DIRECTION DES ÉTUDES ET RECHERCHES

SERVICE ENSEMBLES DE PRODUCTION
DÉPARTEMENT RETOUR D'EXPÉRIENCES
MESURES-ESSAIS

EDF
Electricité de France

Juillet 1996

CLAROTTI C.A.
PROCACCIA H.
VILLAIN B.

METHODES NUMERIQUES POUR L'ANALYSE BAYESIENNE DU VIEILLISSEMENT DE COMPOSANTS OU STRUCTURES

NUMERICAL METHODS FOR BAYESIAN INFERENCE IN THE FACE OF AGING

Pages : 9

96NB00187

Diffusion : J.-M. Lecœuvre
EDF-DER
Service IPN. Département PROVAL
1, avenue du Général-de-Gaulle
92141 Clamart Cedex

ISSN 1161-0611

VOL 28 № 22

SYNTHÈSE :

On s'est beaucoup intéressé ces dernières années aux méthodes bayésiennes pour l'analyse de risque. Elles ont jusqu'à présent fait l'objet d'études théoriques, les chercheurs s'intéressant surtout aux points suivants :

- étude de l'efficacité de ces méthodes pour l'analyse des événements rares ;
- débats sur le problème des a priori, et autres sujets de nature philosophique.

Un des aspects essentiels de l'approche bayésienne est le calcul numérique, étant donné que dans un cadre bayésien, tout problème de Sûreté/Fiabilité aboutit à un problème d'intégration numérique. Cet aspect a jusqu'à présent été négligé parce que la plupart des études de risque supposaient le modèle exponentiel comme modèle probabiliste de base. L'existence d'a priori conjugué rend dans ce cas superflue l'intégration numérique.

Dans le cas du vieillissement, il n'existe pas de famille conjuguée, et l'application de l'intégration numérique devient obligatoire.

EDF, la compagnie électrique nationale française, et l'ENEA, comité italien des technologies nouvelles, de l'énergie et de l'environnement, se sont associés dans un programme de recherche visant à développer des méthodes quadratiques d'inférence bayésienne reposant sur des distributions de Weibull ou gamma.

Cette note expose les principaux résultats obtenus dans ce programme et commente, à l'aide de quelques cas représentatifs, les performances des algorithmes numériques mis en oeuvre. Les cas considérés font partie d'une étude de l'apparition de fissures par corrosion sous contrainte dans les tubes des générateurs de vapeur des centrales REP françaises.

EXECUTIVE SUMMARY :

In recent years, much attention has been paid to Bayesian methods for Risk Assessment. Until now, these methods have been studied from a theoretical point of view. Researchers have been mainly interested in :

- studying the effectiveness of Bayesian methods in handling rare events ;
- debating about the problem of priors and other philosophical issues.

An aspect central to the Bayesian approach is numerical computation because any safety/reliability problem, in a Bayesian frame, ends with a problem of numerical integration. This aspect has been neglected until now because most Risk studies assumed the Exponential model as the basic probabilistic model. The existence of conjugate priors makes numerical integration unnecessary in this case.

If aging is to be taken into account, no conjugate family is available and the use of numerical integration becomes compulsory.

EDF (National Board for Electricity, of France) and ENEA (National Committee for Energy, New Technologies and Environment, of Italy) jointly carried out a research program aimed at developing quadrature methods suitable for Bayesian Inference with underlying Weibull or gamma distributions.

The paper will illustrate the main results achieved during the above research program and will discuss, via some sample cases, the performances of the numerical algorithms which were implemented. The sample cases that will be considered are part of the study on the appearance of stress corrosion cracking in the tubes of Steam Generators of PWR French power plants.

NUMERICAL METHODS FOR BAYESIAN INFERENCE IN THE FACE OF AGING

Carlo Alberto Clarotti
ENEA, INN-TEC, CRE Casaccia,
Via Anguillarese 301, 00100 Rome
Italy

Benjamin Villain & Henri Procaccia
ELECTRICITE DE FRANCE, Research and Development Division,
25, Allée Privée, Carrefour Pleyel, 93206 Saint-Denis Cedex 1
France

1. STATISTICAL vs. NUMERICAL ASPECTS

Every safety/reliability analyst must use censored samples of reliability-component-life-lengths to make inferences. For the sake of brevity, this type of problem will be referred to as "Bayesian inference problem" and "censored sample of reliability-component-life-lengths" will be referred to as "data".

With these terms in mind, we claim that in the Bayesian frame, all inference problems in the face of aging lead to a problem of numerical integration. To support the above statement, let us recall the well known Bayes theorem, according to which the posterior pdf $\Pi(\underline{\theta})$ on the unknown, possible vector parameter $\underline{\theta}$ is :

$$\Pi(\underline{\theta}) = \frac{L(D|\underline{\theta})\,\Pi_0(\underline{\theta})}{\int_\Omega L(D|\underline{\theta})\,\Pi_0(\underline{\theta})\,d\underline{\theta}} \tag{1.1}$$

where :

$\Pi_0(\underline{\theta})$, is the prior pdf on the parameter ;

$L(D|\underline{\theta})$, is the value of the likelihood function corresponding to the observed data D ;

Ω, is the set of all possible values of $\underline{\theta}$.

In the case of Bayesian inference, there is the need to compute the normalization integral in the denominator of eq. (1.1). In addition, there is the need to compute the moments of order 1 and 2 of the posterior pdf for estimating the parameter and assessing its dispersion around the estimated value.

The integrals we evoked are easily calculable only in the case that the reliability-components of concern fail according to an exponential distribution. In this case, one can exploit the fact that the gamma-family is closed under sampling from the exponential distribution (De Groot, 1970b). That is if the prior pdf belongs to the gamma-family, so does the posterior, and the integrals of interest can be easily calculated analytically.

Note that the use of conjugate priors (gamma-priors in the case of the exponential distribution) is not a trick for making computation easy. As a matter of fact, conjugate priors are legitimate reference-priors because, as shown in (Cifarelli, 1991), they represent "vague" prior knowledge.

In the case of non-exponential probability distributions, that is in the case of aging, no family of conjugate priors is available. It follows that the integrals whose values are needed for estimating the unknown parameter must be calculated numerically. The required effort of numerical computation is much stronger if one shares the opinion of the predictive Bayesians. According to the predictive Bayesian approach estimating the parameters of probability distributions leads to non-conservative safety decisions (Clarotti, 1988) ; safety/reliabili-ty-related decision problems must rather be solved by assessing the conditional probability below (Clarotti, 1993) :

$$R(t|D) = \frac{\int_\Omega R(t|\underline{\theta})\, L(D|\underline{\theta})\, \Pi_0(\underline{\theta})d\underline{\theta}}{\int_\Omega L(D|\underline{\theta})\, \Pi_0(\underline{\theta})d\underline{\theta}} \qquad (1.2)$$

where :

t, is the time during which the piece of engineering equipment must keep working ;

$R(t|\underline{\theta})$, is the reliability of the piece of engineering equipment in the case that $\underline{\theta}$ is known ;

$R(t|D)$, is the reliability of the piece of engineering equipment given the observed data.

Numerically computing the right-hand-side of eq. (1.2) is much harder than numerically computing the moments of order 1 and 2 of the posterior pdf, because the function $R(t|\underline{\theta})$ is extremely non linear.

Another difficulty, when a family of conjugate priors does not exist, is that vague prior knowledge must be handled (in a heuristic way) via improper histograms. The prior must be a histogram, i.e. it must be piece-wise flat, because vague prior knowledge about the parameter can, at the most, indicate that some sets of values are more likely than others. The prior must be improper (De Groot, 1970a) because vague prior knowledge cannot confine with certainty the histogram within finite limits.

In summary, for Bayes inference in the face of aging to be practically possible, the following two objectives must be achieved :

- prove that commonly observed samples are such that the improper prior histograms yield proper posteriors in a way that the integrals of interest exist ;
- find quadrature (i.e. numerical integration) methods suitable for estimating the integrals of highly non linear functions over unbounded domains.

It should be noted that from a purely statistical point of view, Bayesian inference is compellingly appropriate for the analysis of field data (i.e. highly censored data coming from the real, operative life of the pieces of engineering equipment) (Bartholomew, 1963), (Barlow, 1988).

2. THE SAMPLE QUADRATURE PROBLEM : THE WEIBULL DISTRIBUTION

The Weibull distribution was selected as the subject of the research on numerical methods suitable to Bayes inference because this distribution is currently used at EDF to model the appearance of stress-corrosion-cracking in a tube of steam generators (Pitner, 1988).

The parametrization will be the one specified below (with reference to the Weibull survival function $\bar{F}(\bullet|\eta,\theta)$) :

$$\bar{F}(\bullet|\eta,\theta) = \exp\left(-\left(\frac{\bullet}{\eta}\right)^\theta\right) \qquad (2.1)$$

where :

θ, is the (unknown) shape parameter ;

η, is the (unknown) scale parameter.

The following, very general sampling plan has been considered : some of the failures of the on-test pieces of engineering equipment are immediately observed as soon as they occur, some others are detected at pre-established inspection times.

With the above assumptions, equation (1.2) must be modified according to eqs (2.2) through (2.6) :

$$\underline{\theta} \equiv (\eta,\theta) \qquad (2.2)$$

$$R(t|\underline{\theta}) = \exp\left(-\left(\frac{t}{\eta}\right)^\theta\right) \qquad (2.3)$$

$$D \equiv \left\{x_1,\dots,x_r;x'_1,\dots,x'_s;\left[t_1,t'_1\right],\dots,\left[t_k,t'_k\right]\right\} \qquad (2.4)$$

where :

$x_i, i = 1,\dots,r$, are the observed failure times ;

$x'_j, j = 1,\dots,s$, are the observed survival times ;

$[t_m,t'_m], m = 1,\dots,k$, are the time intervals where a failure was detected by periodic inspections;

$$L(D|\eta,\theta) = \left(\frac{\theta}{\eta}\right)^r \prod_{i=1}^{r}\left(\frac{x_i}{\eta}\right)^{\theta-1} \exp\left(-\sum_{i=1}^{r}\left(\frac{x_i}{\eta}\right)^\theta - \sum_{j=1}^{s}\left(\frac{x'_j}{\eta}\right)^\theta\right) \times \qquad (2.5)$$

$$\times \prod_{m=1}^{k}\left[\exp\left(-\left(\frac{t_m}{\eta}\right)^\theta\right) - \exp\left(-\left(\frac{t'_m}{\eta}\right)^\theta\right)\right]$$

$$\Omega \equiv (0 < \eta < +\infty) \times (1 \le \theta < +\infty) \qquad (2.6)$$

Note that :

- the lower bound for the shape parameter is 1 because of aging ;
- n the sequel only the improper prior $\Pi(\eta,\theta) \equiv 1$ will be considered because, as one can easily grasp, there is no loss of generality for what concerns numerical computation problems.

As mentioned in section 1, the first problem to face is to determine the types of data for which the integrals of interest exist.

As shown in (Villain, 1996), these integrals exist when :

- at least two failures at two different times were observed as soon as they had occurred ;
- or the intersection of the intervals $\left[t_m,t'_m\right]$, m=1,...,k, is the empty set.

For the latter condition to be satisfied, at least three intervals must be adjacent or disjoint. This is the case when new failures were observed during at least three different inspections in one of the sites which are supplying field data.

We can then conclude that the use of improper flat priors does not really disturb Bayesian inference with underlying Weibull distribution, in the sense that commonly observed data yield proper posteriors.

3. HINT OF THE QUADRATURE STRATEGY

For the right hand side of eq. (1.2) to be practically calculable via quadrature formulas when the identifications specified by eqs. (2.2) through (2.6) are made, some manipulation is needed.

First of all, we need to find a variable transform which will map the unbounded domain Ω onto one of the standard bounded domains.

The standard domains are :

- the multidimensional cube;
- the multidimensional simplex ;
- the multidimensional sphere.

Such a transformation is imposed by the theory of multidimensional numerical integration. Indeed, in practice no multidimensional quadrature formula can be proved to converge to the exact value of the integral of concern for integration domains other than the standard ones (Stroud, 1971).

The variable transform :

$$\begin{cases} u = \dfrac{1}{\theta} \\ v = e^{-\frac{1}{\eta}} \end{cases} \qquad (3.1)$$

maps Ω onto the unit square (i.e. the unit cube in two dimensions), and is then suitable for our purposes.

A sufficient condition for a multidimensional formula to converge in the case of standard domains is : the integrand is bounded and the quadrature formula is a product formula formed from one of the following formulas in one dimension (Stroud, 1971) :

- the mid-point formula ;
- the trapezoidal formula ;
- the Simpson formula ;
- the Gauss formula.

The observed data of the type discussed in section 2 make it certain that the integrand is bounded (Villain, 1996).

The concept of product formula is explained below with reference to the integration over a square.

Simply stated, a quadrature formula in one dimension is a mathematical device which approximates any integral over the interval (a, b) accordingly to eq. (3.2) :

$$\int_a^b f(x)\,dx \approx \sum_{j=0}^{N} a_j f(x_j) \qquad (3.2)$$

where :

N is independent of both the integrand and the interval (a, b) ;

the x_j's and the a_j's are independent of the integrand but depend on the integration interval.

Using eq. (3.2) tantamounts to interpolating f(x) by a linear combination of pre-established polynomials. The integrals of the latter over the interval (a, b) have exactly the value prescribed by eq. (3.2) itself.

Let $\varphi(u,v)$ be the integrand of interest.

Approximating the integral of $\varphi(u,v)$ over (a,b) x (a,b) by the product formula formed from the quadrature formula (3.2) means to assess the integral according to eq. (3.3) :

$$\int_a^b du \int_a^b dv\varphi(u,v) \approx \sum_{j=0}^{N} a_j \int_a^b du\,\varphi(u,x_j) \approx \sum_{j=0}^{N}\sum_{i=0}^{N} a_j a_i \varphi(x_i,x_j) \qquad (3.3)$$

Actually, the strategy for numerically assessing the integrals which appear in the Bayes formula is as follows : after mapping the integration domain onto a standard one, a compound product formula is applied.

Applying a compound formula essentially means to decompose the standard domain at hand into M smaller standard domains of the same type. The integral of interest is obtained by summing up the integrals over the subdomains. The integral over each subdomain is calculated by the selected quadrature formula (e.g. Simpson's one). M is increased until two subsequent iterations give the same result up to the prescribed precision.

Of course, the product formula selected as a basis for the compound formula must be formed from one of the formulas in one dimension which assure that the exact value for the integral will be eventually obtained.

In the case of Bayesian inference, before implementing any quadrature strategy, one has to solve a problem relative to the order of magnitude of the integrands, irrespective of the underlying distribution.

Any computer can process only the numbers whose absolute value lies between an upper bound (overflow limit) and a lower bound (underflow limit). Any number smaller than the underflow limit is dealt by the computer as if it were equal to zero. Each of the factors of the likelihood function corresponding to failure-events is generally quite small because we are interested in reliable pieces of equipment. As a result, the value of the integrand can be smaller than the underflow limit at many points of the integration domain, which can introduce considerable errors.

The solution to this problem is to multiply the likelihood function by a suitable large factor which will disappear when the Bayes theorem is applied. This corrective factor must be obtained via a step-by-step procedure. Let C_i be the corrective factor which is needed at the generic point of the integration domain. The factor C which is suitable all over the domain is clearly the largest of the C_i's. The value of C_i can be calculated as follows :

- arrange the factors of the likelihood function in decreasing order and let $l_1, l_2, \ldots l_q, \ldots$, the decreasing sequence which is obtained ;
- compute the partial products :

$p_1 = l_1.l_2.c_1$

$p_q = p_{q-1}.l_{q+1}.c_q, \; q=2, 3, \ldots$

where c_q is a corrective term such that p_q lies between the underflow and overflow limit of the computer ;

Among all the mathematical processes which permit to determine C, there is just one process which is effective from a computational point of view. We are obliged to be reticent about this process because it is covered by the property rights which protect the numerical computer code we produced for the Bayesian inference.

The effectiveness of this code will be discussed in the next section. This will effectively show that current numerical analysis coupled with a bit of mathematical fantasy (the undisclosed mathematical process) can efficiently solve Bayesian inference problems in the face of aging.

4. BAYESIAN INFERENCE. HOW LONG DOES IT TAKE ?

The prototype integration algorithm we produced according to the quadrature strategy of previous section, underwent a validation program including :

1. Bayesian inference from field data yielding a sharply bell shaped posterior distribution on Weibull parameters ;
2. Bayesian inference from data supplied by Monte Carlo simulation with underlying Weibull distribution.

Step 1 was aimed to challenge the speed of the algorithm. Indeed, the more the integrand is sharply bell-shaped, the longer the computation time will be. This is so because as the bell-shape of the integrand gets sharper, a finer partition of the unit square into subdomains is needed for achieving the prescribed precision.

Step 2 was to ascertain that the integration algorithm was error free and capable of estimating precisely the "true value" of the unknown parameters.

An example of field data used for step 1 is reproduced in table 1. Data are relative to the appearance of stress-corrosion-cracking on tubes of steam generators of EDF PWR power plants. Failures[1] were detected at the inspection times which are listed in the second column. The integer in the fourth column is the number of failures which occurred between the two inspection times on the same row as the integer. The last row refers to components still surviving at the time of the latest inspection.

TABLE 1

i	time intervals where failures occurred (in hours)		ni
1	0	12274	44
2	12274	35151	919
3	35151	49700	544
4	49700	∞	1693

The performances of the integration algorithm, implemented on a SUN workstation, are summarized in table 2; note that two versions of the algorithm are available. The bell-shape of the posterior is quantified by the ratios between the standard deviation and the mean of the two parameters (both versions of the algorithm can assess the moments of first and second order of the posterior).

TABLE 2

Expected value of θ	=	2.160463E+00
Standard Deviation of θ	=	5.854542E-02
Expected Value of η	=	6.069223E+04 (hours)
Standard Deviation of η	=	8.501606E+02 (hours)
Covariance	=	-2.569439E+01 (hours)

	normal algorithm			accelerated algorithm		
Estimate of numerical error	Numerical estimate of reliability at one year	number of nodes used for discretisation	CPU time (seconds)	Numerical estimate of reliability at one year	number of nodes used for discretisation	CPU time (seconds)
10^{-3}	.984	65	8	.984	65	7
10^{-4}	.9848	257	35	.9848	257	9
10^{-5}	.98476	513	121	.98476	513	15
10^{-6}	.984760	513	121	.984760	513	15
10^{-7}	.9847603	1025	469	.9847603	1025	35

Typical data for step 2 are in table 3. Note that failure data reported in table 3 are censored both on the right and on the left. As a matter of fact the simulation experiment supplied us with uncensored failure times. Reporting in a table 59 observed lives would have been impractical. For ease of representation, we then transformed the observed lives into censored failure data. Since the time scale of the Weibull distribution at hand is very large (η = 8 year) with respect to the inspection periodicity, reported data are almost equivalent to those obtained by observing the failures as soon as they occur. Data in table 3 give rise to a sharply bell-shaped posterior. Motivation for sharp bell-shape in this context is twofold, namely :

- checking the speed of the algorithm,
- checking the correctness of the algorithm.

Indeed, if the posterior is close to a Dirac-δ at the "true values" of the parameters, the exact values of the integrals are known. If the value "guessed" by the algorithm are the same as the true values, within one standard deviation of the posterior, we can conclude that the algorithm does the job it is supposed to do well and this is actually the case.

In the light of the CPU times required for numerical computation, it follows that we can afford to make Bayesian inference in the face of aging.

[1] Failure means : appearance of stress corrosion cracking ; the size of the observed crack is much smaller than the safety critical size.

TABLE 3

Seed Weibull Distribution

η = 8 years = 70080 hours
$\theta = 4$
Reliability at one year = 0.999755889

i	time intervals where failures occurred (in hours)		ni
1	4000	8000	2
2	8000	12000	10
3	12000	16000	26
4	16000	17520	21
5	17520	∞	9941

Output of the integration algorithm

Expected value of θ	=	4.312163E+00
Standard Deviation of θ	=	6.097334E-01
Expected Value of η	=	6.013270E+04 (hours)
Standard Deviation of η	=	1.139066E+04 (hours)
Covariance	=	-6.451790E+03 (hours)

	normal algorithm			accelerated algorithm		
Estimate of numerical error	**Numerical estimate of reliability at one year**	**number of nodes used for discretisation**	**CPU time (seconds)**	**Numerical estimate of reliability at one year**	**number of nodes used for discretisation**	**CPU time (seconds)**
10^{-8}	.99967698	257	887	.99967698	257	448

5. CONCLUSIONS

The structural components of power stations are typical examples of components which fail according to IFR distributions (Weibull, etc.). As already mentioned (cf. §1) classical inference from these distributions is practically impossible. The availability of quick Bayesian inference algorithm will make it possible to assess the residual life of structural components of power stations (such as the tube bundle of steam generators in a nuclear power plant) which is one of the main interests of safety engineers.

REFERENCES

Barlow, R. E., Proschan, F., 1988, "Life Distribution Models and Incomplete Data", in Handbook of Statistics, P.R. Krishnaiak and C.R. Rao Eds, Elsevier Science Publishers, Amsterdam, Vol. 7, pp. 225-245.

Bartholomew, D. J., 1963, "The Sampling Distribution of an Estimate Arising in Life Testing", Technometrics, Vol. 5, pp. 361-374.

Cifarelli, M. D., 1991, L'approccio predittivo dell'inferenza in Problemi di Ricerca Operativa, W. Racugno Editor, Societa Italiana di Statistica, pp. 49-77 (in Italian).

Clarotti, C. A., 1988, "Failure Rate Estimation. A Dangerous Non-sense in a Bayesian View", Reliability Engineering & System Safety, Vol. 20, pp. 117-126.

Clarotti, C. A., 1993, "Making Decision via PRA. The Frequentist vs. the Bayesian Approach", Reliability and Decision Making, R.E. Barlow, C.A. Clarotti, and F. Spizzichino Eds, Chapman & Hall, London, pp. 323-345.

De Groot, M. H., 1970a, Optimal Statistical Decisions, McGraw Hill, New York, Chapt. 2.

De Groot, M. H., 1970b, Optimal Statistical Decisions, McGraw Hill, New York, chapt. 9.

Pitner, P., 1988, "Statistical Analysis of Steam Generator Tube Lifetime and Probabilistic Method for Tube Bundle Inspection", Reliability Engineering & System Safety Vol. 21, pp. 211-292.

Stroud, A. H., 1971, Approximate Calculation of Multiple Integrals, Prentice Hall Inc., New Jersey, Chapts. 1 & 2.

Villain, B., 1996, Analyse numérique du logiciel IBW (Inférence Bayésienne d'une loi de vieillissement de Weibull), EDF report, (in French).