Modelling and eliciting expert knowledge with fictitious data.

François BILLY
EDF R&D, site de Chatou, France.

Nicolas BOUSQUET, Gilles CELEUX
INRIA Futurs, Équipe SELECT, Université Paris-Sud Orsay, France.

ABSTRACT : Considering reliability models In a Bayesian context, we propose an approach where experts opinions are supposed to come from fictitious data. Acting in such a way allows the analyst to weight the importance of the expert opinion in regard to the actual sample size in a sensible and reliable way. The control of experts knowledge becomes a control of the Fisher information on the reliability model parameters. Thus, the comparison between Feedback Experience Data (FED) and expert data through the reliability model and the prior distribution is easy and makes simple the calibration of the hyperparameters prior distribution to control the importance of expert contribution compared to the information provided by the observed sample. The presented approach is exemplified with Weibull models.

1 INTRODUCTION

The Bayesian approach is the optimal approach in statistical inference when prior knowledge from experts is available. Moreover, in such a context, the superiority of Bayesian inference over maximum likelihood can be important for small sample sizes (Robert, 1994). But, a reliable Bayesian analysis needs to take into account properly the prior knowledge. Usually, expressing expert opinions as central value of some parameters is not too difficult, but expressing the analyst doubt on expert opinions is more critical.

Considering a parametric reliability model $\mathscr{M}(\theta)$, we show that integrating of expert opinion into prior distribution can be regarded as the problem of integrating the information of a fictitious sample into a posterior distribution from a non informative prior distribution (§4). Thus, the control of expert opinions is a control of the information provided by this fictitious sample on the model parameters. To compare the information of the fictitious data on the parameter θ, we consider the Fisher information on θ provided by FED censored data (§3).

Each step of the elicitation process is exemplified with the Weibull distribution (sometimes simplified to the exponential distribution), which is the most used model in reliability, whose characteristics are reminded in §2. Moreover, the Weibull model is considered in connection with the companion paper of Bousquet *et al.* (2005), which is specifically concerned with Bayesian inference for Weibull models.

2 NOTATION

All the notions we present are illustrated with the Weibull reliability model. The Weibull probability density function (pdf) with shape parameter β and scale parameter η is

$$f_W(t) = \frac{\beta}{\eta^\beta}\, t^{\beta-1}\, e^{-\left(\frac{t}{\eta}\right)^\beta}\, \mathbb{I}_{[0,+\infty)}(t).$$

Denoting $\mu = (\frac{1}{\eta})^\beta$, it becomes

$$f_W(t) \;=\; \beta\mu\, t^{\beta-1}\, e^{-\mu t^\beta}\, \mathbb{I}_{[0,+\infty)}(t).$$

The Weibull cumulative density function (cdf) is

$$F_W(t) \;=\; 1 - e^{-\mu t^\beta}\, \mathbb{I}_{[0,+\infty)}(t).$$

Typically, data available in a reliability context are right censored failure times. The data are denoted as follows.
FED : $\mathbf{x}_n = (x_1, \ldots, x_n)$ including

- r uncensored data $\mathbf{y}_r = (y_1, \ldots, y_r)$
- $n-r$ censored data $\mathbf{c}_{n-r} = (c_1, \ldots, c_{n-r})$.

Notation for a Bayesian analysis are the following.
The prior distribution on θ is denoted $\pi(\theta)$
The posterior distribution on θ knowing data $\mathbf{x}$ is $\pi(\theta|\mathbf{x})$
We make use of a Gamma distribution with parameters a and b, with mean a/b and variance a/b^2, that it is denoted $\mathscr{G}(a,b)$.

3 INFORMATION OF CENSORED DATA

For the Weibull distribution $\mathscr{W}(\mu,\beta)$, we had chosen the prior distributions on parameter $\theta=(\mu,\beta)$ proposed in Bousquet *et al.* (2005), namely a uniform distribution for β on an interval $\Omega_\beta=[\beta_\ell,\beta_r]$ and a Gamma distribution for μ knowing β. The Bayesian model can be described as follows. Denoting X the random failure time, we have

$$\begin{aligned} X &\rightsquigarrow \mathscr{W}(\mu,\beta) \text{ with} \\ \pi(\beta,\mu) &= \pi(\beta)\,\pi(\mu|\beta) \\ &= \mathbb{I}_{[\beta_\ell,\beta_r]}(\beta)/(\beta_r-\beta_\ell)\,\mathscr{G}(a,b(\beta)) \end{aligned}$$

It is worth noting that since the prior distribution on β is weakly informative, the comparison between prior information and FED information is essentially limited to the information on μ.

Because, most of the time, FED are censored, there is the need to extend the notion of Fisher information to this particular context. First, it can be recalled that for a Weibull model, an observed failure time y_i brings the following information on μ :

$$I_\mu(y_i) \;=\; -\frac{\partial f_W^2}{\partial^2\mu} \;=\; \frac{1}{\mu^2} \;\;\forall i\in\{1,\ldots,r\}$$

which implies that an r uncensored Weibull sample of sire r brings the global information

$$I_\mu(y_1,\ldots,y_r) \;=\; \sum_{i=1}^{r} I_\mu(y_i) \;=\; \frac{r}{\mu^2}.$$

Right censored data can be of two types :

1. Fixed values, independent from the observed failure times (type-I censored data). A typical example is the shutdown of working systems at different times of care. The pdf of a type-I right-censored Weibull random variable X with fixed censored value c_j is

$$f_X(t)=[f_W(t)]^{\mathbb{I}_{\{t<c_j\}}}\;[1-F_W(t)]^{\mathbb{I}_{\{t\geq c_j\}}}.$$

It leads to the following definition of its information on μ knowing β :

$$I_\mu^I(c_j) \;=\frac{1}{\mu^2}\,E[P(X<c_j)] \;=\; \frac{1-e^{-\mu c_j^\beta}}{\mu^2}.$$

2. The random value of the r^{th} highest failure time (type-II censored data), that is the value of the r^{th} order statistic of the failure time sample, denoted t_r^*. The number r is fixed. The pdf of such a censored Weibull random variable X can be written

$$f_T(t)=[f_W(t)]^{\mathbb{I}_{\{t\leq t_r^*\}}}\;[1-F_W(t)]^{\mathbb{I}_{\{t>t_r^*\}}}.$$

Thus, such a censored data $c_j=t_r^*$ provides the information on μ

$$I_\mu^{II}(c_j) \;=\frac{1}{\mu^2}\,P\left(F_W(T)\leq\frac{r}{n}\right) \;=\; \frac{r}{n}\frac{1}{\mu^2}.$$

Remarks :

1. *$\forall c\in R^{+*}$, $I_\mu^I(c)<I_\mu$ and $I_\mu^{II}(c)<I_\mu$*
2. *$I_\mu^I(0)=I_\mu^{II}(0)=0$: no information is provided when censoring at time $t=0$.*
3. *$I_\mu^I(c) \xrightarrow{c\to\infty} I_\mu$: no information is lost with an infinite censoring time.*
4. *$I_{\mu,2}^r \xrightarrow{c\to n} I_\mu$: the information is increasing with the r^{th} failure time.*

Finally, the global Fisher information on μ brought by the FED can be written, according to the case, as

$$I_{\mu|\beta}^{I}(FED) \;=\; \frac{r}{\mu^2}+\frac{\sum_{k=1}^{n-r}\left(1-e^{-\mu\,c_k^\beta}\right)}{\mu^2}, \quad (1)$$

$$I_{\mu}^{II}(FED) \;=\; \frac{r}{\mu^2}+\frac{r}{n}\frac{(n-r)}{\mu^2}. \quad (2)$$

In the following we always note

$$I_{\mu|\beta}(FED) \;=\; \frac{\tilde{n}}{\mu^2}. \quad (3)$$

4 THE EXPERT FICTITIOUS DATA

When no prior information is available on the parameter θ of a model $\mathscr{M}(\theta)$, a natural choice for the prior distribution $\pi(\theta)$ is the non informative Jeffreys distribution (see Robert 1994). In many cases, the Jeffreys distribution can be seen as a limit position of a family of conjugate prior distributions. For instance, for an exponential distribution with inverse scale parameter λ, the Jeffreys distribution $J(\lambda)=\frac{1}{\lambda}$ is the limit position denoted $\mathscr{G}(0,0)$ hereafter, of a $\mathscr{G}(a,b)$ distribution with hyperparameters $a,b\to 0$.

In such a case, a Jeffreys distribution can be expressed as a conjugate prior distribution for $\mathscr{M}(\theta)$. We can take profit from this feature to interpret a prior opinion as the information provided by a fictitious data set from a non informative prior distribution. Continuing the exponential example, consider a fictitious failure times sample $(y'_1,\dots,y'_a)$ with $a \in \mathbb{N}^*$ from the $\mathscr{E}(\lambda)$ distribution. With a non informative prior distribution $\mathscr{G}(0,0)$ the posterior distribution of λ knowing the fictitious sample is $\mathscr{G}(0+a, 0+\sum_{i=1}^{a} y'_i)$.

Thus, denoting $b = \sum_{i=1}^{a} y'_i$, an expert opinion modelled with a Gamma $\mathscr{G}(a,b)$ distribution, can be seen as the integration of a fictitious exponential sample with size a and mean b/a to the non informative prior distribution $\mathscr{G}(0,0)$ on λ. It means that an expert opinion on λ can be translated in terms of a fictitious likelihood leading to a posterior distribution $\mathscr{G}(a,b)$ distribution.

Thereby it becomes easy to estimate the impact of the prior distribution on the posterior distribution. For conjugate distributions, the fictitious size a is replaced by the sum of sizes $a+r$ and the information coming from prior mean b is replaced with $b + \sum_{i=1}^{n} x_i$. Finally, the posterior estimation of θ can be regarded as the maximum likelihood estimation on the union of FED and fictitious data.

Note that a corresponding approach on generalized linear models was proposed in Bedrick *et al.* (1996). Moreover, a study of minimal informative priors specifically for model selection, with similar use of fictitious data, has been done in Andrieu *et al.* (2000).

The prior distributions of Weibull parameters (μ,β) have been chosen as in Bousquet *et al.* (2005). The prior distribution $\pi(\mu|\beta) = \mathscr{G}(a,b)$ is a conjugate prior distribution conditionally to β. Hyperparameter b is chosen as $b(\beta)$ and is defined now as $\sum_{i=1}^{a} (y'_i)^{\beta}$ with fictitious Weibull data y'_i.

Note that from this general point of view, combining several expert opinions is easily achieved by concatenating all opinions in a unique fictitious sample. Each one of the fictitious sample provides a Fisher information on μ, which must be compared to the information provided by FED in order to calibrate the final prior distribution. Such considerations are presented in §5.

5 CALIBRATING THE PRIOR DISTRIBUTION

The next step of the elicitation is the calibration of the prior distribution (measuring its importance in comparison to the information provided by FED). The calibration of the prior distribution leads generally a rule for controlling the prior variance of the parameters. In the Weibull example, this prior variance for μ around an estimation $\bar{\mu}$, knowing β, can be written

$$\mathrm{Var}(\mu|\beta) \;=\; \frac{a}{b^2} \;=\; \frac{\bar{\mu}^2}{a}.$$

The control of the prior variance is achieved by controlling the Fisher information of the underlying fictitious data $\mathbf{Y}'$ under the following hypothesis.

Hypothesis : *It is assumed that the expert is able to derive a consistent (asymptotically unbiased) estimator $\bar{\mu}$, knowing β, with minimum variance, for a certain size a of underlying sample.*

Thus, the Cramer-Rao theorem ensures that

$$Var\ (\mu|\beta) \;=\; \frac{1}{I_\mu(\mathbf{Y}')})$$

where $I_\mu(\mathbf{Y}')$ is the Fisher information on μ of the fictitious data $\mathbf{Y}'$. In fact, a represents the quantity of Fisher information of the fictitious data on μ.

In the Weibull example, from relations (1) and (2) we can remark that

- Hyperparameter a can be < 1 when an expert is considered as unreliable. In such a case, its opinion can be regarded as a censoring time.
- Choosing a is scoring the level of confidence of the expert. This score is not an absolute score, but is given in accordance to the information provided by the FED. From that point of view, it is important to note that combining experts opinions has to be achieved in accordance to this common rule (see Bousquet *et al.*, 2005).

To end this paper, it is worth to remark that seeing the information provided by an expert as the information provided by a fictitious data set is useful to control the reliability of his opinion. For instance, when an expert gives a decile on a Weibull distribution, its means that it has integrated in his knowledge the information provided by at least ten observed failure times for the material into study. As a consequence, it seems natural to assume that $a \geq 10$. But, in many cases, it is doubtful that such a large value for a is reasonable in regard to the FED size. The companion paper Bousquet *et al.* (2005) is analyzing such situations.

Andrieu C., Doucet A., Fitzgerald W.J. and Pérez J.-M. (2000) ‘Bayesian computational approaches to model selection’, Nonlinear and Non Gaussian Signal Processing, Cambridge : Newton Institute Series. Cambridge University Press.

Bedrick J., Christensen R. and Johnson W. (1996) ’A new perspective on priors for generalized linear models’ *Journal of the American Statistical Society*, **91**.

Bousquet, N., Celeux G. and Rémy E. (2005) ‘A protocol for integrating FED and expert data in a study of durability using the Weibull distribution’. Proceedings of the Workhsop on the Use of Expert Judgement in Decision-Making, 21-23 June, Cadarache.

Robert, C. P. (1994) *‘The Bayesian Choice’*, Springer-Verlag.