

Approche décisionnelle bayésienne pour estimer une courbe de fragilité

Title: A Bayesian decision approach for estimating a fragility curve

Guillaume Damblin^{1,2}, Merlin Keller¹ Alberto Pasanisi¹ Pierre Barbillon² et Eric Parent²

Résumé : L'étude présentée dans cet article se focalise sur les aspects décisionnels de l'estimation statistique d'une courbe de fragilité sismique. Cet objet utilisé en ingénierie du risque renseigne la probabilité de défaillance d'une structure conditionnellement à un certain niveau de sollicitation. Après avoir détaillé la modélisation statistique utilisée pour sa construction, précisé les motivations industrielles et dressé un panel des méthodes existantes dans la littérature, nous mettons l'accent sur le cadre rigoureux que permet l'analyse décisionnelle bayésienne pour estimer une courbe de fragilité en tenant compte des conséquences socio-économiques du problème. L'estimation statistique est réalisée à partir de données simulant le comportement d'une maquette de bâtiment à échelle réduite. Plusieurs estimateurs sont comparés au regard de la fonction de coût utilisée.

Abstract: This study deals with decisional analysis for estimating a seismic fragility curve, a tool used by risk engineers to provide the probability of a structure to suffer a given damage level conditionally to a given seismic intensity. After having described the statistical model and emphasized both the industrial motivations and the methods usually used to assess fragility curves, we focus on Bayesian decision analysis to estimate it accounting for social-economic consequences. Datasets are collected from numerical simulations and some estimators of the fragility curve are compared with respect to the chosen loss function.

Mots-clés : analyse de risque, courbes de fragilité, théorie bayésienne, analyse décisionnelle, fonction de coût, expériences numériques

Keywords: risk analysis, fragility curves, Bayesian theory, decision analysis, loss function, computer experiments

Classification AMS 2000 : 62c10, 62f15, 62p30

1. Introduction à la modélisation statistique d'une courbe de fragilité sismique

Une structure soumise à une sollicitation sismique s défaille au sens où une quantité d'intérêt, par exemple le *drift* (valeur maximale des déplacements différentiels entre deux étages de la structure), dépasse un seuil limite préfixé. Deux valeurs identiques pour la sollicitation n'occasionnent pas systématiquement les mêmes effets sur le drift du fait des nombreuses sources d'incertitudes inhérentes à l'événement sismique et à sa modélisation. C'est la raison pour laquelle, on s'intéresse à la fonction croissante $F : \mathbb{R}^+ \rightarrow [0, 1]$, supposée continue, qui relie la probabilité de défaillance à la sollicitation s . La fonction F est appelée *courbe de fragilité* et s'interprète comme la fonction de répartition d'une variable aléatoire C à densité par rapport à la mesure de Lebesgue. La variable C

¹ EDF R&D MRI 8 quai Watier, 78400 Chatou

E-mail : alberto.pasanisi@edf.fr

² AgroParisTech 16 rue Claude Bernard, Paris 05

E-mail : eric.parent@agroparistech.fr

que l'on qualifie de *capacité* de la structure modélise en fait la sollicitation limite avant défaillance et n'est pas observable (variable latente). Mathématiquement parlant, la défaillance se traduit par la réalisation de l'événement ($C \leq s$). Ainsi on peut vérifier que la courbe de fragilité s'écrit bien comme la fonction de répartition de C :

$$F(s) = \mathbb{P}(C \leq s). \quad (1)$$

Dans cet article, nous avons recours à un modèle paramétrique pour décrire la courbe de fragilité F en supposant que, conformément à la pratique technique courante, C suit une loi log-normale $\mathcal{LN}(\mu, \tau)$. L'estimation de F repose alors sur l'estimation du paramètre $\theta = (\mu, \tau)$ avec μ (resp. $\tau = \sigma^{-2}$) la moyenne (resp. la précision) de la log-capacité normale $Z := \ln C$. Les données disponibles à cet effet sont issues d'événements sismiques antérieurs (données empiriques) ou le plus souvent obtenues par simulations numériques via un modèle éléments finis reproduisant le comportement de la structure face à l'activité sismique (voir section 5). Elles sont résumées sous la forme de couples $(s_i, y_i)_{1 \leq i \leq n}$ avec s_i la sollicitation et $y_i = \mathbf{1}_{C_i \leq s_i}$ l'indicateur binaire de défaillance (voir figure 1). La densité d'une loi normale étant plus facile à manipuler, on considère $x_i := \ln s_i$ les log-sollicitations. Le paramètre θ peut alors être estimé à partir des données $(x_i, y_i)_{1 \leq i \leq n}$ en maximisant la fonction de vraisemblance avec un algorithme d'optimisation numérique ou via un algorithme EM (Dempster et al., 1977). Celui-ci consiste à estimer la valeur de θ qui maximise l'espérance de la vraisemblance des données complètes (z_i, y_i) relativement à la distribution des log-capacités latentes z_i conditionnellement aux données binaires y_i . Son expression analytique repose sur les deux premiers moments de la loi normale standard tronquée (voir annexe A).

La modélisation adoptée (1) s'apparente à un modèle de régression *probit* (McCullagh and Nelder, 1989) dans lequel la variable binaire de défaillance Y_i suit une loi de Bernoulli de paramètre $\mathbb{P}(\tilde{C} < \beta_0 + \beta_1 s_i)$ avec \tilde{C} la loi normale centrée réduite et β_0 et β_1 à estimer (McCulloch, 1994). En effet, quitte à considérer la variable aléatoire $C := (\tilde{C} - \beta_0)\beta_1^{-1}$, on se ramène à la modélisation de la capacité par une loi normale de paramètres $\mu = -\beta_0\beta_1^{-1}$ et $\tau = \beta_1^2$. Remarquons que le modèle *probit* a déjà été utilisé dans la littérature pour l'estimation de courbes de fragilité (Fabbrocino et al., 2005; Lallemand and Kiremidjian, 2013; Reese et al., 2011).

Nous utilisons tout au long de l'article la notation $[\cdot|\cdot]$ introduite en 1990 par Gelfand and Smith (1990) pour désigner la densité de probabilité d'une variable aléatoire quand le modèle bayésien a été spécifié sans équivoque.

Lorsque les courbes de fragilité sont obtenues par analyse statistique, le problème des incertitudes d'estimation du paramètre θ , ou en d'autres termes la quantification de l'incertitude qui affecte la courbe estimée, est traitée de différentes manières par les praticiens. Par exemple, la méthode américaine EPRI (voir section 3) suggère une modélisation probabiliste qui sépare l'incertitude de la variable $x|\mu$ de l'incertitude (parfois qualifiée d'*épistémique*) qui affecte la log-capacité moyenne μ . Shinozuka et al. (2000) invoquant cette pratique proposent de représenter l'incertitude sur l'estimation par maximum de vraisemblance en se focalisant uniquement sur

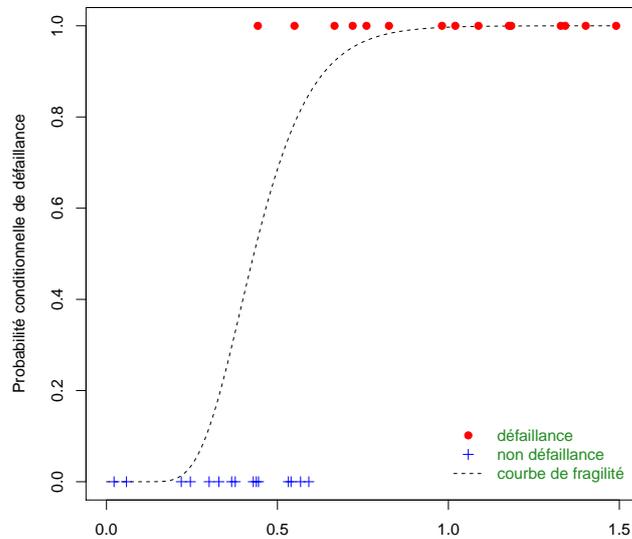


FIGURE 1. Représentation des données binaires de défaillance (en bleu : $y_i = 0$, en rouge : $y_i = 1$) et de la courbe de fragilité à estimer. En abscisse : la sollicitation s (PGA, peak ground acceleration). En ordonnée : la probabilité de défaillance.

l'incertitude qui affecte l'estimateur $\hat{\mu}$. Après avoir obtenu par bootstrap un échantillon de $(\hat{\mu}, \hat{\tau})$, une *enveloppe* à 90% est proposée pour la courbe de fragilité encadrée par les deux lois :

$$F(\cdot | \hat{\mu}_{5\%}, \hat{\tau}_{mv}) \text{ et } F(\cdot | \hat{\mu}_{95\%}, \hat{\tau}_{mv}).$$

où $\hat{\mu}_{5\%}$ et $\hat{\mu}_{95\%}$ sont les quantiles de probabilité à 5% et 95% de $\hat{\mu}$ obtenus par bootstrap et $\hat{\tau}_{mv}$ est l'estimateur de maximum de vraisemblance de τ . Pour la mise en œuvre pratique dans les études industrielles, les auteurs proposent de choisir la courbe suivante (qu'ils qualifient de *moyenne*) :

$$\int F(\cdot | \hat{\mu}, \hat{\tau}_{mv}) [\hat{\mu}] d\hat{\mu}. \quad (2)$$

où $[\hat{\mu}]$ est une loi log-normale calée sur l'échantillon de $\hat{\mu}$ obtenu par bootstrap. La cohérence de l'approche est critiquable : l'incertitude sur la courbe de fragilité F est codée à travers une loi de probabilité uniquement sur le paramètre μ et ceci par commodité mathématique afin d'obtenir des quantiles de probabilité sur la courbe médiane (qui correspond au quantile à 50%).

Nous privilégions dans cet article une approche bayésienne dans laquelle il est possible d'incorporer des informations annexes. La possibilité d'encoder une expertise préalable à l'analyse à travers la loi jointe *a priori* $[\theta]$ plaide en faveur de cette approche, en particulier dans ce type de problèmes, où les données sont par définition rares (données observées) ou coûteuses (simulation numérique). Des méthodes existent pour construire des courbes de fragilité uniquement par jugements d'experts (voir section 3). Les intégrer à notre démarche en tant que loi *a priori* pourrait pallier le nombre limité de données disponibles. [Straub and Der Kiureghian \(2008\)](#)

mènent l'estimation bayésienne afin de s'affranchir des différentes hypothèses sur la modélisation probabiliste des paramètres μ et τ et proposent comme estimateur la courbe *prédictive* :

$$\int \int F(\cdot | \mu, \tau) [\mu, \tau] d\mu d\tau. \quad (3)$$

Nous nous proposons dans un cadre bayésien d'aller au delà de la courbe *prédictive* (3) et préconisons l'utilisation de courbes de fragilité choisies conformément à une règle de décision explicite sur l'espace de toutes les courbes possible. Cette approche s'appuie sur la théorie de la décision (Berger, 1985) afin de pénaliser certains types d'erreur dans l'estimation à travers une fonction de coût. Du point de vue de l'ingénieur, l'approche bayésienne permet non seulement de quantifier proprement l'incertitude sur les courbes de fragilité, mais aussi de justifier le choix d'une courbe en faisant appel à des fonctions de coût qui pénalisent différemment les écarts entre la *vraie* courbe (inconnue à jamais) et la courbe estimée (par exemple pour différentes plages de la sollicitation). Les avantages de cette approche *entièrement* bayésienne dans les études d'incertitudes et de fiabilité basées sur la simulation numérique, par rapport aux méthodes d'estimation basées sur la loi prédictive, sont discutés entre autres dans (Keller et al., 2011; Pasanisi et al., 2012). Nous pourrions également tenir compte de la distribution des sollicitations qui n'est pas nécessairement celle observée lors de l'acquisition des données.

Dans la section 2, nous présentons le contexte industriel dans lequel s'inscrit l'étude des courbes de fragilité puis faisons un tour d'horizon des méthodes de construction proposées en ingénierie (section 3). Nous détaillons en section 4 l'approche décisionnelle bayésienne et proposons des fonctions de coût pour la construction d'estimateurs d'une courbe de fragilité. Les résultats obtenus pour notre cas d'étude sont présentés en section 5 et une comparaison avec le maximum de vraisemblance est effectuée. Une discussion quant aux perspectives à envisager est menée en section 6.

2. Motivations industrielles

Dans le contexte industriel d'EDF, cette étude puise ses motivations dans la problématique de l'évaluation probabiliste de la sûreté des installations nucléaires. Une *étude probabiliste de sûreté* (EPS) a pour but l'analyse systématique des scénarios accidentels à l'origine de certains événements particulièrement redoutés, par exemple, dans le cas des réacteurs nucléaires, la fusion du cœur ou le rejet de matière contaminée en dehors de l'enceinte de confinement (Kalalo and Brenot, 2003). Plus précisément, en partant d'un certain nombre d'événements dits *initiateurs* caractérisés par une certaine fréquence annuelle d'occurrence, sont construits des arbres d'événements menant aux situations critiques évoquées ci-dessus. A chaque événement (sollicitation d'un mécanisme ou action humaine visant à ramener le réacteur vers des conditions sûres) est affectée une probabilité de succès ou d'échec. Un calcul probabiliste permet au final d'attribuer à chaque initiateur une probabilité d'occurrence de la situation redoutée.

Les événements initiateurs (Lanore, 2003) peuvent être des défaillances internes à l'installation (humaines ou matérielles) ou des *agressions* internes (e.g. incendie d'un composant) ou encore externes (séisme, crue, surcôte ...). Dans ce contexte, il est bien sûr important d'évaluer correctement les probabilités de défaillance des composants ou des mécanismes de sauvegarde relativement aux

événements sismiques. Dans la pratique technique (IAEA-TECDOC-724, 1993; Ellingwood and Kinali, 2009; Huang et al., 2011a), cette évaluation est faite à partir de deux ingrédients de base : la probabilité conditionnelle de défaillance de la structure soumise à une sollicitation d'intensité donnée s (*courbe de fragilité*) et la mesure de probabilité P de ladite intensité (*aléa sismique*). La probabilité de défaillance de la structure face au risque sismique est alors définie par :

$$\int F(s)dP(s).$$

Notre étude se focalise exclusivement sur la fragilité sismique.

Dans l'évaluation de la fragilité, le séisme (qui est de par nature une sollicitation dynamique) est habituellement résumé par une ou plusieurs mesures scalaires (sollicitations), parmi lesquelles l'accélération spectrale (Huang et al., 2011b), la vitesse maximale au sol (PGV, *peak ground velocity*) et l'accélération maximale au sol (PGA, *peak ground acceleration*). Dans cette étude, nous caractérisons la sollicitation sismique s par la PGA horizontale, choix usuel dans la pratique du génie civil (Solomos et al., 2008; Zentner, 2010). Un exemple de signal sismique (accélérogramme) enregistré lors d'un événement de particulière intensité est représenté en figure 2. L'accélération sismique étant mesurée selon deux directions orthogonales (typiquement Nord-Sud et Est-Ouest), la PGA de référence pour un séisme donné est généralement prise comme la moyenne des deux PGA.

Pour une valeur donnée de la PGA, nous faisons l'hypothèse que l'occurrence de l'événement *défaillance* de la structure est aléatoire : d'une part à une même valeur de PGA peuvent correspondre plusieurs signaux sismiques (plus ou moins nocifs pour la structure) et d'autre part le comportement même de la structure peut être affecté par des paramètres de nuisance difficilement quantifiables et soumis à de fortes incertitudes (Straub and Der Kiureghian, 2008). Par conséquent, la courbe de fragilité F peut s'interpréter comme la fonction de répartition d'une distribution de probabilité appelée *capacité* de la structure (voir section 1). Au delà de son utilisation dans le contexte spécifique qui motive cette étude, une courbe de fragilité est un résultat intéressant *en soi* pour l'ingénieur, plus complet que l'évaluation faite habituellement de la probabilité de défaillance d'une structure en correspondance d'une sollicitation *de référence* (Schultz et al., 2010).

Par ailleurs, la pratique des courbes de fragilité ne se limite pas à l'analyse du risque sismique. En particulier, le dysfonctionnement de systèmes mécaniques et l'*analyse de survie* d'organismes vivants (Collett, 2003) présentent des traits statistiques communs et partagent des concepts de modélisation identiques. De nombreux exemples d'utilisation de la même notion de courbe de défaillance (ou fiabilité) conditionnelle à un niveau de sollicitation sont disponibles dans d'autres domaines de l'ingénierie : e.g. enveloppes de bâtiments soumises à l'impact de débris emportés par le vent (Herbin and Barbato, 2012), baies vitrées sollicitées par des explosions (Stewart and Netherton, 2008), ponts sous l'effet de vagues de tsunami (Shoji and Moriyama, 2007), ouvrages de protection hydraulique exposés à des niveaux d'eau exceptionnels (comme dans les études menées aux États-Unis dans le cadre des analyses qui ont suivi l'ouragan *Katrina* (Ayyub et al., 2009a,b)). Se reporter également à Schultz et al. (2010) pour une liste commentée d'exemples d'applications. La méthodologie et les considérations que nous présentons ici sont donc transposables à d'autres domaines.

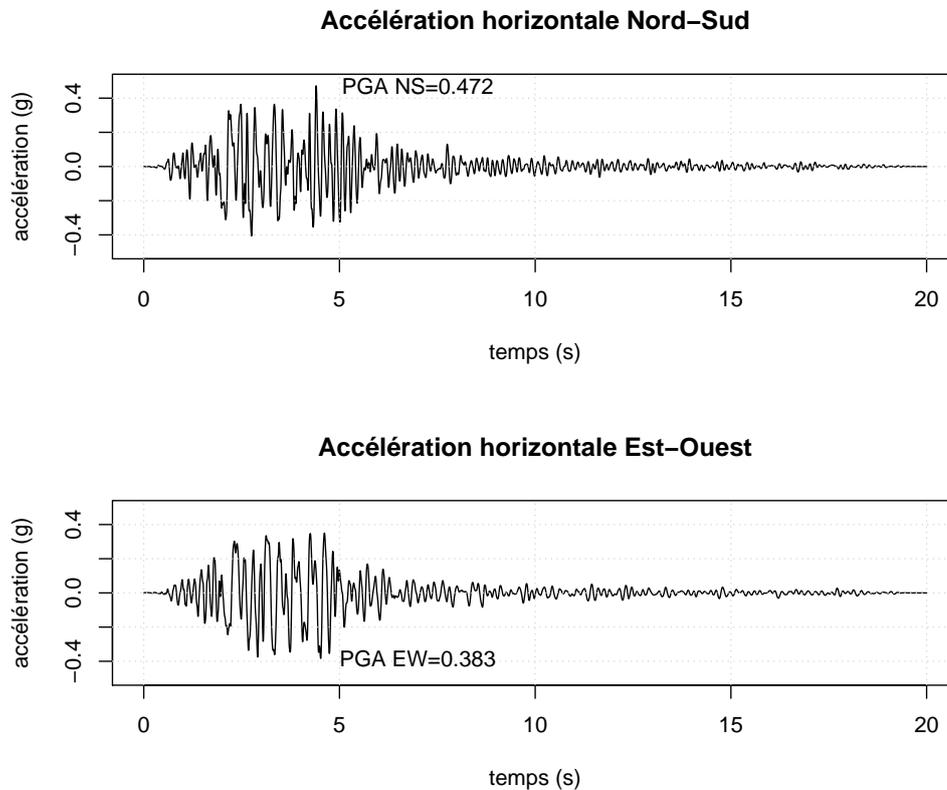


FIGURE 2. Exemple de signal sismique. Accélérogramme enregistré par la station de Nocera Umbra (Italie) lors du séisme du 26/09/1997, 9h40 (PGA moyenne 0.427). Les accélérations horizontales sont mesurées selon les deux directions Nord-Sud et Est-Ouest et sont exprimées en g (9.80 m/s^2). Source des données : SISMA (Site of Italian Strong-Motion Accelerograms), <http://sisma.dsg.uniroma1.it/>.

3. Construction des courbes de fragilité dans la pratique technique

Dans la pratique technique (Zentner, 2010), les courbes de fragilité sont pour la plupart construites par jugement d'experts. La méthode de référence est mise en œuvre au sein de l'institut de recherche américain EPRI (1994, *Electric Research Power Institute*). Elle repose sur l'hypothèse de normalité de la log-capacité $Z = \ln C$ de moyenne μ et de précision τ . Le calcul de $\theta = (\mu, \tau)$ s'effectue au travers de l'expertise du bâtiment et des propriétés d'un séisme de référence (séisme de dimensionnement). La méthode distingue deux niveaux d'incertitudes :

1. le comportement de la structure sous la contrainte sismique dépend de N facteurs de marges M_i , rattachés au signal sismique (la résistance inélastique, le facteur d'incohérence sismique,...) mais aussi à la structure (le facteur de résistance mécanique,...). L'incertitude sur chacune de ces caractéristiques est codée par une loi normale de moyenne μ_i et de précision τ_i . Z suit une loi normale de moyenne $\mu = \sum_i \mu_i$ et de précision $\tau = (\sum_i \frac{1}{\tau_i})^{-1}$ comme somme des lois normales des M_i . Une courbe de fragilité est exhibée, modélisant

l'aléa intrinsèque du phénomène physique (deux séismes de même intensité n'impliquent pas forcément défaillance).

- la moyenne μ_i de chaque M_i est considérée incertaine. Elle est aussi modélisée par une loi normale. On parle alors d'incertitude de méconnaissance (souvent qualifiée d'incertitude *épistémique*). La moyenne μ de Z suit désormais une loi normale $[\mu]$ comme somme des lois normales codant l'incertitude sur les μ_i .

La méthode EPRI par ce double niveau d'incertitude permet le calcul d'une courbe de fragilité caractérisée par un niveau de confiance portée sur μ . La courbe moyenne est obtenue en calculant :

$$\int F(\cdot|\mu, \tau) [\mu] d\mu.$$

Les autres méthodes sont basées sur la simulation numérique. Parmi elles, les méthodes utilisant l'analyse statistique de couples de données (x, y) , i.e. indicateur binaire de défaillance vs. niveau de log-sollicitation, sous l'hypothèse de loi normale pour la log-capacité de la structure, sont les plus répandues. Les données peuvent provenir du retour d'expérience, i.e. l'observation de dégâts effectivement constatés sur des structures lors de séismes ou d'autres agressions externes. C'est la raison pour laquelle, on emploie aussi le terme de méthodes empiriques. Dans les régions à forte activité sismique, des modèles permettant d'actualiser la courbe de fragilité de référence en tenant compte de données récoltées consécutivement à un nouveau séisme, ont été élaborés par [Hoshi et al. \(2011\)](#). Par ailleurs, lorsque l'activité sismique est faible voire inexistante, des données issues de la simulation numérique peuvent être utilisées pour mener à bien ces méthodes (section 5). D'autres techniques dites analytiques s'appuient sur des modèles stochastiques simulant l'action sismique et le comportement de la structure. Pour modéliser le signal sismique, [Marano et al. \(2011\)](#) utilisent un filtre linéaire basé sur l'approche stochastique proposée par [Clough and Penzien \(1977\)](#). Les déformations de la structure sont modélisées au travers d'une équation différentielle stochastique couramment utilisée en vibration, utilisée par [Bouc \(1967\)](#) et [Wen \(1976\)](#). Ces modèles font appel à des bases de données ([HAZUS, 2003](#)) rassemblant les paramètres utilisés dans ces modèles ainsi que les courbes de capacité qui dépendent notamment du type de structure à modéliser. Une autre famille de méthodes, que nous n'aborderons pas ici, est basée sur l'évaluation directe de la probabilité de défaillance de la structure en correspondance avec un certain nombre (discret) de valeurs de la sollicitation s , ce qui permet une construction *point par point* de la courbe. Ces méthodes appelées dans [Zentner \(2010\)](#) se heurtent à des problèmes de temps de calcul, car les modèles physiques des structures (typiquement par différences ou éléments finis) peuvent être coûteux et un unique calcul de la probabilité de défaillance pour une certaine valeur de la sollicitation peut déjà constituer un défi numérique. Les calculs sont alors généralement réalisés après l'approximation du modèle physique par une surface de réponse ([Lagaros et al., 2009](#); [Zentner et al., 2011](#)). Les surfaces de réponses ainsi construites peuvent être également utilisées pour réaliser d'autres calculs coûteux, notamment l'analyse de sensibilité de la courbe de fragilité aux variables d'entrées du code, comme dans l'étude de [Zentner et al. \(2011\)](#).

4. Théorie de la décision en avenir incertain

L'analyse décisionnelle bayésienne est une théorie applicable à la construction d'un estimateur bayésien de la courbe de fragilité F à partir de la loi *a posteriori* $[F|(s, y)]$ et en tenant compte des

conséquences socio-économiques résultant des choix (décisions) possibles. Nous rappelons dans un premier temps les principes du calcul bayésien et les simplifications faites pour l'étude qui nous occupe avant ensuite de développer les aspects décisionnels, en particulier le choix de la fonction de coût.

4.1. Présentation du canevas bayésien

L'estimation bayésienne d'une courbe de fragilité nécessite :

- la construction d'une loi *a priori* $[F]$ sur F .
- l'échantillonnage de la loi *a posteriori* à partir de la formule de Bayes ($[F|s, y] \propto [y|F, s][F]$), que nous effectuons ici par des méthodes de Monte-Carlo qui permettent de s'affranchir du calcul de la constante de normalisation (échantillonnage de Gibbs [Casella and George, 1992](#); [Geman and Geman, 1984](#), algorithme de Metropolis-Hastings [Chib and Greenberg, 1995](#); [Metropolis et al., 1953](#)).
- le choix d'une fonction de coût pour la construction de l'estimateur \hat{F} .

Des modèles non paramétriques à base de processus de Dirichlet ([Ferguson, 1973](#); [Ghosh and Ramamoorthi, 2002](#)) pourraient être utilisés (voir section 6). Dans ce travail, nous traitons de l'estimation paramétrique de la courbe de fragilité $F = F_\theta$ dans la classe des fonctions de répartition des lois log-normales. Il suffit dans ce cas d'élucider la loi *a priori* $[\theta] = [\mu, \tau]$. Soient x les log-sollicitations et y les données binaires de défaillance. La formule de Bayes exprime la loi de probabilité *a posteriori* $[\theta|x, y]$ du paramètre θ comme proportionnelle à la vraisemblance du modèle $[y|x, \theta]$ multipliée par la loi *a priori* $[\theta]$:

$$[\mu, \tau|x, y] \propto [y|x, \mu, \tau] \times [\mu, \tau]. \quad (4)$$

L'expertise EPRI (voir section 3) est un outil précieux pour la construction d'une loi informative *a priori* sur θ (étape *a.*). Par exemple, la loi codant l'incertitude épistémique dans l'approche EPRI peut être choisie en tant que loi *a priori* sur μ . On peut ensuite coder l'incertitude *a priori* sur τ par une loi gamma ayant pour moyenne la valeur de référence donnée par EPRI et avec une variance à spécifier. Ce faisant, on peut définir la densité *a priori* $[\theta] = [\mu][\tau]$. Plus $[\theta]$ est informative, plus elle impacte l'estimation de la loi *a posteriori*, notamment quand peu de couples (s_i, y_i) sont disponibles. En l'absence d'informations *a priori*, des distributions dites non informatives peuvent être utilisées. Nous en donnons deux exemples ci-dessous :

- *Lois conjuguées (pour le modèle normal)* : Il est courant pour faciliter les calculs de la loi *a posteriori* de se restreindre aux lois *a priori* conjuguées. Ici, il est pertinent d'utiliser une loi *a priori* gamma-normale pour la moyenne et la précision de la loi normale sous la forme :

$$[\mu, \tau] = [\mu|\tau][\tau] \text{ avec } \begin{cases} [\mu|\tau] \sim \mathcal{N}(m, \lambda \tau) \\ [\tau] \sim \mathcal{G}(a, b) \end{cases} \quad (5)$$

où (m, a, b, λ) sont des constantes prédéfinies, appelées *hyperparamètres*. En choisissant cette loi pour $[\theta]$, la loi *a posteriori* du modèle latent (données z_i) est aussi gamma-normale. Lorsque ces constantes tendent vers zéro, cette loi *a priori* devient non informative.

- *Lois impropres* : il est souvent d'usage (Box and Tiao, 1973; Gelman et al., 2004) de définir des lois *a priori* non informatives, invariantes par reparamétrisation via une certaine classe de fonctions bijectives. Soit :

$$[\mu, \tau] = [\mu][\tau] \text{ avec } \begin{cases} [\mu] \propto 1, \\ [\tau] \propto \tau^{-1}. \end{cases} \quad (6)$$

Le modèle des données latentes (x_i, z_i) étant gaussien, la loi *a priori* (6) constitue la loi *a priori* de Jeffreys (Bernardo and Smith, 1994) avec $[\mu]$ et $[\tau]$ impropres. Dans ce cas, la loi *a posteriori* $[\mu, \tau|x, y]$ n'est pas une densité de probabilité. En effet, soit :

$$[y|x, \mu, \tau] = \prod_{i=1}^n F_{(\mu, \tau)}(x_i)^{y_i} (1 - F_{(\mu, \tau)}(x_i))^{1-y_i} \quad (7)$$

alors (voir annexe B),

$$\int_{\mathbb{R}} \int_{\mathbb{R}^+} [y|x, \mu, \tau] [\mu, \tau] d\tau d\mu = \int_{\mathbb{R}} \int_{\mathbb{R}^+} [y|x, \mu, \tau] \frac{1}{\tau} d\tau d\mu = +\infty. \quad (8)$$

La loi *a priori* (6) est un cas limite de (5) ($\lambda = 0 = a = b$).

Une autre possibilité non testée dans nos travaux consisterait à calculer pour le modèle correspondant aux données observées (x_i, y_i) et dont la vraisemblance est donnée par (7), la loi *a priori* de Jeffreys $[\theta] \propto \sqrt{|\mathbf{I}(\theta)|}$ où $|\mathbf{I}(\theta)|$ désigne le déterminant de la matrice d'information de Fisher (voir annexe A).

L'étape c) fait appel à la théorie de la décision développée ci-après.

4.2. Théorie de la décision

4.2.1. Les estimateurs bayésiens

La théorie de la décision statistique a été initiée par les travaux de Wald (1950). Cette vision décisionnelle offre une présentation unifiée d'un grand nombre de préoccupations du statisticien : tests d'hypothèses, choix de modèles, estimation ponctuelle ou par intervalles. Parmigiani and Inoue (2008) et Jordaan (2005) en proposent un exposé moderne selon une perspective très générale (Ulmo and Bernier, 1973; Bernier, 2003). Appliqué à notre modélisation paramétrique de la courbe de fragilité, on y retrouve les ingrédients suivants :

- Définir l'ensemble \mathcal{D} des décisions possibles (l'ensemble dans lequel on choisit l'estimateur) et l'ensemble Ψ des états de la nature (les inconnus du problème, ici les courbes de fragilité possibles). Compte-tenu de notre choix de modèle pour F , nous considérons pour les sous-ensembles de décisions \mathcal{D} et des états de la nature Ψ , l'ensemble des fonctions de répartition F_θ des lois log-normales $\mathcal{LN}(\mu, \tau)$. Estimer le paramètre à deux composantes $\theta = (\mu, \tau) \in \Theta \subset \mathbb{R} \times \mathbb{R}^+$ constitue donc l'acte de prise de décision.

- Munir l'espace des états de la nature $\Psi \leftrightarrow \Theta$ (\leftrightarrow désigne l'existence d'une bijection) où évolue la "vraie" courbe de fragilité, elle inconnue, d'une loi de probabilité notée abusivement sous forme d'une densité $[\theta]$. La formule de Bayes permet d'écrire la crédibilité accordée à la valeur θ au vu des données (s, y) , notée $[\theta|s, y]$ et dite loi *a posteriori* de θ .
- Utiliser un critère d'évaluation dépendant de la décision prise et de l'état de la nature. Ce critère est appelé *fonction de coût* notée L , définie de $\mathcal{D} \times \Psi \leftrightarrow \Theta \times \Theta$ dans \mathbb{R}^+ . $L(F_{\hat{\theta}}, F_{\theta}) = L(\hat{\theta}, \theta)$ donne la pénalisation lorsque la courbe estimée $F_{\hat{\theta}}$ est adoptée en lieu et place de la "vraie" courbe inconnue F_{θ} .

La courbe de fragilité retenue $F_{\hat{\theta}}$ est celle qui s'approche au plus des données au sens du coût moyen *a posteriori* :

$$F_{\hat{\theta}} = \operatorname{argmin}_{d \in \mathcal{D}} \int_{F_{\theta} \in \Psi} L(d, F_{\theta}) [F_{\theta}|s, y] dF_{\theta}$$

ou de façon équivalente,

$$\hat{\theta} = \operatorname{argmin}_{\tilde{\theta} \in \Theta} \int_{\theta \in \Theta} L(\tilde{\theta}, \theta) [\theta|s, y] d\theta. \tag{9}$$

Si l'on estime θ à partir des log-sollicitations x et en faisant apparaître la loi *a priori* $[\theta]$, l'équation (9) se spécialise en :

$$\hat{\theta} = \operatorname{argmin}_{\tilde{\theta} \in \Theta} \int_{\theta \in \Theta} L(\tilde{\theta}, \theta) [y|\theta, x] [\theta] d\theta. \tag{10}$$

Dans le paragraphe suivant, nous introduisons des fonctions de coûts L de la forme :

$$L(\tilde{\theta}, \theta) = \int_0^{+\infty} D(F_{\tilde{\theta}}(s), F_{\theta}(s)) dP(s). \tag{11}$$

avec D une pré-métrique entre les probabilités de défaillance $F_{\tilde{\theta}}(s)$ et $F_{\theta}(s)$ ($D(F_{\theta}(s), F_{\theta}(s)) = 0$ et $D(F_{\tilde{\theta}}(s), F_{\theta}(s)) > 0$).

4.2.2. Des fonctions de coûts L_1 et L_2

Pour estimer les paramètres de la courbe de fragilité minimisant (10), les fonctions de coût absolu (12) et quadratique (13) sont les options les plus classiques (Robert, 1992). On les construit comme une mesure intégrée du risque sismique en considérant $P(s)$ une mesure de probabilité sur la demi-droite réelle positive des sollicitations sismiques. Sous condition d'intégrabilité, on définit :

$$L_1(\tilde{\theta}, \theta) = \int_0^{+\infty} |F_{\tilde{\theta}}(s) - F_{\theta}(s)| dP(s), \quad (12)$$

$$L_2(\tilde{\theta}, \theta) = \int_0^{+\infty} (F_{\tilde{\theta}}(s) - F_{\theta}(s))^2 dP(s), \quad (13)$$

$$L_{K_1, K_2}(\tilde{\theta}, \theta) = \int_0^{+\infty} K_1 |F_{\tilde{\theta}}(s) - F_{\theta}(s)| 1_{F_{\tilde{\theta}}(s) > F_{\theta}(s)} dP(s) \\ + \int_0^{+\infty} K_2 |F_{\tilde{\theta}}(s) - F_{\theta}(s)| 1_{F_{\tilde{\theta}}(s) < F_{\theta}(s)} dP(s). \quad (14)$$

La fonction de coût absolu généralisée (14), linéaire par morceaux, permet de pénaliser davantage la surestimation ou la sous-estimation de F selon le choix des constantes K_1 et K_2 telles que par convention $K_1 + K_2 = 1$. Si l'on choisit $K_1 > K_2$, on pénalise d'autant plus la sous-estimation que K_1 est proche de 1.

5. Résultats numériques pour le cas d'étude SMART

Le benchmark international SMART (*Seismic design and best-estimate Methods Assessment for Reinforced concrete buildings subjected to Torsion and non-linear effects*) a été lancé en 2008 par le CEA et EDF en vue de l'analyse dynamique des structures en béton armé. En particulier, il est question de l'évaluation numérique de la courbe de fragilité d'une structure modélisée par une maquette échelle réduite. Nous disposons de données provenant d'un simulateur numérique mimant le comportement de la maquette soumise à un profil d'accélération sismiques sur une durée déterminée. Les profils utilisés sont tirés aléatoirement dans une bibliothèque de profils observés en conditions réelles et sont résumés dans les données disponibles par la seule valeur s de la PGA horizontale moyenne au cours du séisme. Le simulateur numérique est déterministe. Les sources d'aléa proviennent du fait qu'à une même valeur de s correspond différents profils mais aussi du choix de modéliser aléatoirement les incertitudes sur les paramètres du simulateur (par exemple, les valeurs exactes des coefficients d'élasticité de toutes les poutres d'une construction sont inconnues en pratique). Nous considérons comme réponse du modèle une variable binaire indiquant si le *drift* a dépassé le seuil préfixé. Ainsi nous obtenons des couples d'observations $(x, y) := (x_i, y_i)_{1 \leq i \leq n}$ pour $n = 50$, avec $x_i = \ln s_i \in \mathbb{R}$ la log-sollicitation et $y \in \{0, 1\}$. La valeur $y = 1$ indique qu'il y a eu défaillance, c'est-à-dire que l'événement $(C \leq s) = (Z \leq x)$ a été observé.

5.1. Calcul de la loi a posteriori

L'algorithme de Gibbs (détaillé en annexe C) nous permet de simuler des réalisations de la loi *a posteriori* de θ de densité $[\theta|(x, y)]$ grâce à une technique d'augmentation de données (Tanner,

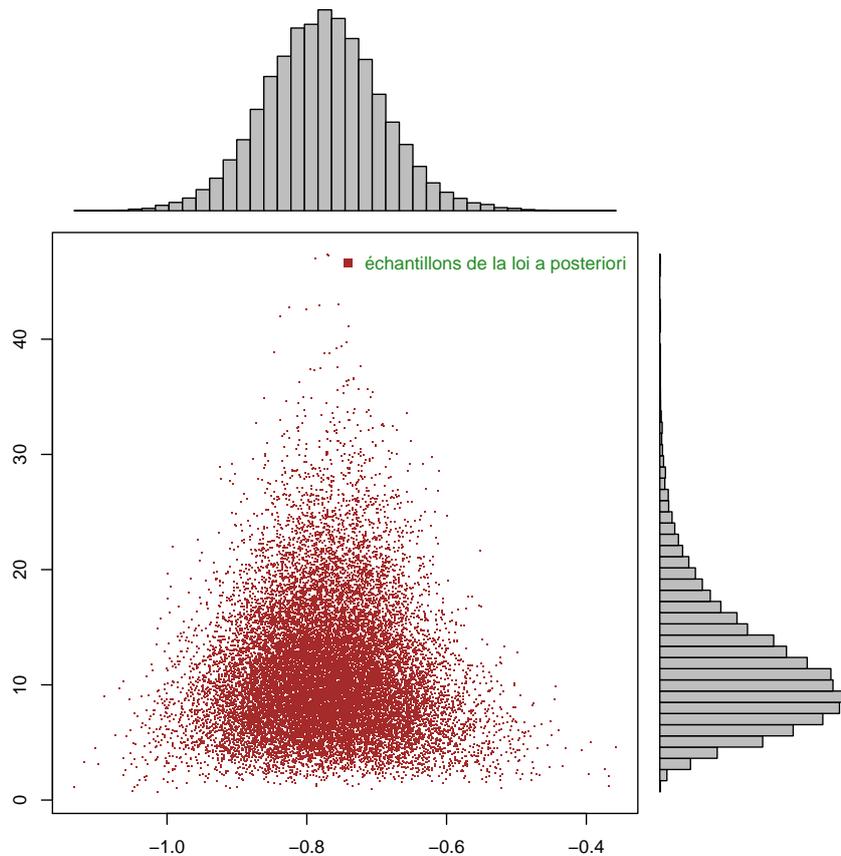


FIGURE 3. *Distribution a posteriori $[\mu, \tau|y,x]$ à partir d'un prior peu informatif Gamma-Normal. 20000 réalisations sont générées (après burn-in) avec l'algorithme de Gibbs (détaillé en annexe). En haut : la loi marginale a posteriori de μ . A droite : la loi marginale a posteriori de τ .*

1992) proposée par Albert et Chib pour le modèle probit (Albert and Chib, 1993). Pour ce faire, la loi *a priori* gamma-normale (5) a été utilisée. Les échantillons obtenus (en fixant ($m = 0$ et $a = b = \lambda = 0.01$)) sont représentés en figure 3. Des enveloppes de crédibilité sur la courbe médiane *a posteriori* sont construites en calculant pour chaque valeur de la PGA un intervalle de crédibilité sur la probabilité de défaillance (voir figure 4). Nous avons aussi estimé $[\theta|(x, y)]$ à partir de la loi *a priori* (6) en appliquant l'algorithme de Gibbs via les lois conditionnelles (33) et (34) explicitées dans Box and Tiao (1973). Les résultats obtenus sont similaires.

5.2. Estimation de la courbe de fragilité

La théorie de la décision nous mène à évaluer les conséquences socio-économiques imputées au choix de chacune des courbes du faisceau *a posteriori*. Dans la pratique cela n'est pas envisageable car débouchant sur des problèmes trop complexes (voir section 6). En général, en fiabilité des structures, on veut se prémunir d'une éventuelle sous-estimation des probabilités de défaillance

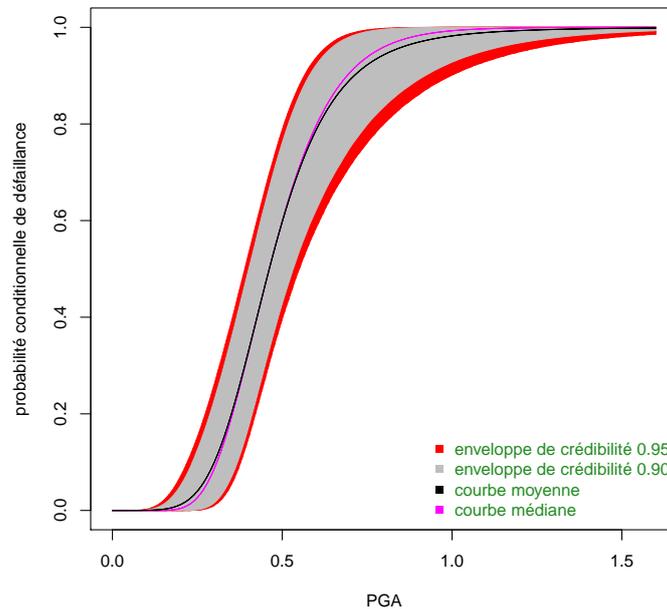


FIGURE 4. *Domaine de crédibilité pour l'inférence bayésienne d'une courbe de fragilité : les enveloppes de confiance à 90% et 95% sont calculées à partir des échantillons a posteriori (voir figure 3).*

$p_a = \mathbb{P}[Z \leq a]$ pour une certaine valeur a de la sollicitation sismique (ici la PGA). Cela justifie la définition d'un coût absolu généralisé (14). De plus, les quantiles de crédibilité construits "PGA par PGA" (voir figure 4), n'appartiennent pas à l'espace des décisions \mathcal{D} : l'espace des fonctions de répartition des lois log-normales. C'est la raison pour laquelle, proposer un estimateur qui tolère des probabilités de défaillance plus élevées (ou moins élevées) et appartenant à \mathcal{D} , nécessite l'introduction de la classe de fonctions de coût (14). Nous détaillons dans cette section la démarche pratique pour l'estimation de la courbe de fragilité minimisant le coût moyen a *posteriori* relativement aux fonctions de coût définies dans la section précédente. Enfin, nous introduisons des fonctions de coût ponctuelles qui peuvent être appréhendées comme des cas particuliers. Plutôt que de moyenniser un critère de coût sur un intervalle de valeurs de la PGA, la pénalité associée au choix d'un estimateur est basée sur un ensemble discret de valeurs de la PGA. Ces fonctions de coût permettent de se prémunir de façon optimale (au sens de (9)) d'un ou plusieurs séismes d'intensités données.

Les premiers estimateurs que nous construisons pour F_θ sont calculés à partir des fonctions de coût définies dans la section précédente. Si l'on suppose l'équiprobabilité de survenue d'un séisme de PGA s jusqu'à une valeur limite s_{max} , alors

$$dP(s) = \frac{ds}{s_{max}} \mathbf{1}_{[0, s_{max}]}(s).$$

Ainsi les coûts absolu (12), quadratique (13) et absolu généralisé (14) sont bien définis. L'estima-

TABEAU 1. Estimateurs bayésiens / maximum de vraisemblance.

	$\hat{\mu}$	$\hat{\tau}$	(17)
(16) à partir du coût L1 (12)	-0.775	10.21	0.01569
(16) à partir du coût L2 (13)	-0.773	8.98	0.01593
(16) à partir du coût (14) $K_1 = 6/7, K_2 = 1/7$	-0.88	10.64	0.02232
maximum de vraisemblance	-0.772	10.22	0.01570

teur (10) qui minimise la moyenne du coût *a posteriori* s'écrit désormais :

$$\hat{\theta} = \operatorname{argmin}_{\theta} \mathbb{E}_{[\theta|y]} \left[\mathbb{E}_{\mathcal{U}(0, s_{max})} [D(F_{\hat{\theta}}(s), F_{\theta}(s))] \right]. \tag{15}$$

On calcule en pratique une approximation Monte-Carlo de (15),

$$\hat{\theta} = \operatorname{argmin}_{\theta} \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{j=1}^M D(F_{\hat{\theta}}(s_j), F_{\theta_i}(s_j)), \tag{16}$$

avec $\theta_i \sim [\theta|x, y]$ et $s_j \sim \mathcal{U}(0, s_{max})$ de loi uniforme sur $[0, s_{max}]$. Pour chacun des coûts introduits, on résout numériquement ce problème d'optimisation sous R en utilisant la routine `optim()`. Nous avons choisi pour limiter les temps de calculs, $N = 5000$ réalisations de la loi *a posteriori* $[\theta|x, y]$ et $M = 2000$ réalisations de la loi uniforme sur $[0, s_{max}]$. Les estimateurs $\hat{\theta}$ obtenus dans chacun des cas sont rapportés dans le tableau 1. La dernière colonne du tableau renseigne l'approximation Monte-Carlo de la moyenne du coût absolu *a posteriori*

$$\frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{j=1}^M |F_{\hat{\theta}}(s_j) - F_{\theta_i}(s_j)| \tag{17}$$

pour ces estimateurs, mais aussi pour le maximum de vraisemblance. Évidemment, l'estimateur (16) calculé avec $D(F_{\hat{\theta}}(s_j), F_{\theta_i}(s_j)) = |F_{\hat{\theta}}(s_j) - F_{\theta_i}(s_j)|$ possède la valeur de (17) la plus faible. On peut remarquer que le maximum de vraisemblance est le plus proche de ce minimum.

La figure 5 illustre quant à elle les estimations des courbes de fragilité $F_{\hat{\theta}}$ pour plusieurs valeurs de K_1 . La courbe de fragilité qui minimise un coût absolu généralisé tolère des probabilités de défaillance d'autant plus élevées sur toute la gamme des PGA que K_1 augmente. Par conséquent, la classe de distances absolu-généralisé (14) est pertinente du point de vue de son utilisation en fiabilité où il est souvent de mise de surestimer les probabilités de défaillance, notamment quand peu de données (s_i, y_i) sont disponibles.

Remarque 1. Une même PGA peut caractériser plusieurs signaux sismiques différents (voir section 2). C'est la raison pour laquelle, dans un contexte industriel, une analyse de sensibilité à l'accélérogramme du séisme devrait être menée.

Nous étudions dans la suite l'estimation d'une courbe de fragilité minimisant un coût ponctuel dans l'espace des décisions. Nous remarquons que la courbe minimisant la probabilité de défaillance relativement à un séisme de sollicitation $a \in \mathbb{R}^+$ n'est pas unique et qu'en fait, deux valeurs distinctes de la PGA sont nécessaires pour définir un estimateur de F_{θ} . Soit $P(s) = \delta_a(s)$

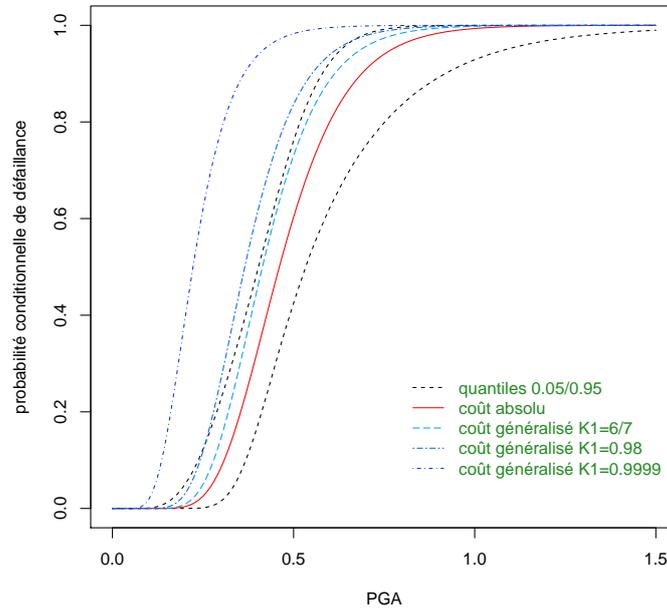


FIGURE 5. Courbes de fragilité bayésiennes calculées à partir de (16) et quantiles à 5% et 95% calculés à partir des échantillons *a posteriori* (voir figure 3).

la mesure de Dirac en a . Alors, les coûts absolu (12), quadratique (13) et absolu généralisé (14) se réécrivent respectivement :

$$L_1^a(\tilde{\theta}, \theta) := |F_{\tilde{\theta}}(a) - F_{\theta}(a)|, \quad (18)$$

$$L_2^a(\tilde{\theta}, \theta) := (F_{\tilde{\theta}}(a) - F_{\theta}(a))^2, \quad (19)$$

$$L_{K_1, K_2}^a(\tilde{\theta}, \theta) := K_1 |F_{\tilde{\theta}}(a) - F_{\theta}(a)| \mathbf{1}_{F_{\tilde{\theta}}(a) > F_{\theta}(a)} + K_2 |F_{\tilde{\theta}}(a) - F_{\theta}(a)| \mathbf{1}_{F_{\tilde{\theta}}(a) < F_{\theta}(a)}. \quad (20)$$

Par suite, l'estimateur qui minimise le coût moyen *a posteriori* (10) s'écrit :

$$\hat{\theta} = \operatorname{argmin}_{\tilde{\theta}} \int_{\Theta} L^a(\tilde{\theta}, \theta) [\theta | x, y] d\theta. \quad (21)$$

Les coûts ponctuels (18), (19) et (20) ne mesurent l'écart entre la courbe estimée et la "vraie" courbe inconnue qu'à travers la probabilité de défaillance conditionnellement à la valeur a de la sollicitation. Par conséquent, le problème de minimisation (21) n'admet pas une unique solution et il est nécessaire de reformuler le problème. Posons :

$$\hat{p} = \operatorname{argmin}_p \int_{\Theta} L^a(p, \theta) [\theta | x, y] d\theta. \quad (22)$$

Ce problème d'optimisation a pour solution une unique probabilité de défaillance \hat{p} . L'équation (22) s'écrit encore :

$$\hat{p} = \operatorname{argmin}_p \mathbb{E}_{[\theta | x, y]} [L^a(p, \theta)]. \quad (23)$$

L'estimateur Monte-Carlo $\frac{1}{n} \sum_{i=1}^n L^a(p, \theta_i)$ calculé à partir de n réalisations de la loi *a posteriori* $[\theta|x,y]$ fournit une bonne approximation de $\mathbb{E}_{[\theta|x,y]}[L^a(p, \theta)]$ pour n suffisamment grand. Soit a, b deux valeurs de la PGA qui caractérisent l'intensité de deux séismes d'intérêts. On pose $P(s) = \delta_a(s) + \delta_b(s)$. Alors, le coût quadratique (13) se réécrit :

$$\begin{aligned} L_2^{a,b}(\tilde{\theta}, \theta) &= L_2^a(\tilde{\theta}, \theta) + L_2^b(\tilde{\theta}, \theta) \\ &= (F_{\tilde{\theta}}(a) - F_{\theta}(a))^2 + (F_{\tilde{\theta}}(b) - F_{\theta}(b))^2. \end{aligned} \tag{24}$$

De la même façon, on peut définir les coûts $L_1^{a,b}$ et $L_{K_1, K_2}^{a,b}$.

Lemme. Soit a, b ; $0 < a < b$ et $p_a \in [0, 1]$, $p_b \in [0, 1]$ vérifiant $p_a < p_b$. Alors, il existe une unique variable aléatoire log-normale C telle que $p_a = \mathbb{P}_{\theta}[C \leq a]$ et $p_b = \mathbb{P}_{\theta}[C \leq b]$.

Démonstration. voir annexe D □

D'après le lemme précédent, l'estimateur (10) calculé à partir du coût (24) est unique.

Exemple. Soit $a = 0.3$ et $b = 0.7$ deux valeurs de la PGA. En utilisant le coût quadratique ponctuel (19) pour a (resp. pour b), on obtient via (23), $\hat{p}_a = 0.098$ (resp. $\hat{p}_b = 0.894$). En utilisant les équations (36), on calcule $\theta = (\mu, \tau) = (-0.77, 8.99)$, qui définit la courbe de fragilité minimisant le coût $L_2^{0.3,0.7}$ (voir figure 6).

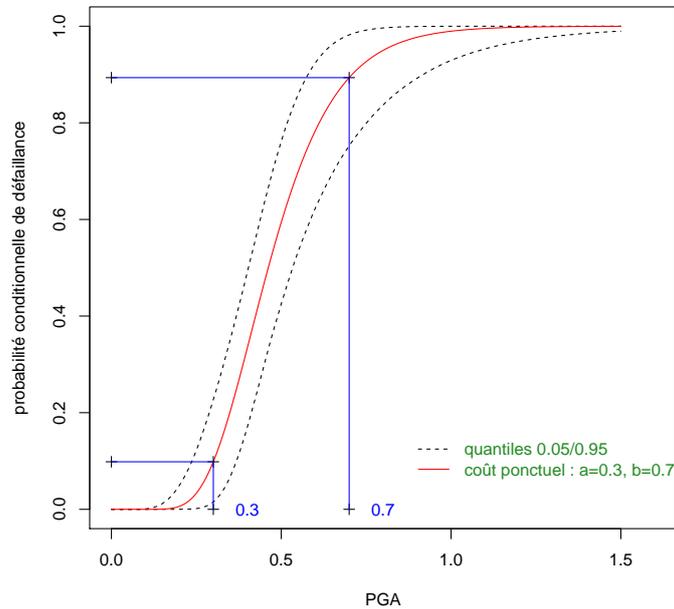


FIGURE 6. Courbe de fragilité minimisant $L_2^{0.3,0.7}$.

Remarque 2. Soit $b < a$ deux valeurs pour la PGA et $dP(s) := \frac{\mathbf{1}_{[b,a]}(s)}{a-b} d\lambda(s)$. Alors,

$$\begin{aligned} \lim_{b \rightarrow a} \int_0^{+\infty} D(F_{\bar{\theta}}(s), F_{\theta}(s)) dP(s) &= D(F_{\bar{\theta}}(a), F_{\theta}(a)) \\ &= \langle D(F_{\bar{\theta}}(s), F_{\theta}(s)), \delta_a \rangle \end{aligned}$$

avec δ_a la distribution de Dirac décentrée en la PGA a . Avec ce choix de P , on peut donc appréhender (18), (19) et (20) comme des cas limites des coûts fonctionnels (12), (13) et (14).

6. Discussion

Dans cette étude, toutes les données proviennent d'un simulateur numérique recevant en entrée le séisme sous la forme d'un profil d'accélération sur une période de temps. Le séisme est ensuite résumé par la seule valeur de l'accélération maximale (PGA). Si des questions d'échantillonnage optimal se posent au regard d'une fonction de coût choisie, la difficulté résidera dans l'échantillonnage des PGA. Nous rappelons que les profils sismiques sont issus d'une bibliothèque de profils observés en conditions réelles. À des fins d'échantillonnage, peut-il être réaliste de construire un profil quelconque ? Il peut également être proposé une analyse de sensibilité aux profils possibles pour une même valeur de la PGA. Si, en plus des données provenant du simulateur, les observations du comportement de la maquette et des observations réelles pour lesquelles la PGA est très partiellement connue sont disponibles, dérouler cette étude de courbes de fragilité en tenant compte des différentes natures d'observations s'inscrirait dans les démarches proposées par Kennedy and O'Hagan (2001); Bayarri et al. (2007).

Les fonctions de coût utilisées dans cet article sont des mesures de risque sismique intégré et, à ce titre elles font intervenir la mesure de probabilité $P(x)$ de survenue d'un séisme d'intensité x . Cette loi est bien sûr spécifique au lieu soumis à sollicitation. Sur la maquette utilisée pour produire les données utilisées dans cet article, il s'agit de tirages dans une bibliothèque de signaux temporels sismiques bivariés (Nord-Sud et Est-Ouest). La sollicitation étant définie comme la moyenne des deux composantes maximales de ces signaux pour chaque événement, il serait sans doute plus approprié d'un point de vue théorique d'utiliser une loi généralisée des extrêmes plutôt que la loi uniforme utilisée ci-dessus comme preuve de concept. Néanmoins, en pratique, à intensité de séisme x fixée, existent des effets de compensations entre probabilité de la sollicitation $P(x)$ et écart entre $\hat{F}(x)$ et $F(x)$, par exemple des rôles complètement duaux dans la fonction de coût (12). Dans une perspective décisionnelle, on ne peut donc demander des informations plus pertinentes sur l'aléa sismique indépendamment d'une réflexion plus approfondie quant aux effets d'adopter une courbe de fragilité erronée dans une gamme d'intensité de séisme.

Cet article présente plusieurs fonctions de coût définies comme des *distances* (au sens intuitif du terme, i.e. de la longueur qui sépare deux points, car au sens strictement mathématique l'axiome de symétrie n'est pas toujours vérifié, en particulier pour le coût absolu-généralisé) dans l'espace des courbes de fragilité. En théorie, cela suppose de faire correspondre une fonction de coût aux types de conséquences socio-économiques occasionnées par l'adoption d'une courbe de fragilité au lieu de la "véritable" courbe inconnue. Ce travail d'ingénieur difficile est pourtant nécessaire pour appliquer rigoureusement la théorie de la décision. La difficulté réside plus dans

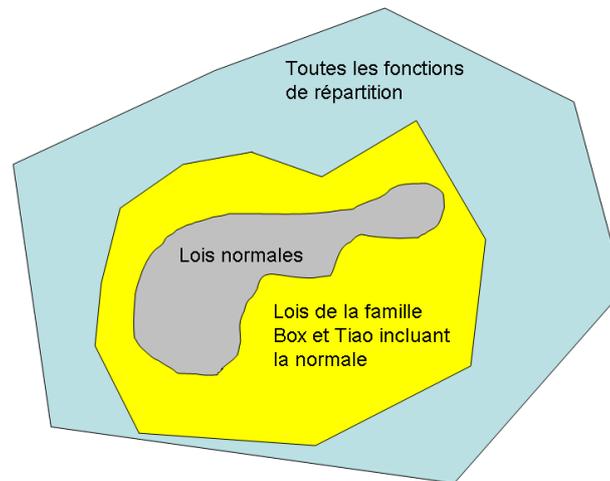


FIGURE 7. La quantification d'une incertitude de modèle nécessite un cadre de référence plus large incluant le modèle d'intérêt.

la traduction et l'encodage des desiderata de l'ingénieur de terrain que dans la disponibilité des fonctions de coût envisageables. Pour mesurer la "distance" entre deux fonctions F et \hat{F} , la classe des divergences d'Ali-Silvey (Ali, 1966) offre par exemple une large palette variée et intéressante, offrant comme cas particuliers les distances de Kullback-Leibler, de Hellinger et de Kolmogorov. En outre, force est de constater que dans la pratique courante de l'ingénierie des risques, il est bien difficile que l'analyste ait à sa disposition une fonction de coût comme donnée d'entrée de l'étude. D'une part, l'élicitation formelle de fonctions de coût par interrogation du (ou des) *décideurs* potentiels ou avérés est un problème difficile et encore peu transposé à la réalité industrielle. D'autre part, dans le cas de notre étude, et plus généralement de toute analyse visant à l'évaluation d'un critère de risque de type *quantile* ou *probabilité de défaillance*, il nous semble irréaliste d'imaginer que l'analyste puisse posséder, à son niveau, une vision très claire des conséquences socio-économiques de l'erreur d'estimation, et ce pour deux raisons. Premièrement, ces analyses ont souvent un caractère qu'on pourrait qualifier de *normatif* ou encore de *conventionnel*, i.e. l'évaluation du critère probabiliste, souvent conditionnellement à des scénarios accidentels imposés d'avance (et déjà très improbables), est une fin en soi et il n'est nullement demandé à l'analyste de mener une analyse poussée des conséquences de la survenue de l'événement redouté. Deuxièmement, les résultats de l'analyse ne sont souvent qu'un des nombreux ingrédients de chaînes de calcul complexes et il est donc pratiquement impossible d'avoir une vision décisionnelle claire pour chacun de ses ingrédients. Néanmoins, nous défendons la validité de l'approche décisionnelle bayésienne dans le cadre industriel pour des raisons de cohérence mathématique mais aussi opérationnelles (voir également Keller et al., 2011; Pasanisi et al., 2012) : la méthode permet à l'analyste de proposer des estimateurs ponctuels et de les justifier auprès de son *commanditaire* au vu, par exemple, de leur caractère conservatif, par des arguments de type "cet estimateur pénalise deux (ou quatre ou dix !) fois plus la sous-estimation que la sur-estimation de la probabilité de défaillance". De cette manière, il est possible d'associer

une représentation mathématique précise au concept de *conservatisme*, cher aux ingénieurs. Dans ce sens, l'approche bayésienne est ici davantage *descriptive* que *décisionnelle* (Keller et al., 2011). Lorsqu'un choix neutre de fonction de coût est fait (coût quadratique ou absolu pour lesquels la pénalisation est symétrique), nous avons observé sous un *a priori* vague que l'estimateur calculé est proche de l'estimateur du maximum de vraisemblance.

Enfin, dans cet article, l'hypothèse de normalité est la pierre angulaire du raisonnement qui n'est jamais remise en cause. Il s'agit certes d'une hypothèse de commodité n'impliquant que des calculs à deux inconnues mais aussi du modèle systématiquement considéré chez les praticiens de l'ingénierie. Pourtant, ce choix d'hypothèse réduit l'incertitude de modélisation à l'incertitude liée à la méconnaissance des paramètres de moyenne et de précision. Pour exprimer cette incertitude de modèle tout en restant dans un cadre paramétrique, il pourrait être judicieux de découper un ensemble plus vaste, comme par exemple le modèle paramétrique à 3 dimensions $\theta = (\mu, \tau, \lambda)$ obtenu par la transformation de Box-Cox :

$$T = \frac{Z^\lambda - 1}{\lambda}$$

$$Z \sim \mathcal{N}(\mu, \tau)$$

$F_{\lambda, \mu, \sigma^{-2}}$, la distribution de la variable T , par continuité pour $\lambda = 1$, redonne la loi normale. D'autres extensions possibles sont suggérées dans Box and Tiao (1973). Dans ce contexte, l'incertitude de modèle peut être évaluée séparément grâce au "poids" de l'hypothèse $\lambda = 1$ dans cette famille de lois à 3 paramètres, comme le visualise la figure 7. De la même façon, l'espace des décisions \mathcal{D} peut aussi être agrandi à cette famille ou même à une famille différente comme les lois de Student (lois à queues plus lourdes). Enfin, doter d'une mesure l'espace fonctionnel des fonctions de répartition est aussi une perspective à envisager, qui conduirait à l'estimation non paramétrique de F . Dans un contexte classique, l'estimateur de référence serait le produit limite de Kaplan-Meier (Kaplan and Meier, 1958), tandis que le paradigme bayésien s'appuie généralement sur une modélisation à base de processus de Dirichlet comme dans Susarla and Van Ryzin (1976). Cependant, l'inconvénient bien connu d'une telle modélisation est qu'elle restreindrait F à un espace de distributions discrètes, composé de sommes pondérées infinies dénombrables de masses de Dirac. Afin de pouvoir modéliser F comme une fonction de répartition continue, une approche plus adéquate serait donc d'employer un mélange de processus de Dirichlet (également introduit dans Ferguson, 1973), à noyau continu, par exemple gaussien. Cette approche peut-être vue dans le cas présent comme l'équivalent bayésien d'une estimation non-paramétrique à noyaux de la densité des log-capacités. Enfin, étant donné la multitude de modélisations possibles d'une courbe de fragilité mentionnées ci-dessus, une autre perspective serait de comparer ces modèles entre eux. Dans un cadre bayésien, ceci impliquerait de faire appel à une procédure de sélection bayésienne de modèle (Wasserman, 2000), voire d'intégration de l'incertitude de modèle (Hoeting et al., 1999), si l'on souhaite tenir compte de l'information apportée par chacun d'entre eux, en les pondérant par leurs probabilités *a posteriori* dans l'espace des modèles considérés. Dans les deux cas, on est amené sur le plan pratique à calculer la vraisemblance marginale sous chacun des modèles, soit l'espérance *a priori* de la vraisemblance. Sauf dans des cas très simples, cette vraisemblance marginale n'a pas d'expression explicite, mais il existe de nombreuses techniques d'approximation asymptotique ou numérique (Kass and Raftery, 1995) dans le cas de modèles

paramétriques. Le calcul de la vraisemblance marginale dans un modèle de processus de Dirichlet est également possible (Basu and Chib, 2003), modulo l'utilisation d'un algorithme de filtrage particulière (Doucet et al., 2001) pour l'échantillonnage *a posteriori*.

Remerciements. Nous remercions Irmela Zentner et Anne-Laure Popelin (EDF R&D) pour avoir partagé leur expertise dans le domaine de la fragilité sismique et pour avoir mis en œuvre les simulations numériques qui nous ont permis d'illustrer nos propos. Ces simulations ont été réalisées à l'aide des logiciels Code_Aster et Open TURNS, au sein de la plateforme informatique Salome-Meca (voir www.code-aster.org et www.openturns.org).

Annexe A: Maximum de vraisemblance

On dispose d'un n -échantillon $(x, y) := \{(x_1, y_1), \dots, (x_n, y_n)\}$ avec $x = (x_1, \dots, x_n)$ les log-sollicitations sismiques et $y = (y_1, \dots, y_n)$ le vecteur de données binaires correspondant (défaillance ou non). La vraisemblance du modèle vaut :

$$\begin{aligned}
 [y|x, \theta] &= \prod_{i=1}^n [y_i|x_i, \theta] \\
 &= \prod_{i=1}^n F_\theta(x_i)^{y_i} \times (1 - F_\theta(x_i))^{1-y_i}
 \end{aligned}$$

avec $F_\theta(x_i) = \Phi((x - \mu)\sqrt{\tau})$ la fonction de répartition de $Z \sim \mathcal{N}(\mu, \tau)$. Par suite,

$$\begin{aligned}
 \log [y|x, \theta] &= \sum_{i=1}^n \log [y_i|x_i, \theta] \\
 &= \sum_{i=1}^n y_i \ln F_\theta(x_i) + (1 - y_i) \ln(1 - F_\theta(x_i))
 \end{aligned}$$

On peut utiliser la fonction `R optim()` pour calculer l'estimateur du maximum de vraisemblance $\hat{\theta} = (\hat{\mu}, \hat{\tau})$. De plus, la matrice d'information de Fisher $\mathcal{I}(\theta)$ peut être calculée. Elle mesure la variabilité de $\hat{\theta}$ et définit la loi *a priori* de Jeffreys $[\theta] \propto \sqrt{|I(\theta)|}$. Soit $l_i := \log [y_i|x_i, \theta]$ et $F_i := F_\theta(x_i)$, alors :

$$\begin{aligned}
 \mathcal{I}(\theta) &= \sum_{i=1}^n \begin{pmatrix} -\mathbb{E}_\theta \left[\frac{\partial^2 l_i}{\partial \mu^2} \right] & -\mathbb{E}_\theta \left[\frac{\partial^2 l_i}{\partial \tau \partial \mu} \right] \\ -\mathbb{E}_\theta \left[\frac{\partial^2 l_i}{\partial \mu \partial \tau} \right] & -\mathbb{E}_\theta \left[\frac{\partial^2 l_i}{\partial \tau^2} \right] \end{pmatrix} \\
 &= \sum_{i=1}^n \begin{pmatrix} \frac{\partial_\mu^2 F_i}{F_i(1-F_i)} & \frac{\partial_\mu F_i \partial_\tau F_i}{F_i(1-F_i)} \\ \frac{\partial_\mu F_i \partial_\tau F_i}{F_i(1-F_i)} & \frac{\partial_\tau^2 F_i}{F_i(1-F_i)} \end{pmatrix} \tag{25}
 \end{aligned}$$

Méthode E.M. Cet algorithme d'estimation itératif est une alternative à l'approche standard lorsque la vraisemblance est celle d'un modèle hiérarchique à variables latentes. Soit Z_i la variable aléatoire log-capacité de paramètre θ à estimer. Dans la modélisation statistique pour l'estimation

de F_θ , nous n'avons pas connaissance du vecteur des log-capacités $z = (z_1, z_2, \dots, z_n)$. Après avoir initialisé l'algorithme à $\theta = \theta_0$, nous présentons les deux étapes qui conduisent de l'itération k à l'itération $k + 1$:

1. **Étape E.** Calculer $Q(\theta|\theta_k) = \mathbb{E}_{Z|y,x,\theta_k}(\ln[y, z|x, \theta])$.
2. **Étape M.** Calcul de $\theta_{k+1} = \underset{\theta}{\operatorname{argmax}} Q(\theta|\theta_k)$.

Notons $f_\theta(\cdot)$ la densité de probabilité de la loi normale de paramètre $\theta = (\mu, \tau)$. La vraisemblance ponctuelle complète s'écrit :

$$\begin{aligned} [y_i, z_i|x_i, \theta] &= [z_i|\theta] \times [y_i|z_i, x_i, \theta] \\ &= f_\theta(z_i) \times (\mathbf{1}_{z_i < x_i})^{y_i} \times (\mathbf{1}_{z_i > x_i})^{1-y_i} \\ &= f_\theta(z_i), \end{aligned}$$

puisque $(\mathbf{1}_{z_i < x_i})^{y_i} \times (\mathbf{1}_{z_i > x_i})^{1-y_i} = 1$. La log-vraisemblance complète peut s'écrire comme le produit d'un vecteur de statistiques $S(z)$ par un vecteur $v(\theta)$ fonction du paramètre $\theta = (\mu, \tau)$ à estimer :

$$\begin{aligned} \ln [y, z|x, \theta] &= \sum_{i=1}^n \ln [y_i, z_i|x_i, \theta] \\ &= S(z)^t v(\theta) \end{aligned} \quad (26)$$

avec,

$$S(z) = \begin{pmatrix} n \\ \sum_{i=1}^n z_i \\ \sum_{i=1}^n z_i^2 \end{pmatrix} \text{ et } v(\theta) = \begin{pmatrix} \frac{-\mu^2\tau}{2} - \frac{1}{2} \ln 2\pi + \frac{1}{2} \ln \tau \\ \mu\tau \\ -\frac{\tau}{2} \end{pmatrix}. \quad (27)$$

En utilisant (26), nous pouvons calculer $Q(\theta|\theta_k)$ (**Étape E.**) :

$$\begin{aligned} Q(\theta|\theta_k) &= \mathbb{E}_{Z|y,x,\theta_k}[S(z)^t v(\theta)] \\ &= \tilde{S}(z)^t v(\theta) \end{aligned} \quad (28)$$

avec $\tilde{S}(z)^t = (n, \sum_{i=1}^n \mathbb{E}_{Z_i|y_i, x_i, \theta_k}[z_i], \sum_{i=1}^n \mathbb{E}_{Z_i|y_i, x_i, \theta_k}[z_i^2])$. Soit $[z_i|y_i, x_i, \theta]$ la densité de probabilité de la donnée manquante conditionnellement à la donnée observée :

$$\begin{aligned} [z_i|y_i, x_i, \theta] &= \frac{[y_i, z_i|x_i, \theta]}{[y_i|x_i, \theta]} \\ &= \frac{f_\theta(z_i)}{F_\theta(x_i)^{y_i} \times (1 - F_\theta(x_i))^{1-y_i}}. \end{aligned}$$

On reconnaît la densité de probabilité de la loi normale tronquée à droite (si $y_i = 1$) resp. à gauche (si $y_i = 0$). On peut maintenant calculer l'espérance $\mathbb{E}_{Z|y,x,\theta_k}(\ln[y, z|x, \theta])$ de la log-vraisemblance des données complètes pour le n -échantillon z . L'**étape M.** est basée sur le calcul du gradient de $Q(\theta|\theta^k)$:

$$\frac{\partial Q(\theta|\theta^k)}{\partial \mu} = \tilde{S}(z)^t \frac{\partial v(\theta)}{\partial \mu}$$

et

$$\frac{\partial Q(\theta|\theta^k)}{\partial \tau} = \tilde{S}(z)' \frac{\partial v(\theta)}{\partial \tau}.$$

D'après (27),

$$\frac{\partial v(\theta)}{\partial \mu} = \begin{pmatrix} -\tau\mu \\ \tau \\ 0 \end{pmatrix} \text{ et } \frac{\partial v(\theta)}{\partial \tau} = \begin{pmatrix} -\frac{\mu^2}{2} + \frac{1}{2\tau} \\ \mu \\ -\frac{1}{2} \end{pmatrix}. \quad (29)$$

Par suite, en utilisant l'expression de $v(\theta)$ (29), on démontre aisément que :

$$\frac{\partial Q(\theta|\theta^k)}{\partial \mu} = 0 \implies \mu_{k+1} = \frac{\sum_{i=1}^n m_i}{n}$$

$$\frac{\partial Q(\theta|\theta^k)}{\partial \tau} = 0 \implies \tau_{k+1} = n \left[\sum_{i=1}^n (m_i - \mu_{k+1})^2 + \sum_{i=1}^n s_i^2 \right]^{-1}$$

avec $m_i = \mathbb{E}_{\mu_k, \tau_k} [Z_i | Z_i < x_i]^{y_i} \mathbb{E}_{\mu_k, \tau_k} [Z_i | Z_i > x_i]^{1-y_i}$ la moyenne de la loi normale tronquée à droite (si $y_i = 1$), resp. à gauche (si $y_i = 0$) et $s_i^2 = (\mathbb{V}_{\mu_k, \tau_k} [Z_i | Z_i < x_i])^{y_i} (\mathbb{V}_{\mu_k, \tau_k} [Z_i | Z_i > x_i])^{1-y_i}$ la variance de la loi normale tronquée à droite (si $y_i = 1$), resp. à gauche (si $y_i = 0$).

Notons ϕ la densité de la loi normale centrée réduite, Φ sa fonction de répartition et $\alpha_i := \sqrt{\tau}(x_i - \mu)$. Nous rappelons ci-dessous les expressions de la moyenne et de la variance de la loi normale tronquée :

$$\mathbb{E}[Z_i | Z_i > x_i] = \mu + \frac{\phi(\alpha_i)}{\sqrt{\tau}(1 - \Phi(\alpha_i))},$$

$$\mathbb{E}[Z_i | Z_i < x_i] = \mu - \frac{\phi(\alpha_i)}{\sqrt{\tau}\Phi(\alpha_i)},$$

et

$$\mathbb{V}[Z_i | Z_i > x_i] = \tau^{-1} \left(1 - \frac{\phi(\alpha_i)^2 - \alpha_i \phi(\alpha_i)(1 - \Phi(\alpha_i))}{(1 - \Phi(\alpha_i))^2} \right),$$

$$\mathbb{V}[Z_i | Z_i < x_i] = \tau^{-1} \left(1 - \frac{\phi(\alpha_i)^2 + \alpha_i \phi(\alpha_i)\Phi(\alpha_i)}{\Phi(\alpha_i)^2} \right).$$

Annexe B: Preuve de (8)

Il suffit de remarquer, qu'à μ fixé, au voisinage de $\tau = 0$:

$$\frac{1}{\tau} \prod_{i=1}^n F_{(\mu, \tau)}(x_i)^{y_i} (1 - F_{(\mu, \tau)}(x_i))^{1-y_i} \sim \frac{1}{2^n \tau}, \quad (30)$$

et, pour tout $a > 0$, $\int_0^a \frac{1}{2^n \tau} d\tau = +\infty$.

Annexe C: Algorithme de Gibbs pour l'échantillonnage de $[\theta|(x, y)]$

Une valeur $\theta_0 = (\mu_0, \tau_0)$ du paramètre est choisi afin d'initialiser l'algorithme. Connaissant θ_k , nous détaillons ci-dessous l'étape $k + 1$:

1. On génère :

$$\{z_{k+1}\}^{i=1 \dots n} \sim [z|y, x, \theta_k] = \prod_{i=1}^n [z_k | z_k < x_i]^{y_i} [z_k | z_k > x_i]^{1-y_i}. \quad (31)$$

2. On génère θ_{k+1} suivant la distribution de probabilité *a posteriori* de densité $[\theta|z_{k+1}, y, x]$ (soit directement, soit en calculant $[\mu|\tau_k, z_{k+1}, y, x]$ puis $[\tau|\mu_k, z_{k+1}, y, x]$).

Soit z le n -échantillon des capacités (variables latentes). La formule de Bayes (4) appliquée aux données complètes (y, z) s'écrit :

$$\begin{aligned} [\theta|z, y, x] &\propto [z, y|\theta, x][\theta] \\ &\propto [z|\theta, x][\theta]. \end{aligned} \quad (32)$$

Soit τ et z supposés connus,

$$[\mu|\tau, z, y, x] \propto [\tau, z, y|\mu, x][\mu]. \quad (33)$$

De même, soit μ et z supposés connus,

$$[\tau|\mu, z, y, x] \propto [\mu, z, y|\tau, x][\tau]. \quad (34)$$

Opter pour une loi *a priori* gamma-normale (5) est commode pour appliquer l'algorithme de Gibbs. En effet, puisque la loi (31) des log-capacités latentes z conditionnellement aux données observées y est gaussienne, la loi *a posteriori* (32) est aussi gamma-normale :

$$[\mu, \tau|z, y, x] = [\mu|\tau, z, y, x][\tau|z, y, x] ; \begin{cases} [\mu|\tau, z, y, x] \sim \mathcal{N}\left(\frac{n\bar{z} + \lambda m}{n + \lambda}, (n + \lambda)\tau\right), \\ [\tau|z, y, x] \sim \mathcal{G}\left(\frac{n}{2} + a, \frac{1}{2}\left(S + 2b + \frac{n\lambda(\bar{z} - m)^2}{n + \lambda}\right)\right) \end{cases} \quad (35)$$

avec $S = \sum_{i=1}^n (z_i - \bar{z})^2$ et $\bar{z} = n^{-1} \sum_{i=1}^n z_i$. Étant donné θ_k , il suffit donc d'échantillonner $(z_i)_{1 \leq i \leq n}$ selon la loi normale tronquée n -dimensionnelle (31), puis d'utiliser (35) pour simuler θ_{k+1} et ainsi de suite.

Si l'on choisit d'utiliser la loi *a priori* non informative (6), l'algorithme de Gibbs requiert le calcul des lois *a posteriori* (33) et (34) dont les expressions analytiques sont connues (voir par exemple Box and Tiao, 1973). Dans ce cas, la loi *a posteriori* $[\mu, \tau|x, y]$ est impropre. Il suffit alors de considérer les lois *a priori* $[\mu] \propto \mathbf{1}_{[-N, N]}$ et $[\tau] \propto \frac{1}{\tau} \mathbf{1}_{[1/N, N]}$, rendant propre la loi *a posteriori*.

Annexe D: Une caractérisation de la loi log-normale

Lemme. Soit a, b ; $0 < a < b$ et $p_a \in [0, 1]$, $p_b \in [0, 1]$ vérifiant $p_a < p_b$. Alors, il existe une unique variable aléatoire log-normale C telle que $p_a = \mathbb{P}_\theta[C \leq a]$ et $p_b = \mathbb{P}_\theta[C \leq b]$.

Démonstration. On rappelle que Φ désigne la fonction de répartition de la loi normale centrée réduite. On suppose que $p_a = \mathbb{P}_\theta[C \leq a]$ et $p_b = \mathbb{P}_\theta[C \leq b]$. Soit $\theta = (\mu, \tau)$, alors :

$$\begin{aligned}
 & \begin{cases} p_a = \mathbb{P}_\theta[Z \leq \ln a] \\ p_b = \mathbb{P}_\theta[Z \leq \ln b] \end{cases} \\
 & \iff \\
 & \begin{cases} p_a = \Phi\left[(\ln a - \mu)\sqrt{\tau}\right] \\ p_b = \Phi\left[(\ln b - \mu)\sqrt{\tau}\right] \end{cases} \\
 & \iff \\
 & \begin{cases} \Phi^{-1}(p_a) = (\ln a - \mu)\sqrt{\tau} \\ \Phi^{-1}(p_b) = (\ln b - \mu)\sqrt{\tau} \end{cases} \\
 & \iff \\
 & \begin{cases} \sqrt{\tau} = [\ln b - \ln a]^{-1} \times [\Phi^{-1}(p_b) - \Phi^{-1}(p_a)] \\ \mu = \ln a - \tau^{-1/2}\Phi^{-1}(p_a) \end{cases} \tag{36}
 \end{aligned}$$

□

Références

- Albert, J. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422) :669–679.
- Ali, S. M. ; Silvey, S. D. (1966). A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society, Series B*, 28(1) :131–142.
- Ayyub, B., Foster, J., and McGill, W. (2009a). Risk analysis of a protected hurricane-prone region. I : Model development. *Natural Hazards Review*, 10(2) :38–53.
- Ayyub, B., Foster, J., McGill, W., and Jones, H. (2009b). Risk analysis of a protected hurricane-prone region. II : Computations and illustrations. *Natural Hazards Review*, 10(2) :54–67.
- Basu, S. and Chib, S. (2003). Marginal likelihood and bayes factors for dirichlet process mixture models. *Journal of the American Statistical Association*, 98(461) :224–235.
- Bayarri, M. J., Berger, J. O., Sacks, P. R., Cafeo, J. A., Cavendish, J., Lin, C.-H., and Tu, J. (2007). A framework for validation of computer models. *Technometrics*, 49 :138–154.
- Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*. In Statistics. Springer-Verlag, New York.
- Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. Wiley, London, 1 edition.
- Bernier, J. (2003). Décisions et comportement des décideurs face au risque hydrologique. *J. Sci. Hydrol.*, 48(3) :301–316.
- Bouc, R. (1967). Forced vibration of mechanical systems with hysteresis. *Proceedings of the 4th conf. on nonlinear oscillation*.
- Box, G. E. P. and Tiao, G. T. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading.
- Casella, G. and George, E. (1992). Explaining the Gibbs sampler. *Am. Stat.*, 46 :167–174.
- Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4) :327–335.

- Clough, R. and Penzien, J. (1977). *Dynamics of structures*. New York : McGraw-Hill, Inc.
- Collett, D. (2003). *Modelling Survival Data in Medical Research*. Chapman & Hall/CRC, Boca Raton, 2 edition.
- Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1) :1–38.
- Doucet, A., de Freitas, N., and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. Springer Verlag, New York.
- Ellingwood, R. and Kinali, K. (2009). Quantifying and communicating uncertainty in seismic risk assessment. *Structural Safety*, 31(2) :179–187.
- EPRI (Final Report TR-103959, 1994). Methodology for developing seismic fragilities.
- Fabbrocino, G., Iervolino, I., Orlando, F., and Salzano, E. (2005). Quantitative risk analysis of oil storage facilities in seismic areas. *Journal of Hazardous Materials*, pages 61–69.
- Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *The American Statistician*, 1 :209–230.
- Gelfand, A. and Smith, A. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85 :398–409.
- Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004). *Bayesian Data Analysis. Second Edition*. Col. Texts in Statistical Science. Chapman & Hall, New-York USA.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of image. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6 :721–741.
- Ghosh, J. and Ramamoorthi, R. (2002). *Bayesian nonparametrics*. Springer.
- HAZUS (FEMA, 2003). Hazus multy hazard.
- Herbin, A. and Barbato, M. (2012). Fragility curves for building envelope components subject to windborne debris impact. *Journal of Wind Engineering and Industrial Aerodynamics*, 107,108(0) :285–298.
- Hoeting, J., Madigan, D., Raftery, A., and Volinsky, C. (1999). Bayesian model averaging : A tutorial. *Statistical Science*, 14(4) :382–417.
- Hoshi, S., Maruyama, Y., and Yamazaki, F. (2011). Reevaluation method of fragility curves of wooden house based on collected damage information. *Procedia Engineering*, 14 :227–232.
- Huang, Y., Whittaker, S., and Luco, N. (2011a). A probabilistic seismic risk assessment procedure for nuclear power plants : (I) methodology. *Nuclear Engineering and Design*, 241(9) :3996–4003.
- Huang, Y., Whittaker, S., and Luco, N. (2011b). A probabilistic seismic risk assessment procedure for nuclear power plants : (II) application. *Nuclear Engineering and Design*, 241(9) :3985–3995.
- IAEA-TECDOC-724 (1993). *Probabilistic safety assessment for seismic events*. International Atomic Energy Agency.
- Jordaan, I. (2005). *Decisions Under Uncertainty : Probabilistic Analysis For Engineering Decisions*. Cambridge University Press. 275 pages.
- Kalalo, E. and Brenot, D. (2003). Rôles et limites des EPS. *Contrôle*, 155 :39–42.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53 :457–481.
- Kass, R. and Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430) :773–795.
- Keller, M., Pasanisi, A., and Parent, E. (2011). Réflexions sur l’analyse d’incertitudes dans un contexte industriel : information disponible et enjeux décisionnels. *Journal de la Société Française de Statistique*, 152(4) :60–77.
- Kennedy, M. and O’Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society, Series B, Methodological*, 63 :425–464.
- Lagaros, D., Tsompanakis, Y., Psarropoulos, N., and Georgopoulos, C. (2009). Computationally efficient seismic fragility analysis of geostructures. *Computers & Structures*, 87(19,20) :1195–1203.
- Lallemant, D. and Kiremidjian, A. (2013). Fitting fragility curves to empirical data. *GEM Technical Report*.
- Lanore, J. (2003). Perspectives sur le développement des EPS de niveau 1,2 et 3. *Contrôle*, 155 :53–57.
- Marano, G., Greco, R., and Morrone, E. (2011). Analytical evaluation of essential facilities fragility curves by using a stochastic approach. *Engineering Structures*, 33 :191–201.
- McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. Monographs on Statistics and Applied Probability. Chapman and Hall.
- McCulloch, C. (1994). Maximum likelihood variance components estimation for binary data. *Journal of the American Statistical Association*, 89(425) :330–335.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., and Teller, A. Hand Teller, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.*, 21 :1087–1091.
- Parmigiani, G. and Inoue, L. (2008). *Decision Theory : Principles and Approaches*. Wiley.

- Pasanisi, A., Keller, M., and Parent, E. (2012). Estimation of a quantity of interest in uncertainty analysis : Some help from bayesian decision theory. *Reliability Engineering and System Safety*, 100(0) :93–101.
- Reese, S., Bradley, B., Bind, J., Smart, G., Power, W., and Sturman, J. (2011). Empirical building fragilities from observed damage in the 2009 south pacific tsunami. *Earth-Science Reviews*, 107 :156–173.
- Robert, C. (1992). *L'analyse statistique bayésienne*. Economica.
- Schultz, T., Gouldby, P., Simm, D., and Wibowo, L. (2010). Beyond the factor of safety : Developing fragility curves to characterize system reliability. Technical Report ERDC SR-10-1, US Army Corps of Engineers.
- Shinozuka, M., Feng, M., Lee, J., and Naganuma, T. (2000). Statistical analysis of fragility curves. *Journal of Engineering Mechanics*, 126(12) :1224–1231.
- Shoji, G. and Moriyama, T. (2007). Evaluation of the structural fragility of a bridge structure subjected to a tsunami wave load. *Journal of Natural Disaster Science*, 29(2) :73–81.
- Solomos, G., Pinto, A., and Dimova, S. (2008). A review of the seismic hazard zonation in national building codes in the context of eurocode 8. Technical Report EUR 23563 EN-2008, JRC European Commission.
- Stewart, G. and Netherton, D. (2008). Security risks and probabilistic risk assessment of glazing subject to explosive blast loading. *Reliability Engineering and System Safety*, 93(4) :627–638.
- Straub, D. and Der Kiureghian, A. (2008). Improved seismic fragility modeling from empirical data. *Structural Safety*, 30(4) :320–336.
- Susarla, J. and Van Ryzin, J. (1976). Nonparametric bayesian estimation of survival curves from incomplete observations. *Journal of the American Statistical Association*, 71(356) :897–902.
- Tanner, M. H. (1992). *Tools for Statistical Inference : Observed Data and Data Augmentation Methods*. Springer-Verlag, New York.
- Ulmo, J. and Bernier, J. (1973). *Eléments de Décision Statistique*. PUF.
- Wald, A. (1950). *Statistical Decision Functions*. Wiley.
- Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology*, 44 :92–107.
- Wen, Y. (1976). Method for random vibration of hysteretic systems. *J Eng Mech Div, ASCE*, 4 :102–150.
- Zentner, I. (2010). Numerical computation of fragility curves for NPP equipment. *Nuclear Engineering and Design*, 240(6) :1614–1621.
- Zentner, I., Boronovo, E., and Tarantola, S. (2011). Use of hdmr metamodel for seismic fragility analysis. In Nishijima, K., editor, *Applications of Statistics and Probability in Civil Engineering, ICASP 2011*, pages 653–660. CRC Press.