

Estimation de probabilité d'événement rare sous contraintes de monotonie

Journée MAS

27-29.08.2014

Vincent Moutoussamy

Nicolas Bousquet, Bertrand Iooss - EDF R&D - MRI

Fabrice Gamboa, Thierry Klein - Université Toulouse Paul Sabatier

Contexte

- ▶ Fournir des indicateurs de fiabilité d'un composant par des méthodes statistiques
- ▶ on dispose de peu d'observations, cette étude de fiabilité passe par un modèle numérique simulant le comportement physique du composant
- ▶ ce code permet d'obtenir une grandeurs physique représentant la fiabilité du composant étudié
- ▶ on considère qu'il y a événement indésirable lorsque la sortie du code dépasse un seuil fixé
- ▶ de plus ce code est
 - coûteux en temps de calcul
 - déterministe
 - de type boîte noire
 - présence possible d'effets-falaises

Notations

▶ On note

- $\mathbf{X} = (X_1, \dots, X_d)$ le vecteur aléatoire d'entrée
- g le code numérique

▶ Sans perte de généralité

- $\mathbf{X} \sim \mathcal{N}_d(0, I_d)$
- $G(\mathbf{X}) = g(\mathbf{X}) - \textit{seuil}$
- on veut estimer $P_f = \mathbb{P}(G(\mathbf{X}) \leq 0)$

Quelques méthodes classiques

- ▶ Méthode de Monte Carlo, on estime P_f par

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{G(\mathbf{x}_i) \leq 0\}} \text{ où } (\mathbf{X}_i)_{i \geq 1} \text{ est une suite de variables}$$

aléatoires indépendantes de même loi que \mathbf{X}

- ▶ facile à estimer et aucune hypothèse sur le code n'est nécessaire
- ▶ mais pour obtenir un coefficient de variation de 10% lorsque $P_f = 10^{-k}$ il faut environ 10^{k+2} appels au code numérique
- ▶ si un appel au code dure une seconde, il faut pour
 - $k = 3$ attendre 1 jour
 - $k = 4$ attendre 11 jours
 - $k = 5$ attendre 4 mois
 - $k = 6$ attendre 3 ans

Quelques méthodes classiques

- ▶ Méthode de simulation préférentielle (ou tirage d'importance), on estime P_f par $\tilde{p}_n = \frac{1}{n} \sum_{i=1}^n \frac{f_{\mathbf{X}}(\mathbf{V}_i)}{\tilde{f}(\mathbf{V}_i)} \mathbb{1}_{\{G(\mathbf{V}_i) \leq 0\}}$ où $(\mathbf{V}_i)_{i \geq 1}$ est une suite de variables aléatoires indépendantes ayant pour fonction de densité \tilde{f}
- ▶ \tilde{f} doit être facile à calculer et \mathbf{V}_i facile à simuler
- ▶ obtenir \tilde{f} qui minimise la variance de \tilde{p}_n revient à connaître P_f

Quelques méthodes classiques

- ▶ Méthode de *Subset Simulation*, consiste à décomposer l'ensemble $\mathbb{U}^- = \{\mathbf{x} : G(\mathbf{x}) \leq 0\}$ en plusieurs ensembles $A_0 = \mathbf{R}^d \supset A_1 \supset \dots \supset A_{k-1} \supset A_k = \mathbb{U}^-$ et

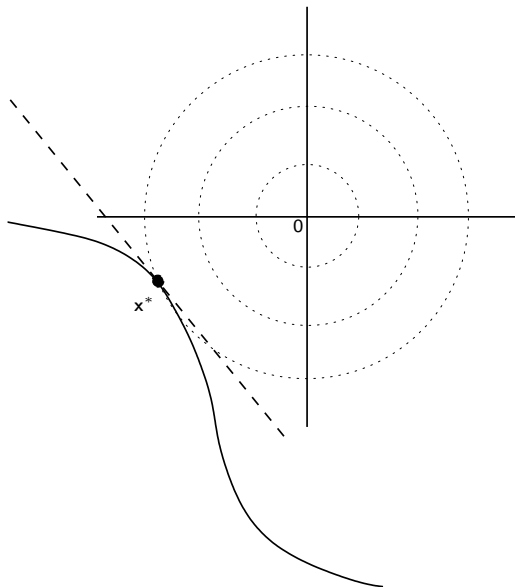
$$P_f = \prod_{i=1}^k \mathbb{P}(\mathbf{X} \in A_i | \mathbf{X} \in A_{i-1})$$

- ▶ on estime $\mathbb{P}(\mathbf{X} \in A_i | \mathbf{X} \in A_{i-1})$ par une méthode de Monte Carlo
- ▶ Si $P_f = 10^{-k}$ et $\mathbb{P}(\mathbf{X} \in A_i | \mathbf{X} \in A_{i-1}) = 10^{-1}$, avoir un coefficient de variation de 10% sur chacune des quantités nécessitera 10^3 appels au code numérique
- ▶ si un appel au code dure une seconde, il faut pour $k = 7$ attendre 2 heures

Quelques méthodes classiques

- ▶ Méthode FORM est une méthode "ingénieur"
- ▶ consiste à approcher l'état limite $\Gamma = \{\mathbf{x} : G(\mathbf{x}) = 0\}$ par une surface linéaire
- ▶ Deux étapes
 - 1) trouver le point \mathbf{x}^* le plus proche de l'origine appartenant à Γ
 - 2) estimer P_f par $\Phi(-\|\mathbf{x}^*\|)$
- ▶ nécessite peu d'appels au code, facile à coder
- ▶ il faut que le code soit régulier et l'approximation dépend fortement de la forme de Γ
- ▶ aucun contrôle sur l'estimation (pas d'intervalle de confiance)
- ▶ peut être couplé au tirage d'importance en prenant $\mathbf{V} \sim \mathcal{N}_d(\mathbf{x}^*, I_d)$

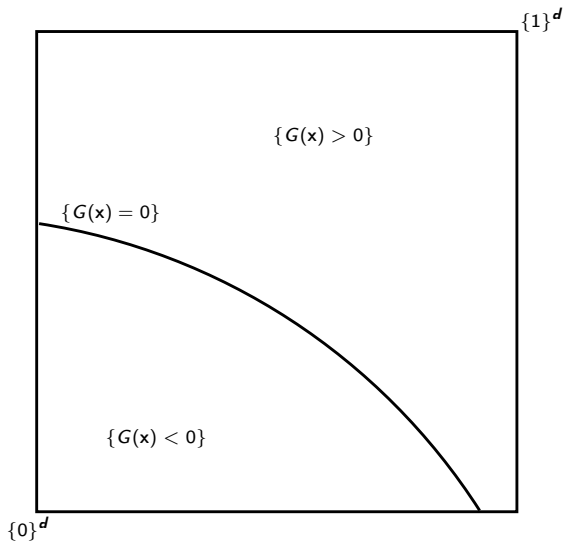
Quelques méthodes classiques



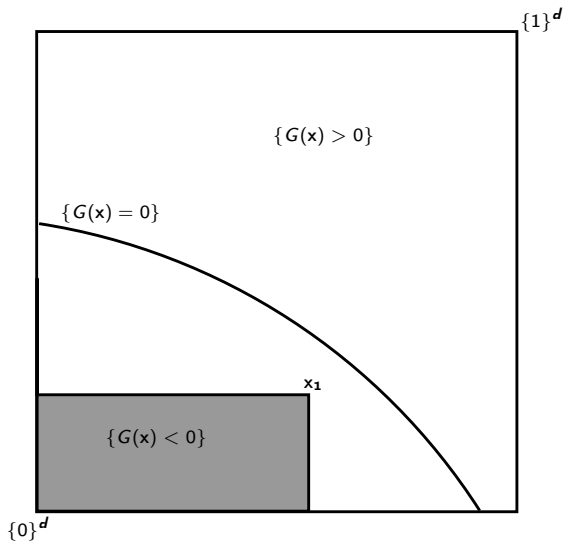
Apport de la monotonie

- ▶ Connaissance partielle du phénomène étudié : hypothèse de monotonie
- ▶ permet d'encadrer les probabilités de façon sûre à 100%
- ▶ obtenir une délimitation fine de l'espace des entrées menant à l'événement indésirable
- ▶ on considère maintenant que G renvoi une valeur binaire $-1 \setminus +1$

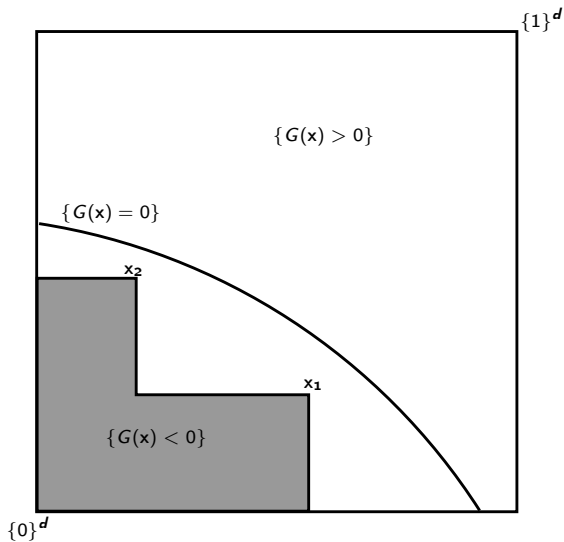
Principe général 1/5



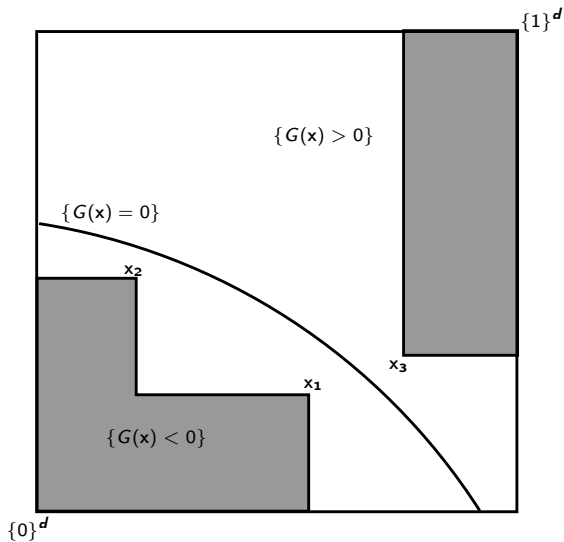
Principe général 2/5



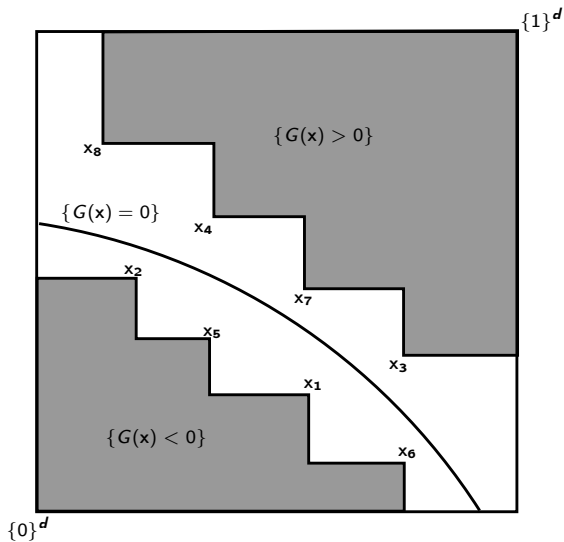
Principe général 3/5



Principe général 4/5



Principe général 5/5



Première approche: tirage uniforme

- ▶ A l'étape n , on simule \mathbf{x}_n uniformément dans l'espace non dominé [Bousquet, 2012]

$$\xi_{\mathbf{x}_n} = \mathbb{1}_{\{G(\mathbf{x}_n) \leq 0\}} \sim \text{Ber} \left(\frac{P_f - p_{n-1}^-}{p_{n-1}^+ - p_{n-1}^-} \right)$$

- ▶ l'estimateur du maximum de vraisemblance conditionnel \hat{p}_n^u maximise

$$L_n(p) = \prod_{k=1}^n \left(\frac{p - p_{k-1}^-}{p_{k-1}^+ - p_{k-1}^-} \right)^{\xi_{\mathbf{x}_k}} \left(\frac{p_{k-1}^+ - p}{p_{k-1}^+ - p_{k-1}^-} \right)^{1 - \xi_{\mathbf{x}_k}}$$

et l'information de Fisher associée est égale à

$$J_n(P_f) = \sum_{k=1}^n \mathbb{E} \left[\left[(P_f - p_{k-1}^-) (p_{k-1}^+ - p) \right]^{-1} \right]$$

sous certaines conditions théoriques:

$$J_n^{1/2}(P_f)(\hat{p}_n^u - P_f) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

Tirage d'importance séquentiel

- ▶ On propose de revenir dans l'espace gaussien: $\mathbb{U} = \mathbf{R}^d$ et $\mathbf{X} \sim \mathcal{N}_d(0, I_d)$
- ▶ à l'étape i , on pose $\mathbf{V}_i \sim \mathcal{N}_d(\mathbf{x}_i, I_d)$ de fonction de densité \tilde{f}_{i-1} et $\mathbf{x}_i \in \mathbb{U}_{i-1}$

- ▶ on estime P_f par $\hat{p}_{n,1} = \frac{1}{n} \sum_{i=1}^n \frac{f_{\mathbf{X}}(\mathbf{V}_i)}{\tilde{f}_{i-1}(\mathbf{V}_i)} \xi_{\mathbf{V}_i}$

- ▶ on a $\frac{(\hat{p}_{n,1} - P_f)}{\sqrt{v_{n,1}}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$ avec

$$v_{n,1} = \frac{1}{n^2} \sum_{i=1}^n \text{Var} \left(\frac{f_{\mathbf{X}}(\mathbf{V}_i)}{\tilde{f}_{i-1}(\mathbf{V}_i)} \xi_{\mathbf{V}_i} \right)$$

Tirage d'importance séquentiel

▶ Mais à l'étape i , $\mathbb{P}(\mathbf{V}_i \notin \mathbb{U}_{i-1}) > 0$. Si $\mathbf{V}_i \notin \mathbb{U}_{i-1}$ alors $\xi_{\mathbf{V}_i}$ est connu sans faire d'appel au code numérique

▶ On pose $T_0 = 0$ et pour tout $k \geq 1$, $T_k = \inf \{j \geq 1 : \mathbf{V}_j^{(k-1)} \in \mathbb{U}_{T_{k-1}}\}$

▶ pour un budget de n appels au code, on a donc fait

$$N_n = \sum_{k=1}^n T_k \text{ simulations.}$$

▶ on propose d'estimer P_f par

$$\hat{p}_{n,2} = \frac{1}{N_n} \sum_{k=1}^n \sum_{j=1}^{T_k} \frac{f_{\mathbf{X}}(\mathbf{V}_j^{(k-1)})}{\tilde{f}_{k-1}(\mathbf{V}_j^{(k-1)})} \xi_{\mathbf{V}_j^{(k-1)}}$$

Résultats numériques

▶ Un exemple test en dimension d . Soit $\mathbf{X} = (X_1, \dots, X_d)$ avec $X_i \sim \Gamma(i + 1, 1)$

▶ on note

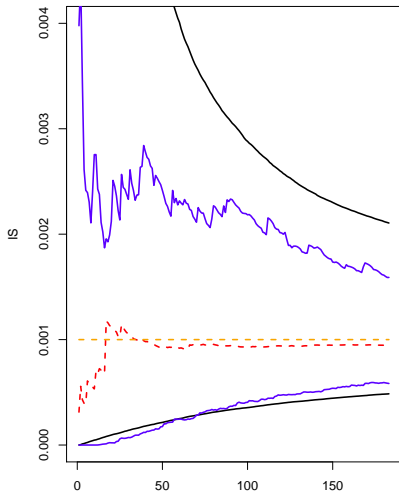
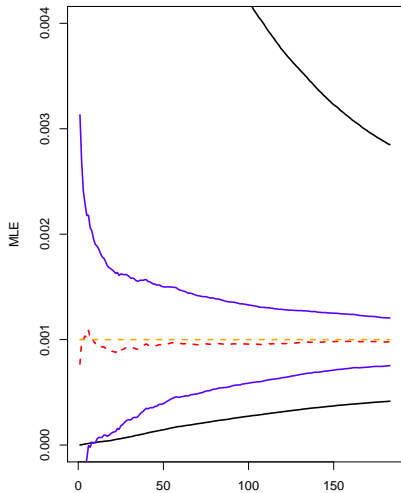
$$Z_d = \frac{X_1}{\sum_{i=1}^d X_i} \sim \text{Beta}(2, (d + 1)(d + 2)/2 - 3)$$

▶ $G(\mathbf{X}) = Z_d - q_{d, P_f}$ où $q_{d, p}$ est le quantile d'ordre p de Z

▶ on cherche à estimer $P_f = \mathbb{P}(G(\mathbf{X}) \leq 0)$

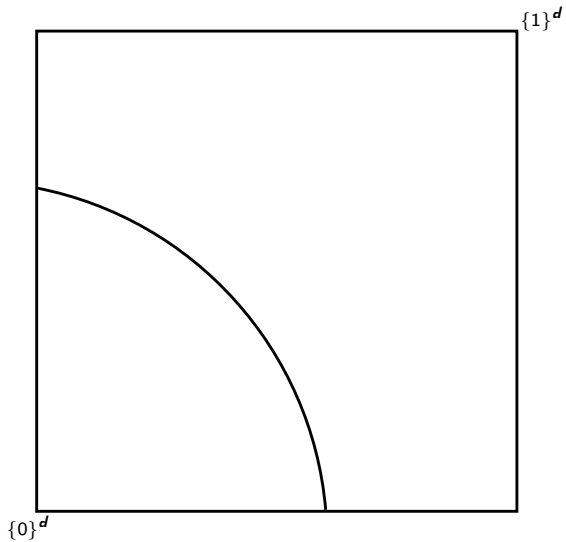
Méthode d'estimation de probabilité - $d = 3$;

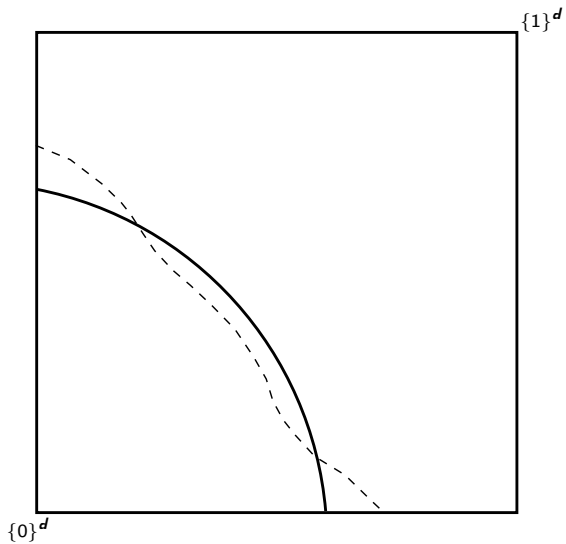
$$P_f = 10^{-3}$$

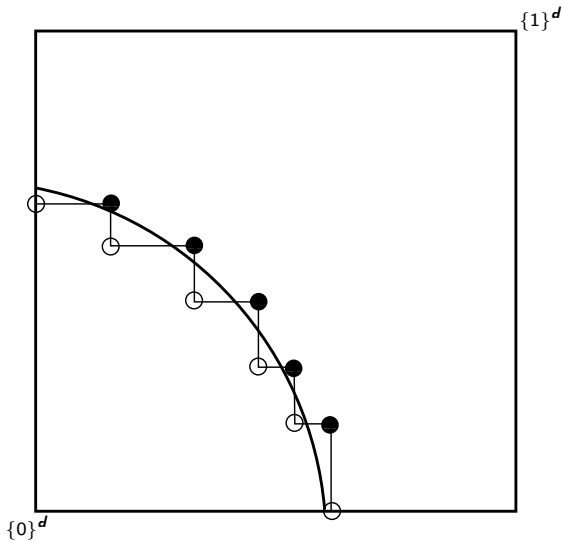


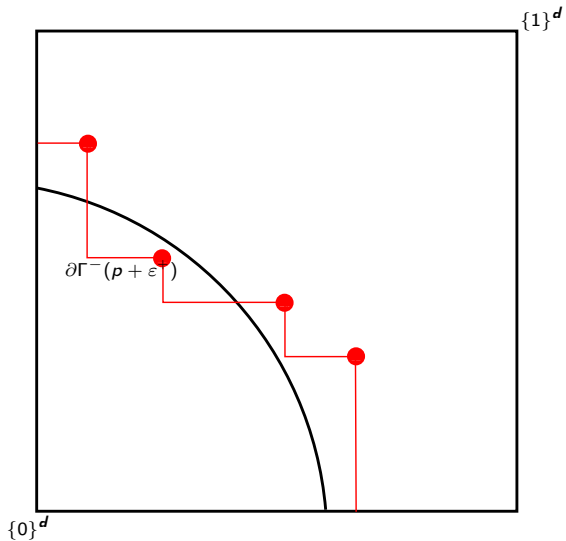
Encadrement de quantiles

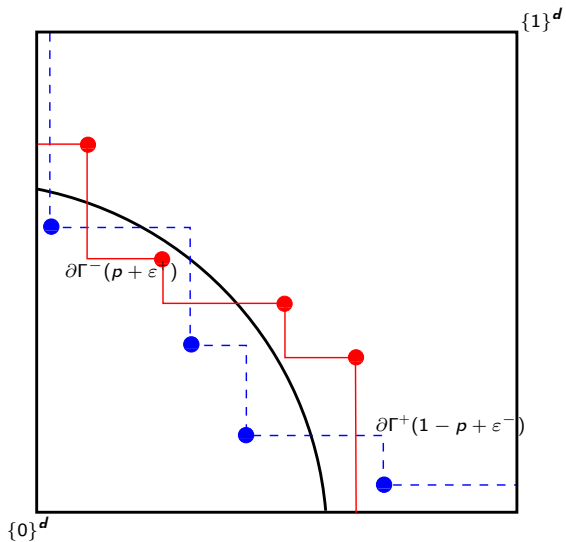
- ▶ Etude connexe à l'estimation de probabilité d'événement indésirable
- ▶ pour une valeur de P_f fixée on cherche à estimer le seuil q_{P_f} définissant une enveloppe de sécurité
- ▶ comme pour l'estimation de probabilité on peut obtenir deux bornes sûres à 100% pour q_{P_f}
- ▶ travaux en cours: méthode géométrique











Résultats numériques

- ▶ Un un exemple test en dimension d . Soit $\mathbf{X} = (X_1, \dots, X_d)$ avec $X_i \sim \Gamma(i + 1, 1)$
- ▶ on note

$$G(\mathbf{X}) = \frac{X_1}{\sum_{i=1}^d X_i} \sim \text{Beta}(2, (d + 1)(d + 2)/2 - 3)$$

- ▶ soit $q_{d,p}$ le quantile d'ordre p de $G(\mathbf{X})$
- ▶ soit P_f la probabilité de l'événement indésirable
- ▶ on cherche à estimer $q_{P_f} = \inf_{y \in \mathcal{R}} (\mathbb{P}(G(\mathbf{X}) \leq y) > P_f)$

Résultats numériques

$n = 1000, P_f = 10^{-2}$			
Dimension	q_n^-	q_n^+	q_{P_f}
2	-	0.44927246	0.04199864
3	-	0.18560746	0.01965787
4	-	0.23161791	0.01182391

$n = 1000, P_f = 10^{-3}$			
Dimension	q_n^-	q_n^+	q_{P_f}
2	-	0.03320955	0.01302295
3	-	0.192127468	0.006048936
4	-	0.195553283	0.003628516

Références

- [1] Bousquet N. (2012). Accelerated monte carlo estimation of exceedance probabilities under monotonicity constraints. *Annales de la Faculté des Sciences de Toulouse*. XXI(3), 557 – 592
- [2] De Rocquigny, E. (2009). Structural reliability under monotony: A review of properties of form and associated simulation methods and a new class of monotonous reliability methods (MRM). *Structural Safety* 31, 363 – 374
- [3] Hall P. & Heyde C. C. (1980). *Martingale limit theory and its application*. Academic press New York.
- [4] Kroese D. & Rubinstein R. (2007). *Simulation and the Monte Carlo Method* (Wiley ed.)
- [5] Lemaire M. (2009). *Structural reliability*. Wiley
- [6] Madsen H. & Ditlevsen O.(1996). *Structural reliability methods* (Wiley ed.)
- [7] V. Moutoussamy, N. Bousquet, B. looss, P. Rochet & F. Gamboa (2013). Comparing conservative estimations of an exceedance probabilities using sequential designs of experiments. ICOSAR.