TPK4120 - Lecture summary

Jørn Vatn
eMail: jorn.vatn@ntnu.no

Updated 2021-11-17

Chapter 15 - Bayesian Reliability Analysis

Introduction

In classical estimation approaches the main idea is that we believe in “true” reliability parameters. The objective of the statistician is to “reveal” these true parameters in an “objective” manner. The statistician makes assumption regarding the observed data in terms of for example independent and identical distributed life times from some distribution class, for example the Weibull distribution. The more data available, the better will be the result.

Bayesian methods takes another starting point. The Bayesian statistician treats the parameters as stochastic variables. Before he or she looks into the data, a subjective judgement is made about the parameters. This judgement is denoted the *prior* distribution, i.e., prior to observing the data. The prior distribution for each of the relevant parameters are described by some distribution class, for example the normal distribution, the gamma distribution and so on.

There are various ways to establish the prior distribution. The statistician may utilize statements from experts having domain knowledge relevant for the problem at hand, he or she might utilize data from similar components or systems and so forth. In this presentation we will not elaborate on how to establish the prior distribution. To find out more the key words “expert judgement” would be a starting point.

When the prior distribution is established, the statistician consider the data, $\mathbf{t}$ as *evidence*. This means that he or she will update the prior distribution to what is called the *posterior* distribution which also takes the evidence into account.

Procedure

The procedure for Bayesian estimation could briefly be described as follows:

1. Specify a *prior* uncertainty distribution of the reliability parameter, $\pi(\Theta)$.

2. Structure reliability data information into a likelihood function, $L(\theta;\mathbf{t})$ (The likelihood function was discussed in Chapter 14 in the textbook).

3. Calculate the *posterior* uncertainty distribution of the reliability parameter vector, $\pi(\theta|\mathbf{t})$. The posterior distribution is found by $\pi(\theta|\mathbf{t}) \propto L(\theta;\mathbf{t})\pi(\theta)$, and the proportionality constant is found by requiring the posterior to integrate to one.

4. The Bayes estimate for the reliability parameter is given by the posterior mean, which in principle could be found by integration.

Note that the relation $\pi(\theta|\mathbf{t}) \propto L(\theta;\mathbf{t})\pi(\theta)$ follows from Bayes' theorem and the law of total probability: If $B_1, B_2, \ldots, B_r$ (corresponding to the θ-vector) represent a division of the sample space, and A is an arbitrary event (corresponding to $\mathbf{t}$ = the data vector), then:

$$\Pr(B_j|A) = \frac{\Pr(A|B_j)\cdot\Pr(B_j)}{\Pr(A)} = \frac{\Pr(A|B_j)\cdot\Pr(B_j)}{\sum_{i=1}^{r}\Pr(B_i)\cdot\Pr(A|B_i)}$$

Since we in the denominator sum over all possible B_i values (corresponding to the θ-vector) it will not contain θ, hence it may be regarded as a constant wrt θ. Further $\Pr(B_j)$ corresponds to the prior distribution, and $\Pr(A|B_j)$ corresponds to the likelihood function (in terms of the the product of the pdf's for each data point).

It is not obvious that we in step 4. should use the posterior *mean*. But if we aim for a single parameter estimate, and we have a posterior uncertainty distribution, it is reasonable to choose the mean value in this distribution. It might be proven that the posterior mean is the optimal value under "quadratic loss".

Exponential distribution

In the following we give an example showing the main elements of the procedure. In the example we will estimate the failure rate in the constant failure rate situation. Assume that we express our prior believe[1] about the failure rate λ of a certain component (gas detector used on an oil and gas platform), in terms of the mean value $\mu = 0.7\cdot 10^{-6}$ (failures / hour), and the standard deviation $\sigma = 0.3\cdot 10^{-6}$. For mathematical convenience, it is common to choose a gamma distribution[2] with parameters α and ξ for the prior distribution. The expected value and the variance in the gamma distribution are given by $\mu = \alpha/\xi$ and $\sigma^2 = \alpha/\xi^2$ respectively, and we obtain the following expressions for α and ξ:

[1] This could be based on statements from experts, see Øien et.al (1998), or by analysis of similar components (empirical Bayesian analysis).

[2] $\pi(\lambda) \propto \lambda^{\alpha-1}e^{-\xi\lambda}$ for the gamma distribution.

$$\xi = \mu/\sigma^2 = (0.7\cdot 10^{-6})/(0.3\cdot 10^{-6})^2 = 7.78\cdot 10^6$$
$$\alpha = \xi\mu \approx (7.78\cdot 10^6)\cdot(0.7\cdot 10^{-6}) \approx 5.44$$

To establish the likelihood function, we look at the data. In this example we assume that we have observed identical units for a total time in service, t, equal to 525 600 hours (e.g., 60 detector years). In this period we have observed $n = 1$ failure. If we assume exponentially distributed time-to-failures, we know that the number of failures in a period of length $t, N(t)$, is Poisson distributed with parameter $\lambda \cdot t$. The probability of observing n failures is thus given by:

$$L(\lambda; n, t) = \Pr(N(t) = n) \propto \lambda^n \mathrm{e}^{-\lambda\cdot t}$$

and we have an expression for the likelihood function $L(\lambda; n, t)$.

The posterior distribution is found by multiplying the prior distribution with the likelihood function:

$$\pi(\lambda|n) \propto L(\lambda; n, t)\cdot\pi(\lambda) \propto \lambda^n \mathrm{e}^{-\lambda\cdot t}\cdot\lambda^{\alpha-1}\mathrm{e}^{-\xi\lambda} \propto \lambda^{(\alpha+n)-1}\mathrm{e}^{-(\xi+t)\lambda}$$

and we recognize the posterior distribution as a gamma distribution with new parameters $\alpha' = \alpha + n$, and $\xi' = \xi + t$. The Bayes estimate is given by the mean in this distribution:

$$\hat{\lambda} = \frac{\alpha+n}{\xi+t} \approx \frac{5.44+1}{7.78\cdot 10^6 + 525600} \approx 0.78\cdot 10^{-6}$$

We note that the maximum likelihood estimate gives a much higher failure rate estimate ($t/n = 1.9\cdot 10^{-6}$), but the “weighing procedure” favours the prior distribution in our example. Generally we could interpret α and ξ here as “number of failures” and “time in service” respectively for the “prior information”. Note that as more and more data becomes available, the data will dominate, and the effect of the prior distribution will be wiped out.

In Bayesian probability theory, if the posterior distribution $\pi(\theta|\mathbf{t})$ is in the same probability distribution family as the prior probability distribution $\pi(\theta)$, the prior and posterior are then called conjugate distributions, and the prior is called a conjugate prior for the likelihood function $L(\theta; \mathbf{t})$.

A conjugate prior is an algebraic convenience, giving a closed-form expression for the posterior. If we cannot establish a conjugate prior we usually need numerical integration to solve the denominator in Bayes' theorem. The conjugate priors may also give some intuition because it shows how the data updates the prior distribution. In the example we had $\alpha' = \alpha + n$, and $\xi' = \xi + t$.