Quelques principes de modélisation bayésienne pour le traitement des incertitudes

Nicolas Bousquet

EDF R&D

CEA-DAM, 1er octobre 2014

Basically, we are interested in explaining the behavior of a random vector of interest

$$Y = g(X, d)$$

- g is (most often) a deterministic function (computer model)
- X is a random vector of inputs living in a probability space χ, often associated to a parametric density $X \sim f(x|\theta)$
- d is a set of environmental fixed parameters in $\boldsymbol{D}$

Bayesian statistical modelling of uncertainties

Technical tools of assessment and exploration allowing for:

- agglomerating several sources of information (*legacy data, simulated data, expert opinions, constraints...*)
- differentiating aleatory and epistemic uncertainties in assessment and simulation
- decision-helping under uncertainty

Main assumption

Quantities usually considered as fixed $(\theta, g(x_0, d))$ but unknown are given the sense of realizations of random variables associated to so-called prior distributions

Conditioned to data (observed *in situ* or produced by numerical means) $(x_1, \ldots, x_n, y_1, \ldots, y_n)$, the uncertain prior information is updated and summarized by the posterior distribution

Example on θ

Given a prior measure (or density) $\pi(\theta)$, the posterior density is given by [Bayes rule]

$$\pi(\theta|x_1, \ldots, x_n) = \frac{\ell(x_1, \ldots, x_n|\theta)\pi(\theta)}{\int_\Theta \ell(x_1, \ldots, x_n|\theta)\pi(\theta)\ d\theta}$$

where $\ell(x_1, \ldots, x_n|\theta)$ is the data likelihood

Consequences:

1. the posterior distributions "traduce" less uncertainty on the true $(\theta, g(x_0, d))$ than the prior distributions
2. rather than focusing on single estimators $(\hat{\theta}_n, \hat{g}_n(x_0, d))$, one focuses on estimating the whole posterior distribution

A practical view of mind, but reinforced by the de Finetti theorem (1931), generalized by Hewitt, Savage (1955), Diaconis, Freedman (1980)

Let $X_1, \ldots, X_n, \ldots$ be an exchangeable sequence of 0-1 random variables with joint probability P. Then there exists a unique probability measure $\pi(\theta)$ such that

$$P(X_1 = x_1, |dots, X_1 = x_n, \ldots) \quad = \quad \int_\Theta f(x_1, \ldots, x_n, \ldots |\theta)\pi(\theta)\ d\theta$$

where $f(x_1, \ldots, x_n|\theta)$ is the likelihood of iid Bernoulli observations

Consequences:

- Bayesian modelling appears as a natural statistical modelling for correlated but exchangeable data
- Formal existence of a prior $\pi(\theta)$ defined by the sampling mechanism
  $= \{$ uncertain information about the state of nature θ $\}$

In facts:

- The sampling model $f(x|\theta)$ is determined by statistical testing, physical reasoning... (aleatory part)
- Need to define a prior (epistemic part)

Subjectivist view: a prior/posterior probability is a degree of belief

1 - Modelling of inputs: $X \sim f(x|\theta)$ with $\theta \in \Theta \subset \boldsymbol{R}^d$

- (a) estimate the posterior distribution $\pi(\theta|\mathbf{x_n})$
- (b) predict the next input x_{n+1} according to the posterior predictive distribution

$$f(x_{n+1}) = \int_\Omega f(x_{n+1}|\theta)\pi(\theta|\mathbf{x_n})d\theta$$

2 - Emulating a time-consuming code g: a stochastic prior is placed on g through the choice of a random process (e.g., Gaussian)

$$\forall z = (x, d) \in \chi \times \boldsymbol{D}, \qquad g(z) \sim m(z)^T\theta_1 + G(z)$$

with $\mathrm{E}_f[H(z)] = 0$ and $\mathrm{Cov}(H(z), H(z')) = \theta_2^2 R_{\theta_3}(\|z - z'|)$

- Ex: Bayesian kriging for estimating $\theta = (\theta_1, \theta_2, \theta_3)$ using a numerical DOE (Berger et al. 2001, Paulo 2005, Helbert et al. 2009, Deng et al. 2011...)

3 - Mixing both frameworks for **Bayesian inversion**

Possibly, the data $\mathbf{x_n}$ are missing and the posterior is $\pi(\theta|\mathbf{y}_\mathbf{n}^*)$ where

$$y_i^* = g(x_i, d_i) + \epsilon_i$$

Prospective questions about prior elicitation of parameters of input distributions

expert strenght quantifying the ratio "prior information" / "data information"

- needs for an understandable definition

practicity is π easy to handle? [Rios Insua and Ruggeri 2000]

- explicit if possible (sensitivity studies are simplified)
- easy to sample (comparisons a posteriori-a priori)
- defined by formal rules in a unique way

coherence w.r.t. consensual qualitative knowledge on Σ, is π coherent?

equitability Do the complete Bayesian models $(f_i(.|\boldsymbol{\theta}_i), \pi_i(\boldsymbol{\theta}_i))$ be equitable?

- a model should not be arbitrarily favorized a priori
- the prior of a nested sampling model should be itself nested in the prior of a more complex model

Remark: Incorporation of subjective information

Subjective degrees of belief (*gambles*) are biased and should be partially

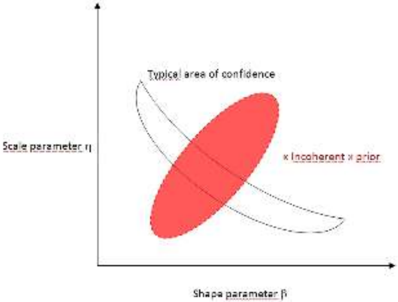
- corrected from empirical studies and meta-analyses in similar situations (Lannoy & Procaccia 2003)
- accounted for in the decision process via game theory (Green 2002)
- reduced by theories of evidence and knowledge representation (e.g., Dempster-Shafer theory)

Weibull distribution in lifetime data analysis

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

A prior $\pi(\theta) = \pi(\beta,\eta)$ with strongly positive correlation threatens to be incoherent with the meaning of the model:

- high $\beta \Leftrightarrow$ strong ageing $\Rightarrow$ short lifetime $\Leftrightarrow$ small η



Σ = Steam turbine rotor from a fossil-fuel power station

Feedback experience lifetime

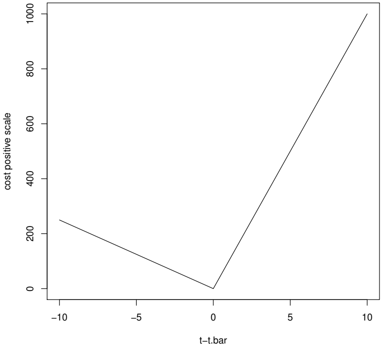
real failure times:	134.9, 152.1, 133.7, 114.8, 110.0 129.0, 78.7, 72.8, 132.2, 91.8
right-censored times :	70.0, 159.5, 98.5, 167.2, 66.8 95.3, 80.9, 83.2

To account for technical improvement, two independent experts can express their feeling about X and the qualitative behavior of Σ (aging)

A decision-theoretic elicitation in reliability (1/2)

1. Consider a replacement of Σ planned at time $X = \bar{t}$
2. Denote x the unknown true lifetime of Σ
3. It x were known, there would be the approximate costs (at first order)
   - $C_2(\|x - \bar{x}\|)$ to have been too pessimistic if $x > \bar{x}$
   - $C_1(\|x - \bar{x}\|)$ to have been too optimistic if $x < \bar{x}$

Question: can you plot how the costs evolve in a signed scale?



Denote $\delta = C_2(\|x-\bar{x}\|)/C_1(\|x-\bar{x}\|)$ and

$$L_\delta(x,\bar{x}) = \frac{1}{1+\delta}(\bar{x}-x)\mathbb{1}_{\{x\leq\bar{x}\}} + \frac{\delta}{1+\delta}(\bar{x}-x)\mathbb{1}_{\{x>\bar{x}\}}$$

the total cost function

the mean cost due to carrying out the management decision at time $\bar{x}$ is

$$\ell_\delta(\bar{x}) = \int_\Theta \left[\int_0^\infty L_\delta(f,\bar{f}) f(x|\theta)\, dx\right] \Pi(d\theta)$$

The aim of the dialog between expert and analyst is to estimate a couple

$$(\bar{x},\delta)^* = \arg\min_{x,\delta} \ell_\delta(x)$$

(what is the best decision and how much are the associated costs?)

Given δ, $\bar{x}^*$ is the prior predictive percentile of order α:

$$P(X<\bar{x}^*) = \int_0^{\bar{x}^*} P(X<\bar{x}^*|\theta)\Pi(d\theta) = \alpha = \delta/(1+\delta)$$

Typically, a maximum entropy approach can be conducted to elicit a prior density $\pi(\theta) = \Pi(d\theta)$ under this kind of constraint

$$\textbf{Principle}: \quad \pi^{ME}(\theta) \quad = \quad \arg \max_{\pi(\theta) \geq 0} H^J(\Theta)$$

with $H^J(\Theta)$ the relative entropy

$$H^J(\Theta) \quad = \quad -\int_\Theta \pi(\theta) \log \frac{\pi(\theta)}{\pi^J(\theta)} \, d\theta$$

- where $\pi^J(\theta)$ is a benchmark (noninformative) prior measure
- under linear constraints $\int_\Theta g_i(\theta)\pi(\theta) \, d\theta = c_i$ for $i = 1, \ldots, q$

$$\Rightarrow \quad \pi^{ME}(\theta) \quad \propto \quad \pi^J(\theta) \exp \left\{ -\sum_{i=1}^{q} \lambda_i g_i(\theta) \right\}$$

Probably the most popular systematic elicitation method

- allowing to progress carefully from noninformativeness to informativeness
- usage quite broad for many applications (Jaynes 2003,...)

Example: θ lives in $[1,2] \Leftrightarrow \pi^{ME}(\theta) = \pi^J(\theta)$ is uniform

$$\text{assuming} \quad \theta \sim \mathcal{U}[1,2] \quad \text{is not equivalent to assuming} \quad \theta^{-1} \sim \mathcal{U}[1/2,1]$$

Invariance rule for defining noninformative priors (Jeffrey priors)

$$\pi^J(\theta) \quad \propto \quad \sqrt{\det I(\theta)}$$

For any bijective variable change $\eta = h(\theta)$, one has

$$\pi^J(\eta) \quad \propto \quad \sqrt{\det I(\eta)}$$

Besides, we want that the data remain much more conclusive than the prior $\Rightarrow$ need for a "broader" criterion of non-informativeness than $H^J(\Theta)$

Definition: Maximal data information priors (Zellner 1977, 1991)

$$\pi^{MDI}(\theta) = \arg \max_{\pi(\theta) \geq 0} G(\Theta)$$

with $G(\Theta) = \mathrm{E}_\theta \left[H^J(\Theta) - Z(\theta) \right]$ where $Z(\theta)$ = entropy of the sampling model

$$Z(\theta) = \int f(x|\theta) \log f(x|\theta) \, dx$$

$G(\Theta)$ gives "the total information provided by an experiment over and above the prior" (Zellner 1997)

Maximizing gain $G(\Theta)$ implies to minimize the information carried by $\pi(\theta)$ through the inference (Soofi 2000)

$$\pi^{MDI}(\theta) \propto \pi^J(\theta) \exp\left(Z(\theta) - \sum_{i=1}^{p} \lambda_i g_i(\theta) \right)$$

A second issue

The obvious propriety constraint

$$\int_\Theta \pi^{ME}(\theta)\; d\theta \quad = \quad \int_\Theta \pi^{MDI}(\theta)\; d\theta \;=\; 1$$

remains inoperative in the (Lagrange) resolution

Doing better: an indirect prior form constraint (Soofi 2000, Soofi et al. 2007, B. 2010)

The integrability of π over Θ implies

$$\int_\Theta Z(\theta)\pi(\theta)\; d\theta \quad = \quad c < \infty \tag{1}$$

The ME prior under (1) encompasses usual ME and MDI priors

$$\pi^{MEH}(\theta) \quad \propto \quad \pi^J(\theta)\exp\left(-\gamma_0 Z(\theta) - \sum_{i=1}^{p} \lambda_i g_i(\theta)\right).$$

Coming back to our industrial example: Weibull lifetimes

Use the noninformative prior $\pi^J(\eta,\beta) \propto (\eta\beta)^{-1}$

The MEH prior takes the conditional form

$$\begin{aligned} \eta|\beta &\sim \mathcal{G}(\gamma_0, \lambda_1\Gamma(1+1/\beta)), \\ \pi^{MEH}(\beta) &\propto \frac{\beta^{-1}}{\Gamma^{\gamma_0}(1/\beta)} \exp(-\gamma_0\gamma/\beta) \end{aligned}$$

It is proper iif $\gamma_0 > 0$

γ_0 is the solution of

$$\Pi(\beta > \beta_e) = \frac{\displaystyle\int_{\beta_e}^{\infty} \frac{\beta^{-1}}{\Gamma^{\gamma_0}(1/\beta)} \exp\left(-\frac{\gamma\gamma_0}{\beta}\right) \, d\beta}{\displaystyle\int_{0}^{\infty} \frac{\beta^{-1}}{\Gamma^{\gamma_0}(1/\beta)} \exp\left(-\frac{\gamma\gamma_0}{\beta}\right) \, d\beta} = 1 - \alpha_e.$$

and

$$\lambda_1 = \gamma_0/x_e \qquad \text{where } x_e \text{ is the prior median lifetime}$$

Default case : $\Pi(\beta > 1) = 1/2$ (the experts have no opinion on aging...) $\Rightarrow \log\gamma_0 \simeq -5 \Rightarrow$ extremely flat but integrable prior

Assume $X|\theta$ lives in the natural exponential family:

$$f(x|\theta) = h(x)\exp(\theta \cdot x - \psi(\theta))$$

then the ME prior defined by

$$\pi(\theta|x_0, m) = K(a,b)\exp(\theta \cdot x_0 - m \cdot \psi(\theta)) \quad (2)$$

is conjugated: given $\mathbf{x_n} = (x_1, \ldots, x_n)$, then

$$\pi(\theta|x_0, m, \mathbf{x_n}) = \pi(\theta|t_0 + \sum_{i=1}^{n} x_i, m+n)$$

The posterior predictive mean is

$$\mathrm{E}[X|\mathbf{x_n}] = \frac{x_0 + n\bar{x}_n}{m+n} \quad (3)$$

Under continuity conditions, (3) $\Rightarrow$ (2) from Diaconis & Ylvisaker (1979)

The prior can be interpreted as a posterior based on both virtual data and a noninformative prior

Some conjugate prior/posterior distributions for some usual exponential families

Likelihood functions	Information	Estimated parameter	Prior	Posterior
Multinomial	$s_1, s_2, \ldots s_k$ successes in k categories	Probabilities $p_1, p_2, \ldots p_k$	Dirichlet$(\alpha_1, \alpha_2, \ldots \alpha_k)$	$\alpha_k' = \alpha_k + s_k$
Binomial	s successes in n trials	Probability p	Beta(α_1, α_2)	$\alpha_1' = \alpha_1 + s$ $\alpha_2' = \alpha_2 + n - s$
Exponential	n "times" x_i	mean$^{-1} = \lambda$	Gamma(α, β)	$\alpha' = \alpha + n$ $\beta' = \dfrac{\beta}{1 + \beta \sum_i x_i}$
Normal (with known σ)	n data values with mean $\bar{x}$	Mean μ	Normal(μ_μ, σ_μ)	$\mu_\mu' = \dfrac{\mu_\mu(\sigma^2/n) + \bar{x}\sigma_\mu^2}{\sigma^2/n + \sigma_\mu^2}$ $\sigma_\mu' = \sqrt{\dfrac{\sigma_\mu^2\sigma^2}{n\sigma_\mu^2 + \sigma^2}}$
Poisson	α observations in time t	Mean events per unit time λ	Gamma(α, β)	$\alpha' = \alpha + x$ $\beta' = \dfrac{\beta}{1 + \beta t}$

courtesy of VS-RSF

Imagine the "true" prior information is $\tilde{\mathbf{x}}_{\mathbf{m}} \sim f(.|\theta)$ of size m

A nice (and logical) prior is $\pi(\theta) = \pi^J(\theta|\tilde{\mathbf{x}}_{\mathbf{m}})$ where π_i^J is noninformative

It answers to most of our requirements (unicity, assessing correlations within θ...)

Construction principle of conjugate models, with π entirely explicit only in those cases

Nonconjugate cases

For a given pdf $f(x|\theta)$

1. select $\pi^J(\theta)$
2. assume there exists a "hidden" (virtual) sample $\tilde{\mathbf{x}}_\mathbf{m}$ of size m
3. give a unique form choosing $\pi(\theta) \equiv \pi^J(\theta|\tilde{\mathbf{x}}_\mathbf{m})$, ie.

$$\pi(\theta) = \pi(\theta|\boldsymbol{\Delta}_m)$$

   with $\boldsymbol{\Delta}_m$ a set of virtual statistics

4. estimate $\boldsymbol{\Delta}_m$ by $\widehat{\boldsymbol{\Delta}}_m = \arg\min_{\boldsymbol{\delta}_m} \mathcal{D}\left(\boldsymbol{\Lambda}_e, \boldsymbol{\Lambda}(\boldsymbol{\delta}_m)\right)$
   - $\boldsymbol{\Lambda}_e$ are wished prior features elicited (e.g., from expert knowledge)
   - $\boldsymbol{\Lambda}(\boldsymbol{\delta}_m)$ are features of the effective prior distribution
   - $\mathcal{D}$ is some kind of distance

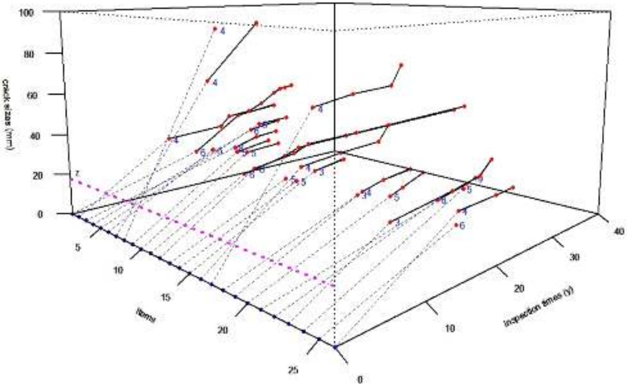
   under homogeneity constraints

Calibrating priors with information-theoretic distances: methodological works by Cooke (1991).Clarke (1996), Neal (2001), Liu & Clarke (2004), Lin et al. (2007), Morita et al. (2007)

crack size $Z_{k,t}$ monotonously increasing

independent increments $X_{k,i} = Z_{k,t_i} - Z_{k,t_{i-1}}$ assumed to follow gamma distributions

$$f_{\alpha(t-s),\beta}(x) = \frac{1}{\Gamma(\alpha_i(t-s))} \cdot \frac{x^{\alpha(t-s)-1} e^{-\frac{x}{\beta}}}{\beta^{\alpha(t-s)}} \mathbb{1}_{\{x \geq 0\}}$$



Building an informative prior on (α, β) [B. et al. 2014]

Noninformative prior (Jeffreys) $\pi^J(\alpha,\beta) \propto \frac{1}{\beta}\sqrt{\alpha\Psi_1(\alpha)-1}$

Posterior prior of a virtual sample of crack increases $\tilde{\mathbf{x}}_\mathbf{m} = (\tilde{x}_1, \ldots, \tilde{x}_m)$ observed at times $\tilde{\mathbf{t}}_\mathbf{m} = (\tilde{t}_1, \ldots, \tilde{t}_m)$

$$\begin{aligned} \beta|\alpha &\sim \mathcal{IG}\left(\alpha m \tilde{t}_{e,1}, m\tilde{x}_e\right) \\ \alpha &\sim \mathcal{G}\left(m/2, m\tilde{t}_{e,2}\right) \end{aligned}$$

the meaning of which being given by

$$\begin{aligned} \tilde{t}_{e,1} &= \frac{1}{m}\sum_{i=1}^m \tilde{t}_i && \textit{(average time of observation)} \\ \tilde{x}_e &= \frac{1}{m}\sum_{i=1}^m \tilde{x}_i && \textit{(average crack increase)} \\ \tilde{t}_{e,2} &= \frac{1}{m}\sum_{i=1}^m \tilde{t}_i \log \frac{\sum_{j=1}^m \tilde{x}_j/\tilde{x}_i}{\sum_{j=1}^m \tilde{t}_j/\tilde{t}_i} && \textit{(tuning hyperparameters)} \end{aligned}$$

Actually, approximation at the first order of the real posterior (Taylor-Edgeworth expansion)

Finding nice properties to calibrate from expert opinions

The mean crack increasing during the time interval Δ_i admits as its most likely a priori value

$$\hat{r}(\Delta_i) \quad = \quad \frac{\tilde{x}_e \Delta_i}{\tilde{t}_{e,1}}.$$

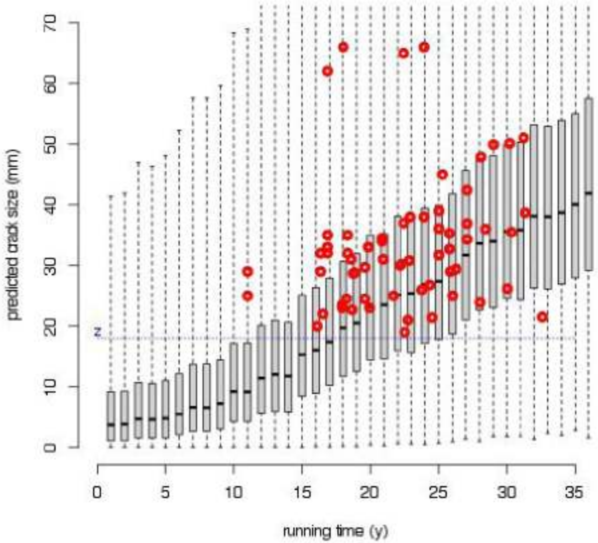
Questioning an expert. During the next 15 then 30 years (ie., the value of $m\tilde{t}_{e,1}$), what will be the chances $(1-\delta_1, 1-\delta_2)$ that any crack of size Z appearing on the device be upper than $(z_1, z_2) = (5, 10)$ mm? ie., for $i = \{1, 2\}$,

$$P\left(Z_{m\tilde{t}_{e,1}} < z_i\right) \quad = \quad \delta_i \;=\; \int_0^{z_i} \int_0^{\infty} \frac{x^{\alpha m\tilde{t}_{e,1}-1}(m\tilde{x}_e)^{\alpha m\tilde{t}_{e,1}}\Gamma\left(2\alpha m\tilde{t}_{e,1}\right)}{(m\tilde{x}_e + x)^{2\alpha m\tilde{t}_{e,1}}\,\Gamma^2\left(\alpha m\tilde{t}_{e,1}\right)}\pi(\alpha)\; d\alpha dx$$

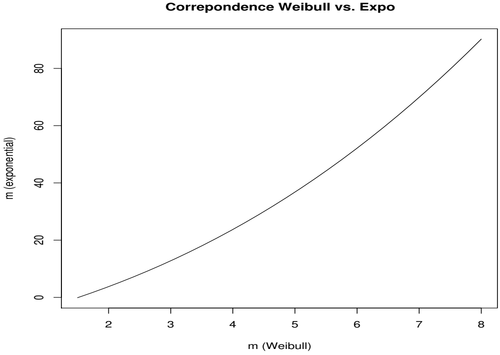
Calibrating the prior form by minimizing in $(m, \tilde{t}_{e,2})$ the L_2 relative distance

$$\sum_{i=1}^{2}\left\{1 - \delta_i^{-1} P\left(Z_{m\tilde{t}_{e,1}} < z_i\right)\right\}^2$$

Agreement between prior and data



.


Correpondence Weibull vs. Expo
m (exponential)
0
20
40
60
80
2
3
4
5
6
7
8
m (Weibull)

A final remark: Bayesian inversion of a computer code (Fu et al. 2013)

$$\begin{aligned} Y &= g(X) + \epsilon, \\ X &\sim \mathcal{N}(m, C) \\ \epsilon &\sim \mathcal{N}_p(0, R) \qquad \textit{(known observation noise)} \end{aligned}$$

where $Y = (Y_1, \ldots, Y_d)$, $X = (X_1, \ldots, X_q)$

Given the Gaussian structure of the missing data X, the Jeffrey noninformative prior can be elicited

$$\pi^J(\theta) = \frac{\mathbf{I}_{\Omega_m}(m)}{\text{Vol}(\Omega_m)} \cdot \frac{\Delta_C}{|C|^{\frac{q+2}{2}}} \mathbf{I}_{\Omega_C}(C)$$

with $\Omega_m \times \Omega_C$ the prior domain

This prior should be constrained by a basic condition of relevance (well-posed problem in Hadamard's sense)

- solving the inversion problem (getting the posterior of θ from information on Y) makes sense only if the uncertainty on Y is mainly explained by the uncertainty on X
- If g is linear, it is equivalent to ANOVA or the result of Sobol' analysis

Rewrite

$$\mathbf{Y} = \mathbf{\Gamma X} + \epsilon$$

(with $\mathbf{\Gamma}$ associated to the Jacobian of g in linearized cases)

A condition of relevance is for instance

$$\mathrm{Cov}[\mathbf{\Gamma X}] > \mathrm{Cov}[\epsilon]$$

Another one is based on entropy

$$H(\mathbf{\Gamma X}) > H(\epsilon)$$

Both lead to a similar kind of constraint

$$|C| > \mathcal{C}_{R,\mathbf{\Gamma}}$$

Then

$$\Omega_C \in \left\{C \in \mathcal{S}_q^+, \ |C| > \mathcal{C}_{R,\mathbf{\Gamma}}\right\}$$

and the Jeffrey prior can be integrable (proper)

Other constraints should be added if g is replaced by an emulator estimated from a finite design of numerical experiments (e.g., Gaussian process prior)

Uniform priors should not be used with purely artificial parameterizations

Formal elicitation of priors:

- seminal review by Kass and Wasserman (1996)
- theoretic arguments from decision and information theories
- virtual data posterior prior approach = emerging methodology
  - virtual sizes = levers of sensitivity analysis
  - clear weighs of subjectivity within the model
  - useful to defend Bayesian choices in an objective world

Constraints specific to uncertainty problems can help to elicit useful priors

Towards **Robust Bayesian analysis** in industrial applications (Rios Insua and Ruggeri 2000)

- critics of prior guesses: generic approaches
- links to develop with global sensitivity analysis

Noninformative priors

- Berger, J.O & Bernardo, J.M. (1992). On the development of reference priors (with discussion). *Bayesian Statistics 4*
- Kass, R.E., Wasserman, L. (1996). The selection of prior distributions by formal rules. J. Amer. Stat. Assoc.
- Sun, D. (1997). A note on noninformative priors for Weibull distributions, *J. Statist. Plan. Inf.*

Eliciting expert's beliefs

- Green, K.C. (2002). Forecasting decisions in conflict situations: a comparison of game theory, role-playing, and unaided judgement. *Intern. J. Forecasting*
- Lannoy, A., Procaccia, H. (2003). *L'utilisation du jugement d'expert en sûreté de fonctionnement*. Tec & Doc
- 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.
- Tversky, A., Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*
- Cooke, R.M. (1991). Experts in Uncertainty: Opinion and Subjective Probability in Science. Oxford University Press.
- Morita, S., Thall, P.F., Mueller, P. (2007). Determining the Effective Sample Size of a Parametric Prior, J. Biol. Stat.

Bayesian kriging

- Deng, H., Shao, W., Ma, Y., Wei, Z. (2011). Bayesian metamodeling for computer experiments using the Gaussian Kriging Models. *Qualit. Reliab. Engin. Intern.*
- Helbert, C., Dupuy, D., Carraro, L. (2009). Assessment of uncertainty in computer experiments, from Universal to Bayesian Kriging. *Appl. Stoch. Mod. Busin. Ind.*

Maximum entropy priors

- Bousquet, N. (2010). Eliciting vague but proper maximal entropy priors in Bayesian experiments. *Stat. P.*
- Friedlander, M.P., Gupta, M.R. (2006). On minimizing distortion and relative entropy, IEEE Trans. Inform. Theory
- Jaynes, E. T. (2003). Probability Theory: The Logic of Science. Cambridge University Press
- Soofi, E.S. (1992). Information theory and Bayesian statistics. In: *Bayesian analysis in statistics and econometrics in honor of Arnold Zellner*. Wiley
- Soofi, E.S. (2000). Principal information theoretic approaches, J. Amer. Stat. Assoc.
- Zellner, A., (1991). Bayesian methods and Entropy in Economics and Econometrics. In: *Maximum entropy and Bayesian Methods*. Kluwer
- Zellner, A. (1996). Models, prior information and Bayesian analysis, J. Econ.

Denote $\mathcal{GIG}(a, b, \gamma)$ the generalized inverse gamma distribution with density

$$\frac{b^a \gamma}{\Gamma(a)} \frac{1}{\eta^{a\gamma+1}} \exp\left(-\frac{b}{\eta^\gamma}\right) \mathbb{1}_{\{\eta \geq 0\}}.$$

Use Jeffrey's prior $\pi^J(\eta, \beta) \propto \eta^{-1} \mathbb{1}_{\{\eta \geq 0\}} \mathbb{1}_{\{\beta \geq 0\}}$

Posterior prior

$$\begin{aligned} \eta|\beta &\sim \mathcal{GIG}\left(m, b(\tilde{\mathbf{x}}_\mathbf{m}, \beta), \beta\right), \\ \pi(\beta) &\propto \frac{\beta^{m-1}}{b^m(\tilde{\mathbf{x}}_\mathbf{m}, \beta)} \exp\left(m \frac{\beta}{\beta(\tilde{\mathbf{x}}_\mathbf{m})}\right) \mathbb{1}_{\{\beta \geq 0\}} \end{aligned}$$

$$\text{with} \quad \mathbf{\Delta}_m = \left\{ b(\tilde{\mathbf{x}}_\mathbf{m}, \beta) = \sum_{i=1}^m \tilde{x}_i^\beta, \quad \beta(\tilde{\mathbf{x}}_\mathbf{m}) = m \left(\sum_{i=1}^m \log \tilde{x}_i \right)^{-1} \right\}$$

To satisfy the prior predictive assumption

$$P(X < x_\alpha^{(e)}) \quad = \quad \int_{-\infty}^{x_\alpha^{(e)}} f(x|\boldsymbol{\delta}_m)\, dx \;=\; \alpha$$

replace $b(\tilde{\mathbf{x}}_{\mathbf{m}}, \beta)$ by

$$b(\tilde{\mathbf{x}}_{\mathbf{m}}, \beta) \quad = \quad \left((1-\alpha)^{-1/m} - 1\right)^{-1} \left(x_\alpha^{(e)}\right)^\beta$$

Consequence: $\pi(\beta)$ is gamma with shape parameter m and mean parameter $\beta_e(m) = (\log x_\alpha^{(e)} - \beta^{-1}(\tilde{\mathbf{x}}_{\mathbf{m}}))^{-1}$

Important points

- The substituted joint prior is proper for any virtual size m extended on half-line $\boldsymbol{R}_+$
- A supplementary information is necessary to calibrate $\pi(\beta)$ given m

How choosing $\mathcal{D}$?

1. Ex: Cooke's method of discrete Kullback loss (1991).
   - assume prior information $(t_{i,e}, \alpha_{i,e})$ such that $P(X < x_{i,e}) = \alpha_{i,e}$

$$\mathcal{D}\left(\mathbf{\Lambda}_e, \mathbf{\Lambda}(\boldsymbol{\delta}_m)\right) = \sum_{i=0}^{q} (\alpha_{i+1,e} - \alpha_{i,e}) \log \frac{\alpha_{i+1,e} - \alpha_{i,e}}{\alpha_{i+1}(\boldsymbol{\delta}_m) - \alpha_i(\boldsymbol{\delta}_m)}$$

   with $\alpha_{0,e} = \alpha_0 = 0$, $\alpha_{q+1,e} = \alpha_{q+1} = 1$ and

$$\alpha_i(\boldsymbol{\delta}_m) = \int_{-\infty}^{x_{i,e}} f(t|\boldsymbol{\delta}_m)\; dx$$

2. One may weight the Kullback loss such that the most important constraints $\lambda_{i,e}$ are nearly fully respected

One cannot hope all expert specifications are simultaneous coherent with the Bayesian model

Issue 1: other uncertainties in expert opinions

Observable values x_e are provided by the Bayesian analyst (= the statistician)

Expert subjectivity is mainly expressed through an estimate α_e of α

Assume to have q sorted prior estimates $(\alpha_{e,i})_{1\leq i\leq q}$

Pursuing the **virtual sample** idea, one has a priori

$$(\alpha_1, \alpha_2 - \alpha_1, \ldots, \alpha_q - \alpha_{q-1}, 1 - \alpha_q) \quad \sim \quad \mathcal{D}_{ir}\left(\nu_1, \ldots, \nu_{q+1}\right)$$

with $\nu_j - 1 =$ number of virtual "past" observations of event $x_{e,j-1} \leq X \leq x_{e,j}$, ie.,

$$\sum_{j=1}^{q+1} \nu_j \quad = \quad m + q + 1$$

A simple choice is

$$\nu_j \quad = \quad (m + q + 1)(\alpha_{e,j} - \alpha_{e,j-1})$$

How calibrating m?

Let $\tilde{\mathbf{x}}_{\mathbf{m_1}}$ and $\tilde{\mathbf{x}}_{\mathbf{m_2}}$ be two virtual samples associated to two experts $\mathcal{E}_1$ and $\mathcal{E}_2$

$\mathcal{E}_1$ and $\mathcal{E}_2$ dependent

In our view, $\tilde{\mathbf{x}}_{\mathbf{m_1}}$ and $\tilde{\mathbf{x}}_{\mathbf{m_2}}$ are "generated" dependently (possibly share common data)

Following O'Hagan et al. (2006), obtaining a consensus virtual sample

Same methodology:

1. looking for a most trustworthy specification;
2. minimizing a loss function w.r.t marginal specifications

Supra-Bayes approach?

Replacing $\pi_1(\theta) = \pi^J(\theta|\tilde{\mathbf{x}}_\mathbf{m})$ by

$$\pi_2(\theta) \;=\; \pi_2(\theta|m,\theta_e) \;=\; \int \pi^J(\theta|\tilde{\mathbf{t}}_\mathbf{m}) f(\tilde{\mathbf{x}}_\mathbf{m}|\theta_e)\; d\tilde{\mathbf{x}}_\mathbf{m}$$

For a fixed, m calibrating θ_e using a loss function w.r.t. prior predictive specifications

Total variance theorem: $\text{Var}_{\pi_2}[\theta] \geq \text{E}_{\theta_e}\left[\text{Var}_{\pi_1}[\theta|\tilde{\mathbf{x}}_\mathbf{m}]\right]$

Can appear more cautious (but somewhat difficult to work with in non-conjugate cases)

Ex: exponential model $\mathcal{E}(\lambda)$ with $\lambda_e = x_e^{-1} \log 2$ and $x_e =$ prior median

$$\pi_2(\lambda) \;=\; \frac{\lambda_e^m \lambda^{m-1}}{(\lambda_e+\lambda)^{m+1}} \mathbb{1}_{\{\lambda \geq 0\}}$$