DIRECTION DES ÉTUDES ET RECHERCHES

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

EDF
Electricité de France

Juin 1996

CLAROTTI C.A.
PROCACCIA H.
VILLAIN B.

ANALYSE BAYESIENNE DES DONNEES DE DEFAILLANCE GENERALES PROVENANT D'UNE DISTRIBUTION DE VIEILLISSEMENT : PROGRES ACCOMPLIS DANS LES METHODES NUMERIQUES

BAYESIAN ANALYSIS OF GENERAL FAILURE DATA FROM AN AGEING DISTRIBUTION : ADVANCES IN NUMERICAL METHODS

Pages : 10

96NB00186

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 :

EDF et l'ENEA ont conduit un programme commun de développement de méthodes numériques et de codes informatiques nécessaires à l'analyse bayésienne des durées de vie des composants dans le cas du vieillissement. Les premiers résultats de cette étude ont été présentés à l'ESREL 94. Les étapes suivantes ont été franchies depuis :

- les données d'entrée ont été étendues au cas où les durées de vie observées sont tronquées à la fois à gauche et à droite ;

- les distributions de fiabilité possibles sont de types Weibull et gamma : dans les deux cas, leurs paramètres sont inconnus, et peuvent être statistiquement dépendants ;

- les a priori possibles sont des histogrammes relatifs à différentes paramétrisations de la distribution de fiabilité étudiée ;

- les moments de premier et de second ordre des distributions a posteriori peuvent être calculés. La covariance donne notamment certaines informations importantes sur le degré de dépendance statistique entre les paramètres étudiés.

On présente une expérience d'application du code à l'apparition de fissures par corrosion sous contrainte sur un tube de générateur de vapeur de REP.

EXECUTIVE SUMMARY :

EDF and ENEA carried out a joint research program for developing the numerical methods and computer codes needed for Bayesian analysis of component-lives in the case of ageing. Early results of this study were presented at ESREL'94. Since then the following further steps have been gone :

- input data have been generalized to the case that observed lives are censored both on the right and on the left ;
- allowable life distributions are Weibull and gamma ; their parameters are both unknown and can be statistically dependent ;
- allowable priors are histograms relative to different parametrisations of the life distribution of concern ;
- first-and-second-order-moments of the posterior distributions can be computed. In particular the covariance will give some important information about the degree of the statistical dependence between the parameters of interest.

An application of the code to the appearance of a stress corrosion cracking in a tube of the PWR Steam Generator system is presented.

Bayesian Analysis of General Failure Data from an Ageing Distribution : Advances in Numerical Methods

C.A. Clarotti
ENEA-INN-TEC, CR Casaccia
Rome (Italy)

H. Procaccia, B. Villain
Electricité De France, Research and Development Division
St Denis (France)

1 Introduction

In Bayesian analysis of general failure data for the sake of assessing the predictive reliability [2], engineering aspects and mathematical ones are intimately interrelated. For example, think of the mechanical components in the steam generator system of a nuclear power plant. Engineers claim that these components age ; this claim raises a nested set of mathematical problems. To begin with, most of conditional probability distributions which can represent ageing do not belong to the exponential family [3]. If the parameters of a component-life-distribution not belonging to the exponential family are unknown, there is no sufficient statistics of finite dimension for making inference about them [3] ; this makes it impossible to define a reference prior on the unknown parameters [4]. Indeed, a reference prior is the prior which, among those satisfying some (weak) restrictions, maximizes the degree of statistical dependence between past and future observations from a given distribution, according to a pre-established reasonable mathematical criterion. The (weak) restrictions reflect the statistician's (vague) prior knowledge.

The maximization criterion is the mathematical counterpart of the heuristic idea : let data do most of the work in deriving the posterior distribution. If the observations from a given distribution cannot be summarized by a sufficient statistics of finite dimension there is no mathematical tool for maximizing the degree of statistical dependence between past and future observations from that distribution : consequently there is no possibility of defining a reference prior on the related unknown parameters. Since in the case of ageing, there is not the possibility of defining any "reference" prior, the impact of prior knowledge on the final decision about system safety must be assessed by performing sensitivity analyses starting from different histograms as priors on the unknown parameters ; as histograms are piece-wise-constant-pdf's the intuitively represent vague prior knowledge. Yet one has to be able to manage histograms which are improper distributions [5]. This is so because, according to De Groot [6], any Bayesian has always to leave a "pinch of probability" on the event : "I was completely wrong". In the context of assessing priors, the event "I was completely wrong" may be the event "the unknown parameters have very large values", which can entail that the histogram is an improper distribution. In view of the above, for the reliability assessment relative to systems containing ageing components to be possible, one has to devise and implement numerical methods capable of quickly computing the posterior reliability of a component, the posterior estimate of parameters, etc., in the case of life-distributions such as Weibull, Gamma and other IFR distributions, the prior being an improper histogram. Devising and implementing such methods is actually the main goal of the research program on Bayesian inference [7] jointly carried out by EDF and ENEA.

2 The Numerical Integration Strategy for Bayesian Inference

Bayes theorem answers any question it is asked in the form of the ratio between two integrals.

No matter what the life-distribution of concern, the computation of the integrals of interest for the sake of Bayesian inference entails, in general, numerical integration over unbounded domains.

The following rules must be obeyed when dealing numerically with unbounded integration domains [8] :

1. the unbounded integration domain U must be mapped onto a standard domain S by some variable transform ;
2. the numerical integration over S must be carried out by some quadrature formula which is a Riemann sum ; this is generally done by applying a suitable compound product formula to S ; "suitable" means that the "factors" of the formula must be : mid point formulas, trapezoidal formulas, Simpson formulas, or Gauss formulas.

Exegesis :

- the standard domains are : the multidimensional cube, the multidimensional simplex, and the multidimensional sphere ;

- the Riemann sums are nothing but the sums which in the limit define the very concept of integral ;

- a compound product formula is a prank as follows :

  - suppose that equation (2.1) is a quadrature formula for numerically estimating integrals in one dimension :

$$\int_a^b f(v)\,dv \cong \sum_{j=0}^{N} a_j\, f(x_j) \tag{2.1}$$

where : N is independent of both the integrand and of the interval [a,b] ;
the x_j's and the a_j's are independent of the integrand but depend on the integration interval ;

  - the product formula in two dimensions obtained from equation (2.1) is then the formula which approximates the integral of any function defined over the square [a,b]×[a,b] according to equation (2.2) :

$$\int_a^b du \int_a^b dv\, \varphi(u,v) \cong \sum_{j=0}^{N} a_j \int_a^b du\, \varphi(u,x_j) \cong \sum_{j=0}^{N}\sum_{i=0}^{N} a_j\, a_i\, \varphi(x_i,x_j) \tag{2.2}$$

  - if the square [a,b]×[a,b] is subdivided into sub-squares, eq.(2.2) is applied to each sub-square, and the integrals over the sub-squares are summed up, we get a compound product formula over the square [a,b]×[a,b] ;

- prescriptions 1 and 2 stem from the following circumstances :

  - in the case of multidimensional quadrature formulas it is impossible to derive an estimate of the error in closed form and to prove that the error goes to zero as N in equation (2.2) goes to infinite ;
  - the proof that a given multidimensional quadrature formula does converge rests on the fact that the quadrature formula gives rise to a Riemann sum ;
  - one is capable of proving that a quadrature formula, for instance in two dimensions, gives rise to a Riemann sum only if the formula is a product formula formed from factors which are Riemann sums in one dimension ;
  - product formulas are applicable only to standard domains ;
  - the formulas in one dimension which are Riemann sums are : the mid- point formula, the trapezoidal formula, the Simpson formula, and the Gauss formulas ;
  - we know that Riemann sums converge in the limit, we are to decide whether or not the limiting value has been reached in practice ; this is feasible only if the quadrature formula is a compound formula.

3 The Life-Distributions and The Samples of Concern

For the sake of validation of the numerical-integration algorithms we developed, the following life distributions were selected : the Weibull distribution, the Gamma distribution.

Motivation for this choice is that these two distributions are typical in the following sense :

- both the survival function and the pdf of the Weibull distribution are available in closed form ;
- the pdf of the Gamma distribution is available in closed form but the survival function must be computed numerically.

In view of the above, if a numerical integration algorithm performs well both in the case of the Weibull distribution and in the case of the Gamma distribution, then it has satisfactory performances for any life distribution.

The following, very general sampling plan will be considered : some of the failures[1] of the observed components, appearances of a critical event, are immediately detected as soon as they occur ; some others are detected at pre-established inspection times.

3.1 The Weibull Distribution - Any Trouble is Over

As already mentioned in the introduction, improper priors are generally of concern.

In the case of Weibull distribution we proved that [9] improper priors yield proper posteriors provided only that the intersection of all time-intervals where failures were observed is empty ; e.g. we observed two failures at two different times, as soon as the failures had occurred.

Let $R(\bullet|\theta,\eta)$ be the conditional survival function of the Weibull distribution and let the parametrisation be as follow : $R(t|\eta,\theta) = e^{-\left(\frac{t}{\eta}\right)^{\theta}}$.

In the case of ageing, for the sake of Bayes inference we have to handle the unbounded integration domain $1 \le \theta < +\infty,\ 0 \le \eta < +\infty$. This domain is mapped onto the unit square by the variable transform : $v = \frac{1}{\theta},\ u = e^{-\frac{1}{\eta}}$.

In the worst case, the quadrature algorithm we implemented on a SUN workstation, by making use of a Simpson product formula, takes several tens of seconds for analyzing Weibull data [9].

3.2 The Gamma Distribution - We Are Gonna Make It

The problem with the gamma distribution is that the related conditional survival function must be computed numerically ; more precisely one has to compute numerically the integral on the left-hand-side of either eq. (3.2.1) or eq. (3.2.2).

$$\int_0^t e^{-\beta x} x^{\gamma}\, dx = -\frac{t^{\gamma}}{\beta} e^{-\beta t}\left[\sum_{j=0}^{k-1} \frac{1}{(\beta t)^j} \frac{\Gamma(\gamma+1)}{\Gamma(\gamma-j+1)}\right] + \frac{\Gamma(\gamma+1)}{\beta^k \Gamma(\gamma-k+1)} \int_0^t e^{-\beta x} x^{\gamma-k}\, dx \qquad (3.2.1)$$

$$\int_0^t e^{-\beta x} x^{\gamma}\, dx = \frac{t^{\gamma}}{\beta} e^{-\beta t}\left[\sum_{j=1}^{k} \frac{\Gamma(\gamma+1)}{\Gamma(\gamma+j+1)}\right] + \frac{\Gamma(\gamma+1)\beta^k}{\Gamma(\gamma+k+1)} \int_0^t e^{-\beta x} x^{\gamma+k}\, dx \qquad (3.2.2)$$

where : $\Gamma(\bullet)$ is the gamma function, and k is smaller then j in eq.(3.2.1).

---

[1] Failure in this context means appearance of a pre-established critical event.

The above equations transform the integral of concern in a way that numerical quadrature needed for deriving the survival function, will take a shorter time. The two equations are being studied for stating which one is the more convenient in terms of computation time.

4 Application to Ageing of Steam Generator Tube Bundle

Steam Generators of pressurized nuclear plants are equipped with alloy 600 tube bundle (inconel 600). This type of material is affected by stress corrosion phenomena, initiated from the primary side of the tubes, leading to essentially longitudinal cracks in the rolled zone of the tube just above the tube sheet.

Cracks size increases as time goes by and tubes are preventively plugged when the size of the cracks reaches 13mm, the critical size being 25mm. The Steam Generators are replaced when 15% of the tubes are plugged.

IBW code has been used to evaluate the probability of crack initiation in the first generation of nuclear plant Steam Generators. Another code COMPROMIS[10] (COde de Mécanique PRObabiliste pour la Maintenance et l'Inspection en Service) assesses the distribution of crack-size at different times in a way that one is capable of foreseeing at what date 15% of tubes have been plugged in. The integrated use of IBW and COMPROMIS makes it possible to assess the date at which the Steam Generator must be replaced.

An example of application of IBW code to the problem of cracking initiation in a particular Steam Generator is given in table 1.

We remark that :

- Input data came from the steam generator system of a French nuclear power station which has been supplying data during the whole validation phase for the EDF-ENEA code ;
- Data reported in the table made the posterior distribution sharply bell shaped ;
- The more sharply the posterior is bell shaped, the longer is the CPU time required for computing.

Taking into account the above remarks one can conclude that the developed algorithms on a SUN Sparc20 workstation are suitably fast.

<u>Selected prior</u> : Uniform, Π(p, teta) = constant, with p=exp(-((t/eta)**teta)), where t is the mission time,

<u>Mission time</u> : 8760 hours (1 year),

<u>Type of input data</u> : right and left censored data, in the time intervals $[t_{i-1}, t_i]$, crack initiation appeared in n_i tubes of a SG system,

i	t_{i-1} (hours)	t_i (hours)	n_i
1	0	12274	44
2	12274	35151	919
3	35151	49700	544
4	49700	∞	1693

Numerical estimation of component reliability :

	gauss algorithm			simpson algorithm		
error	estimate of component reliability	number of nodes used for discretisation	CPU time (sec.)	estimate of component reliability	number of nodes used for discretisation	CPU time (sec.)
10^{-3}	0.984	33	4	0.984	65	1
10^{-4}	0.9847	65	6	0.9847	257	3
10^{-5}	0.98476	129	9	0.98476	513	7
10^{-6}	0.984760	129	9	0.984760	513	6
10^{-7}	0.9847602	129	10	0.9847602	1025	22

Table 1 : Sample output of EDF/ENEA computer code

5 Conclusion

From a mathematical stand point, the main achievement of the research program jointly carried out by EDF and ENEA is the implementation of numerical methods capable of quickly computing the posterior reliability of components which fail according to IFR distributions (Weibull, gamma). This results is of value from an engineering point of view because it enables the safety analysts to assess the residual life of structural components of power stations (e.g. : the tube bundle of Steam Generators in a nuclear power station ; the reheated steam pipes of conventional power stations).

References

1. Clarotti C.A. et al. : A numerical-Integrated-based approach to reliability decision problems with underlying Weibull distribution. In : Proc. of the 9th International Conference on Reliability&Maintenability. UN press, Evry, 1994, pp 479-487. (in French)
2. Clarotti C.A., Spizzichino F. The Bayesian predictive approach in reliability. IEEE Transactions on Reliability 1989 ; 38 (n°3) : 378-382.
3. De Groot M.H. Optimal Statistical Decisions, Mc Graw-Hill, New York, 1970, chap. 9.
4. Cifarelli M.D. L'approccio predittivo dell'inferenza, in problemi di ricerca operativa. W. Racugno Editor, Società Italiana di Statistica 1991 : 49-77. (in Italian)
5. De Groot M.H. Optimal Statistical Decisions, Mc Graw-Hill, New York, 1970, chap. 2.
6. De Groot M.H. « Lindley's paradox » (Comment), Journal of American Statistical Association 1982 ; 77, 378 : 336-339.
7. Villain B. EDF/ENEA Contract n°P16L01/2M0458/EP707 : « Développement de méthodes bayésiennes pour le calcul de la fiabilité des matériels », 1995. (in French)
8. Stroud A.H. Appoximate calculation of multiple integrals, Prentice Hall Inc., New Jersey, 1971.
9. Villain B., Clarotti C.A. Analyse numérique du logiciel IBW (Inférence Bayésienne d'une loi de vieillissement de Weibull), EDF report, January 1996. (in French)
10. Pitner P. et al. Application of Probabilistic Fracture Mechanics to Optimize the Maintenance of PWR Steam Generator Tubes, Nuclear Engineering and Design 1993 ;