Bayesian Methods in Reliability Analysis

A. Colombo
Commission of the European Communities, J.R.C. Ispra Establishment, I-21020 Ispra, Italy

Abstract

In this paper we will discuss Bayesian methods for the estimation and the combination of failure rates, repair rates and unavailabilities. To estimate a constant reliability parameter we present a parametric method based on a binomial likelihood and a beta prior. To estimate a time-dependent failure or repair distribution we propose a non-parametric method that gives a point and interval estimate of the percentiles of the distribution for both complete and censored samples. To combine constant reliability parameters we present three different methods; the choice of a method, in a specific application, depends on the origin and quality of the data.

1. Introduction

The estimation and the combination of failure and repair parameters are crucial problems in reliability. From the mathematical point of view, the same model can be used to estimate a failure or a repair parameter. The same is true for the combination of failure or repair parameters, and thus we will only consider failure models.

We will discuss a parametric method for the estimation of a constant failure rate, a non-parametric method for the estimation of a time-dependent failure distribution, and three different methods for the combination of constant failure rates. All the methods are Bayesian.

To estimate a constant failure rate we propose a simple procedure [1] which assumes a binomial likelihood and a beta prior. The beta as prior has been rationally justified by Colombo and Costantini [2]. The procedure can also be used to estimate a constant unavailability. To estimate a time-dependent failure distribution we consider the approach proposed by Ferguson [3]. We developed, on the basis of this approach, a method for point and interval estimation of the percentiles of a failure distribution which can also be applied to censored samples [4]. To combine constant failure rates, or unavailabilities, we discuss three different methods [5] and suggest the choice of one of them according to the origin and the quality of the data available.

2. Estimation of a Time-Independent Failure Rate, and a Constant Unavailability

Notation

N number of demands to operate an item which works on demand, or operation time of an item working continuously;

r number of failures on demand, or on operation;

θ failure parameter of the item (unavailability or failure rate);

a,b parameters of a beta prior.

Failure Process

Unavailability. The failure process is modelled by a binomial distribution with parameters θ and N.

Failure Rate. This failure process is usually modelled by the Poisson distribution (exponential failure probability). Identifying each unit of operation time with a trial, it follows that this process can also be modelled by a binomial distribution with parameters θ and N. See [1].

Prior Distribution

Law. We assume a beta prior, as some simple and reasonable hypotheses imply this distribution as prior for unavailability and failure rate. See [2].

Parameters. The parameters a and b of a beta prior can be determined in various ways, e.g.:

- give the parameters directly (a can be seen as the number of failures in a hypothetical sample in which the number of demands (or the operation time) is a+b);
- give the mean μ and standard deviation σ of the distribution;
- give two percentiles such as the 50th (M) and the 95th (U);
- give the values s and t defined in [2].

Posterior Distribution

In a binomial process, the beta distribution is conjugate. It follows that, applying the Bayes theorem to a binomial likelihood, one obtains as posterior a beta distribution. The parameters of the beta posterior are r+a and N-r+b.

Considerations

Prior Information. Often the experimental data on the specific type of item of interest is very poor. Nevertheless, information is available on "similar" items working in "similar" operation conditions. In this situation the Bayesian approach is particularly convenient as all the information available can be used in defining the prior distribution of the failure rate of the specific item of interest. It is worth while noting that other distributions can be used as prior, e.g. the loguniform, the lognormal, or a histogram. In these cases the posterior distribution is obtained by numerical integration. The report Wash-1400

[6] and Mosleh and Apostolakis [7] support the lognormal distribution as prior.

Updating. The Bayesian approach allows an easy updating of the posterior distribution. In fact, each time a new experimental datum becomes available, a new posterior can be computed assuming the previous posterior as prior. This is particularly simple in the case of a beta prior, as in this case the posterior is also beta, and is then analytically defined.

Computer Programs

We developed two conversational computer programs for Bayesian estimation of a constant reliability parameter: BIDIPES [8] and BAESNUM [9]. BIDIPES computes point and interval estimates of constant failure rates, repair rates and unavailabilities assuming a beta prior. BAESNUM computes point and interval estimates of the same parameters assuming a non-conjugate prior. It is based on the concept of EQPTABLE (a distribution with equiprobable intervals [10]). Any distribution can be assumed as prior. At present the following distributions are implemented: uniform, loguniform, lognormal, gamma. The prior knowledge can also be described by a histogram or an EQPTABLE.

3. Estimation of a Time-Dependent Failure Distribution

In the case where a constant failure rate cannot be assumed, the problem of estimating a time-dependent distribution arises. In this situation, the Bayesian non-parametric method proposed by Ferguson [3] is of great interest. This method has two main advantages: it allows the use of all the information available about the item of interest and does not imply the choice of a parametric model for the (time-dependent) failure distribution.

Notation

$\hat{F}_n(t)$	predictive distribution of the time to failure of an item;
$F_0(t)$	initial distribution, defined on the basis of the available information;
$F_n(t)$	sample distribution of a sample of size n;
p	weight assigned to the initial distribution, $(0 < p < 1)$.

Results

Ferguson's result is summarized in eq.(1) which follows:

$$\hat{F}_n(t) = p\, F_0(t) + (1-p)\, F_n(t)\,. \tag{1}$$

Simply speaking, the predictive distribution is given by a weighted average of the initial and sample distributions.

If $F_0(t)$ is expressed in the form of a hypothetical sample of size m, then $\hat{F}_n(t)$ can be interpreted as derived from a single sample of size m+n. On the basis of this consideration we developed a method for point and interval estimation of the quantiles of a time-dependent failure distribution, for both complete and censored samples. See [4] and [11] for details.

4. Combination of Time-Independent Failure Rates

To combine failure rates of different data sources, some engineers use a "personal" criterion based on some weighted averaging of the available values. The IEEE Study 500 [12] proposed a procedure based on an unweighted geometric averaging of some point estimates which they obtained by the Delphi method. Martz and Waller [13] compared thirteen methods (the method of [12] included) and recommend some of them, depending on the number of data sources. We [5] propose three alternative methods, and suggest that one of them should be chosen according to the origin and the quality of the data available. A Bayesian approach is taken (as in [13]), so that the failure rate λ is assumed to be a random variable.

Notation

λ failure rate of a "family" of "similar" items operating in "similar" operating conditions (λ is assumed to be a random variable);

λ_M 50th percentile (median) of the distribution of λ;

λ_U 95th percentile of the distribution of λ;

μ_λ mean value of the distribution of λ;

σ_λ standard deviation of the distribution of λ;

a,b parameters of the distribution of λ (in the case where λ is assumed to be beta-distributed);

n number of data sources;

w_i $(i = 1,2,...,n)$ weight of the i-th data source (when available);

$\lambda_{i,M}$ $(i = 1,2,...,n)$ "central" estimate of λ given by the i-th data source (here assumed as an estimate of λ_M);

$\lambda_{i,U}$ $(i = 1,2,...,n)$ "pessimistic" estimate of λ given by the i-th data source (here assumed as an estimate of λ_U).

Methods

We consider three situations: homogeneous data, non-homogeneous data and weighted data sources, non-homogeneous data and unweighted data sources.

Homogeneous data. In this situation we assume a binomial failure process and a beta prior. Each pair of values $(\lambda_{i,M}, \lambda_{i,U})$ defines a beta distribution with parameters (a_i,b_i). Any of these distributions can be assumed as prior. Each pair (a_i,b_i) determines a hypothetical sample of $r_i = a_i$ failures in operation time $T_i = a_i+b_i$ (Ref. [14], p. 463).

We can thus combine the data by simply pooling the samples, and obtain

$$r = \sum_{i=1}^{n} r_i, \qquad T = \sum_{i=1}^{n} T_i . \tag{2}$$

Then we obtain $a = r$, $b = T-a$ which are the parameters of the posterior beta distribution of λ.

Non-homogeneous data and weighted data sources. In this case we propose a weighted arithmetic averaging method and a beta distribution for λ. The shape of the distribution of λ is defined by

$$\lambda_M = \frac{\sum_{i=1}^{n} \lambda_{i,M} w_i}{\sum_{i=1}^{n} w_i}, \qquad \lambda_U = \frac{\sum_{i=1}^{n} \lambda_{i,U} w_i}{\sum_{i=1}^{n} w_i} . \tag{3}$$

Non-homogeneous data and unweighted data sources. In this situation we refer to Chebychev's inequality. As is well known, Chebychev's inequality states that, given the random variable X with mean μ and standard deviation σ, then

$$\Pr \{|X-\mu| \geq k\sigma\} \leq \frac{1}{k^2} \qquad (k \geq 1) . \tag{4}$$

In this distribution-free method we assume:

$$\mu_\lambda = \lambda_M = \frac{\sum_{i=1}^{n} \lambda_{i,M}}{n}, \qquad \lambda_U = \max_{(i = 1,2,\ldots,n)} \{\lambda_{i,U}\} . \tag{5}$$

By taking in (4) $1/k^2 = .10$ one obtains a symmetrical probability interval of level $a = .90$ for λ, and the standard deviation of the distribution of λ. The standard deviation σ_λ is given by

$$\sigma_\lambda = \frac{\lambda_U - \lambda_M}{k} \simeq \frac{\lambda_U - \lambda_M}{3.18} . \tag{6}$$

Considerations

The binomial-beta model is well supported from a theoretical point of view; nevertheless it must be considered critically, as with an increasing number of data sources the distribution of λ may become very tight. The weighted arithmetic averaging method allows us, through the weights w_i, to take into account the opinions of the specialists concerned. In the case where these weights are not available, one could use the operation times as relative weights. The Chebychev inequality method is reasonably conservative.

5. Applications

These methods have been developed for the European Reliability Data Banks that we are implementing at Ispra. Some examples of application of the methods are discussed in Refs. [4,5,8,9,11].

References

[1] COLOMBO, A.G., JAARSMA, R.J., "Bayesian estimation of constant failure rate and unavailability", *IEEE Trans. Rel.*, R-31, 84-86 (1982).

[2] COLOMBO, A.G., COSTANTINI, D., "Ground-hypothesis for beta distribution as Bayesian prior", *IEEE Trans. Rel.*, R-29, 17-21 (1980).

[3] FERGUSON, T.S., "A Bayesian analysis of some non-parametric problems", *Ann. Statist. 1*, 209-230 (1973).

[4] COLOMBO, A.G., COSTANTINI, D., JAARSMA, R.J., "Bayesian non-parametric estimation of time-dependent failures", submitted to *IEEE Trans. Rel.*

[5] COLOMBO, A.G., JAARSMA, R.J., "Combination of reliability parameters from different data sources", Proc. 4th EuReDatA Conf., Venice, Italy, 23-25 March 1983.

[6] U.S. NUCLEAR REGULATORY COMMISSION, "Reactor safety study: an assessment of accident risks in U.S. commercial nuclear power plants", WASH-1400 (NUREG/75-014), October 1975.

[7] MOSLEH, A., APOSTOLAKIS, G., "Models for the use of expert opinions", Proc. Workshop on Low-Probability/High Consequence Risk Analysis, Arlington, Virginia, 15-17 June 1982, Plenum Press, New York, 1983.

[8] COLOMBO, A.G., JAARSMA, R.J., "BIDIPES. A conversational computer program for point and interval estimation of constant failure rate, repair rate and unavailability", Report EUR 7645EN (1982).

[9] COLOMBO, A.G., JAARSMA, R.J., "BAESNUM. A conversational computer program for the Bayesian estimation of a parameter by a numerical method", Report EUR 7900EN (1982).

[10] COLOMBO, A.G., JAARSMA, R.J., "A powerful numerical method to combine random variables", *IEEE Trans. Rel.*, R-29, 126-129 (1980).

[11] COLOMBO, A.G., JAARSMA, R.J., "BANPES. A conversational computer program for point and interval estimation of time-dependent failures", to appear as an EUR report.

[12] NUCLEAR POWER ENGINEERING COMMITTEE OF THE IEEE POWER ENGINERING SOCIETY, "IEEE Guide to the Collection and Presentation of Electrical, Electronic and Sensing Component Reliability Data for Nuclear Power Generation Stations", IEEE STD-500, June 1977.

[13] MARTZ, H.F., WALLER, R.A., "An exploratory comparison of methods for combining failure-rate data from different data sources", Report LA-7556-MS. Los Alamos Scientific Laboratory (1977).

[14] HAYES, W.L., WINKLER, R./., "Statistics: probability, inference and decision", Vol.1, New York, Holt, Rinehart and Winston Inc., 1970.