Calibrating Weibull priors using virtual data

Nicolas Bousquet
EDF Research and Development
Department of Industrial Risks Management
6 quai Watier, 78401 Chatou
France
nicolas.bousquet@edf.fr

Abstract

A Bayesian reliability framework is considered where an expert questioned on the prospective behavior of a device Σ answers in terms of lifetime values and gives crude qualitative information about its ageing. Our desirated procedure consists in weighting the expert information by the size m of a virtual sample yielding a similar information when working with the nonconjugate Weibull distribution for modelling the lifetime T. The prior is also seen as the posterior of this virtual sample. An attractive result is the full tractability of the prior in mild conditions. The interest of such a procedure is to make easier the calibration task where the subjective information has to be moderated compared to the data information since m is a practical focus point for a discussion between Bayesian analysts and experts. Besides such elicited priors are interesting for sensitivity studies. Some methods for calibrating the prior are considered, and exemplified on a real case.

1 Presentation

1.1 Context

The versatile Weibull $\mathcal{W}(\eta,\beta)$ distribution, with probability density function (pdf)

$$f_W(t|\eta,\beta) \quad = \quad \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \exp\left\{-\left(\frac{t}{\eta}\right)^{\beta}\right\} \mathbb{1}_{\{t\geq 0\}}$$

$(\eta,\beta)\in[0,\infty)^2)$, is often chosen for modelling T in industrial reliability studies. Frequently, observed lifetime data $\mathbf{t_n}=t_1,\ldots,t_n$ are small-sized and contain missing or censored values. When the framework is Bayesian since expert knowledge is available, one has to pay great attention to the elicitation of the prior pdf $\pi(\eta,\beta)$ that formalizes the expert knowledge. Indeed, it inserts subjectivity into decision-making functions integrated over the joint posterior distribution of pdf $\pi(\eta,\beta|\mathbf{t_n})$. These posterior decisions must also be defendable face to control authorities as tools of decision-making, which implies that the prior pdf must be itself clearly defendable.

There are two main difficulties to work with Weibull distributions: (a) its only conjugate prior distribution is continuous-discrete (Soland 1969) and remains difficult to justify and use in real problems; (b) the meanings of scale parameter η and shape parameter β greatly differ. Their correlation remain little intuitive to assess by non-statistician experts. Therefore Kaminskiy and Krivtsov (2005) recently provided a simple procedure to elicit a prior $\pi(\eta,\beta)$ using expert knowledge on the mean and standard deviation of the cumulative distribution function (cdf) $F_W(t|\eta,\beta)$. They insist on the fact that these values are easier to assess than parameter values. In our alternative view, the approach introduced here can prevent from the remaining difficulty to elicit standard deviations.

1.2 Prior information

Because of various factors, especially subjective ones, expert information on lifetime T is summarized through questioning processes (O'Hagan et al. 2006). We assume that basically, an expert can yield

1. a *quantitative* information, under the form of an estimation $t_\alpha^{(e)}$ of the α-order percentile, namely $P(T<t_\alpha^{(e)})=\alpha\in]0,1[$, defined with respect to the marginal prior pdf

$$f(t) \quad = \quad \iint_{\mathbb{R}_+^2} f_W(t|\eta,\beta)\pi(\eta,\beta)\; d\eta d\beta. \tag{1}$$

Experts are thus assumed to be independent from any parametrization choice (and even from any sampling model choice) made by the Bayesian analyst.

2. a *qualitative* information on the nature of ageing of Σ: we assume that an δ-order percentile $\beta_\delta^{(e)}$ of $\pi(\beta)$ can be set through questioning process or using databasis (*http://www.barringer1.com*). Especially, if the aim of the inference is ageing detection, it requires $\beta_\delta^{(e)} = 1$ and $\delta = 0.5$ *a priori*; if one is more interested in detecting ageing acceleration, it requires $\beta_\delta^{(e)} = 2$ and $\delta = 0.5$. Otherwise, specifying an extreme quantile $(1-\delta \ll 1)$ is technically not difficult since a large β (> 5) reflects inconceivable kinetics of ageing in industrial applications.

2 Prior modelling

2.1 Central idea

The elicitation central idea comes from the simple "frequentist" vision of an informative expert opinion. We suggest to consider that a perfect expert opinion should be, roughly speaking, similar to a real data survey, and provide an i.i.d sample $\tilde{\mathbf{t}}_{\mathbf{m}} = (\tilde{t}_1, \ldots, \tilde{t}_m)$ of lifetime data. The virtual size m is a practical, intuitive lever for calibrating the prior uncertainty. This will be discussed further. Assuming to know $\tilde{\mathbf{t}}_{\mathbf{m}}$, an ideal prior should be the posterior distribution with density $\pi^J(\eta, \beta|\tilde{\mathbf{t}}_{\mathbf{m}})$, where π^J is a consensual, formal representation of ignorance (namely, a noninformative prior). Jeffrey's prior

$$\pi^J(\eta,\beta) \quad \propto \quad \eta^{-1} \mathbb{1}_{\{\eta \geq 0\}} \mathbb{1}_{\{\beta \geq 0\}}$$

is chosen here (instead of the classic reference prior) since its posterior is proper for any small virtual size m.

2.2 Single expert prior modelling

Replacing possibly the virtual sample by sufficient statistics, this kind of elicitation is often practical for conjugate distributions but not for Weibull. As said before, besides, the complete information $\tilde{\mathbf{t}}_{\mathbf{m}}$ is summarized as assumed in § 1.2. Fortunately, after some work (Bousquet 2009) one can prove next proposition.

Proposition 1. *The only possible choice of a continuous prior $\pi(\eta, \beta)$ respecting the frequentist view and the prior constraints in (1.2) is*

$$\eta|\beta \quad \sim \quad \mathcal{GIG}\left(m, b_\alpha^{(e)}(m,\beta), \beta\right), \quad \beta \ \sim \ \mathcal{G}\left(m, \frac{m}{\widetilde{\beta}^{(e)}(m)}\right) \tag{2}$$

where $\mathcal{G}$ and $\mathcal{GIG}$ refer to the gamma and generalized inverse gamma distributions,

$$b_\alpha^{(e)}(m,\beta) \quad = \quad \left((1-\alpha)^{-1/m} - 1\right)^{-1} \left(t_\alpha^{(e)}\right)^\beta, \tag{3}$$

$$\widetilde{\beta}^{(e)}(m) \quad = \quad 2m\beta_\delta^{(e)} / \chi^2_{2m}(\delta), \tag{4}$$

and $\chi^2_{m/2}(q)$ is the $q-$order percentile of the $\chi^2_{m/2}$ distribution.

2.3 Multiple experts prior modelling

In our view, two experts are dependent (resp. independent) if they share (do not share) virtual data in their past experience. We only treat here independent experts. In this case, the aggregation of $i = 1, \ldots, p$ opinions yields a similar information to the information carried by a global virtual sample, which is the concatenation of all expert samples $\tilde{\mathbf{t}}_{\mathbf{m_i}}$. Technically, the global prior is $\pi^J(\eta, \beta|\tilde{\mathbf{t}}_{\mathbf{m_1}}, \ldots, \tilde{\mathbf{t}}_{\mathbf{m_p}})$.

Because the complete virtual sample is not known, one case use an observed concatenation of samples $\tilde{\mathbf{s}}_{\mathbf{m_1}}, \ldots, \tilde{\mathbf{s}}_{\mathbf{m_p}}$ from another model $\mathcal{M}(\eta,\beta)$ such that its replacement parametric likelihood $(\eta, \beta) \mapsto \ell(\tilde{\mathbf{s}}_{\mathbf{m_1}}, \ldots |\eta, \beta)$ leads to the same inference than the virtual sample. Indeed, we can show easily that

$$\pi^J_{\mathcal{W}}\left(\theta|\tilde{\mathbf{t}}_{\mathbf{m_1}}, \ldots, \tilde{\mathbf{t}}_{\mathbf{m_p}}\right) \quad = \quad \pi^J_{\mathcal{M}}\left(\theta|\tilde{\mathbf{s}}_{\mathbf{m_1}}, \ldots, \tilde{\mathbf{s}}_{\mathbf{m_p}}\right) \ \propto \ \pi^J(\theta) \prod_{i=1}^p \ell(\tilde{\mathbf{s}}_{\mathbf{m_i}}|\eta, \beta).$$

Next proposition gives an example of replacement likelihood for one expert opinion. The resulting final prior for all expert opinions is (2) for which $m = \sum_{i=1}^{p} m_i$, $b^{(e)}_{\alpha}(m,\beta) = \sum_{i=1}^{p} b^{(e_i)}_{\alpha_i}(m_i,\beta)$ and

$$\widetilde{\beta}^{(e)}(m) \quad = \quad m\left(\sum_{i=1}^{p}\frac{m_i}{\widetilde{\beta}^{(e_i)}(m_i)}\right)^{-1}$$

Proposition 2 . Consider $\tilde{\mathbf{s}}_{\mathbf{m}} = (k_{\alpha,m}, \beta_{t_\alpha,\delta})$, where $k_{\alpha,m} = ((1-\alpha)^{-1/m}-1)^{-1}$ and $\beta_{t_\alpha,\delta} = \widetilde{\beta}^{(e)}(m)/(1+\widetilde{\beta}^{(e)}(m)\log t^{(e)}_{\alpha})$, as a sample whose components follow independently the $\mathcal{G}(m, m(t^{(e)}_{\alpha}/\eta)^{\beta})$ and $\mathcal{IG}(m, m\beta)$ distributions, respectively. Then its defines a replacement likelihood for the expert opinion summarized by $(t^{(e)}_{\alpha}, \alpha, \beta^{(e)}_{\delta}, \delta)$.

2.4 Calibrating virtual size m

A discussion between the expert(s) and the Bayesian analyst is needed to calibrate m. Default values can easily be chosen to open this discussion. For instance, histogram methods (van Noortwijk et al. 1992) or bisection methods (Garthwaite et al. 2005) can be used to assess m in a logical correspondence with the quantitative specification: specifying a 90%-order percentile requires to have "perceived" at least 10 past values. These first values of m can rapidly increase in function of the precision of α: this is a welcome effect since it favors a cautious reaction of the expert, whose self-confidence is often overoptimistic in reliability and risk assessment (Lannoy and Procaccia 2001).

Posterior conservatism. Before the inference, the final elicited m should be compared to the size n of the real data $\mathbf{t_n}$. Choosing $m < n$ can be regarded as necessary for the posterior results to be accepted by a control authority since it avoid a posterior decision mainly guided by the subjective expert information. Sensitivity studies are interesting too, that can easily be led in function of $\rho = m/n$. The aim can be to understand how many independent expert opinions, each similar to one i.i.d. data, are necessary to get a particular posterior result.

In experiments when real data can be strongly censored, the size r of the uncensored data can be considerably lower than n. But censored data yield partial information and forcing $m < r$ might restrict abusively the prior information. A Bayesian interpretation of the effective sample "size" (ESS) $\tilde{n}$ of $\mathbf{t_n}$ can be given, based on a more general work by Lin et al. (2007): on average, $\tilde{n}$ is the minimal mean size of i.i.d samples $\mathbf{t_s}$ yielding the closest data information than $\mathbf{t_n}$. This closeness is measured with the Kullback-Leibler distance between the reference posteriors $\pi^J_c(\eta,\beta|\mathbf{t_n})$ and $\pi^J(\eta,\beta|\mathbf{t_s})$, π^J_c being a noninformative Jeffreys' prior dedicated to censored cases with better posterior coverage properties than π^J (De Santis et al. 2001). Namely, denoting $\Theta = I\!R_+ \times I\!R_+$ and $\theta = (\eta,\beta)$,

$$\tilde{n} \quad = \quad \arg\min_{s} \iint_{R^s_+\times\Theta} \pi^J_c(\theta|\mathbf{t_n}) \log\frac{\pi^J_c(\theta|\mathbf{t_n})}{\pi^J(\theta|\mathbf{t_s})}\ d\theta d\mathbf{t_s}.$$

From Lin et al. (2007) existence and unicity of $\tilde{n}$ are ensured and $r \leq \tilde{n} \leq n$. Besides, $\tilde{n}$ can be defined on the real line, which offers more versatility to the prior.

3 Extended prior modelling

Consider now the case where an expert provides p non-independent specifications $\{t^{(e)}_{\alpha_i}, \alpha_i\}_{i\in\{1,\ldots,p\}}$. They lead to p marginal basic priors $\pi_1,\ldots,\pi_p$ which can be agglomerated as non-independent p expert opinions.

This case strongly differs from the previous case (Section 1) since the expert uncertainty is indirectly assessed through the choices of the α_i. Select the most trustworthy prior π_i as a "envelope" prior. Then (m,δ) are the parameters to calibrate. Following Cooke (1991) we choose for it a discrete Kullback-Leibler loss function between desirated and elicited marginal features:

$$(m,\delta) \quad = \quad \arg\min \sum_{j\neq i}^{p} (\alpha_{j+1}-\alpha_j) \log\frac{(\alpha_{j+1}-\alpha_j)}{(\alpha^{(i)}_{j+1}-\alpha^{(i)}_j)} \tag{5}$$

where $\alpha_0 = \alpha_0^{(i)} = 0$, $\alpha_{p+1} = \alpha_{p+1}^{(i)} = 1$ and $\alpha_j^{(i)} = \alpha_j^{(i)}(m,\delta) = \iint_{I\!R_+^2} F_W\left(t_{\alpha_j}^{(e)}|\eta,\beta\right)\ \pi_i(\eta,\beta)\ d\eta\beta$.

4 Application: quantifying prior ageing

Among possible applications of previous results, one can elaborate a strategy where m is fixed or constrained as described in the previous section, after a dialog with the expert, and δ has to be calculated, fixing a reasonable value $\beta_\delta^{(e)}$. In most industrial applications (Lannoy and Procaccia 2001), β evolutes between 1 and 5. Hence underlying ageing placed into the marginal expert information can be quantified.

The common expert median from Table 1 being the most trustworthy prior specification, we assume $\beta_\delta = 2$ (being sure of ageing). Then the underlying ageing from expert $\mathcal{E}_1$ is surely accelerated since the solution of (5), m for instance being set to 5, is $\hat{\delta} = 6.10^{-4}$. Effective percentiles orders for $(200, 300)$ are $(0.05, 0.96)$. About expert $\mathcal{E}_2$, we get $\hat{\delta} = 0.13$ which is logical for this wider opinion. Effective percentiles orders for $(100, 500)$ are $(0.05, 0.96)$ again.

The two experts opinions being independent, one can finally build a common prior of virtual size $m = 10$, which nearly yields as much inferential information as the data since the ESS of data given in Table 1 is $\tilde{n} \simeq 11$. Prior and posterior distributions are finally plotted on Figures 1 and 2.

Table 1: Lifetimes (months) of EDF components (left) and two expert opinions (right).

		expert	cred.intervals (5%,95%)	median value
real failure times:	134.9, 152.1, 133.7, 114.8, 110.0, 129.0, 78.7, 72.8, 132.2, 91.8	$\mathcal{E}_1$	[200,300]	250
right-censored times :	70.0, 159.5, 98.5, 167.2, 66.8, 95.3, 80.9, 83.2	$\mathcal{E}_2$	[100,500]	250

References

Bousquet, N. (2009). An easily-calibrated prior family for the Bayesian reliability analysis of the Weibull distribution, *in progress*.

Cooke, R.M. (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science*. New York: Oxford University Press.

De Santis, F., Mortera, J. and Nardi, A. (2001). Jeffreys priors for survival models with censored data, *J. Statis. Planning and Inference*, **99**, pp. 193-209.

Garthwaite, P.H., Kadane, J.B. and O'Hagan, A. (2005). Statistical methods for eliciting probability distributions, *J. Amer. Statist. Assoc.*, **100**, pp. 680-701.

Kaminskiy, M.P., and Krivtsov, V.V. (2005). A Simple Procedure for Bayesian Estimation of the Weibull Distribution, *IEEE Trans. Reliability, 54*, pp. 612–616.

Lannoy, A. and Procaccia, H. (2001). *L'utilisation du jugement d'expert en sûreté de fonctionnement*, Tec & Doc (in French).

Lin, X., Pittman, J. and Clarke, B. (2007). Information Conversion, Effective Samples, and Parameter Size, *IEEE Trans. Inform. Theory*, **53**, pp. 4438-4456.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., Rakow, T. (2006), *Uncertain Judgements: Eliciting Expert Probabilities.* John Wiley and Sons, Chichester.

Soland, R. (1969). Bayesian analysis of the Weibull process with unknown scale and shape parameters, *IEEE Transactions on Reliability, 18*, pp. 181–184.

van Noortwijk, J.M., Dekker, R., Cooke, R.M. and Mazzuchi, T.A. (1992). Expert judgment in Maintenance Optimization, *IEEE Transactions on Reliability*, **41**, pp. 427-431.

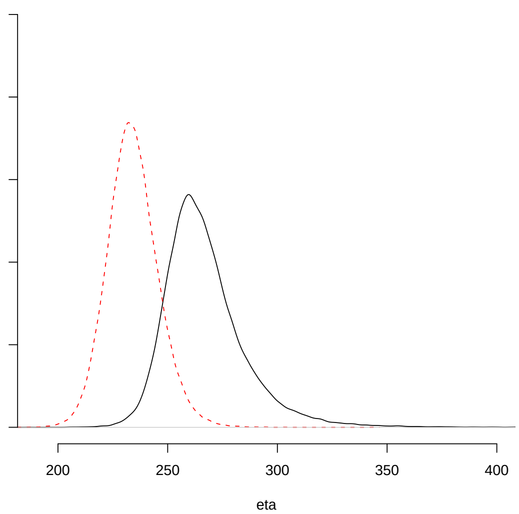

Figure 1: Prior and posterior pdf of η, for a full virtual size $m = 10$.

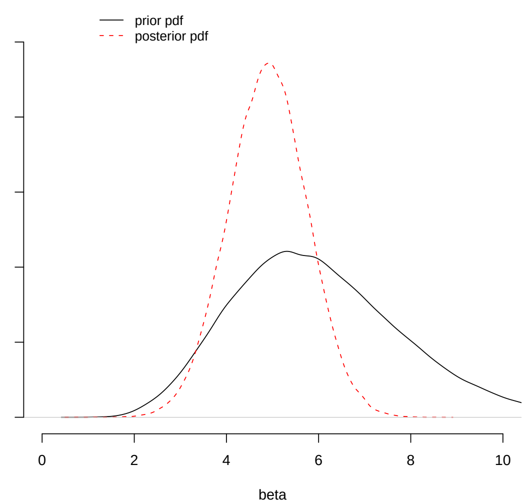

Figure 2: Prior and posterior pdf of β, for a full virtual size $m = 10$.