

## The Design Aspect of the Bruce-ton Test for Pyrotechnics Sensitivity Analysis

C. D. Fuh<sup>1</sup>, J. S. Lee<sup>2</sup> and C. M. Liaw<sup>2</sup>

<sup>1</sup>*Academia Sinica*, <sup>2</sup>*Chung-Shan Institute of Science and Technology*

*Abstract:* We start with a data set obtained from a study of the CS-M-3 ignitor in a military experiment and is based on the classical up-and-down method of Dixon and Mood (1948). Since the Bruce-ton tests are actively employed in pyrotechnical sensitivity studies, we reexamine this method based on the view that it is designed for data-collection. Two different aspects are addressed: as a design for parameter estimation and as a design for giving clues about the goodness of fit. Two sets of data are employed to illustrate our points. For the estimation of  $(\mu, \sigma)$ , the location and the scale parameters, we show that a properly selected up-and-down design is quite informative; for the estimation of  $x_p$ , the 100 $p$ %-th quantile, however, the best selected up-and-down method is only about 50% effective as compared with the corresponding  $c$ -optimal design. Although not particularly useful, the up-and-down method does judge the proper selection of underlying model. In any case, all the quantal response models are rather poor in terms goodness of fit.

*Key words:*  $A$ -optimal,  $c$ -optimal,  $D$ -optimal; Markov chains, maximum likelihood estimate, probit model, sequential design, up-and-down method.

### 1. Introduction

Table 1.1 summarizes the test results for the CS-M-3 ignitor obtained from an experiment carried out at the Chung-Shan Institute of Science and Technology in Taiwan. The experiment is carried out by a Bruce-ton test,

which is based on the classical up-and-down method of Dixon and Mood (1948). Bruceton is the name of a military installation in Pennsylvania, USA, that tests weapons and ammunition among other things (Mood 1998). Since this method is specifically documented in certain military standards (e.g., MIL-STD-331B, MIL-STD-322B), 50 years after its initial introduction, it is still actively used in Pyrotechnics Sensitivity Analysis (PSA).

An implicit assumption underlying this quantal response model is that there is a latent random variable  $X$  which represents the critical value of the testing item in the sense that  $Y(x) = 1$  if and only if  $X \leq x$ . Although  $F$ , the distribution function of the critical stimulus levels, can be any distribution function, in practice, a parametric family is often assumed. In particular, two distributional forms have been commonly used, the probit and the logit models, perhaps after a transformation of  $x$ . For the probit model,  $F(x)$  is modeled as an integrated normal distribution function. Hence, for the probit model,

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

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

$$F(x) = P(Y(x) = 1) = \frac{1}{1 + \exp\{-\frac{x-\mu}{\sigma}\}}. \quad (1.2)$$

An up-and-down test 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 level by one unit at the next stage; if the response is 1 (explosive), we decrease the stimulus level by one unit in the next stage. The process continues until a fixed number of observations is obtained.

When we are presented with the given data set, the main objective of the experiment is to estimate  $x_p$ , the 100 $p$ %-th quantile of  $F$ , where  $p$  is close to unity. In pyrotechnical terms, this is called a sensitivity test. For the case  $p$  is close to 0, it is called a safety test. We shall not address the question of safety tests in this paper since a dual argument is sufficient.

Table 1.1 Experiment of CS-M-3 ignitor

.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 salient feature of the Bruceton test is that the observations are concentrated around the stimulus level that will produce the median value of  $F$ , and this holds true irrespective of the choice of the initial stimulus level and without prior knowledge about the values of the unknown parameters. When the data in Table 1.1 is projected on the verticle-axis, we can see a histogram-like picture and will intuitively expect that estimation of the mean of  $F$  can be effectively carried out. This was an important feature in the late 1940's when the value of the location parameter was of primary interest. However, for the data set in Table 1.1 (and Table 5.4 presented later), the main objective of the experiment is to estimate the high quantiles of  $F$ . This task is usually achieved by first estimating the parameters and then estimating  $x_p$  by using its corresponding parametric form. The classical method estimates the parameters of the underlying distribution first and then estimates the quantile  $x_p$  by means of extrapolation. For  $p \in (.2, .8)$ , several authors have noted that the probit and logit models have good agreement for quantiles in this range (Wetherill 1963, Wu 1985); on the other hand, the choice of model to be used when estimating quantiles in this range is up to the investigator. For  $p$  close to 1, however, the proper estimation of  $x_p$  depends on the correct choice of the parametric model, probit or otherwise. Hence, the issue of goodness of fit becomes more and more important as  $p$  gets closer to unity.

Although there is no lack of good designs for estimating  $x_p$  based on the binary quantal response data (see, for example, Young and Easterling 1993), we were rather surprised when we found that the 1948 method is still in active use at PSA today. One of the main objectives of this study was to take a close look at this classical method and examine it based on what we know about current binary data analyses. Emphasis was placed on, but

not limited to, the estimation of  $x_p$ .

Since for the problem at hand, the data set had already been collected, we made no attempt to improve the Bruceton test as far as the data collection part is concerned. Rather, we examined the method with respect to two distinct issues: design and goodness of fit. For analysis of binary data with a parametric model, the known optimal designs are all 1-, 2- or 3-point designs (Wu 1988, Sitter and Wu 1993). However, the Bruceton test usually takes values at 6 or more different levels of stimulus. How does the Bruceton test compare with a corresponding optimal design? As a data-collection design, in Section 3, we calculate and compare its efficiency with that of the  $D$ - and  $A$ -optimal designs. In most cases, the Bruceton test demonstrates a relative efficiency of 70% or more as compared with the corresponding optimal design. Section 4 is devoted for the  $c$ -optimal design since it is the optimal design when the estimation of  $x_p$  is of major concern. For  $p$  close to 1, the Bruceton test is less than 50% effective as compared with the corresponding  $c$ -optimal designs.

For goodness of fit, it is clear that a 2-point design, even when it is optimal in some other sense, cannot give us any clue about the unknown functional form of  $F$ , which is important, particularly when we need to extrapolate it into the tail part to estimate a high quantile. For comparison with the optimal estimator of  $F$  with the same amount of data collected by means of fixed designs, we examine the corresponding estimator of  $F$  based on the Bruceton test (Section 5). Even with a moderately large data set ( $n = 240$ ), the best estimator of  $F$  we can construct is far from satisfactory. This is typical of all quantal response models, however. Inference at the extreme tail of the distribution is risky unless the sample size is extremely large.

## 2. Markov chain representation

To analyze the performance of the up-and-down test insofar as estimating  $\mu$ ,  $\sigma$  and  $x_p$  is concerned, we first represent the process based on the data presented in Table 1.1 as a Markov chain and then perform point and interval estimation for the parameters of interest. The idea of the Markov chain representation is not new, see for example Derman (1957), Wetherill

(1963), Wetherill and Glazebrook (1986), and more recently Thomas (1994). The advantage of using the Markov chain representation can be described as follows. For the up-and-down method, since we can reproduce the whole ordered data set  $\{(X_i, Y_i), i = 1, \dots, n\}$  from the ordered set of stimulus levels  $\{X_i, i = 1, \dots, n\}$ , and since the latter can be formulated as a Markov chain with a specific transition probability matrix  $P$  as in (2.1) below, in the analysis that follows, we can completely ignore the  $Y$  values and concentrate on the  $X$  values only. Therefore, the classical method of maximum likelihood in parametric Markov chains can be applied to perform point estimation. The statistical inferences for ergodic Markov chains have been well documented by Anderson and Goodman (1957), Billingsley (1961) and Basawa and Prakasa Rao (1980).

For ease of exposition, let us assume that there are five stimulus levels  $a_0, a_1, \dots, a_4$  with  $a_i < a_j$  for  $i < j$ . It is easy to see that  $\{X_t, t = 1, \dots, n\}$  forms a Markov chain on a state space  $\{a_0, \dots, a_4\}$  with the following transition probability matrix:

$$P = \begin{pmatrix} p_0 & 1 - p_0 & 0 & 0 & 0 \\ p_1 & 0 & 1 - p_1 & 0 & 0 \\ 0 & p_2 & 0 & 1 - p_2 & 0 \\ 0 & 0 & p_3 & 0 & 1 - p_3 \\ 0 & 0 & 0 & p_4 & 1 - p_4 \end{pmatrix} \quad (2.1)$$

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

### 3. Design aspect

The data collection process for the up-and-down method described earlier is a sequential design in the sense that the stimulus level of the next observation depends on the stimulus level and the outcome of the current observation. This design has the effect that the stimulus levels are clustered around the median value of  $F$ . In addition, this happens despite the fact that the parameters, or even the functional forms, of  $F$  are unknown.

In contrast, let  $\mathbf{a} = (a_0, a_1, \dots, a_k)$  be fixed stimulus levels, and let  $n_j$  independent binary observations  $Y_t(a_j), t = 1, 2, \dots, n_j$  be taken at level  $a_j$ .

This is a so-called fixed-design, and an optimal fixed design may be used as a benchmark with which the up-and-down methods can be compared. We assume that  $n$  observations are taken using an optimal fixed design, and that another  $n$  observations are taken using the up-and-down method. In this section, we shall derive numerical values, which in effect represent the relative efficiency of the Bruceton test. We shall consider three cases for  $F$ : probit, logit and double exponential. The  $D$ -,  $A$ - and  $c$ -optimalities will also be considered.

Suppose that among the  $n_j$  observations taken at  $a_j$ ,  $r_j$  of the  $Y$ 's take the value 1. Then, the likelihood function is of the form

$$L(F) = \text{constant} \times \prod_{j=0}^k [F(a_j)]^{r_j} [1 - F(a_j)]^{n_j - r_j}. \quad (3.1)$$

It is easy to see that the above equation holds for the fixed design case. It also holds true for the up-and-down method, thanks to its special data collection procedure, despite the fact that the  $X$ 's follow a Markov chain. The difference is that, for the fixed design, the  $n_j$ 's are constants and the  $r_j$ 's follow a binomial distribution  $b(n_j; F(a_j))$  whereas for the up-and-down method, the  $n_j$ 's are random variables.

We shall introduce the location-scale family  $F((\cdot - \mu)/\sigma)$  to facilitate likelihood analysis. Omitting the details, the Fisher information produced by a binary observation taken at  $a_j$  is

$$I(a_j) = \frac{1}{F_j(1 - F_j)} \begin{bmatrix} \left(\frac{\partial F_j}{\partial \mu}\right)^2 & \left(\frac{\partial F_j}{\partial \mu}\right)\left(\frac{\partial F_j}{\partial \sigma}\right) \\ \left(\frac{\partial F_j}{\partial \mu}\right)\left(\frac{\partial F_j}{\partial \sigma}\right) & \left(\frac{\partial F_j}{\partial \sigma}\right)^2 \end{bmatrix},$$

where  $F_j = F((a_j - \mu)/\sigma)$ . Let  $\mathbf{n} = (n_0, n_1, \dots, n_k)$ . A fixed design can be represented by the two vectors  $\mathbf{a}$  and  $\mathbf{n}$ , where  $n_j$  data points are collected at level  $a_j$ ,  $j = 0, 1, \dots, k$ . Let  $\mathbf{w} = (w_0, w_1, \dots, w_k)$ , where  $w_j = n_j/n$ ,  $n = \sum n_j$ . In optimal design terms, the pair  $(\mathbf{w}, \mathbf{n})$  constitutes an approximate design since when analytical methods are employed to find the optimal  $\mathbf{w}$ , we cannot assure that  $nw_j$  are integers.

Let  $I = \sum n_j I(a_j)$  be the Fisher information matrix of the whole data set. Then  $I^{-1}$  is the variance-covariance matrix for the MLE's of  $\mu$  and  $\sigma$ . A

Table 3.1  $D$ -optimal designs

model	$F(\mathbf{a})$	weights
$F_1$	(.18, .82)	(.5, .5)
$F_2$	(.13, .87)	(.5, .5)
$F_3$	(.10, .50, .90)	(.28, .44, .28)
$F_4$	(.21, .50, .79)	(.26, .48, .26)

Table 3.2  $A$ -optimal designs

model	$F(\mathbf{a})$	weights
$F_1$	(.21, .79)	(.5, .5)
$F_2$	(.14, .86)	(.5, .5)
$F_3$	(.10, .50, .90)	(.36, .28, .36)
$F_4$	(.21, .50, .79)	(.40, .20, .40)

$D$ -optimal design is a design that maximizes the determinant of  $I$ , and an  $A$ -optimal design is a design that minimizes the sum of  $Var(\hat{\mu}) + Var(\hat{\sigma})$ . The exact solution of the approximate  $D$ - and  $A$ -optimal designs for 4 different distributions were found by Sitter and Wu (1993):

The four distribution forms are as follows:

- (i) Logit:  $F_1(z) = \frac{1}{1 + e^{-z}}$ ;
- (ii) Probit:  $F_2(z) = \Phi(z)$ ;
- (iii) Double exponential:  $F_3(z) = \int_{-\infty}^z h(u)du$ ,  $h(u) = \frac{1}{2}e^{-|u|}$ ;
- (iv) Double reciprocal:  $F_4(z) = \int_{-\infty}^z h(u)du$ ,  $h(u) = \frac{1}{2(1 + |u|)^2}$ .

We shall consider not only the most popular cases, i.e., the logit and probit, but also the double exponential distribution. To use the above

optimal designs, one needs to guess the approximate values of  $\mu$  and  $\sigma$  in order to obtain the optimal  $a_j$ 's; hence they cannot be strictly employed in practice. Nevertheless, these results give us a benchmark with which the up-and-down method can be compared. Let  $D_{opt}$  and  $A_{opt}$  denote the determinant and the sum of variances (trace of  $I^{-1}$ ) for the corresponding optimal fixed designs. We can compute  $D_{ud}$  and  $A_{ud}$ , where the subscript  $ud$  stands for "up-and-down"; then, the ratios  $D_{ud}/D_{opt}$  and  $A_{opt}/A_{ud}$  will give us an efficiency measurement of the Bruceton test.

The  $D$ - and  $A$ -values are of the order of  $O(n)$  and  $O(1/n)$ , respectively, for a fixed design of size  $n$ . Table 3.3 lists the benchmark values of  $D_{opt}/n$  and  $nA_{opt}$ . For the up-and-down test, we divide the interval  $(-b, b)$  into  $k$  equal parts ( $k = 4, 6, \dots, 12$ ) and use  $\Delta = 2b/k$  as the step size.

Table 3.3 Benchmark values of certain distributions

model	$D_{opt}/n$	$nA_{opt}$
$F_1$	0.05	9.4615
$F_2$	0.1986	4.1527
$F_3$	0.0809	7.6035

Let  $\underline{\pi} = (\pi_0, \pi_1, \dots, \pi_k)$  be the corresponding stationary distribution of the Markov chain; then, it can be shown that

$$\pi_j \propto \prod_{i < j} (1 - F(a_i)) \prod_{i > j} F(a_i), \quad (3.2)$$

and the Fisher information matrix reduces to

$$\begin{aligned} I_{ud} &= n \sum_{j=0}^k \pi_j I_j \\ &:= \sum_{j=0}^k \frac{\phi(a_j)}{\Phi(a_j)(1 - \Phi(a_j))} \begin{bmatrix} 1 & a_j \\ a_j & a_j^2 \end{bmatrix}, \end{aligned} \quad (3.3)$$

for the probit model. Now  $I_{ud}$  is a  $2 \times 2$  matrix, and its determinant and inverse are easy to obtain. It is in this way that we can calculate the values

of  $D_{ud}$  and  $A_{ud}$  for the case in which  $F = F_2$ . For the optimal  $A$ -design, we simply use  $k = 3$ ,  $\tilde{\pi} = (.5, .25)$  and  $\mathbf{a} = (-z, z)$  in (3.3) according to Table 3.2, where  $z = \Phi^{-1}(0.86)$ . For the optimal  $D$ -design, we employ Table 3.1 and use (3.3) but with  $z = \Phi^{-1}(.87)$  and the same  $\tilde{\pi}$ . The calculation for  $F_1$  and  $F_3$  is similar and will not be repeated here. Table 3.4 and Table 3.5 list the values of  $D_{ud}/D_{opt}$  and  $A_{ud}/A_{opt}$  for  $F_1, F_2$  and  $F_3$ .

Table 3.4 Efficiency of  $D$ -optimal design  
vs up-and-down

$k$	$F_1$	$F_2$	$F_3$
4	.1894	.4970	.2883
6	.4834	.6827	.6545
8	.6624	.6520	.8568
10	.7337	.6000	.9405
12	.7475	.5500	.9648
14	.7359	.5030	.9613

Table 3.5 Efficiency of  $A$ -optimal design  
vs up-and-down

$k$	$F_1$	$F_2$	$F_3$
4	.4450	.6978	.3246
6	.7296	.8135	.6645
8	.8445	.7720	.8246
10	.8936	.7172	.8848
12	.9090	.6635	.8996
14	.9065	.6144	.8938

For the probit model, if our main concern is to obtain a minimum volume confidence ellipsoid for  $(\mu, \sigma)$ , then the best up-and-down design will use 7 levels ( $k = 6$ ) whereas for the logistic model, we need  $k = 12$ . In all these cases, the design points are equally spaced by  $k$  intervals, essentially covering the whole range of  $F$ .

If, on the other hand, our main concern is to minimize  $Var(\hat{\mu}) + Var(\hat{\sigma})$ , then  $k = 12, 6$  and  $12$  is the right choice for logit, probit, and the double

exponential cases respectively. In most of these cases ( $D$ - and  $A$ - optimal alike), the up-and-down method exhibits an efficiency level of 70% or more compared to the best design and this is achieved **without** prior knowledge of  $\mu, \sigma$ . With respect to estimation, we can conclude that the up-and-down design, 50 years after its invention, still remains efficient in terms of data collection.

#### 4. $c$ -optimal designs

In this section, we will investigate the efficiency of the up-and-down method in estimating  $x_p$ , the  $100p\%$ -th quantile of  $F((\cdot - \mu)/\sigma)$ , in comparison with the  $c$ -optimal design for the same  $x_p$ . As mentioned earlier, one of the main objectives in obtaining our data set is to estimate the high quantiles.

We write  $x_p = \mu + z_p\sigma$ , where  $z_p$  satisfies  $F(z_p) = p$  and is the  $100p\%$ -th quantile of the “standard  $F$ ”. The proper estimate is  $\hat{x}_p = \hat{\mu} + z_p\hat{\sigma}$ , where  $(\hat{\mu}, \hat{\sigma})$  is the MLE of  $(\mu, \sigma)$ . It follows that

$$Var(\hat{x}_p) = Var(\hat{\mu}) + 2z_pCov(\hat{\mu}, \hat{\sigma}) + z_p^2Var(\hat{\sigma}),$$

and using the result of Section 3, we can find the Fisher information matrix of  $x_p$  for both the fixed design and the up-and-down method. For selected values of  $p$  and  $k = 4, 6, \dots, 14$ , Tables 4.1 and 4.2 give the values of  $Var(\hat{x}_p)$  of  $F_1$  and  $F_2$ , respectively.

Table 4.1  $Var(\hat{x}_p)$ , with the up-and-down method:  
logistic model

$k$	$p = .93$	.95	.97	.99	.994	.998
4	80.919	102.538	139.833	238.511	293.104	429.505
6	41.944	52.216	69.936	116.821	142.760	207.568
8	36.027	44.744	59.781	99.567	121.578	176.573
10	35.756	44.541	59.696	99.795	121.979	177.406
12	37.201	46.519	62.592	105.122	128.651	187.439
14	39.278	49.286	66.550	112.228	137.498	200.638

Table 4.2  $Var(\hat{x}_p)$ , with the up-and-down method:  
probit model

$k$	$p = .93$	.95	.97	.99	.994	.998
4	10.957	12.968	16.139	23.284	26.711	34.232
6	9.481	11.243	14.021	20.281	23.284	29.873
8	10.355	12.376	15.561	22.738	26.181	33.735
10	11.450	13.761	17.404	25.613	29.549	38.189
12	12.634	15.247	19.367	28.650	33.102	42.873
14	13.872	16.795	21.405	31.792	36.773	47.706

A  $c$ -optimal design is a fixed design that minimizes  $Var(\hat{x}_p)$ . For  $F = F_i, i = 1, 2$ , the exact solution of the approximate  $c$ -optimal design was obtained by Wu (1988) and can be described as follows. With  $F = F_i$ , there exists a  $p^* = p_i^*$  such that for  $p > p_i^*$ , the  $c$ -optimal design is a 2-point design which assigns probabilities  $(1 - \alpha, \alpha)$  to  $(-x_0, x_0)$ ,  $x_0 = F_i(p^*)$ , respectively. Table 4.3 lists the values of  $p_i^*$  (collected from Wu (1988)) and  $x_0$  for various  $F$ 's.

Table 4.3  $Var(\hat{x}_p)$ , with the up-and-down method:  
double exponential model

$k$	$p = .93$	.95	.97	.99	.994	.998
4	84.818	115.591	171.581	329.88	421.09	655.16
6	38.498	52.060	76.736	146.50	186.70	289.85
8	29.924	40.301	59.182	112.56	143.32	222.25
10	27.520	37.007	54.269	103.07	131.19	203.35
12	27.003	36.301	53.219	101.05	128.61	199.34
14	27.255	36.652	53.750	102.09	129.94	201.42

For  $X \sim F_i, i = 1, 2$ , it can be shown that  $E(X) = 0$ , but  $E(X^2) \neq 1$ , so that the corresponding  $F$ 's are not, strictly speaking, "normalized". We can adjust the  $F$ 's using a scale factor to  $F_i^*$  such that if  $X^* \sim F_i^*$ , then  $E(X^{*2}) = 1$ . The last row in Table 4.4 gives the value of  $x_0^*$  such that  $F_i(x_0^*) = p_i^*, i = 1, 2$ . For various  $p$ 's, the corresponding weight that the  $c$ -optimal design assigns to  $(-x_0, x_0)$  is  $(1 - \alpha, \alpha)$ . Table 4.4 lists the values

of  $\alpha$  for different values of  $p$  (also complied by Wu (1988)) as well as the corresponding values of  $Var(\hat{x}_p)$ .

Table 4.4 Support points for  $c$ -optimal designs

	$F_1$	$F_2$	$F_3$
$p_i^*$	.917	.942	.921
$x_0$	2.402	1.572	1.845
$x_0^*$	1.324	1.572	1.304

We leave the cell for  $p = .93$  and  $F = F_2$  open since for  $p = .93$ , the corresponding  $c$ -optimal design for the probit case is a one-point design.

The values of  $Var(\hat{x}_p)$  presented in Table 4.4 serve as benchmarks with which the up-and-down method can be compared when the estimation of  $x_p$  is a major concern.

Comparing Table 4.1 with the first row of Table 4.2, we can see that for the logit case, the up-and-down method behaves poorly compared to the corresponding  $c$ -optimal design. The main reason is, of course, that the up-and-down method was designed mostly with estimation of the median in mind. In contrast with the  $A$ - or  $D$ -optimal case, our result shows that the up-and-down method is far from being optimal when estimation of  $x_p$  is employed. The same conclusion is also drawn when the probit is compared with its respective  $c$ -optimal designs. For the highest quantile  $x_{.998}$  considered by Wu (1988), the best choices for the up-and-down method are  $k = 8, 6$  and  $12$  for  $F_1, F_2$  and  $F_3$ , respectively. The relative efficiency obtained for  $F_1$  and  $F_2$  are 49.8% and 45.6% respectively. It is quite clear that, in general, the up-and-down method is rather inefficient in terms of data collection when the estimation of high quantiles is our main concern.

## 5. Goodness of fit issues

The likelihood function of any fixed design as well as the likelihood function for the up-and-down method depends on the unknown distribution function  $F$  through the points in  $\mathbf{a} = (a_0, a_1, \dots, a_k)$  only. This implies that if  $F$  and  $G$  are two distinct cdf's that agree on the points of  $\mathbf{a}$ , then there is no way, other than by simple pure guessing, for the statistician to check

whether the underlying model is  $F$  or  $G$ . This non-identifiability problem exists and cannot be fixed unless we have  $n \rightarrow \infty$  and at the same time, let the bandwidth  $\Delta$  go to 0.

In this section, we will keep  $\Delta$  fixed and study the goodness of fit issue for the up-and-down method by looking at the points of  $\mathbf{a}$  only. We take  $\hat{F}(a_j) = r_j/n_j$  as the estimator for  $F(a_j)$  and study the behavior of

$$MSE = \int \sum_j \left[ \frac{r_j}{n_j} - F(x) \right]^2 dx, \quad (5.1)$$

and compare it with a similar estimator for the best fixed design. We note here that since the likelihood functions are of the same form for either the fixed design or the up-and-down method,  $\hat{F}(a_j)$  is the MLE of  $F(a_j)$  in either case although their statistical behaviors are different. To keep things simple, we adopt the same  $k$  equally-spaced intervals.

For the fixed design case, since  $n_j$  is fixed and  $r_j$  is binomial, it can be shown that the MSE can be expressed as the sum of a variance term and bias squared:

$$MSE = \sum_j \frac{\Delta}{n_j} F(a_j)(1 - F(a_j)) + \sum_j \int_{a_j}^{a_{j+1}} [F(a_j) - F(x)]^2 dx. \quad (5.2)$$

It is clear that except by letting  $\Delta \rightarrow 0$ , there is nothing we can do to reduce the bias term. But for the variance, using a Lagrange multiplier, it can be shown that the best choice is

$$n_j \propto \sqrt{F(a_j)(1 - F(a_j))}$$

subject to the condition that  $\sum_j n_j = n$ .

Using this optimal fixed design, we have

$$\begin{aligned} MSE &\geq \frac{\Delta}{n} \left[ \sum_j \sqrt{F(a_j)(1 - F(a_j))} \right]^2 + \sum_j \int_{a_j}^{a_{j+1}} [F(a_j) - F(x)]^2 dx \\ &:= Var_{opt} + Bias^2. \end{aligned} \quad (5.3)$$

We can see that  $Var_{opt} = O(1/n)$  and

$$Bias^2 \doteq \frac{\Delta^3}{3} \sum_j f^3(a_j),$$

where  $f$  is the pdf of  $F$ .

For the up-and-down method, when  $n$  is large,

$$\begin{aligned} MSE &\geq \frac{\Delta}{n} \left[ \sum_j \frac{F(a_j)(1 - F(a_j))}{\pi_j} \right]^2 + \sum_j \int_{a_j}^{a_{j+1}} [F(a_j) - F(x)]^2 dx \\ &:= Var_{ud} + Bias^2. \end{aligned} \tag{5.4}$$

In Tables 5.1–5.3, we list values for  $Var_{opt}$ ,  $Var_{ud}$  for the probit, the logit, and the double exponential cases and various values of  $k$ .

Table 5.1 MSE: optimal fixed design  
vs up-and-down: logit

$k$	$nV_{opt}$	$nV_{ud}$	bias <sup>2</sup>	$\Delta$
4	2.186	2.834	103.725	5.18
6	2.840	3.473	43.770	3.45
8	3.716	25.085	23.791	2.59
10	4.638	1072.072	14.873	2.07
12	5.569	78965.15	10.159	1.73

Table 5.2 MSE: optimal fixed design  
vs up-and-down: probit

$k$	$nV_{opt}$	$nV_{ud}$	bias <sup>2</sup>	$\Delta$
4	1.274	1.378	.4400	2
6	1.916	2.059	.1684	1.33
8	2.564	5.6295	.0940	1
10	3.211	49.301	.060	.8
12	3.858	1010.31	.0417	.6667

Table 5.3 MSE: optimal fixed design  
vs up-and-down: double exponential

$k$	$nV_{opt}$	$nV_{ud}$	bias <sup>2</sup>	$\Delta$
4	1.891	2.578	84.39	4.83
6	2.244	3.133	35.79	3.22
8	2.788	39.154	19.554	2.416
10	3.397	2596.03	12.265	1.993
12	4.032	268590.7	8.394	1.161

For the up-and-down method,  $\pi_0 = \pi_k \rightarrow 0$  as  $k \rightarrow \infty$ , and this is reflected in the estimation of  $F(a_j), j = 0, \dots, k$ . This means that very few resources are allocated to estimate the tail part of  $F$  if  $k$  is large. Its effect can be seen from Tables 5.1-5.2 for the  $nV_{ud}$  entries when  $k$  is more than 8. In practical terms, while the up-and-down method can provide some clues about the goodness of fit of certain models, this analysis can only be trusted, roughly speaking, for the central part of the distribution  $F$ . From these tables, it is clear that  $k = 6$  to 8 is about optimal, and that data points can be used to construct a proper QQ-plot. A formal goodness of fit test can be suggested based on the up-and-down data, but its null distribution is difficult to derive. In conclusion, we suggest a simple QQ-plot of  $\hat{F}(a_j)$  against  $F(a_j)$ , which should provide us with some clues about the underlying model. Since it is fair to say that the up-and-down method does not include the goodness of fit feature in its original design, we should be satisfied with the limited side benefit of a rough QQ-plot. Although not powerful enough, in fact not powerful at all, we can at least examine the shape of  $F$  at points  $F(a_j), j = 0, 1, \dots, k$  for  $k = 6$  to 8. None of the optimal designs we have discussed so far enables us to do the same.

To illustrate, let us consider the data set presented in Table 5.4, which consists of  $n = 240$  observations taken from a Bruceton test conducted by the 4-th Quality Assurance Group of the Chung-Shan Institute of Science and Technology.

In Table 5.5 below, we calculate the MLE for  $F(a_j)$ , which is listed as  $\hat{F}_j(a_j)$ . Note that the MLE of  $F$  is not even monotonically non-decreasing.

Table 5.4 Data from a Bruceton test

0.51	0.49	0.47	0.45	0.47	0.45	0.43	0.41	0.39	0.41
0.43	0.45	0.47	0.45	0.47	0.49	0.47	0.49	0.51	0.53
0.51	0.53	0.51	0.49	0.51	0.49	0.47	0.49	0.47	0.45
0.43	0.45	0.47	0.49	0.51	0.53	0.51	0.53	0.51	0.49
0.51	0.49	0.47	0.45	0.43	0.45	0.43	0.45	0.47	0.45
0.43	0.45	0.47	0.45	0.47	0.45	0.43	0.41	0.43	0.41
0.43	0.45	0.43	0.45	0.43	0.41	0.43	0.45	0.47	0.45
0.43	0.45	0.43	0.41	0.39	0.41	0.43	0.45	0.43	0.41
0.39	0.41	0.39	0.41	0.39	0.37	0.39	0.37	0.35	0.37
0.39	0.41	0.43	0.45	0.47	0.49	0.47	0.49	0.47	0.45
0.47	0.49	0.51	0.49	0.51	0.53	0.51	0.49	0.47	0.49
0.51	0.53	0.51	0.49	0.51	0.49	0.51	0.49	0.51	0.53
0.55	0.57	0.59	0.57	0.55	0.53	0.51	0.49	0.51	0.49
0.51	0.49	0.47	0.49	0.47	0.49	0.51	0.53	0.51	0.49
0.47	0.49	0.51	0.53	0.51	0.53	0.51	0.53	0.51	0.49
0.51	0.49	0.51	0.53	0.51	0.49	0.51	0.49	0.51	0.49
0.51	0.53	0.51	0.49	0.51	0.49	0.51	0.49	0.51	0.49
0.51	0.49	0.51	0.53	0.51	0.49	0.47	0.49	0.51	0.53
0.51	0.53	0.55	0.53	0.51	0.53	0.51	0.53	0.51	0.53
0.51	0.49	0.47	0.49	0.51	0.53	0.51	0.53	0.51	0.53
0.51	0.53	0.51	0.53	0.51	0.53	0.51	0.49	0.51	0.53
0.51	0.49	0.47	0.45	0.47	0.49	0.47	0.45	0.43	0.45
0.47	0.45	0.43	0.41	0.43	0.41	0.39	0.41	0.39	0.41
0.43	0.45	0.47	0.45	0.47	0.45	0.47	0.45	0.47	0.49

We employed the “smooth” function in S-plus of  $\hat{F}(a_j)$  and obtained the values listed in last row of Table 5.4. These smoothed values are now monotonic and the corresponding QQ-plot against the normal distribution (not shown here) appeared to be a straight line.

Does this imply that we can safely use the probit as the proper model? Hardly. Take, for example, the first difference in the last row of Table 5.4; we should, in principle, see an estimate of the corresponding density function. But this one is U-shaped, instead of being bell-shaped, as one

Table 5.5 Estimation of  $F(a_j)$

$j$	0	1	2	3	4	5	6	7	8	9	10	11	12
$n_j$	1	3	9	15	21	29	31	43	54	28	3	2	1
$r_j$	0	1	2	7	8	13	16	16	28	26	2	1	1
$\hat{F}(a_j)$	0	.33	.22	.47	.38	.45	.52	.37	.52	.93	.67	.5	1
smoothed	0	.10	.19	.26	.31	.33	.37	.41	.47	.54	.66	.83	1

would wish. Curve fitting with the quantal response data is an extremely difficult problem, and using data from a Bruceton test does not help either. Even for the case of a fixed design, a very large data set is needed to carry out a reasonable goodness-of-fit study. The only decent effort we know of is the study done by of Xie, Ren and Wang (1987), where  $n = 8440$  and a logistic curve was properly fitted with  $\alpha = 0.05$ .

## 6. Conclusion

Although the up-and-down test is a classical sequential procedure, it is still very much in use in the 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 have investigated the behavior of the up-and-down design from the point of view of data collection and data analysis. For the estimation of  $(\mu, \sigma)$ , it has been shown that a properly selected up-and-down design is quite informative, with an efficiency level of 70% or more for most of the cases we have considered. For the estimation of  $x_p$ , however, the best selected up-and-down method is only about 50% effective as compared with the corresponding  $c$ -optimal design. However, although not particularly useful, the up-and-down method does judge the proper selection of underlying model. In any case, all the quantal response models are rather poor in terms goodness of fit.

## Acknowledgements

This paper is parially supported by a National Science Council grant

under contract NSC 91-CS-7-001-001.

## References

- Anderson, T. W. and Goodman, L. A. (1957). Statistical inference about Markov chains, *Annals of Mathematical Statistics* **28**, 89-110.
- 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.
- Derman, C. (1957). Non-parametric up-and-down experimentation. *Annals of Mathematical Statistics* **30**, 601-605.
- Dixon, W. J. and Mood, A. M. (1948). A method for obtaining and analyzing sensitivity data. *Journal of the American Statistical Association* **43**, 109-126.
- Mood, A. (1998). Personal communication.
- Sittter, 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., Dutten, R. C. and Smith, N. TY. (1996). A measure of association for assessing prediction accuracy that is a generalization of non-parametric ROC area. *Statistics in 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). *Journal of the Royal Statistical Society, 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. *Statistics in Medicine* **14**, 885-893.

- Wu, C. F. J. (1985). Efficient sequential designs with binary data. *Journal of the American Statistical Association* **80**, 974-984.
- Wu, C. F. J. (1988). Optimal designs for percentile estimation of a quantal response curve. In *Optimal Designs and Analysis of Experiments*, edited by Y. Dodge, V.V. Fedorov and H.P. Wynn, North Holland, 213-223.
- Xie, G., Ren, M. and Wang, S. (1987). Critical current distribution of hotwire EED and estimation of distribution parameters by the up-and-down method. *Proceedings of the International Symposium on Pyrotechnics and Explosives* 136-141.
- Young, L. and Easterling, R. (1994). Estimation of extreme quantiles based on sensitivity tests: A comparative study. *Technometrics* **36**, 48-60.

Received December 14, 2001; accepted March 19, 2002

C. D. Fuh  
Institute of Statistical Science  
Academia Sinica  
Taipei, Taiwan

J. S. Lee  
Chemical Systems Research Division  
Chung-Shan Institute of Science and Technology  
Lungtan, Taiwan

C. M. Liaw  
Electronic Systems Research Division  
Chung-Shan Institute of Science and Technology  
Lungtan, Taiwan