationary distribution π. We formulate the problem of approximating confidence intervals for μ, σ and x_p by a bootstrap method with importance resampling as follows. Let f be a bounded real-valued function defined on the state space S and let $S_n = \sum_{t=1}^{n} f(X_t)$. Let $\beta = \sum_{j\in S} f(j)\pi_j$ be the stationary mean and $\tau^2 = E_\pi \bar{f}^2(X_0) + 2\sum_{t=0}^{\infty} E_\pi \bar{f}(X_0)\bar{f}(X_{t+1})$ be the asymptotic variance, where $\bar{f}(X_t) = f(X_t) - \beta$, and E_π refers to the expectation of $\{X_k\}$ when the initial state $\{X_0\}$ has the stationary distribution π. Here, we want to estimate the probability of a typical tail event

$$u = P_\pi\{g(S_n/n) \leq a_n\} \tag{4.3}$$

by simulation, for $a_n = g(\beta) + a\tau/\sqrt{n}$ with $a < 0$, where $g : \mathbf{R} \to \mathbf{R}$ is a sufficiently smooth function in a neighborhood of β. Note that in the bootstrap setting, we simple replace P by $\hat{P}_n$. The appropriate exponential tilting formula (see Chapter 23 of Efron and Tibshirani (1993) and Chapter 9 of Davison and Hinkley (1997) for a general discussion) is as follows:

$$q_{ij} = p_{ij}\exp\{-\frac{c_{ij}}{\sqrt{n}}\}/\sum_{j=1}^{k} p_{ij}\exp\{-\frac{c_{ij}}{\sqrt{n}}\}, \tag{4.4}$$

where $c_{ij} = (a_j + \delta_i - \delta_j)C$, $a_j = (f(j) - \beta)/(\tau g'(\beta))$, and $C = C(a) > 0$ is chosen to minimize $\exp(C^2)\Phi(a - C)$. Note that $C \cong -a$. The vector $(\delta_0, \dots, \delta_k)^t$ is the solution of the following Poisson equation:

$$\begin{bmatrix} p_{00} & p_{01} & \cdots & p_{0k} \\ p_{10} & p_{11} & \cdots & p_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k0} & p_{k1} & \cdots & p_{kk} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} + \begin{bmatrix} 1-p_{00} & -p_{01} & \cdots & -p_{0k} \\ -p_{10} & 1-p_{11} & \cdots & -p_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ -p_{k0} & -p_{k1} & \cdots & 1-p_{kk} \end{bmatrix} \begin{bmatrix} \delta_0 \\ \delta_1 \\ \vdots \\ \delta_k \end{bmatrix} = 0. \tag{4.5}$$

It is known that the matrix $P = (p_{ij})$ in (4.5) is of rank k, and $\delta_i - \delta_j$ is uniquely determined for $i \neq j$. The rational of the tilting formula (4.4) will be explained in detail in Appendix 2.

To demonstrate the power of the importance resampling technique, we use the following example. Let $X = \{X_t, t \geq 0\}$ be a Markov chain on a finite space $S = \{1, 2, 3\}$, with transition probability matrix

$$P = \begin{bmatrix} .2000 & .2000 & .6000 \\ .3000 & .4000 & .3000 \\ .5000 & .3000 & .2000 \end{bmatrix},$$

and stationary distribution $\pi_P = (.3391,\ .2957,\ .3652)^t$. We are interested in estimating the probability $u = u(a) = P_\pi\{(S_n - n\beta)/\sqrt{n}\tau \leq a\}$, where $S_n = \sum_{t=1}^n X_t$. Here, $\beta = 2.0261$, and $\tau = 0.5807$.

For instance, consider $a = -2.5758$ so $\Phi(a) = .005$ and $C(a) = 2.6564$. By using (4.4), we obtain the tilting transition probability matrix

$$Q_a = \begin{bmatrix} .3979 & .2641 & .3381 \\ .4612 & .4081 & .1306 \\ .6616 & .2635 & .0750 \end{bmatrix},$$

with stationary distribution $\pi_{Q_a} = (.4746,\ .3084,\ .2170)^t$.

Let $\hat{u}$ be the estimator of u and $r(a) = Var_P(\hat{u})/Var_{Q_a}(\hat{u})$ the relative efficiency of the Monte-Carlo simulation under original probability P with respect to the tilting formula Q_a. Table 4.1 reports values of $r(a)$. The time horizon is $n = 20$ and the simulation number is 3000.

Table 4.1. Estimated relative efficiencies $r(a)$ of P relative to Q_a.

$\Phi(a)$	$C(a)$	$r(a)$
.005	2.6561	51.31
.010	2.5704	29.56
.025	2.1787	10.74
.050	1.8940	5.92
.100	1.5751	3.82
.500	0.6120	1.73
.900	0.1150	1.11
.950	0.0602	1.11
.975	0.0320	1.02
.990	0.0139	1.03

Note that $r(a)$ is a strictly decreasing function in a, so the importance sampling can be considerably more efficacious for negative a (small probability) than positive a.

Next we compare the ordinary bootstrap-t and the importance resampling bootstrap-t when generating various confidence intervals by the up-and-down method. We do the comparison for all the parameters μ, σ and x_p in the probit model (1.1). Linearized maximum likelihood estimates of $\hat{\mu}$, $\hat{\sigma}$ and $\hat{x}_p$ from (5.3), (5.5) and (5.6) are used in the tilting formula (4.4).

In this simulation study, we take $\mu = 0,\ \sigma = .02,\ \Delta = 1.5\sigma = .03$ and initial point $.8\sigma = .016$. The nominal coverage probability for the confidence intervals is .95 in all cases. The average sample size is $n = 100$ and the results are summarized in Table 4.2, where B denotes the number of bootstrap replications and IR refers to importance resampling. Other notations are the same as in Tables 3.1-3.4.

Table 4.2. Comparison of the approximate confidence intervals for $n = 100$ with bias correction.

	μ		σ		$x_{.999}$	
	CP	AL	CP	AL	CP	AL
IR ($B = 200$)	.9205	.0182	.9102	.0141	.9101	.0412
IR ($B = 400$)	.9574	.0195	.9502	.0158	.9497	.0436
Bt ($B = 4000$)	.9438	.0175	.9502	.0168	.9398	.0495

Examination of Table 4.2 shows that compared with the ordinary bootstrap resampling of Section 3, the importance resampling bootstrap method permits a reduction in replication size of at least 10 to 1 without significant loss of performance.

5. CS-M-3 Ignitor

In this section, we illustrate the proposed procedure by applying it to the data set of Table 1.1. For the test, three choices must be made in advance: the proper dosage-stimulus transform, the initial step, and the step size.

Test levels are equally spaced stimulus intensities. For best efficiency, the initial level should be close to the mean. As has been brought out in a previous study, an initial level some distance away from the mean also tends to bias the estimate of the mean in small samples.

The choice of the step size Δ depends on what kind of information is needed and often entails a compromise. If an estimate of the mean is more crucial than the estimate of the standard deviation, then Δ might be $\sigma/4$ to $\sigma/2$ if the probit model is assumed, or $\sigma/2$ to σ if the logit model is assumed. A better estimate of the standard deviation with a corresponding sacrifice of an accurate estimate of the mean would require the use of Δ about 2σ. Comparable treatment of both parameters leads to a choice of Δ about equal to 1.4σ (see Chao and Fuh (1997)).

Only the data set with 4 to 7 different levels will be used for analysis, since, if the step size is large compared to the variability parameter, an alternating two-level pattern may be obtained. On the other hand, a three-level pattern may evolve when the middle level is close enough to μ that a mixed response can be expected there, and the other levels are so remote from μ that the probability of observing all responses on one and all non-responses on the other is very large. In either of these cases (where the step size is large) we can deduce that we have found the limits within which μ is located. We may also guess at an upper limit from the size of the variability parameter. But these estimates and guesses may not be very informative unless the size of the step is small enough, compared to the desired accuracy. If the number of different levels is more than eight, it means that the variability of the experimental ignitor is too large for it to be used.

For the interval estimation of the unknown parameter x_p, we first estimate the sampling distribution of $\sqrt{n}(\hat{x}_p - x_p)$. The bootstrap version of this is $P^*\{\sqrt{n}(\hat{x}_p^* - \hat{x}_p) \le a|\mathbf{x}\}$, which can be estimated by simulation. In order to apply (4.4) for more efficient Monte-Carlo simulation, we need to express $\hat{\mu}$ and $\hat{\sigma}$ in linear form. Since x_p is a linear combination of μ and σ, we first study the case of μ and σ and consider the linear approximation of the MLE $\hat{\mu}$ ($\hat{\sigma}$) of μ (σ). By (1) of Dixon and Mood (1948), a linear approximation of $\hat{\mu}$ is

$$\hat{\mu} = x' + \Delta\Big(\frac{\sum_{i=1}^k i\, n_i}{N} + .5\Big), \tag{5.1}$$

where x' is the normalized level corresponding to the lowest level on which the less frequent event occurs, i is the stimulus level, n_i is the number of observed $0's$ in i, and $N = \sum_{i=1}^k n_i$ is the total number of $0's$ in the sample. Therefore,

$$\begin{aligned}\hat{\mu} &= x' + \Delta\Big(\frac{\sum_{i=1}^k i\Big(\sum_{t=1}^n I_{\{X_{t-1}<X_t, X_{t-1}=i\}}\Big)}{\sum_{t=1}^n I_{\{X_{t-1}<X_t\}}} + .5\Big)\\ &= x' + \Delta\Big(\frac{\sum_{t=1}^n X_{t-1} I_{\{X_{t-1}<X_t\}}}{\sum_{t=1}^n I_{\{X_{t-1}<X_t\}}} + .5\Big).\end{aligned} \tag{5.2}$$

Since the difference between the total number of 1 and the total number of 0 in the Bruceton test is at most k (cf. Wetherill and Glazebrook (1986)), we substitute $\sum_{t=1}^n I_{\{X_{t-1}<X_t\}}$ by $n' = n/2$ and get

$$\hat{\mu} = x' + \Delta\Big(\frac{1}{n'}\sum_{t=1}^n X_{t-1} I_{\{X_{t-1}<X_t\}} + .5\Big). \tag{5.3}$$

Similarly, from (2) of Dixon and Mood (1948), a linear approximation of $\hat{\sigma}$ is

$$\hat{\sigma} = 1.620\Delta\Big(\frac{\sum_{i=1}^k i^2\, n_i}{N} - \Big(\frac{\sum_{i=1}^k i\, n_i}{N}\Big)^2\Big). \tag{5.4}$$

By a standard Taylor expansion, we approximate $(\sum_{i=1}^{k} i\ n_i/N)^2$ by $a_0^2 + 2a_0(\sum_{i=1}^{k} i\ n_i/N - a_0)$, where $a_0 = E(\frac{1}{n'}\sum_{t=1}^{n} X_{t-1}I_{\{X_{t-1}<X_t\}}) = \frac{1}{n'}\sum_{i=1}^{k} i\ (1-p_i)$. Hence,

$$\hat{\sigma} = 1.620\Delta\Big(\frac{1}{n'}\sum_{t=1}^{n}\Big(X_{t-1}^2 - 2a_0X_{t-1}\Big)I_{\{X_{t-1}<X_t\}} + a_0^2 + .029\Big). \tag{5.5}$$

Now by definition of $\hat{x}_p$ (1.1), we have

$$\begin{aligned}\hat{x}_p &= \hat{\mu} + \Phi^{-1}(p)\hat{\sigma}\\ &= x' + \Delta\frac{1}{n'}\sum_{t=1}^{n}\Big[\Big(X_{t-1}^2 - 2a_0X_{t-1}\Big)\,\Phi^{-1}(p)1.620 + X_t\Big]\,I_{\{X_{t-1}<X_t\}}\\ &\quad +.5\Delta + \Phi^{-1}(p)1.620\Delta(a_0^2 + .029).\end{aligned} \tag{5.6}$$

By considering $S_n = \sum_{t=1}^{n}\left[(X_{t-1}^2 - 2a_0X_{t-1})\,\Phi^{-1}(p)1.620 + X_t\right] I_{\{X_{t-1}<X_t\}}$, we can apply (4.4) to bootstrapping the sampling distribution of $\sqrt{n}(\hat{x}_p - x_p)$.

By using the linear approximate MLE $\hat{\mu}$, $\hat{\sigma}$ and $\hat{x}_p$ from (5.3), (5.5) and (5.6) respectively, together with the computer program provided by McMains (1984) (see also MIL-STD-1512 and MIL-STD-1756), we get $\hat{\mu} = .8585$, $\hat{\sigma}$(with bias correction)$= .0213$ and $\hat{x}_p = .9145$ for $p = .9990$ for the data from Table 1.1. For bootstrap-t confidence intervals, because of the small sample size (n=43) and the Markovian binary data structure, $B_1 = 2,000$ bootstrap replications to achieve stability of the bias estimate for σ. We took $B_2 = 10,000$ bootstrap replications to achieve stability of the interval estimate for x_p. For a total of $B = B_1 \times B_2 = 20,000,000$ bootstrap replications. By using (4.4), a total of $B = 30,000$ bootstrap replications gives the approximate confidence interval $(.8515, .8841)$ for μ, $(.0081, .0290)$ for σ and $(.8909, .9479)$ for x_p. In application, one uses .9479 plus a safety factor specific to the mission.

6. Concluding Remarks and Future Research

Although the up-and-down method is a classical sequential procedure, it is still very much in use in pyrotechnics sensitivity analysis. Applications to biological statistics are in Storer (1989), Whitehead and Brunier (1995) and Smith, Dutton and Smith (1996). In this paper, we provide a computationally intensive method, a bias-corrected bootstrap with importance resampling, to construct approximate confidence intervals for the parameters of interest. Our experience suggests the following.

(1) The bootstrap method provides a more accurate interval estimation of the parameters of interest. An importance resampling technique facilitates the

reduction of variance of the Monte-Carlo simulation for the bootstrap algorithm. Our simulation results indicate that the bootstrap-t (with bias correction) confidence interval is better than the classical normal approximation in many respects.

(2) Although the probit model and logit model are in good agreement for quantiles in the .2 to .8 range, extreme quantiles are rather sensitive to model misspecification (see Wu (1985)). In practice, we can use one of four models (logit, probit, double exponential and double reciprocal) according to the engineering experience about the tail behavior of the response curve. As an alternative, a nonparametric approach is required in which the spacing can be chosen according to a continuous distribution. In this case, the observed data is a general state Markov chain with interval censoring.

(3) The problem of degradation is important in the stockpile of pyrotechnics. We intend to propose a procedure to handle this type of problem.

(4) The Bruceton test is only for a single component. In the development and evaluation of explosive trains that function by transfer of detonation from component to component, we need to predict the probability with which detonation is transferred from the donor, across the interface, to the acceptor, and the confidence that can be associated with the estimated transfer probability. Further investigations are needed in VARICOMP (VARIation of explosive COMPosition), which is a method for determining detonation-transfer probabilities. This will involve a multi-dimensional up-and-down method.

Acknowledgement

This paper is partially supported by the National Science Council of ROC under contract CS-88-2623-D-001-001.

Appendix 1

We first summarize the results from Fuh and Lai (1998) on second-order efficiency for bootstrapping Markov chains. Consider an ergodic Markov chain $\{X_n, n \geq 0\}$ on a finite state space D, with transition probability P and invariant distribution π. Let f be an additive functional from D to R^d. With $S_n = \sum_{t=0}^{n} f(X_t)$, to establish Edgeworth expansions for bootstrapping Markov chains, we strengthen Cramér's (strongly nonlattice) condition to

$$\limsup_{|\theta|\to\infty} \left| \int E_x\{\exp(i\theta S_1)\} d\pi(x) \right| < 1. \tag{A.1}$$

Let

$$\mu = \int E_x S_1 d\pi(x), \tag{A.2}$$

and

$$V =: \lim_{n\to\infty} n^{-1}E_\nu\{(S_n - n\mu)(S_n - n\mu)^t\}. \quad \text{(A.3)}$$

Throughout the sequel we let P_ν denote the probability measure under which X_0 has initial distribution ν.

Proposition 1. (Fuh and Lai (1998))

Let $\lambda > 0$ and let $r \geq 3$ be an integer. Assume $E_\pi S_1^{r-2} < \infty$ and that S_1 is strongly nonlattice. For $0 < \alpha \leq 1$ and $c > 0$, let $\mathcal{B}_{\alpha,c}$ be the class of all Borel subsets B of R such that $\int_{(\partial B)^\varepsilon} \phi_V(y)dy \leq c\varepsilon^\alpha$ for every $\varepsilon > 0$, where ϕ_V is the density function of the d-variate normal distribution with mean 0 and covariance matrix V, ∂B denotes the boundary of B and $(\partial B)^\varepsilon$ denotes its ε-neighborhood. Suppose that $g : R^d \to R^p$ has continuous derivatives of order r in some neighborhood of μ. Let $J_g = (D_j g_i(\mu))_{1\leq i\leq p, 1\leq j\leq d}$ be the $p \times d$ Jacobian matrix and let $V(g) = J_g V J_g'$. Then

$$\sup_{B\in\mathcal{B}_{\alpha,c}} |P_\nu\{\sqrt{n}(g(n^{-1}S_n^*) - g(n^{-1}S_n)) \in B|\mathbf{x}\} \quad \text{(A.4)}$$

$$- \int_B \{\phi^*_{V(g)}(y) + \sum_{j=1}^{r-2} n^{-j/2}\phi^*_{j,V,g}(y)\}dy| = o(n^{-(r-2)/2}) \quad a.s.,$$

where S_n^* is the bootstrap analogy of S_n and $\phi^*_{j,V,g}$ is $\phi_{j,V,g}$ with population moments replaced by sample moments (Fuh and Lai (1998)).

Now, under the assumed probit model (1.1), the strongly nonlattice condition (A.1) is automatically satisfied. Next we want to show that $\hat\mu$ and $\hat\sigma$ are smooth functions of S_n/n. From (9) and (10) of Dixon and Mood (1948), the maximum likelihood estimates of μ and σ are the roots of

$$\sum n_i(\frac{z_{i-1}}{q_{i-1}} - \frac{z_i}{p_i}) = 0, \quad \text{(A.5)}$$

$$\sum n_i(\frac{x_{i-1}z_{i-1}}{q_{i-1}} - \frac{x_i z_i}{p_i}) = 0, \quad \text{(A.6)}$$

where $n_i,\ p_i,\ q_i$ are defined as before, $z_i = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}$ and $x_i = (y_i - \mu)/\sigma$. By the Implicit Function Theorem, there exists a neighborhood N of (μ, σ) and $\eta > 0$ such that for S_n/n in N, a thrice continuously differentiable solution of (A.5) and (A.6)

$$T_n = g(S_n/n), \quad ||T_n - (\mu, \sigma)|| < \eta, \quad \text{(A.7)}$$

can be found. This solution is unique. Therefore, by Proposition 1, bootstrap estimators of $\sqrt{n}(\hat\mu - \mu)$ and $\sqrt{n}(\hat\sigma - \sigma)$ are second-order efficient. Since $x_p =$

$\mu + \Phi^{-1}(p)\sigma$ is a linear combination of μ and σ, the bootstrap estimator of $\sqrt{n}(\hat{x}_p - x_p)$ is also second-order efficient. In Sections 3 and 5, we use the linear approximation of $\hat{\mu}$ and $\hat{\sigma}$ as an initial point. By the classical method of scoring, a one-step Newton-Rapson method provides an efficient estimator.

Appendix 2

Here we use a contiguity argument to derive the tilting formula (4.4) for the simulation of bootstrapping finite state Markov chains. Following the notation in Section 4, let $\{X_n, n \geq 0\}$ be a finite ergodic Markov chain, f be a bounded measurable function defined on the state space and $S_n = \sum_{t=0}^{n} f(X_t)$. We are concerned with the Monte-Carlo simulation of the event

$$P\{\frac{1}{n}S_n \leq a_n\}, \tag{A.8}$$

with $a_n - \beta < 0$ and $a_n - \beta = O(1)$. It is known (cf. Bucklew, Ney and Sadowsky (1990)) that if there exists α belonging to some interval in R which contains the origin, such that $T_\alpha(h)(i) = \sum_{j=1}^{k} \exp(\alpha f(j))h(j)p_{ij}$ is finite with $T_0 = P$, then the standard technique to estimate (A.8) is through exponential tilting by first embedding the distribution of p_{ij} in the following martingale family $Q_\alpha(\cdot) = \{q_{ij}^\alpha\}$ with $q_{ij}^\alpha = p_{ij}\exp(\alpha f(j))r_\alpha(j)(\lambda(\alpha)r_\alpha(i))^{-1}$, where $\lambda(\alpha)$ is the largest eigenvalue of the operator T_α, and $r_\alpha(\cdot)$ is the associated eigenvector for $\lambda(\alpha)$. Let $\psi(\alpha) = \log\lambda(\alpha)$. The point α^* given by $\psi'(\alpha^*) = a_n$ is the optimal point in the sense of minimizing the speed factor defined in Bucklew, Ney and Sadowsky (1990). Therefore the tilting measure at the optimal point α^* becomes

$$q_{ij}^{\alpha^*} = p_{ij}exp\{\alpha^* f(j) - \psi(\alpha^*)\}\frac{r_{\alpha^*(j)}}{r_{\alpha^*(i)}}. \tag{A.9}$$

In statistical applications, one often works with $g(S_n/n)$, where g is a smooth real-valued function, and considers the event

$$u =: P\{g(S_n/n) \leq a_n\}, \tag{A.10}$$

for $a_n = g(\beta) + a\tau/\sqrt{n}$, with $a < 0$.

To obtain a tilting formula in this setting, use $g(S_n/n) - g(\beta) \cong (S_n/n - \mu)g'(\beta)$, $g'(\mu)$ bounded. From $\psi'(\alpha^*) = a_n$ and a one-term Taylor expansion of $\psi'(\alpha^*)$, we have $\psi'(0) + \alpha^*\psi''(0) \cong a_n$, which implies that $\alpha^* \cong (a_n - \mu)/\sigma^2 = a/(\sqrt{n}\sigma)$, $\psi(\alpha^*) \cong (a\mu)/\sqrt{n}\sigma$, and $r_{\alpha^*}(j) = 1 + (ar_0'(j))/\sqrt{n}\sigma$, where $r_0'(i)$ denotes the first derivative of $r_\alpha(i)$ with respect to α at the point $\alpha = 0$.

Hence (A.9) becomes

$$q_{ij} \cong \frac{p_{ij}e^{\frac{a}{\sqrt{n}g'(\beta)\sigma}(f(j)-\mu)+\frac{a}{\sqrt{n}g'(\beta)\sigma}(r_0'(j)-r_0'(i))}}{\sum_{j=1}^{k} p_{ij}e^{\frac{a}{\sqrt{n}g'(\beta)\sigma}(f(j)-\mu)+\frac{a}{\sqrt{n}g'(\beta)\sigma}(r_0'(j)-r_0'(i))}}$$

$$= \frac{p_{ij} e^{\frac{a}{\sqrt{n} g'(\beta)\sigma}(a_j - r_0'(i) + r_0'(j))}}{\sum_{j=1}^{k} p_{ij} e^{\frac{a}{\sqrt{n} g'(\beta)\sigma}(a_j - r_0'(i) + r_0'(j))}}. \tag{A.11}$$

Assume without loss of generality that $g'(\beta) = 1$. Since the vector $(r_{\alpha^*}(1), \ldots, r_{\alpha^*}(k))^t$ is the eigenvector of $\lambda(\alpha^*)$ associated with the operator $T_{\alpha^*}(\cdot)$, we have

$$\begin{bmatrix} p_{11} e^{\alpha^* f(1)} \cdots p_{1k} e^{\alpha^* f(k)} \\ \vdots \quad \ddots \quad \vdots \\ p_{k1} e^{\alpha^* f(1)} \cdots p_{kk} e^{\alpha^* f(k)} \end{bmatrix} \begin{bmatrix} r_{\alpha^*}(1) \\ \vdots \\ r_{\alpha^*}(k) \end{bmatrix} = \lambda(\alpha^*) \begin{bmatrix} r_{\alpha^*}(1) \\ \vdots \\ r_{\alpha^*}(k) \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} p_{11}(1 + af(1)/\sqrt{n}\tau) \cdots p_{1k}(1 + af(k)/\sqrt{n}\tau) \\ \vdots \quad \ddots \quad \vdots \\ p_{k1}(1 + af(1)/\sqrt{n}\tau) \cdots p_{kk}(1 + af(k)/\sqrt{n}\tau) \end{bmatrix} \begin{bmatrix} 1 + ar_0'(1)/\sqrt{n}\tau \\ \vdots \\ 1 + ar_0'(k)/\sqrt{n}\tau \end{bmatrix}$$

$$\cong (1 + a\beta/\sqrt{n}\tau) \begin{bmatrix} 1 + ar_0'(1)/\sqrt{n}\tau \\ \vdots \\ 1 + ar_0'(k)/\sqrt{n}\tau \end{bmatrix}$$

$$\Rightarrow P \begin{bmatrix} (1 + af(1)/\sqrt{n}\tau)(1+ar_0'(1)/\sqrt{n}\tau) \\ \vdots \\ (1+af(k)/\sqrt{n}\tau)(1+ar_0'(1)/\sqrt{n}\tau) \end{bmatrix} = \begin{bmatrix} (1+a\beta/\sqrt{n}\tau)(1+ar_0'(1)/\sqrt{n}\tau) \\ \vdots \\ (1+a\beta/\sqrt{n}\tau)(1+ar_0'(k)/\sqrt{n}\tau) \end{bmatrix}$$

$$\Rightarrow P \begin{bmatrix} af(1)/\sqrt{n}\tau + ar_0'(1)/\sqrt{n}\tau \\ \vdots \\ af(k)/\sqrt{n}\tau + ar_0'(k)/\sqrt{n}\tau \end{bmatrix} = \begin{bmatrix} ar_0'(1)/\sqrt{n}\tau + a\beta/\sqrt{n}\tau \\ \vdots \\ ar_0'(k)/\sqrt{n}\tau + a\beta/\sqrt{n}\tau \end{bmatrix}$$

$$\Rightarrow (I - P) \begin{bmatrix} r_0'(1)/\tau \\ \vdots \\ r_0'(k)/\tau \end{bmatrix} = P\left(\begin{bmatrix} f(1)/\tau \\ \vdots \\ f(k)/\tau \end{bmatrix} - \begin{bmatrix} \beta/\tau \\ \vdots \\ \beta/\tau \end{bmatrix} \right) = P \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}.$$

Now, let $r_0'(i)/\sigma = -\delta_i$ and get

$$(I - P) \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_k \end{bmatrix} = -P \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}. \tag{A.12}$$

This is just equation (4.5). Hence we have (4.4). Note that in this derivation we consider the asymptotic expansion of the large deviation exponential tilting formula (A.9), where the Perron-Frobenius eigenvalue of the operator $T_\alpha(h)(\cdot)$ and its associated eigenvector play an important role. By a Taylor expansion of the eigenvector $r_{\alpha^*}(j)$ for α^* around 0, it turns out that the optimal parameter reduces to the solution of (A.12). When the state space has only one element, this reduces to the case of independent and identically distributed random variables. The tilting formula (4.4) is exactly the same as that in Johns (1988), and Do and Hall (1991). By using the idea of minimizing the (asymptotic) variance of the Monte-Carlo estimate, we have another argument to derive (4.4). A full discussion of this importance sampling technique will be published in a separate paper.

References

Anderson, T. W. and Goodman, L. A. (1957). Statistical inference about Markov chains. *Ann. Math. Statist.* **28**, 89-110.

Asmussen, S. and Rubinstein, R. Y. (1995). Steady state rare events simulation in queueing models and its complexity properties. In *Frontiers in Queueing* (Edited by J. Dhashalow). CRC Press, Boca Raton, Florida.

Athreya, K. B. and Fuh, C. D. (1992). Bootstrapping Markov chains: countable case. *J. Statist. Plann. Inference* **33**, 311-331.

Basawa, I. V. and Prakasa Rao, B. L. S. (1980). *Statistical Inference for Stochastic Processes.* Academic Press, Inc., London.

Billingsley, P. (1961). *Statistical Inference for Markov Processes.* The University of Chicago Press, Chicago.

Bucklew, J. A., Ney, P. and Sadowsky, J. (1990). Monte Carlo simulation and large deviations theory for uniformly recurrent Markov chains. *J. Appl. Probab.* **27**, 44-59.

Chao, M. T. and Fuh, C. D. (1997). Statistical analysis of the Bruceton test. Technical Report.

Chao, M. T. and Fuh, C. D. (1999). On Bruceton test for pyrotechnics sensitivity analysis. *Proceedings of the 17th Symposium on Explosives and Pyrotechnic.* 17 E&PS 5.1-5.29.

Datta, S. and McCormick, W. P. (1993). On the first order Edgeworth expansion for a Markov chain. *J. Multivariate Anal.* **44**, 345-359.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application.* Cambridge University Press, Cambridge.

Derman, C. (1957). Non-parametric up and down experimentation. *Ann. Math. Statist.* **30**, 601-605.

Dixon, W. J. and Mood, A. M. (1948). A method for obtaining and analyzing sensitivity data. *J. Amer. Statist. Assoc.* **43**, 109-126.

Do, K. A. and Hall, P. (1991). On importance resampling for the bootstrap. *Biometrika* **78**, 161-167.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1-26.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap.* Chapman & Hall, London.

Fuh, C. D. (1993). Statistical inquiry for Markov chains by bootstrap method. *Statist. Sinica* **3**, 53-66.

Fuh, C. D. and Lai, T. L. (1998). Characteristic functions and Edgeworth expansions for Markov random walks with applications to bootstrap. Technical Report

Glynn, P. W. and Iglehart, D. L. (1989). Importance sampling for stochastic simulations. *Management Sci.* **35**, 1367-1391.

Hampton, L. D., Blum, G. D. and Ayres, J. N. (1973). *Logistic Anal. Bruceton Data.* NOLTR, 73-91.

Johns, M. V. Jr. (1988). Importance sampling for bootstrap confidence intervals. *J. Amer. Statist. Assoc.* **83**, 709-714.

Kulperger, R. J. and Prakasa Rao, B. L. S. (1990). Bootstrapping a finite state Markov chains. *Sankhyā A* **51**, 178-191.

McMains, F. L. (1984). A digital computer program for a Bruceton analysis. Technical Memorandum 1254, U.S. Department of Commerce, Springfield, VA.

Mood, A. (1998). Personal communication.

Sitter, R. R. and Wu, C. F. J. (1993). Optimal designs for binary response experiments: Fieller, D, and A criteria. *Scand. J. Statist.* **20**, 329-341.

Smith, W. D., Dutton, R. C. and Smith, N. TY (1996). A measure of association for assessing prediction accuracy that is a generalization of non-parametric ROC area. *Statist. Medicine* **15**, 1199-1215.

Storer, B. E. (1989). Design and analysis of Phase I clinical trials. *Biometrics* **45**, 925-937.

Thomas, E. V. (1994). Evaluating the ignition sensitivity of thermal-battery heat pellets. *Technometrics* **36**, 273-282.

Wetherill, G. B. (1963). Sequential estimation of quantal response curves (with discussion). *J. Roy. Statist. Soc. Ser. B* **25**, 1-48.

Wetherill, G. B. and Glazebrook, K. D. (1986). *Sequential Methods in Statistics.* Chapman and Hall, London.

Whitehead, J. and Brunier, H. (1995). Bayesian decision procedures for dose determining experiments. *Statist. Medicine* **14**, 885-893.

Wu, C. F. J. (1985). Efficient sequential designs with binary data. *J. Amer. Statist. Assoc.* **80**, 974-984.

Institute of Statistical Science, Academia Sinica, Taipei, Taiwan.

E-mail: mtchao@stat.sinica.edu.tw

E-mail: stcheng@stat.sinica.edu.tw

(Received September 1999; accepted June 2000)