PoD-Curves, Sensitivity Analysis and Kriging

BROWNE Thomas
Université Paris-Descartes, Paris, France ; EDF Lab Chatou-MRI, Chatou, France

Supervisor(s): Pr. J-C FORT (Université Paris-Descartes), Dr. B. IOOSS (EDF Lab Chatou-MRI & IMT Toulouse) and Dr. L. LE GRATIET (EDF Lab Chatou-MRI)

Ph.D. expected duration: 2014-2017

Adress: EDF Lab Chatou-MRI, 6 quai Watier, 78401 Chatou, France

Email: thomas.ga.browne@gmail.com

Abstract:

Probability of Detection-curves (PoD-curves) are standard tools to evaluate the performance of Non Destructive Testing (NDT) procedures. A new definition for PoD-curves, through random cumulative distribution functions (rcdf), is introduced in this talk. Sensitivity indices are also defined in order to assess the respective influence of the involved parameters over the quality of detection. Besides Kriging estimates, based on the first approaches in [2], are introduced.

1 Probability of Detection Curves

The mathematical framework for defect detection consists in: $Y \in \mathbb{R}$: measure of the signal obtained after a NDT procedure which is function of the following quantities; $a > 0$, size of defect; $X \in E \subset \mathbb{R}^d$, all the other influential parameters over Y such as the structure's conductivity, permeability. X is an independent random vector; δ, the observation noise (in the case of a regression over Y, δ also includes the model error); t_s, the signal threshold so that $Y(a, X, \delta) > t_s$ means "the defect is detected". The random PoD-curve writes: $\forall a > 0 \quad \pi_X(a) := \mathbb{P}\left(Y(a, X, \delta) > t_s \mid X\right)$. With the following assumption: $(a \to Y(a, X, \delta))$ is increasing a.s., so is π_X. Therefore, π_X is identical to a random cumulative distribution function (rcdf) as we see in Figure 1.

2 PoD-Curves through Contrast Functions

Contrast functions offer a supplementary definition for probability features. Indeed, given φ a positive convex function, one can get: $\theta^* := \underset{\theta \in \mathbb{R}}{\arg\min}\, \mathbb{E}\left[\varphi\left(Y - \theta\right)\right]$, providing that such θ^* exists and is unique. The physical meaning of θ^* is induced by the choice of φ, as: if $\forall y, \theta \in \mathbb{R}$
$\varphi(y-\theta) := m(y-\theta) = (y-\theta)^2$, then $\theta^* := \mathbb{E}[Y]$; for any $\alpha \in]0, 1[$, if $\forall y, \theta \in \mathbb{R}$,
$\varphi(y-\theta) := c_\alpha(y-\theta) = (\alpha - \mathbf{1}_{y \leq \theta}) . (y-\theta)$, then $\theta^* := q^\alpha(Y)$, the α-quantile of Y. Based on the 2-Wasserstein distance, which is relevant to compare two probability distributions (or cdf's), let us introduce cdf-contrast functions, for F, G two cdf's:

$$\psi_\varphi(G - F) := \underset{(X \sim F,\ Y \sim G)}{\arg\min}\ \mathbb{E}\left[\varphi\left(Y - X\right)\right]. \tag{1}$$

The Cambanis theorem states that the couple, (X, Y), that minimizes the previous quantity is $\left(F^{-1}(U), G^{-1}(U)\right)$, with $U \sim \mathcal{U}([0,1])$. One can notice that cdf-contrast functions depend on the choice for the real contrast function, φ. Since the probability feature, θ^*, of a real random variable is introduced as the real value that minimizes the mean "distance", regarding φ, to Y, the rcdf-probability feature F^* is the cdf that minimizes the mean "distance", regarding ψ_φ, to the

rcdf, π_X: $F^* := \underset{F \text{ cdf}}{\arg\min} \mathbb{E}\left[\psi_\varphi(\pi_X - F)\right]$. One gets to the definition of the PoD-mean, $\mathcal{E}(\pi_X)$ (resp. α-PoD-quantile, $\mathcal{Q}^\alpha(\pi_X)$) by substituting φ by m (resp. c_α) in (1). The pointwise expressions, that come through their reciprocal functions for $\mathcal{E}(\pi_X)$ and $\mathcal{Q}^\alpha(\pi_X)$, write:
$\forall u \in]0,1[\ \mathcal{E}(\pi_X)^{-1}(u) = \mathbb{E}\left[\pi_X^{-1}(u)\right]$ and $\mathcal{Q}^\alpha(\pi_X)^{-1}(u) = q^\alpha\left(\pi_X^{-1}(u)\right)$. $\mathcal{E}(\pi_X)$ denotes the quality of detection that is the most in the middle while $\mathcal{Q}^\alpha(\pi_X)$ denotes the quality of detection that is worst than only $\alpha\%$ of all realizations. We display these curves in Figure 1.

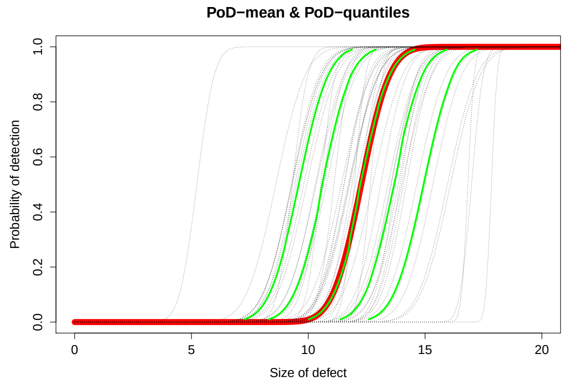

Figure 1: 50 realizations $(\pi_{x_i})_i$ of π_X in black dotted-lines, PoD-mean in red thick line and PoD-quantiles in green lines for $\alpha = 0.1, 0.25, 0.5, 0.75, 0.9$..

3 Sensitivity Analysis for PoD-Curves

Let us recall the sensitivity indices with respect to a contrast [1], for any $i \in \{1, \ldots, d\}$:
$S_\varphi^{X_i}(Y) = \underset{\theta\in\mathbb{R}}{\arg\min}\ \mathbb{E}\left[\varphi\left(Y-\theta\right)\right] - \mathbb{E}\left[\left(\underset{\theta\in\mathbb{R}}{\arg\min}\ \mathbb{E}\left[\varphi\left(Y-\theta\right) \mid X_i\right]\right)\right]$, which quantify the variability of the conditional probability feature $\theta(Y \mid X_i) := \underset{\theta\in\mathbb{R}}{\arg\min}\ \mathbb{E}\left[\varphi\left(Y-\theta\right) \mid X_i\right]$. In this talk we extend the previous indices to cdf-indices with respect to a contrast by using cdf-contrasts (1):

$$\mathcal{S}_\varphi^{X_i}(\pi_X) = \underset{F \text{ cdf}}{\arg\min}\ \mathbb{E}\left[\psi_\varphi\left(\pi_X - F\right)\right] - \mathbb{E}\left[\left(\underset{F \text{ cdf}}{\arg\min}\ \mathbb{E}\left[\psi_\varphi\left(\pi_X - F\right) \mid X_i\right]\right)\right]. \tag{2}$$

We show that (2) quantify the variability of the conditional PoD-mean, $\underset{F \text{ cdf}}{\arg\min}\ \mathbb{E}\left[\psi_\varphi\left(\pi_X - F\right) \mid X_i\right]$ for $\varphi = m$ (resp. conditional α-PoD-quantile, for $\varphi = c_\alpha$).

References

[1] J-C. Fort, T. Klein, and N. Rachdi. New sensitivity analysis subordinated to a contrast. *Communication in Statistics : Theory and Methods*, 2016.

[2] L. Le Gratiet, B. Iooss, G. Blatman, T. Browne, S. Crodeiro, and B. Goursaud. Model Assisted Probability of Detection Curves: New statistical tools and progressive methodology. *Journal of Nondestructive Evaluation*, In Press, 2017.

Short biography – I graduated from Université Paris Descartes in 2014 with a MSc in Probability and Statistics . I started my PhD at EDF Lab Chatou in November 2014 which is about building PoD-curves for non-destructive evaluation .