HERCMA 2001
E. A. Lipitakis (Editor)
2001, L.E.A.
Printed in Hellas

A CLASSIFICATION OF STATISTICAL APPROACHES TO EXPERIMENTAL CARCINOGENESIS

Aristea Limakopoulou and Evdokia Xekalaki*

Abstract – **The interest in the assessment of cancer risk has led to various developments mainly based on biological basis and supported by statistical analysis. The aim of this paper is to discuss the bioassay of experimental carcinogenesis from a statistical point of view. Biological aspects (e.g., cell proliferation, mechanisms of inhibition in mutagenesis and carcinogenesis) of a cancer risk assessment are not taken up. The problem is viewed as a dose response problem and different models are assumed. In particular, several of the statistical models considered in the literature in the area of experimental carcinogenesis are discussed with an emphasis on dose dependent models. Moreover, an optimal experimental design approach for this particular bioassay is examined considering D – optimality as a design criterion and employing a stochastic method of approximation. Further, as the main problem in experimental carcinogenesis is the low – dose extrapolation and prediction due to the fact that animal experiments can not be applied directly to study low concentrations, an optimal sequential design approach is developed to estimate the parameter under investigation. Estimates of the appropriate percentiles of the risk function are obtained via simulation.**

Keywords and Phrases – **Experimental Carcinogenesis; Tolerance distribution models; Optimal designs; Stochastic approximation; Risk assessment.**

I. Introduction

The quantitative description of the process of carcinogenesis is a rational basis for the assessment of cancer risk due to exogenous agents. Research in carcinogenesis is a field of interest for biological and medical scientists and also for mathematicians and statisticians who analyse data dealing with pre – malignant and malignant lesions. There are two main reasons for formulating models of carcinogenesis. One is to provide a framework for evaluating the consequences of proposed mechanisms of carcinogenesis. The other is to help determine allowable concentrations of known carcinogens in the environment, and to estimate the consequences of exceeding them. This is necessary because animal experiments must be done at concentrations high enough to cause some of the animals to develop tumours, while environmental concentrations must be low enough to produce very few tumours in man. Thus, apart from the great difficulties due to interspecies differences, animal experiments cannot be used directly to study low concentrations. Therefore, some model theory is needed to extrapolate the dose – response relationships downward from the high doses used in animal experiments to the low doses to be allowed in the environment.

The aim of this paper is to discuss the bioassay of experimental carcinogenesis from a statistical point of view. Biological aspects (e.g., cell proliferation, mechanisms of inhibition in mutagenesis and carcinogenesis) of a cancer risk assessment are not taken up. The problem is viewed as a dose response problem and different models are assumed. In particular, the paper is organised as follows: in section II the so called *tolerance distribution models*, which allow for an empirical approach to analyse the modest amount of data in a carcinogenic bioassay, are presented. Tolerance distribution models assume that each individual in a population has an exposure threshold above which cancer will result. These thresholds vary according to some distributions across individuals. Different models assume different shapes of the tolerance distribution. As an example, in section III the Weibull model is used to estimate tumorigenic potency based on the results from laboratory experiments [4]. In section IV, the experimental design problem when the underlying model describing the physical phenomenon is assumed to be non – linear in terms of the parameters involved, is discussed and the sequential principle of design has been adopted to overcome this parameter dependence. Moreover, this section aims at evaluating the percentiles of the risk function derived through a dose – response relationship in a multistage model. These percentiles are known as *virtual safe dose* levels or *risk specific dose* levels. Therefore, the optimal design theory is applied to estimate the appropriate percentile and the sequential approach of design is adopted through a stochastic approximation scheme. If the initial design is D – optimal, the limit design is D – optimal as well and it is the one with the minimum entropy. Finally, in section V, a simulation study is performed. Binary response data are generated following the one – hit model, with known parameter value. A static design is constructed. All the observations are obtained on the basis of one selected point. Different sample sizes are used and an initial value of the parameter, being "far" from the true value is chosen. The corresponding percentiles of the risk function are evaluated. The obtained results produce strong evidence that the method considered yields values close to the true ones.

* ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS, Patission 76, 104 34, ATHENS, GREECE

II. STATISTICAL MODELS IN CARCINOGENESIS.

A number of models have been suggested to describe the process of carcinogenesis, i.e. the process by which a normal cell can be transformed to a malignant one. Cancer is affecting the growth rate of affected tissues in which the control mechanisms of cells become altered and the cells divide to form neoplasmatic growths or tumours. The cause of cancer is studied through exposure to carcinogens, i.e. substances which are cancer – causing agents. Typical examples of carcinogens are the nicotine and mustard gas in the presence of certain micro - organisms.

In principle, there are two classes of models depending on whether dose levels or time are treated as random variables. Both classes are based on the fundamental assumption that the probability of developing cancer is likely to increase with increasing doses of carcinogen. That is if X and T represent the random variables of the dose level of carcinogen and the time at which an individual develops a tumour, respectively, then the probability of developing cancer $P(x,t)$ can be represented either as

$$P(x,t) = P(X \le x \mid t)$$

or as

$$P(x,t) = P(T \le t \mid x) \qquad (2.1)$$

This dual representation implies that the probability that an individual can develop tumour, at dose level x by a particular time t, is restricted either on time or on dose. In this study the interest has been focused on restriction on time.

III. DOSE DEPENDENT MODELS

In the preceding section the class of statistical models in experimental carcinogenesis is the union of two subclasses of models:

Those which are considering the given dose level for a given time period.

Those which are considering the time of effect and the given dose.

Consider that a nutshell (i.e., a tumour) occurs at dose $X = x$ if the individual's resistance (tolerance) is broken at x. Then the excess tumour risk is given by

$$F(x) = P(X \le x) = P(\text{"Tolerance"} \le x)$$

This probability is precisely what one seeks to model. In particular, it is assumed that there is a statistical model that approximates the cumulative distribution function $F(.)$, which is termed the *tolerance distribution function*. Then, the dose level x is linked with a binary response variable (success or failure), thus

$$Y_i = \begin{cases} 1, & \text{success with probability } F(x) \\ 0, & \text{failure with probability } 1 - F(x) \end{cases}$$

The tolerance distribution function is usually indexed by a vector of parameters θ, i.e. $F(x;\theta)$, $\theta \in \Theta \subseteq \Re^n$. (For a detailed discussion on dose response models, see [10]).

The *Weibull model* is an alternative model proposed for the tolerance distribution with a shape parameter. It is also based on a set of assumptions on how cancer develops: a tissue sustains "hits" at random; cancer occurs when a portion of the tissue sustains a fixed number of "hits"; cancer is observed when the first such portion has sustained the required number of "hits". The resulting distribution function is

$$F(x) = 1 - e^{-(\theta x)^{\kappa}} \qquad (3.1)$$

The Weibull model can exhibit a dose-response relationship that is either sub-linear (shape parameter $\kappa > 1$) or supra-linear ($\kappa < 1$), and has a point of inflection at $x = \left(\frac{\kappa - 1}{\kappa}\right)^{\frac{1}{\kappa}}$. In practice, both parameters may be unknown. In such cases maximum likelihood estimates are commonly used on the basis of a censored sample. In particular, these are obtained through maximizing the corresponding likelihood function that can be described as

$$L(\kappa,\theta) = \prod_{i=1}^{n} h(t_i) \prod_{i=1}^{n} s(t_i), \qquad (3.2)$$

where $h(.)$ denotes the hazard function associated with $F(.)$ defined as the conditional density function at time t given survival up to time t, i.e., $h(t) = \kappa\, \theta^{\kappa}\, t^{\kappa-1}$, and $s(.)$ denotes the corresponding survivor function, i.e.,

$$s(t) = \exp(-(\theta t)^{\kappa}).$$

Then, the corresponding log likelihood is given by

$$\begin{aligned} \ell = \ell(\kappa,\theta) &= \log L(\kappa,\theta) \\ &= d \log \kappa + \kappa d \log \theta + (\kappa - 1)\sum_{i=1}^{n} \log t_i - \theta^{\kappa} \sum_{i=1}^{n} t_i^{\kappa}, \end{aligned} \qquad (3.3)$$

where n is the number of observations, d is the total number of failures (uncensored) and $\sum \log t_i$ and $\sum t_i^{\kappa}$ are functions of the censored and uncensored failure times. Often $\sum t_i$ is called *the total time at risk*.

Even in the absence of censoring, there is no fixed dimensional sufficient statistic for (κ,θ); the Weibull distribution does not belong to the exponential family of models, where the total number d of failures and the total $\sum t_i$ of the censored and uncensored failure times form a minimal sufficient statistic for θ. This is so as the maximum likelihood estimate θ^* of θ is $\theta^* = \frac{d}{\sum t_i}$, namely the total number of failures divided by the total time at risk and censored failure times contribute to the denominator but not to the numerator of this ratio. Thus, when there is no censoring, the log likelihood for the exponential distribution becomes $l = n \log \theta - \theta \sum t_i$, , and the curved exponential family collapses to a full one-dimensional family with a single minimal sufficient statistic $\sum t_i$ for θ.

The first derivatives of the form (3.3) are

$$U_{\theta} = \frac{\partial \ell}{\partial \theta} = \frac{\kappa d}{\theta} - \kappa \theta^{\kappa - 1} \sum t_i^{\kappa}. \qquad (3.4)$$

$$U_\kappa = \frac{\partial \ell}{\partial \kappa} = \frac{d}{\kappa} + d\log\theta + \sum \log t_i - \theta^\kappa \sum t_i^\kappa \log(\theta t_i). \quad (3.5)$$

If κ is specified, the maximum likelihood estimator θ^* of θ can be found explicitly by solving the equation $U_\theta = 0$ thus

$$U_\theta = 0 \Leftrightarrow \frac{\kappa d}{\theta} - \kappa\theta^{\kappa-1}\sum t_i^\kappa = 0 \Leftrightarrow \frac{d}{\theta} = \theta^{\kappa-1}\sum t_i^\kappa$$

$$\Leftrightarrow d = \theta^\kappa \sum t_i^\kappa \Leftrightarrow \theta^\kappa = \frac{d}{\sum t_i^\kappa}.$$

Hence the maximum likelihood estimator (MLE) of θ is

$$\theta^* = \left(\frac{d}{\sum t_i^\kappa}\right)^{\frac{1}{\kappa}}. \quad (3.6)$$

(This result could be alternatively be derived as an immediate consequence of the fact that T^κ has an exponential distribution with parameter θ^κ, where T is the random variable of the exponentially distributed failure times). Substitution of the value of θ as given by (3.6) in the equation $U_\kappa = 0$ yields

$$0 = \frac{d}{\kappa} + \sum \log t_i - d\frac{\sum t_i^\kappa \log t_i}{\sum t_i^\kappa}, \quad (3.7)$$

which leads to the maximum likelihood estimator κ^* of κ. Equation (3.7), though non linear, does not contain θ and therefore can be solved by a one-dimensional iterative scheme in κ.

The second derivatives of the log-likelihood l are given by

$$I_{\theta\theta} = \frac{\partial^2 \ell}{\partial\theta^2} = \frac{\partial}{\partial\theta}\left(\frac{\kappa d}{\theta} - \kappa\theta^{\kappa-1}\sum t_i^\kappa\right) \quad (3.8)$$

$$= -\frac{\kappa d}{\theta^2} - \kappa(\kappa-1)\theta^{\kappa-2}\sum t_i^\kappa,$$

$$I_{\theta\kappa} = \frac{\partial^2 \ell}{\partial\theta\partial\kappa} \quad (3.9)$$

$$= \frac{d}{\theta} - \theta^{\kappa-1}(\ell + \kappa\log\theta)\sum t_i^\kappa - \kappa\theta^{\kappa-1}\sum t_i^\kappa \log t_i,$$

$$I_{\kappa\kappa} = \frac{\partial^2 \ell}{\partial\kappa^2} = -\frac{d}{\kappa^2} - \theta^\kappa \sum t_i^\kappa (\log(\theta t_i))^2. \quad (3.10)$$

Therefore, the information matrix $I(\kappa, \theta)$ can be evaluated as follows:

$$I = I(\kappa,\theta) = \begin{bmatrix} I_{\kappa\kappa} & I_{\kappa\theta} \\ I_{\theta\kappa} & I_{\theta\theta} \end{bmatrix} \quad (3.11)$$

As an example, consider the data set of *Table I* on the times of remission (weeks) of leukemia patients.

Sample 0 (drug 6-MP)	6*, 6, 6, 6, 7, 9*, 10*, 10, 11*, 13, 16, 17*, 19*, 20*, 22, 23, 25*, 32*, 32*, 34*, 35*
Sample 1 (control)	1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

*censored

TABLE I
TIMES OF REMISSION (WEEKS) OF LEUKAEMIA PATIENTS ([4]).

For the treated group (Sample 0), the following results are obtained:

$d = 9, \sum t_i = 359, \sum \ln t_i = 55.883, \sum \log t_i = 24.27,$

$\sum t_i \log t_i = 1077.228, \sum t_i (\ln t_i)^2 = 3334.778.$

Following Limakopoulou and Xekalaki [10], it follows from (3.7) that the maximum likelihood estimate of κ is $\kappa^* = 1.35$.

Then, the corresponding maximum likelihood estimate of θ is

$$\theta^* = \left(\frac{d}{\sum t_i^{1.35}}\right)^{\frac{1}{1.35}} = \left(\frac{9}{1043.2}\right)^{\frac{1}{1.35}} =$$

$$(0.008627)^{\frac{1}{1.35}} = 0.029581 = 0.03.$$

Figure I depicts the situation for this particular data set.

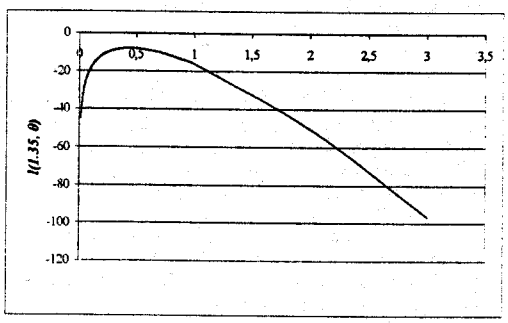

FIGURE I
LEUKEMIA DATA, 6-MP GROUP. LOG-LIKELIHOOD PROFILE FOR THE PARAMETER θ OF THE WEIBULL TOLERANCE DISTRIBUTION WHEN $\kappa = 1.35$.

Thus, the corresponding information matrix is

$$I = I(1.35, 0.03) = \begin{bmatrix} -8.689 & 2717.859 \\ 2717.859 & -14004.453 \end{bmatrix}.$$

The hazard function can be evaluated as

$$h(t) = 0.03 \times 1.35 t^{1.35-1} = 0.04 t^{0.35}.$$

At this point, it should be noticed that the above application does not correspond to a designed experiment. It merely comprises an illustrative example of the use of the Weibull model that might provide initial estimators for a designed experiment in the context of a Weibull model. In

the sequel, the main aspects of the optimal experimental design approach are described and applied in the case of the one – hit model, that can be considered as a special form of the Weibull model with a shape parameter $\kappa = 1$.

IV. AN OPTIMAL DESIGN APPROACH FOR BIOASSAYS.

In the present section, the experimental carcinogenesis is considered from the angle of a *designed* experiment, known as *bioassay*. It is not always certain that a designed experiment has been *optimally*. Clearly, one has to clarify what is meant by the term *optimal*. Therefore, the main points of optimal experimental design refer to the binary problem, which is of central importance in survival analysis, and, thus in experimental carcinogenesis.

In the sequential design approach for non – linear models, the average per observation information matrix depends on ξ, the design measure, and θ, the unknown parameter vector. Therefore, $M(\theta,\xi)$, the average per observation information matrix, cannot be a reliable approximation of the variance – covariance matrix, unless an estimate of θ, say θ^*, is obtained so that $M(\theta,\xi)$ to be approximated by $M(\theta^*,\xi)$. In order to perform an experiment, a "good" estimate of θ might have been known or might have not been suggested. Thus, the experiment should be designed in an optimal sense with an initial guess about θ. For such an initial estimate, a D–optimal design may have to be performed. The initial estimate is used to redefine the design points, which in the non – linear case depend on θ. After the experiment has been performed, a new estimate is obtained. The procedure is repeated until a given degree of accuracy is attained either with respect to the estimate of θ, e.g.

$$\left|\theta_n^* - \theta_{n+1}^*\right| < \varepsilon, \qquad (4.1)$$

or with respect to the measure of $M(\theta, \xi)$, e.g.

$$\left|\log\det M(\theta_n,\xi) - \log\det M\left(\theta_{n+1}^*,\xi\right)\right| < \varepsilon. \qquad (4.2)$$

For D – optimality and other optimality criteria, this method seems to be satisfactory and has been widely adopted in practice. There exists also a variety of other sequential procedures of which the most commonly used is that introduced by Robbins and Monro [13]. In particular, they proposed a recursive scheme to estimate the root, r say, of a real valued function Q defined on the real axis $\Re$, i.e. $Q(r)=p$ when only observations on Q, denoted by $y_n = y_n(x)$, are provided. Actually, $y_n = Q(x_n) + e_n$, with e_n being the error terms with a distribution with zero mean and variance σ^2. The Robbins and Monro recursive scheme known as *Stochastic Approximation* is the statistical analogue of a Newton – Raphson iterative scheme. Iterations are made though the sequence

$$x_{n+1} = x_n - \alpha_n(y_n - p), \quad n = 1,2,... \qquad (4.3)$$

with x_1 having an arbitrary value, $\alpha_n > 0$ a fixed sequence of real numbers (step sizes) and y_n being binary responses. The sequence x_n converges to r in mean square, where a typical step size of $\alpha_n = cn^{-\alpha}$, $\alpha \le 1$, $n = 1, 2, \ldots$ is assumed. The minimum variance estimate is obtained with $c = C_{opt} = (Q'(r))^{-1}$, and $\alpha = 1$ with value $\sigma^2_{min} = (Q'(r))^{-1}$.

If the function Q is not known, it can be approximated. Consider that the function Q given through one of the models defined in the previous sections. The equation $Q(x) = F(x) - p$ with $p \in [0, 1]$ has to be solved. The root of this equation is the p th percentile L_p, of the dose response curve F(x). Moreover, as was pointed out by Ford et al.[3], stochastic approximation leads to a fully sequential design, which is *D - optimal* in the limit.

In the sequel, the problem of estimating the low – dose percentiles is considered by combining the sequential nature of the design with a stochastic approximation scheme. This is demonstrated in the framework of a static design approach: i.e. an experiment performed only once. Assuming a Weibull model as defined by (3.1), one can easily obtain

$$F(L_p) = 1 - \exp\left(\left(-\theta L_p\right)^{\kappa}\right),$$

or, equivalently,

$$L_p = \left(-\frac{1}{\theta^{\kappa}}\ln(1-p)\right)^{\frac{1}{\kappa}}. \qquad (4.4)$$

The value defined by (4.4) is referred to as the static estimate of the percentile point L_p for the Weibull distribution. It should be noted that knowledge of the values of both parameters θ and κ is required.

The procedure may be illustrated using the data of *Table I* by recalling that the maximum likelihood estimates of the parameters κ and θ are $\kappa^* = 1.35$ and $\theta^* = 0.03$, respectively. Applying formula (4.4), the values of L_p are computed and the corresponding results are summarized in *Tables II* and *III*.

The sensitivity of the estimators of L_p to small changes in the values of θ and κ, within the range of their respective confidence intervals, has been studied by Limakopoulou and Xekalaki [10].

P	$\kappa = 1.35$			$\theta = 0.03$		
	θ_1=0.029	θ=0.030	θ_2=0.031	κ_1 1.30	κ = 1.35	κ_2 =1.40
0.001	0.08824	*0.08429*	0.08064	0.07345	*0.08429*	0.09685
0.002	0.17656	*0.16866*	0.16136	0.14698	*0.16866*	0.19381
0.003	0.26497	*0.25312*	0.24216	0.22058	*0.25312*	0.29085
0.004	0.35348	*0.33766*	0.32304	0.29426	*0.33766*	0.38800
0.005	0.44207	*0.42229*	0.40400	0.36801	*0.42229*	0.48524
0.006	0.53075	*0.50700*	0.48505	0.44183	*0.50700*	0.58259
0.007	0.61951	*0.59180*	0.56617	0.51573	*0.59180*	0.68002
0.008	0.70837	*0.67668*	0.64738	0.58970	*0.67668*	0.77756
0.009	0.79732	*0.76165*	0.72867	0.66375	*0.76165*	0.87520
0.010	0.88636	*0.84671*	0.81004	0.73787	*0.84671*	0.97293
0.015	1.33290	*1.27327*	1.21814	1.10960	*1.27327*	1.46309
0.020	1.78172	*1.70201*	1.62831	1.48323	*1.70201*	1.95574
0.025	2.23283	*2.13294*	2.04058	1.85877	*2.13294*	2.45092
0.030	2.68626	*2.56609*	2.45497	2.23624	*2.56609*	2.94863
0.035	3.14203	*3.00147*	2.87150	2.61565	*3.00147*	3.44892
0.040	3.60017	*3.43912*	3.29020	2.99704	*3.43912*	3.95181
0.045	4.06070	*3.87905*	3.71108	3.38043	*3.87905*	4.45733
0.050	4.52366	*4.32129*	4.13417	3.76582	*4.32129*	4.96550

TABLE II

COMPUTATION OF THE L_p (4.4) FOR DIFFERENT VALUES OF P AND FOR DIFFERENT VALUES OF θ AND κ.

p	$\kappa_1 = 1.30$		$\kappa_2 = 1.40$	
	$\theta_1 = 0.029$	$\theta_2 = 0.031$	$\theta_1 = 0.029$	$\theta_2 = 0.031$
0.001	0.07676	0.07039	0.10156	0.09251
0.002	0.15360	0.14085	0.20323	0.18511
0.003	0.23052	0.21138	0.30499	0.27780
0.004	0.30752	0.28198	0.40686	0.37059
0.005	0.38459	0.35265	0.50883	0.46347
0.006	0.46174	0.42339	0.61090	0.55645
0.007	0.53897	0.49421	0.71308	0.64951
0.008	0.61627	0.56509	0.81536	0.74267
0.009	0.69365	0.63605	0.91774	0.83593
0.010	0.77112	0.70708	1.02022	0.92928
0.015	1.15960	1.06330	1.53421	1.39744
0.020	1.55006	1.42133	2.05080	1.86799
0.025	1.94252	1.78120	2.57005	2.34095
0.030	2.33700	2.14292	3.09196	2.81633
0.035	2.73351	2.50650	3.61656	3.29418
0.040	3.13208	2.87198	4.14390	3.77450
0.045	3.53274	3.23936	4.67398	4.25734
0.050	3.93550	3.60867	5.20685	4.74271

TABLE III
COMPUTATION OF THE L_p (4.4) FOR DIFFERENT VALUES OF p AND FOR DIFFERENT VALUES OF θ AND κ.

Furthermore, using (4.4), a sequential estimator of L_p can be constructed. In particular, setting $C_{opt}^{-1} = F'(L_p)$, where F'(.) denotes the first derivative of F(.) as defined by (3.1), one obtains from (4.4)

$$C_{opt_1}^{-1} = F'(L_p) = \exp\left(-(\theta L_p)^{\kappa}\right)\theta^{\kappa}\kappa L_p^{\kappa-1} = (1-p)\theta^{\kappa}\kappa L_p^{\kappa-1}. \quad (4.5)$$

Thus, an iterative procedure leading to an estimator of L_p in the case of the Weibull model can be formulated on the basis of the recurrence relationship

$$L_{p,n+1} = L_{p,n} - (nr_{\kappa})^{-1}(y_n - p), \quad n = n_0 + 1, n_0 + 2, \ldots \quad (4.6)$$

where y_n denotes the number of respondents (binary response data y_n = 1 or 0) among the individuals tested (usually animals) after a predetermined time when treated at dose level x and $r_{\kappa} = F'(L_p) = (1-p)\theta^{\kappa}\kappa L_p^{\kappa}$.

In order to get the estimates of the Weibull model and to proceed, n_0 observations can be denoted. So, for example, in the case of the one – hit model (κ = 1), assuming a typical step size of the form $a_n = cn^{-a}$, the iterative scheme (4.6) yields

$$C_{opt_1}^{-1} = F'(L_p) = \theta(1-p) \quad (4.7)$$

Letting $a = 1$ and using (4.6) one obtains

$$L_{p,n+1} = L_{p,n} - \left(n\theta_{n_0}(1-p)\right)^{-1}(y_n - p) \quad (4.8)$$

The sequence $\{L_{p,n}\}$ of the values of L_p in the n iterations converges, in mean square, to L_p, the root of the equation $Q(L_p) = F(L_p) - p$. As θ is not known, the n_0 initial observations can be used at the first stage to estimate θ using θ_{n_0}.

Therefore, (4.8) is reduced to

$$L_{p,n+1} = L_{p,n} - \left(n\theta_{n_0}(1-p)\right)^{-1}(y_n - p), n = n_0 + 1, n_0 + 2, \ldots \quad (4.9)$$

and $L_{p,n}$ converges to L_p in mean square.

V. SOME SIMULATIONS RESULTS.

In the sequel, the sequential design discussed in the previous section for estimating the percentile L_p is illustrated on simulated data from an one – hit model. In particular, using an algorithm developed by Limakopoulou and Xekalaki [10], N = 1000 samples of binary response data y_i = 1 or 0, i = 1,2, ..., n were generated from an one – hit model, with a *"true"* value of θ equal to 3.18 for various values of n in the framework of a static design. All the observations (x, y_i), i = 1, 2, ..., n have been based on the D – optimal point x = 1.59 / θ. The value 1.7 was considered as an initial guess for θ in order to be *"far"* from the true value. Different values of p have been selected, p = 0.01, 0.02, 0.04 and the sample sizes considered were n = 100, 500.

	n = 100			n=500		
Values of p	0.01	0.02	0.04	0.01	0.02	0.04
*Average value of θ (θ^*)	3.34	3.29	3.30	3.25	3.24	3.25
10^{-4} x Average value of L_p ($10^{-4}L_p^*$)	37	70	120	34	64	115
Mean Square Error of θ^*	0.18	0.14	0.15	0.05	0.04	0.05
10^{-4} x Mean Square Error of L_p^*	6.80	7.16	6.93	2.84	2.83	2.85

* Out of N simulations.

TABLE IV
RESULTS OF A SIMULATION STUDY FOR THE ONE – HIT MODEL.

The first stage estimate, θ_{n_0}, of the parameter θ, was based on n_0 = 50 observations. The estimates of θ and L_p, for various values of p along with their mean square errors are reported in *Table IV*. For each value of p, the estimates of θ and L_p were obtained as the averages of their values in the 1000 replications. The corresponding mean square errors denoted by MSE(θ^*) and MSE(L_p^*) respectively were evaluated through the formulae

$$MSE(\theta^*) = \frac{1}{N}\sum_{i=1}^{N}(\theta^{(i)} - \theta^*)^2$$

and

$$MSE(L_p^*) = \frac{1}{N}\sum_{i=1}^{N}(L_p^{(i)} - L_p^*)^2.$$

The *"true"* value of L_p can be evaluated from

$$L_p = -\frac{1-p}{\theta},$$

for any given value of p and θ = 3.18. From *Table IV*, one may note that, when n is increased, the mean square errors of θ^* and L_p^* decrease.

Although samples of size 100 may at times be considered as *"small"*, they are in general regarded as reasonable in a biological context where their elements

usually are responses from an experiment on rats or from the treatment of patients.

VI. CONCLUDING REMARKS

As the assessment of cancer risk has been studied in detail from the medical, toxicological and biological points of view, the main objective of the paper has been to provide an insight into this problem from the statistical point of view proposing a sequential bioassay for the estimation of low – dose exposure, i.e. low – dose percentiles. The optimal experiment design has been adopted and the sequential principle has been considered.

Of course, there exists a large number of chemicals. At least 50000 have been introduced in the human environment and only a few hundred have been submitted to thorough carcinogenic testing in experimental animals. Moreover, the mechanisms of tumour promotion are far from being completely unravelled, although significant progress has been made in this direction. Nowadays, interest is shifting from initiating towards promoting agents in carcinogenesis and, therefore, the mechanisms involved in the promotion of cancer are also becoming more and more obvious. Promotion proceeds through a wide array of different pathways although resulting in the same outcome. Thus, in carcinogenesis and in its promotion the existent underlying mechanism is sometimes attempted to be interpreted through different statistical models. For this reason, these models are working as the "right hand" for the experimenter who tries to fit the appropriate model to the collected data set, usually generated through a simulation study as it is rather difficult to obtain (or even impossible to "create") a real data set. The experimental situation provides data that hopefully can be transformed to humans.

The inability to measure carcinogenic risks at very low exposure levels (doses) of chemicals precisely and accurately has prompted the use of mathematical models for extrapolating risks from the high range that is observable in humans or animals to the low range that is of general public health concern. This low – dose extrapolation continues to be a highly controversial issue in quantitative risk assessment.

REFERENCES

[1] Ford, I., Kitsos, C. and Titterington, D.M. (1989). Recent Advances in Nonlinear Experimental Design. *Technometrics*, 31, 49-60.

[2] Gehan, E.A. (1965). A generalized Wilcoxon test for comparing arbitrarily single-censored samples. *Biometrika*, 52, 203-223.

[3] Limakopoulou, A. and Xekalaki, E. (2001). A Classification of Statistical Approaches to Experimental Carcinogenesis. *Technical Report*, 127, Department of Statistics, Athens University of Business and Economics, February 2001.

[4] Robbins, H. and Monro, S. (1951). A Stochastic Approximation Method. *Annals of Mathematical Statistics*, 22, 400-407.