LEARNING CURVE ESTIMATION TECHNIQUES FOR NUCLEAR INDUSTRY

Jussi K. Vaurio

Fast Reactor Safety Technology Management Center
Argonne National Laboratory
9700 South Cass Avenue
Argonne, IL USA 60439

ABSTRACT

Statistical techniques are developed to estimate the progress made by the nuclear industry in learning to prevent accidents. Learning curves are derived for accident occurrence rates based on actuarial data, predictions are made for the future, and compact analytical equations are obtained for the statistical accuracies of the estimates. Both maximum likelihood estimation and the method of moments are applied to obtain parameters for the learning models, and results are compared to each other and to earlier graphical and analytical results. An effective statistical test is also derived to assess the significance of trends. The models used associate learning directly to accidents, to the number of plants and to the cumulative number of operating years. Using as a data base nine core damage accidents in electricity-producing plants, it is estimated that the probability of a plant to have a serious flaw has decreased from 0.1 to 0.01 during the developmental phase of the nuclear industry. At the same time the frequency of accidents has decreased from 0.04 per reactor year to 0.0004 per reactor year.

I. INTRODUCTION

Several statistical techniques are developed in this paper for estimating and predicting the likelihood of severe core damage accidents in nuclear power plants. The learning curves are based on actuarial data. Compact analytical expressions are obtained for the statistical accuracies of the estimated accident rate parameters.

The first set of models assumes that learning is associated with the total experience so that the accident rate is a function of the accumulated operating time of all plants. A statistical test is developed to verify that learning actually has taken place, and then both maximum likelihood and moments methods are used to calculate event rate parameters and their accuracies from the time intervals between accidents. This part can be viewed as an analytical extension of the graphical technique of Ref. 1.

The second model assumes that the accident rate is constant between observed events, but the rate is reduced at the time of an event by a certain factor due to lessons learned from the event. This part extends the techniques of Refs. 2 and 3.

The third model assumes that learning is not directly related to time but to the number of plants designed and built. This view is supported by the fact that severe accidents have occurred early in the life of a plant rather than uniformally at a random time. Results obtained with this model indicate somewhat less profound learning than time related models.

II. LEARNING WITH TIME

A number of accidents with serious damage to the plant but benign off-site consequences have occurred at nuclear power plants in the U.S. as well as worldwide. Core damage accidents that have occurred in commercial experimental and developmental electricity-producing plants are listed in Table I together with the cumulative time of occurrence.[1] The total number of plants built by the time of each event is also given in Table I.[5]

Table I. Core Damage Accidents in Electricity-producing Plants

Reactor Incident	Cumulative Reactor Years	Incremental Reactor Years	Cumulative Number of Plants Built	Incremental Number of Plants
EBR-I	5	5	3	3
SRE	45	40	24	21
SL-1	81	36	33	9
FERMI-1	383	302	79	46
LUCENS	561	178	96	17
ST. LAURENT A-1	627	66	97	1
BOHUNICE	1658	1031	223	126
TMI-2	2125	467	252	29
ST. LAURENT A-2	2385	260	270	18
Time of Analysis	2920	535	290	20

Considering this experience as a data base, let the accumulated total operating time of all plants prior to accident i be T_i, and the time between accidents i-1 and i is $t_i = T_i - T_{i-1}$ $(i = 1,2,\ldots,n;\ T_0 = 0)$. Thus, the cumulative number of accidents is a step function in time with increments of one at times T_i and the final value n at the time of observation and analysis T $(T > T_n)$. From this information we wish to estimate what the accident rate (frequency) $\Lambda(T)$ has been as a function of time, and predict what it will be in the future.

Statistical Testing

The incremental reactor years between accidents seem to be increasing "in the average", judging from the data in Table I. However, this appearance could be accidental due to statistical variations of the intervals, even if the accident rate parameter actually is a constant. The first problem is to develop a statistical test that indicates how likely the observed sequence can occur if the actual rate is a constant.

One such test was developed in Ref. 2. Another effective statistic appears to be the ratio of upper and lower sums of accident intervals,

$$z = s_2/s_1 , \tag{1}$$

where

$$s_1 = \sum_{i=1}^{k} t_i ,$$

$$s_2 = \sum_{i=n-m+1}^{n} t_i, \text{ and } m+k \le n.$$

The "null hypothesis," H_0, is that the accident rate is constant, i.e., no learning has taken place. If this were true, the statistic z defined above would tend to be roughly of the order of m/k, the ratio of the number of terms in the sums. However, if significant learning occurs, the later intervals tend to be longer than the early ones and $z >> m/k$.

The event intervals t_i are assumed to be exponentially distributed, with the rate parameter λ constant under the hypothesis H_0. Consequently, the upper and lower sums of t_i's obey gamma probability densities

$$p_\nu(s) = \frac{\lambda(\lambda s)^{\nu-1}}{\Gamma(\nu)} e^{-\lambda s} , \tag{2}$$

where $s = s_1$ with $\nu = k$, and $s = s_2$ with $\nu = m$.

The cumulative distribution function of z can be obtained as

$$P_{k,m}(z) = \int_0^\infty p_m(s_2)ds_2 \int_{s_2/z}^\infty p_k(s_1)ds_1$$

$$= \sum_{j=0}^{k-1} \frac{\Gamma(m+j)}{\Gamma(m)\Gamma(j+1)} \cdot \frac{z^m}{(1+z)^{m+j}} . \tag{3}$$

The complementary cumulative probability $1 - P_{k,m}(z)$ indicates how likely it is to have, under the null hypothesis H_0, a ratio larger than the value observed, z. Thus, if

$$Q_{k,m}(z) = 1 - P_{k,m}(z) \tag{4}$$

is very small, it is taken as an indication that H_0 is not true and real learning has taken place. The probability function $Q_{k,m}(z)$ is given in Fig. 1 for k = m = 1,2,3,4 and 5. For example, if a significance level of 1% is used ($Q_{k,m} = 0.01$), then with k = m = 4 the observed value of z should be larger than 6 to indicate that significant learning has taken place. The worldwide core damage data from Table I with k = m = 4 yields $s_1 = 383$, $s_2 = 1824$ and z = 4.76. Figure 1 indicates that there is only a 2% possibility that no learning has taken place.

For larger values of k and m the null hypothesis (constant event frequency) can be approximately tested by using

$$F = \frac{s_1(m+0.5)}{s_2(k+0.5)} \quad . \tag{5}$$

Under the null hypothesis this F would obey F-distribution with (2k+1, 2m+1) degrees of freedom.

Maximum Likelihood Estimation

After concluding that significant learning has taken place, one can estimate learning parameters and predict accident rates for the future.

A typical time-related learning model assumes that the expected cumulative number of accidents, n, is a function of the total cumulative operating time T,

$$\langle n(T)\rangle = AT^B \quad , \tag{6}$$

with constants A and B determined by the maximum likelihood or moments methods. The accident rate is defined as the derivative of Eq. (6),

$$\Lambda(T) = \frac{d\langle n\rangle}{dT} = ABT^{B-1} \quad . \tag{7}$$

Under the assumption of a nonhomogeneous Poisson process, the probability of the observed sequence of events is proportional to

$$F = \prod_{i=1}^{n} \Lambda(T_i)e^{-\langle n(T)\rangle} \quad . \tag{8}$$

The values of A and B that maximize this are the maximum likelihood (ML) estimates

$$\tilde{B} = 1/\left[\ln(T) - \frac{1}{n}\sum_{i=1}^{n}\ln(T_i)\right] \quad , \tag{9}$$

$$\tilde{A} = n/T^{\tilde{B}} \quad . \tag{10}$$

It is well-known that the ML estimates are asymptotically (with increasing T) unbiased, consistent, minimum variance estimates. It is easy to see from Eqs. (6) and (10) that the ML estimates $\tilde{A}$ and $\tilde{B}$ yield an unbiased estimate of the cumulative number of accidents.

The statistical accuracy of the estimates $\tilde{A}$ and $\tilde{B}$ is of major interest in predicting accident rates for the future. Using the Fisher information matrix[4] the relative variances of $\tilde{A}$ and $\tilde{B}$ can be obtained as

$$\frac{\mathrm{Var}\ (\tilde{A})}{A^2} = [1 + B^2 \ln^2(T)]\,/\langle n(T)\rangle, \tag{11}$$

$$\frac{\mathrm{Var}\ (\tilde{B})}{B^2} = 1/\langle n(T)\rangle\ . \tag{12}$$

The values observed for $\tilde{A}$, $\tilde{B}$ and n(T) can be used for A, B and <n(T)>, respectively, in Eqs. (11) and (12), to estimate the accuracies for practical purposes.

Using the experience summarized in Table I, the ML estimates $\tilde{A}$ = 0.267 ± 0.326 and $\tilde{B}$ = 0.441 ± 0.147 can be obtained with the uncertainties given as ± one standard deviation, obtained from Eqs. (11) and (12).

An estimate or prediction of $\Lambda(T')$ for an arbitrary time T' can be obtained by substituting the estimated values of A and B to Eq. (7). This is presented as the straight line in Fig. 2. The relative variance of the rate $\Lambda(T')$ at time T' is

$$\frac{\mathrm{Var}\ [\tilde{\Lambda}(T')]}{\Lambda^2(T')} = \frac{1 + \left[1 + B\ \ln\left(\frac{T'}{T}\right)\right]^2}{\langle n(T)\rangle}\ . \tag{13}$$

Methods of Moments

The idea in the method of moments is to determine the free fitting parameters A and B so that certain integrals of the analytical rates and numbers of events equal the observed values of the same integrals. For the data of Table I, the observed quantities to be used are

$$T = 2920\ , \tag{14}$$

$$n(T) = 9\ , \tag{15}$$

$$N(T) = \int_0^T n(t)dt = n(T)T - \sum_{i=1}^{n(T)} T_i = 18410\ , \tag{16}$$

and

$$NL(T) = \int_0^T n(t)d\ \ln(t) = n(T)\ \ln(T) - \sum_{i=1}^{n(T)} \ln(T_i) = 20.43\ . \qquad (17)$$

The analytical counterparts of these in terms of A and B are

$$\langle n(T)\rangle = AT^B\ , \qquad (18)$$

$$I(T) = \int_0^T \langle n(t)\rangle dt = \frac{A}{B+1}\ T^{B+1}\ , \qquad (19)$$

and

$$IL(T) = \int_0^T \langle n(t)\rangle d\ \ln(t) = \frac{A}{B}\ T^B\ . \qquad (20)$$

Method 1: The simultaneous conditions $IL(T) = NL(T)$ and $\langle n(T)\rangle = n(T)$ happen to yield results identical to the ML estimates, Eqs. (9) and (10), i.e., $\tilde{A} = 0.267$, $\tilde{B} = 0.441$.

Method 2: The conditions $I(T) = N(T)$ and $\langle n(T)\rangle = n(T)$ yield $\tilde{A} = 0.297$, $\tilde{B} = 0.427$.

Method 3: The previous method used the condition that the expected number of events equals the observation, $\langle n(T)\rangle = n(T)$. However, if the last event happened very close to T, or long before T, the result could be misleading. One can argue that the most correct estimate of n at the time T_n of the last observed event is $n(T) - 1/2$ because just before T_n the number is $n(T) - 1$, and just after T_n it is $n(T)$. Based on this reasoning one could ignore the time after T_n and use the conditions $I(T_n) = N(T_n)$ and $\langle n(T_n)\rangle = n(T) - 0.5$, yielding $\tilde{A} = 0.186$ and $\tilde{B} = 0.491$.

Method 4: Instead of neglecting the time after T_n one can argue that if T is very close to T_n, the correct estimate for $\langle n(T)\rangle$ is $n(T) - 1/2$, while if T is much larger than T_n, the estimate should be $n(T) + 1/2$. This is taken into account in the condition $\langle n(T)\rangle = n(T) - 1/2 + (T - T_n)/T$. Combining this with condition $I(T) = N(T)$ yields $\tilde{A} = 0.428$ and $\tilde{B} = 0.377$.

Method 5: Combining the condition for $\langle n(T) \rangle$ in Method 4 with the logarithmic integral condition $IL(T) = NL(T)$ yields $\tilde{A} = 0.293$ and $\tilde{B} = 0.425$, which are very close to the ML estimates.

Method 6: Finally, one can use the two integral conditions $I(T) = N(T)$ and $IL(T) = NL(T)$ to obtain $\tilde{A} = 0.259$ and $\tilde{B} = 0.446$, again very close to the ML estimates.

Table II contains a summary of these results. All methods yield values that are within one standard deviation from the maximum likelihood estimates. The equations for the estimates are given in Table III, and the cumulative accident curves are presented in Fig. 3.

Table II. Parameters Estimated by the Methods of Moments

Method	$\tilde{A}$	$\tilde{B}$
ML	0.267 ± 0.326	0.441 ± 0.147
Moment 1	0.267	0.441
Moment 2	0.297	0.427
Moment 3	0.186	0.491
Moment 4	0.428	0.377
Moment 5	0.293	0.425
Moment 6	0.259	0.446

III. LEARNING FROM ACCIDENTS

One can argue that lessons are learned rapidly after an accident occurs, reducing the accident rate and that the rate then remains constant until the next significant event occurs. Assuming a geometric reduction in the rate after each event, one obtains a model with

$$\Lambda = \Lambda_i = \lambda\alpha^i \text{ for } T_i \le T < T_{i+1} , \quad (21)$$

$i = 0,1,\ldots,n$ where T_i is the cumulative time to incident i. The common factor λ is the accident rate before the first accident and α is the fraction to which the rate is reduced, on the average, when lessons learned from an event are implemented.

A likelihood equation consistent with Eq. (8) with Λ values substituted from Eq. (21) is maximized with values $\lambda = 0.0439$ and $\alpha = 0.599$ for the data of Table I. These values yield the step function [Eq. (21)] rate presented in Fig. 2.

It can be shown, by making the integral of Eq. (21) equal to n(T), that a smooth function approximating the step function is $\lambda/[1 + \lambda T \ln(1/\alpha)]$. This is also presented in Fig. 2. It can also be used for predictions, yielding an asymptotic estimate $\Lambda \simeq 2/T$.

Table III. Equations for the Methods of Moments

Method	$\tilde{A}$	$\tilde{B}$
Method 1 (and ML)	$\dfrac{n^*}{T^{\tilde{B}}}$	$\dfrac{1}{\ln(T) - \frac{1}{n}\sum_{i=1}^{n}\ln(T_i)}$
Method 2	$\dfrac{n}{T^{\tilde{B}}}$	$\dfrac{\sum_{i=1}^{n} T_i}{nT - \sum_{i=1}^{n} T_i}$
Method 3	$\dfrac{n - 0.5}{T_n^{\tilde{B}}}$	$\dfrac{\sum_{i=1}^{n} T_i - 0.5\, T_n}{nT_n - \sum_{i=1}^{n} T_i}$
Method 4	$\dfrac{n + 0.5 - T_n/T}{T^{\tilde{B}}}$	$\dfrac{\sum_{i=1}^{n} T_i + 0.5\, T - T_n}{nT - \sum_{i=1}^{n} T_i}$
Method 5	$\dfrac{n + 0.5 - T_n/T}{T^{\tilde{B}}}$	$\dfrac{n + 0.5 + T_n/T}{n\ln(T) - \sum_{i=1}^{n}\ln(T_i)}$
Method 6	$\dfrac{\tilde{B}[n\ln(T) - \sum_{i=1}^{n}\ln(T_i)]}{T^{\tilde{B}}}$	$\dfrac{nT - \sum_{i=1}^{n} T_i}{T[n\ln(T) - \sum_{i=1}^{n}\ln(T_i)] - [nT - \sum_{i=1}^{n} T_i]}$

The Fisher information matrix yields compact equations for the statistical accuracies of the estimates. For example, the relative variance of Λ_n is $2(2n+1)/[n(n-1)]$, while that of α is $12/[(n-1)n(n+1)]$. The prediction of Λ from the step function is then $\lambda\alpha^9 = 0.00044 \pm 0.00027$ (per year).

IV. CORRELATIONS WITH THE NUMBER OF PLANTS

Most of the core damage accidents have occurred early in the life of a plant, all but one during the first year of operation.[1] This may indicate a flaw in the design, installation, procedures, training, etc. It may be appropriate to assume that there is a probability P (per plant) that a plant (or the crew) when commissioned contains a serious flaw that will lead to a core damage accident, and study how this probability has decreased as a function of the number of plants built, N, rather than the cumulative time T.

Let the cumulative number of plants completed by the time of accident i be N_i ($i = 1,2,\ldots,n$; $N_0 = 0$), and at the time of the analysis N ($\geq N_n$).

If no learning is explicitly assumed, the probability of the experience of n accidents is given by the bionomial law, $F = \binom{N}{n} P^n (1 - P)^{N-n}$, where $\binom{N}{n}$ is the bionomial coefficient $N!/[n!(N-n)!]$. From this function the maximum likelihood estimate of P is $P_{ML} = n/N$. This is presented in Fig. 4 for the worldwide data from Table I. The decreasing trend of this sawtooth curve indicates that some learning has taken place. If a true learning model

$$P = p\beta^i \text{ for } N_i \leq N < N_{i+1} \ , \quad (i=0,\ldots,n) \ , \tag{22}$$

is assumed, the maximum likelihood estimation problem becomes very similar to that of Section III, with λ, α and t_i replaced by p, β and $(N_i - N_{i-1} - 1)$, respectively (when $n \ll N$). Maximizing the likelihood yields $p = 0.10 \pm 0.06$ and $\beta = 0.77 \pm 0.1$ for data obtained from Table I, with the ± standard deviations obtained from the Fisher information matrix. A smooth approximation of the estimate Eq. (22) is presented in Fig. 4, indicating the asymptotic behavior $\sim 4/N$.

V. SUMMARY

In summary, the features of this paper are (1) a statistical significance test for learning, based on the ratio of gamma variates, (2) models associating learning to accidents and to the number of plants in addition to the cumulative number of operating years, (3) use of the maximum likelihood and several moments methods for estimating parameters, and (4) analytical estimation of the statistical accuracies of the parameters. Using a sample of core damage accidents in electricity producing plants it is estimated that the probability of a plant to have a serious flaw has decreased from 0.1 to 0.01 during the developmental phase of the nuclear industry.

At the same time, the frequency of accidents has decreased from 0.04 per reactor year to 0.0004 per reactor year. This is partially due to the fact that the number of operating years by mature standardized plants grows faster than the number of new plants coming on-line. The relative standard deviation of these estimates is about 60%.

Further conclusions from this study are:

1. Insurance premiums of the order of 1% of the plant cost should cover the economic risk due to core damage accidents.

2. Shutting down nuclear power plants that have operated safely for a year or more is not likely an effective way to reduce risk.

ACKNOWLEDGEMENT

The author wishes to thank Mr. James A. Hartung of Rockwell International for providing plant operating history information used in Table I. The support and encouragement of Dr. Donald R. Ferguson is also gratefully acknowledged.

This work was performed under the auspices of the U.S. Department of Energy.

REFERENCES

1. J. A. Hartung and P. Rutherford, *Trans. Am. Nucl. Soc.*, *43*, 676-677, (1982).

2. G. Campbell and K. O. Ott, *Nucl. Sci. Engng.* *71*, 267-279, (1979).

3. K. O. Ott and J. F. Marchaterre, *Nucl. Tech*, *52*, 179-188, (1981).

4. J. K. Vaurio, *Nucl. Instruments and Methods*, *99*, 373-378, (1972).

5. Estimated from data provided by J. A. Hartung, Rockwell International, private communication, 1982.

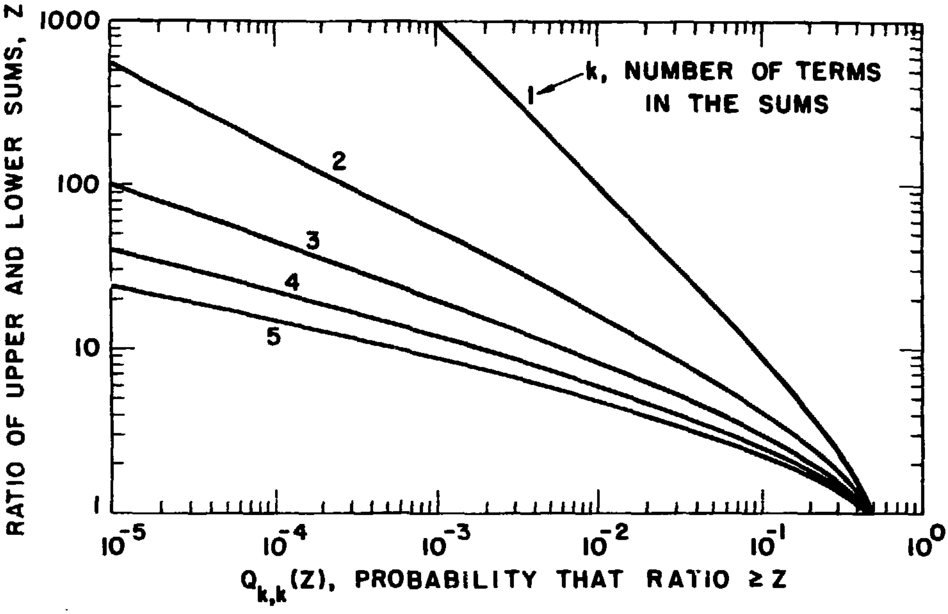

Fig. 1. The Probability of Exceeding the Observed Ratio under the Hypothesis of Constant Accident Rate.

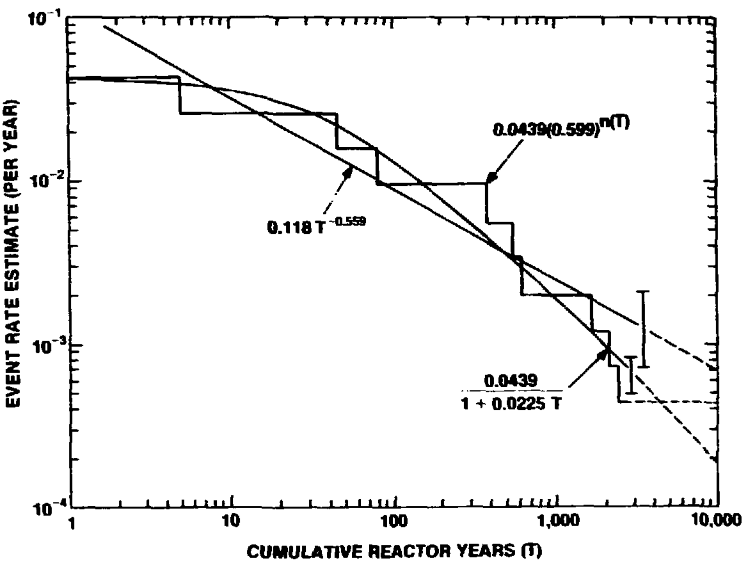

Fig. 2. Time-related Event Rate Estimates and Predictions.

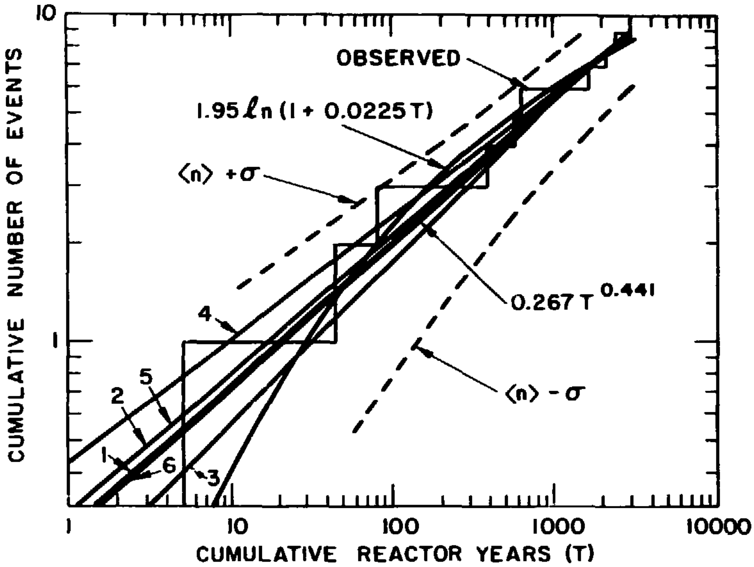

Fig. 3. Observed and Estimated Number of Events. Numbers 1 through 6 refer to the six different methods of moments defined in the text.

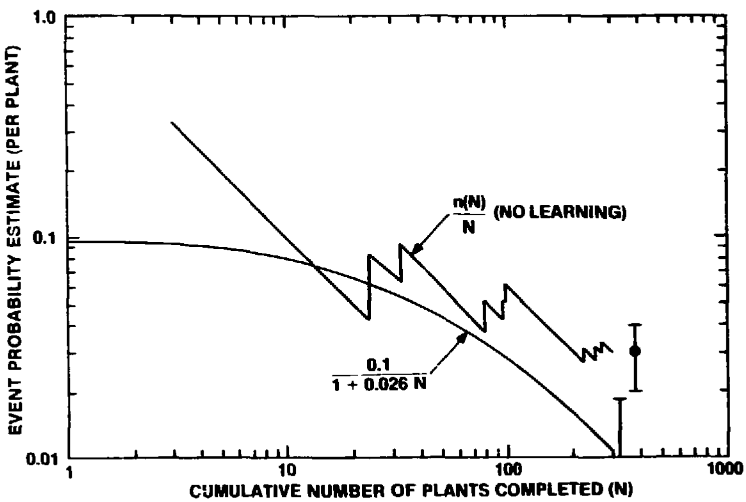

Fig. 4. Plant-related Severe Flaw Probability Estimates.