

Empirical Bayes Methodology for Estimating Equipment Failure Rates with Application to Power Generation Plants

K. Hutchison¹, J. Quigley², L. Walls², M Raza¹

¹ALSTOM Power, Switzerland

²Department of Management Science, University of Strathclyde, Glasgow, Scotland

(lesley.walls@strath.ac.uk)

Abstract – Many reliability databases are used to pool event data for components across power plants, both within and between organizations. The intention is to share data across common equipment and similar operating environments to provide better estimates of reliability and availability. However frequentist methods for such estimation can be poor when few events occur, even when the equipment are observed over long exposure times. This paper proposes an alternative estimation approach based upon Empirical Bayes and illustrates its application to failure data from power generation plants. A comparison is made between the proposed and classical methods, showing that Empirical Bayes is capable of generating more accurate estimates.

Keywords – Power plant event database, Empirical Bayes, Failure rate, Reliability

I. INTRODUCTION

Many organizations and industrial sectors have created databases containing the event histories of critical issues related to items of equipment. The intention is to record and maintain data about the exposure of this equipment with respect to non-performance and hazards. Also identified are significant events, i.e. failures, inspection, repairs and overhauls. In some cases, the databases may be company specific, for example, commercial organizations such as electronic manufacturers [1]. Some industrial sectors, particularly those operating safety-critical equipment, have a legacy of pooling and sharing data across plants, organizations and even nations [2].

While the raw data may reside in such databases in the form of a historical sequence of events, it is common for summaries such as failure and repair, or restoration, rates plus uncertainties to be computed for each class of equipment operating under nominally similar conditions. The information generated through analysis can be used in a variety of ways. For example, to monitor the level and patterns in reliability and availability of the equipment from which the data has been collected and hence used to flag areas where improvements are required. Other uses include providing input data to models of new processes or designs. For example, use of failure and repair rates within process simulation models or within fault tree models for new system design.

There exist methods for computing summaries, such as point and interval estimates of failure rates [3]. Often in

practice, these summaries are based upon classical statistical methods and use only the observed data concerning the number of failures and the relevant exposure time. While easy to implement and well understood, these frequentist approaches may not provide reliable estimates if there are few events and indeed will fail to provide useful estimates if there are no observed events. Yet this corresponds to the very conditions that the plant operators want to achieve. In some cases Bayesian approaches have been adopted. These have the advantage that the epistemic uncertainty in the parameter of interest can be expressed through a prior distribution and combined with the observed data through the likelihood function to obtain an improved *posterior* estimate. Hence Bayesian methods are capable of dealing with the no observed events scenario by relying upon engineering judgment to inform the prior. However, the prior distribution should be specified by an expert *prior* to observing any data that will be used to obtain the posterior distribution. As such, specifying a subjective prior for a parameter can challenge the cognitive abilities of engineering experts.

In this paper we propose to develop and apply a method based upon the principles of Empirical Bayes [4, 5]. This is a statistical philosophy that combines frequentist and Bayesian ideas that is regarded [6] as being a powerful approach for dealing with real data problems in science and engineering. Section II describes the proposed Empirical Bayes method for the problem of estimating failure rates using data across classes of equipment such as we find in reliability databases. Section III illustrates the application of the method to de-sensitized data from power generation plants. An empirical comparison of the proposed approach against the more conventional frequentist based estimates is given in Section IV. We conclude by identifying the potential strengths and limitations of Empirical Bayes for this type of problem.

II. EMPIRICAL BAYES METHODOLOGY FOR ESTIMATING FAILURE RATES

The prior distribution describes the variability in the rate of occurrence of failure for an item of equipment within a pool of plants, prior to observing any data. The prior is denoted by $\pi(\lambda)$, which is the probability density function measuring the likelihood of a failure, chosen at random having a rate of occurrence denoted by λ . Data are used for the item of interest from each specific plant to update the prior, refining the estimate uniquely for each plant. The updated prior is referred to as the posterior distribution.

It is assumed that failures occur at a constant rate through time, which implies a Homogeneous Poisson Process for the times between failures. The distribution of N_i , the number of failures that are realized when k_i operational hours have been accumulated, is Poisson:

$$P(N_i = n_i | \lambda_i) = \frac{(\lambda_i k_i)^{n_i} e^{-\lambda_i k_i}}{n_i!}, \lambda_i > 0, k_i > 0, n_i = 0, 1, \dots \quad (1)$$

An assumption about the parametric form of the prior distribution is made for numerical convenience. Since the number of failures is assumed to follow a Poisson Process, a computationally convenient distribution to describe the prior is a Gamma distribution with density function:

$$\pi(\lambda_i | \alpha, \beta) = \frac{\beta^\alpha \lambda_i^{\alpha-1} e^{-\beta \lambda_i}}{\Gamma(\alpha)}, \quad \lambda_i, \alpha, \beta > 0 \quad (2)$$

The rate of occurrence for any particular failure is not known, only that the rate has been selected at random from a Gamma distribution. We take an average of the Poisson distributions; weighted against the prior distribution. This provides the probability distribution of the number of failures that will occur for plant i , based only on our knowledge of the pool, i.e. the prior distribution. The following result due to [7] shows that the distribution of N_i is Negative Binomial:

$$P(N_i = n_i | \alpha, \beta) = \int_0^\infty \frac{(\lambda_i k_i)^{n_i} e^{-\lambda_i k_i}}{n_i!} \frac{\beta^\alpha \lambda_i^{\alpha-1} e^{-\beta \lambda_i}}{\Gamma(\alpha)} d\lambda_i \\ = \frac{\Gamma(n_i + \alpha) \left(\frac{\beta}{\beta + k_i}\right)^\alpha \left(\frac{k_i}{\beta + k_i}\right)^{n_i}}{\Gamma(\alpha) n_i!}, \quad \alpha > 0, \beta > 0, n_i = 0, 1, 2, \dots \quad (3)$$

As Poisson random variables are closed under convolutions the aggregate observed failure data on each hazard can be treated as though independent Negative Binomial random variables. This will be used to construct a likelihood function to support inference on the parameters α and β .

A. Point Estimate of Failure Rate

Once an estimate of the prior distribution is obtained, Bayes Theorem is used to update the prior for each individual precursor to obtain the posterior distribution. The posterior distribution for the i^{th} event is:

$$\pi(\lambda_i | N_i = n_i, \alpha, \beta) = \frac{(\beta + k_i)^{\alpha + n_i} \lambda_i^{\alpha + n_i - 1} e^{-(\beta + k_i)\lambda_i}}{\Gamma(\alpha + n_i)} \quad (4)$$

The Empirical Bayes estimate of λ_i is the mean of the posterior distribution. It is easily shown that the Empirical Bayes estimate of the i^{th} rate of occurrence of failure is can be expressed as (5).

$$E(\lambda_i | N_i = n_i) = \int_0^\infty \lambda_i \pi(\lambda_i | N_i = n_i, \alpha, \beta) d\lambda_i \\ = \frac{\hat{\alpha} + n_i}{\hat{\beta} + k_i} \\ = \frac{\hat{\alpha}}{\hat{\beta}} (1 - z) + \frac{n_i}{k_i} z \quad (5)$$

$$\text{where: } z = \frac{k_i}{\hat{\beta} + k_i}.$$

The Empirical Bayes estimate is a weighted average between the estimates from the pool (i.e. $\hat{\alpha}/\hat{\beta}$) and the traditional estimate of the individual plant (i.e. n_i/k_i). As more data are obtained, k_i increases, and more weight is applied to the observed frequency.

B. Interval Estimate for Failure Rate

If the true values of α and β are known, then the posterior distribution could be used to assess the uncertainty in estimating λ_i . However, α and β have been estimated and so the uncertainty in these estimates must be considered when developing true confidence intervals. Not accounting for the variability in the estimation of the posterior results in “naïve” intervals.

Quigley et al [8] propose the following method for estimating confidence intervals, which makes use of the limiting distribution of -2 times the natural logarithm of the relative likelihood function, which has a χ^2 distribution with 2 degrees of freedom. This is expressed in (7) where $(\hat{\alpha}, \hat{\beta})$ are the Maximum Likelihood Estimators. Alternative methods such as Bootstrapping are considered in [9].

$$-2 \ln \left(\frac{\prod_{i=1}^m \frac{\Gamma(n_i + \alpha) \left(\frac{\beta}{\beta + k_i}\right)^\alpha \left(\frac{k_i}{\beta + k_i}\right)^{n_i}}{\Gamma(\alpha) n_i!}}{\prod_{i=1}^m \frac{\Gamma(n_i + \hat{\alpha}) \left(\frac{\hat{\beta}}{\hat{\beta} + k_i}\right)^{\hat{\alpha}} \left(\frac{k_i}{\hat{\beta} + k_i}\right)^{n_i}}{\Gamma(\hat{\alpha}) n_i!}} \right) \sim \chi_{(2)}^2 \quad (7)$$

This result can be used to construct a joint confidence region for the parameters. To construct a tolerance interval about (4) first determine the locus of point points for α and β such that the Cumulative Distribution Function (CDF) of (7) does not exceed a specified value assuming only one degree of freedom rather than two.

There is no closed form solution for this approach to obtaining intervals and computationally it can be intensive. However, this approach uses the correlation between the estimates in estimating the intervals. If we are interested in obtaining intervals estimates on a one dimensional metric, such as failure rate at each plant, then we use one degree of freedom, which corresponds to a χ^2 value of 3.84. A region

is then obtained from finding the set of all points (α, β) that bound (7) to be less than or equal to 3.84. We then evaluate the maximum and minimum value of the metric of interest using the points in the region to obtain the 95% Confidence Interval. The acceptable region in an α - β plane can be obtained using standard software packages such as the solver tool in Excel. It is likely that the exposure time of an item of equipment will be different at each plant, hence the maximum and minimum confidence bounds for the failure rates have to be calculated at different α and β values. The solver tool in Excel can be used to maximize and minimize the condition that the parameters must sit within the specified elliptical region. For metrics which are monotonic transformations of the input parameters the extreme values will lie on the perimeter of the acceptable α - β region. The likelihood ratio test for the failure rate bounds finds those values that are not significantly different from the mean at a 95% level.

III. APPLICATION TO FAILURE DATA FROM POWER GENERATION PLANT

The methods described in section II have been applied to data from several power generation plants of the same type and have been implemented in models that draw upon the organization's reliability database. The intention is to obtain a pooled estimate of the failure rate for the equipment of interest and to assess whether there is any variation in the rate of failure between different plants.

A. Equipment Types and Data Structures

With its large product portfolio, ALSTOM has decades of specialized turnkey experience in Engineering Procurement and Construction (EPC) contracting. This includes developing constructing, erecting and commissioning efficient power plants worldwide. Further, ALSTOM does modernizing (i.e. retrofitting) and servicing power plants with the main aim of providing the market with the best technology in terms of efficiency and output. The focus of this study involved Combined Cycle Power Plants (CCPP), which are at the front of technology due to their increasing efficiency (www.alstom.com). The main source of power output for a CCPP is from the Gas Turbine (GT) that operates on the ignition of the input fuel. The exhaust heat that is emitted on completion of this cycle is used to produce steam for the Steam Turbine (ST) cycle.

The data that are analyzed have been collected from event logs recorded at each of the plants. The data recorded for the CCPP includes, for example, the times and dates when the plants are first disconnected from the grid and the time when they are up and running at required capacity again. From this log entry it is possible to know the observed downtimes, type of failure, action taken. A textual description of the outage is also provided by the engineer. It provides valuable experience and information, especially to an engineering department. Note that the system refers to the whole Combined Cycle Power Plant; sub-system refers to a

number of major arrangements, which combine to make the whole system (e.g. Gas Turbine, Generator, Steam Turbine, Heat Recovery Steam Generator); and component represents a single part or the smallest breakdown that is analyzed. The data selected for this study represents records that have been verified for 19 power plants which have a total of 76 unit years of operating experience.

B. Validity of Empirical Bayes Assumptions

The Empirical Bayes method described in section II makes three main assumptions that we seek to assess for the power plant data. Typically, for each piece of equipment, there are few observed events relative to exposure time; hence there is insufficient information to reject the assumption of a Poisson process at this level of analysis. However it is possible to compare the distribution of the total number of events per plant with the assumed Poisson. If it is accepted that these two distributions are not significantly different then the Gamma distribution is assumed a reasonable prior. Graphical analysis and formal Chi-squared goodness-of-fit tests indicate that this Poisson assumption is reasonable. The assumption of independence requires to be considered from the engineering perspective as well as statistically. For example, if a component is replaced on failure due to a fault or breakage then it may be considered 'as new'. However in many cases a component, and certainly not an entire sub-system, will not be replaced when an outage occurs. Engineering insight was used to screen data, as far as possible, to extract any dependent failures and statistical analysis implied that the assumption of independence was acceptable.

C. Empirical Bayes Estimates of Failure Rate

An example of the application of Empirical Bayes is given for one component of type X. Table 1 provides the raw data plus the Empirical Bayes estimates of the failure rate. Note that z is a measure of credibility of the estimate obtained from the individual plant failure rate rather than the pooled estimate.

The MLE estimate of the parameters of the prior gamma distribution for component X are found to be $\hat{\alpha} = 0.6956$ and $\hat{\beta} = 12235$. As z is a function of the exposure time, each plant has a different z_i . This differs from Quigley et al. (2005), whose study considered an exposure period of the same length. A large value of z implies that the variability within the pool is large and the estimate for the individual plant should place more weight on the individual data observed at that plant. A low value of z will occur when there is less exposure time observed and hence the failure rate estimate will be drawn towards the pooled data estimate.

Table 1 shows the expected number of failures estimated under Empirical Bayes and can be compared with the observed number. Note that those plants with no observed failures but accumulated operating experience have failure rate estimated informed by the pool.

A confidence interval can be computed using the procedure outlined in section IIB. Fig. 1 shows an oval-shaped region arising from the use of the limiting distribution given in equation (7). All values of α and β which result in the limiting distribution being less than the critical value lie within this 95% confidence region. The lower and upper bounds, which intuitively should lie on this boundary, can be computed for each plant since the optimal estimates of α and β will change due to the different exposure times. Table 2 gives the upper and lower failure rate bounds and the values of α and β where they occur.

Fig. 2 shows the comparison between the initial estimates of the failure rate and the Empirical Bayes estimates, with 95% confidence bounds, for component X across the 19 plants. It is shown that the initial estimate lies outwith the Empirical Bayes confidence band when there have been no observed failures (e.g. plants 1, 2, 4, 5, 6, 11, 14, 15, 16) and a relatively high number of failures (e.g. plants 10, 18). This is a feature of the Empirical Bayes estimates being pulled towards the pooled mean.

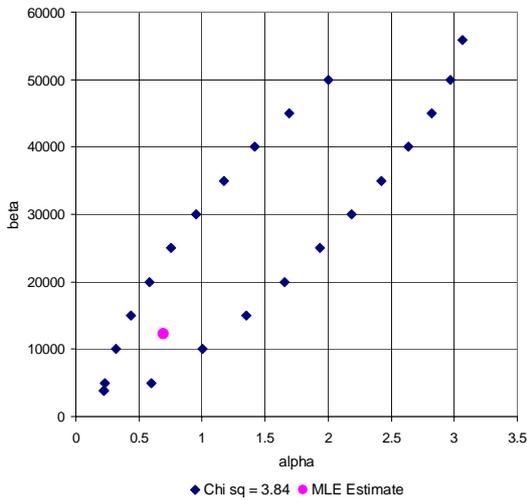


Fig. 1. 95% confidence region for component type X

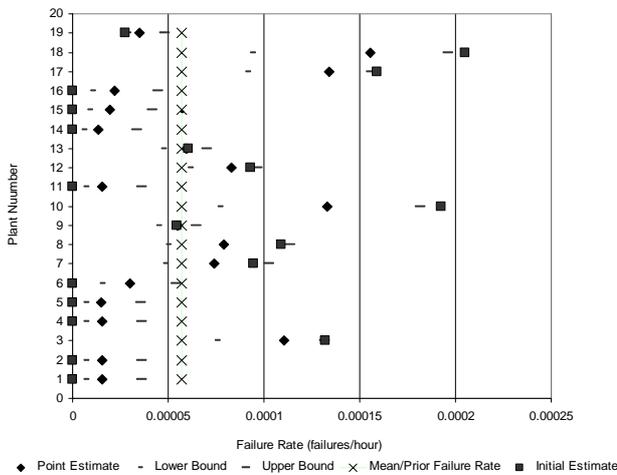


Fig. 2 Comparison of estimates for component type X across plants

TABLE I
DATA & EMPIRICAL BAYES ESTIMATES FOR COMPONENT X

Plant	Total Exposure Time	Observed Number of Failures	z	Empirical Bayes Failure Rate	Expected Number of Failures
1	32976	0	0.729	1.539E-05	0.5074
2	33027	0	0.730	1.537E-05	0.5076
3	30312	4	0.712	1.104E-04	3.3453
4	33323	0	0.731	1.527E-05	0.5088
5	33926	0	0.735	1.507E-05	0.5112
6	10866	0	0.470	3.011E-05	0.3272
7	10610	1	0.464	7.422E-05	0.7875
8	9171	1	0.428	7.921E-05	0.7264
9	36794	2	0.750	5.498E-05	2.0229
10	15568	3	0.560	1.328E-04	2.0704
11	33146	0	0.730	1.533E-05	0.5081
12	32232	3	0.725	8.311E-05	2.6788
13	33076	2	0.730	5.949E-05	1.9677
14	38899	0	0.761	1.360E-05	0.5292
15	23463	0	0.657	1.949E-05	0.4572
16	19326	0	0.612	2.204E-05	0.4260
17	37774	6	0.755	1.339E-04	5.0575
18	24408	5	0.666	1.554E-04	3.7938
19	36176	1	0.747	3.503E-05	1.2671

TABLE 2
EMPIRICAL BAYES 95% INTERVAL ESTIMATES FOR COMPONENT X

Plant	Lower			Upper		
	Alpha	Beta	Failure Rate	Alpha	Beta	Failure Rate
1	0.2249	4355	6.023E-06	2.7620	43371	3.618E-05
2	0.2249	4357	6.015E-06	2.7617	43363	3.615E-05
3	2.1832	52526	7.464E-05	0.4106	3287	1.313E-04
4	0.2248	4351	5.968E-06	2.7645	43441	3.601E-05
5	0.2248	4340	5.874E-06	2.7740	43706	3.573E-05
6	0.2447	6166	1.437E-05	1.9009	24428	5.386E-05
7	1.2154	35950	4.758E-05	0.5387	4414	1.024E-04
8	1.2853	37396	4.907E-05	0.4721	3794	1.135E-04
9	1.0587	32530	4.412E-05	1.3270	14674	6.464E-05
10	2.2130	52922	7.609E-05	0.3439	2842	1.815E-04
11	0.2249	4354	5.996E-06	2.7648	43450	3.610E-05
12	1.7818	46513	6.073E-05	0.5881	4911	9.660E-05
13	1.1706	34998	4.658E-05	1.1150	11515	6.986E-05
14	0.2244	4265	5.198E-06	2.8310	45348	3.360E-05
15	0.2268	4633	8.073E-06	2.5664	38351	4.152E-05
16	0.2289	4866	9.461E-06	2.4258	35089	4.458E-05
17	2.4266	55520	9.032E-05	0.3669	2979	1.562E-04
18	2.4534	55813	9.291E-05	0.3353	2796	1.961E-04
19	0.4655	16134	2.802E-05	2.2354	30986	4.817E-05

IV. COMPARISON OF FREQUENTIST AND EMPIRICAL BAYES ESTIMATES

We extend the comparison of traditional and Empirical Bayes estimation through further analysis of component X and others from the 19 plants to develop an understanding of their relative performance. For example, we have found that that 95% confidence intervals for the failure rate based on

Empirical Bayes are consistently within the 95% bounds computed using the usual frequentist Chi-Square intervals assuming exponentially distributed times between failures. This is expected given the latter are usually computed for 60% confidence levels for small samples. For example, the EIReDA database implements this for equipment with 5 or less failures otherwise the width of the computed intervals tends not to be practically credible.

It is also interesting to compare the rankings of the plants in terms of failure rates to examine where reliability is best or worst. If there is no difference in ranking between traditional and Empirical Bayes estimates then it could be argued that the EB method had added no real value to the analysis. A shortcoming of the traditional approach to estimation is that it will rank the hazards the same that have no failures regardless of the exposure accrued by each; Empirical Bayes does not suffer from this. Fig. 3 shows the comparative ranking for component Y across the 19 plants and provides an example of major changes, unlike component X which has only a marginal shift in ranks, partly because of the high percentage of plants with no observed failures.

The Mean Square Error (MSE) is a conventional measure of the accuracy of an estimator and examines the average deviation between the true and the estimated value. We assume that the failure rate for each plant is plant is selected from a gamma prior with parameters α and β . A random failure rate is simulated from this prior and converted to the number of failures expected within the exposure period for each plant. This value is treated as the mean of the plant Poisson Process from which a random number of events are simulated. For the data generated, three estimates of the expected number of failures for each plant are computed: the MLE; the pooled average of the dataset for the plant; and the Empirical Bayes estimate, which uses a combination of the pooled estimate and the individual estimate at each plant. The different weight values (z) calculated for the point estimation of the failure rates is again used at each plant. Table 3 shows the averaged MSE values based on 500 simulation runs for five components across the 19 plants. This illustrates that the accuracy is best for the Empirical Bayes estimator and summarizes our findings at the disaggregated level when examining the MSE of the estimates for each component at each plant.

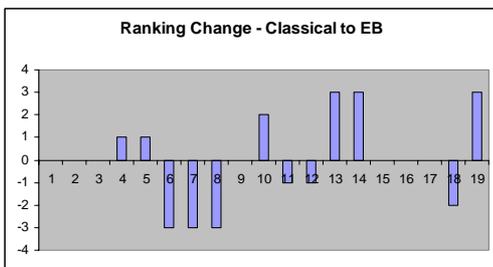


Fig. 3 Differences in frequentist and Empirical Bayes rank order for unreliability of plants for component Y

TABLE 3
AVERAGE MSE FOR ALTERNATIVE ESTIMATORS FOR SELECTED COMPONENTS ACROSS ALL PLANTS

Component	Average MSE		
	MLE	Pool	EB
A	2.091943	1.714102	1.046302
X	1.487919	4.178242	1.293668
B	1.676045	2.362794	1.133338
C	2.652285	5.51873	2.187533
D	1.271349	0.838775	0.564449

V. SUMMARY & CONCLUSION

This paper has implemented an Empirical Bayes method for estimating the failure rate of equipment whose operational experience is pooled across different environments. The method has been implemented for field data for power plants and has given practical insights into the reliability of components and a simulation study has shown that the Empirical Bayes estimates are more accurate than the usual frequentist methods. However all analysis has been implemented in a spreadsheet model and the software limitations become apparent in the complexity of the calculations. For example, although credible solutions have been found, the application of the solver to obtain the maxima and minima involved trial and error. While elaborations in modeling will result in more computational challenges, accurate approximations continue to be developed (Quigley and Walls 2007).

REFERENCES

- [1] P. A. Molenaara, A. J. M. Huijbenb, D. Bouwhuisc and A. C. Brombacher (2002) Why do quality and reliability feedback loops not always work in practice: a case study, *Reliability Engineering and System Safety*, 75, 3, 295-302.
- [2] <http://www.sintef.no/static/tl/projects/oreda/>
- [3] S.E. Rigdon and A.P. Basu (2000) *Statistical Methods for the Reliability of Repairable Systems*, Wiley
- [4] J.S. Maritz (1970) *Empirical Bayes Methods*, Methuen & Co Ltd, London.
- [5] B. Carlin B and T. Louis (2000) *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman & Hall
- [6] B. Efron (2005) *Bayesians, Frequentists and Scientists*, *JASA*, 100, pp 3-5.
- [7] M. Greenwood and G.U. Yule (1920) An inquiry into the nature of frequency distributions representative of multiple happening with particular reference to the occurrence of multiple attacks of disease or of repeated accidents, *Journal of the Royal Statistical Society A*, 83, pp255-279.
- [8] J. Quigley, T. Bedford T and L. Walls (2007) Estimating Rate of Occurrence of Rare Events with Empirical Bayes: A Railway Application *Reliability Engineering and System Safety* 92(5), pp 619-627.
- [9] T Bedford, J Quigley and L Walls (2006) 'Fault Tree Inference Through Combining Bayes and Empirical Bayes Methods' *Proceedings of the ESREL Conference*, Estoril, Portugal pp 859-865.