Rémi LANDRI[a,b,#], Gilles DUCHARME[a], Roseline BONNARD[b]

[a] Laboratory of Probability and Statistics, Montpellier II University, 34095 Montpellier, FRANCE
[b] National institute for the industrial environment and risks (INERIS), Parc Technologique Alata, BP 2, 60550 Verneuil-en-Halatte, FRANCE
[#] landri@math.univ-montp2.fr

INERIS
maîtriser le risque | pour un développement durable |

ENTROPY-BASED SENSITIVITY ANALYSIS IN HEALTH RISK ASSESSMENT

Entropy as sensitivity analysis measure

- Human exposure to environmental pollution and its associated adverse health consequences are recognized as posing major threats to occupational and public health. For realistic decision-making in risk management, the reliability and quality of exposure and risk model outputs, which depend on uncertainty, must be known. Usually, output uncertainty is described in terms of variance. Variance is popular in sensitivity analysis because of its simplicity and its historical development; statisticians also use it as a reference measure of the dispersion.
- Other measures may be used to characterize the uncertainty of a model output *Y* related to input variables. We criticize the use of variance *Var*(*Y*) as a measure of output uncertainty and proposed to use entropy, $H(Y)=-\int_{IR} f(y)\log f(y)dy$. Entropy is an information criterion which measures the amount of uncertainty and information content that is implied by a probability distribution.

Entropy can be prefered to variance?

- Unlike the variance, entropy has the advantage of depending on many more parameters than just the 2[nd] moment which allows only to measure a dispersion around the mean. Consequently, entropy depends on much more information about a vector of random variables than its variance.
- « Variance is finite » (well defined and not infinite) implies « entropy is finite » but the converse may not hold.
- To justify the use of entropy in sensitivity analysis, we examine the role of variance and entropy in ordering distributions in univariate, bivariate and multivariate case.

How to compare entropy and variance in dimension 2 and superior ?

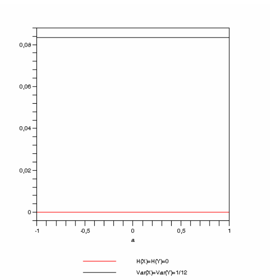
- The entropy is a scalar measure of a distribution. On the contrary, the variance approach in dimension 2 or more leads to a matrix of variance-covariance *Cov*(.), and comparisons with entropy must repose on summary measures, such as the generalized variance, the determinant |*Cov*(.)|, or the total variance, the trace *tr*(*Cov*(.)), of the matrix *Cov*(.).

Univariate distributions : Ratio of two normal variables

- Let $X \sim N(0,\sigma_X)$ and $Y \sim N(0,\sigma_Y)$ two independent normal variables. The ratio $Z = X/Y$ follows a Cauchy distribution *Cauchy* $(0,\sigma_X/\sigma_Y)$ where the primer parameter is a location parameter and the second σ_X/σ_Y is a scale parameter. The Cauchy distribution has no mean, no variance nor higher moments.
- However, the entropy is well defined ($H(Z) = \log(4\pi\sigma_X/\sigma_Y)$) and depends only on the two normal law variances. Equally, H(*X,Y*) is a scalar value : $H(X,Y) = 1+\log(2\pi)+\log(\sigma_X/\sigma_Y)$**.**
- Consequently, variance does not permit to compare vector (*X,Y*) and variable ratio *X* /*Y* whereas entropie does.
- Other example, the variance of the T-Student distribution for some degrees of freedom does not exist, while its entropy always does.

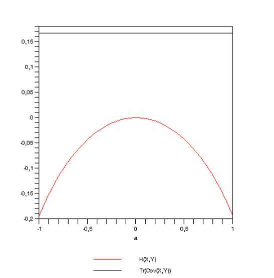
Bivariate distributions : Farlie-Gumbel-Morgenstern copula

- Dependence functions are used to construct joint distributions with fixed marginals. A bivariate copula *c*(*X,Y*) is a bivariate probability density function defined on the unit square $[0,1]^2$ (each marginal law is a uniform *U*[0,1] dependent of the other) which connects marginal distributions in a way giving as a result a joint probability. It has the advantage of characterizing, mixing and modelling the whole dependence structure between random variables, whatever the marginal distributions.
- We compare graphically marginal, conditional and bivariate entropies and (total or generalized) variances of copulas. We take there the Farlie-Gumbel Mogenstern (FGM) copula $c(X,Y) = 1+a(2x-1)(2y-1)$ with *a* in [-1,1] in terms of entropy or variance reduction, this copula is symetric in *x* and *y*.



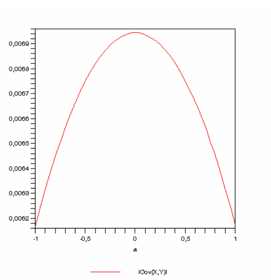
FGM marginal entropies and variances

- Marginal entropies *H*(*X*), *H*(*Y*) and variances *V*(*X*), *V*(*Y*) are constant for all value of the parameter *a*.



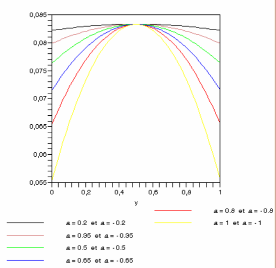
FGM bivariate entropy and total variance

- Bivariate entropy *H*(*X,Y*) varies with the parameter *a* whereas the trace *tr*(*Cov*(*X,Y*)) is constant.



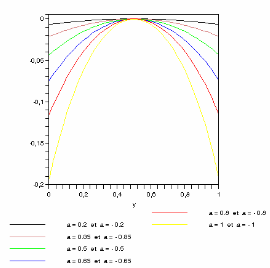
FGM generalized variance

- Likewise, |*Cov*(*X,Y*)| increases when *a* is in [-1,0] and decreases when *a* is in [0,1] in the same manner than H(*X,Y*).



FGM conditional variance

- The two conditional variances *Cov*(*X*|*Y*) and *Cov*(*Y*|*X*) rises for *a* in [-1,0] and falls for *a* in [0,1], in the same manner than |*Cov*(*X,Y*)|.



FGM conditional entropy

- The two conditional entropies *H*(*X*|*Y*) and *H*(*Y*|*X*) behave likewise to conditional variances.

- Entropy and variance orderings of the FGM copula hold (only the trace has a different behavior than bivariate entropy).
- So, we can class entropy and variance similarly for the Farlie-Gumbel-Morgenstern copula.

a	-1 0 1
$\mathbb{V}ar(X), \mathbb{V}ar(Y), tr(\Sigma_{\mathbf{XY}}), \mathbb{H}(X), \mathbb{H}(Y)$	$\longrightarrow$ $\longrightarrow$
$\lvert\Sigma_{\mathbf{XY}}\rvert, \mathbb{H}(X,Y), \mathbb{V}ar(X\|Y), \mathbb{V}ar(Y\|X), \mathbb{H}(X\|Y), \mathbb{H}(Y\|X)$	$\nearrow$ $\searrow$

Entropy and variance rankings

Multivariate distributions : *n*-variate normal law

- Noting $\boldsymbol{\mu}_p = (\mu_1 \mu_2,\ldots,\mu_p)$ and $\xi_p = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1p} \\ \sigma_{21} & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ \sigma_{p1} & \ldots & \ldots & \sigma_{pp} \end{pmatrix}$.
- Let $\boldsymbol{X} \equiv (X_1, X_2,\ldots,X_p) \sim \boldsymbol{N}(\boldsymbol{\mu}_p, \xi_p)$ and $\boldsymbol{Y} \equiv (Y_1, Y_2,\ldots,Y_{p-m}) \sim \boldsymbol{N}(\boldsymbol{\mu}_{p-m}, \xi_{p-m})$ where $p > m > 1$.
- We want to compare the effect of dimension reduction from $\boldsymbol{X}$ to $\boldsymbol{Y}$ in terms of entropy and variance.
- The matrix ξ_p can be rewritten as $\xi_p = \begin{pmatrix} \xi_{p-m} & \boldsymbol{\sigma}_{12}^* \\ \boldsymbol{\sigma}_{21}^* & \boldsymbol{\sigma}_{22}^* \end{pmatrix}$ and we note $\alpha = \left|\boldsymbol{\sigma}_{22}^* - \boldsymbol{\sigma}_{21}^* \xi_{p-m}^{-1} \boldsymbol{\sigma}_{12}^*\right|$.
- We show that : (1) $H(\boldsymbol{X}) \geq H(\boldsymbol{Y}) \Leftrightarrow \alpha \geq \exp(-m - m\log(2\pi))$ and (2) $Var(\boldsymbol{X}) \geq Var(\boldsymbol{Y}) \Leftrightarrow \alpha \geq 1$
- As $\alpha \geq \exp(-m - m\log(2\pi))$ decreases rapidly as m increases, we conclude that $H(\boldsymbol{X}) \geq H(\boldsymbol{Y})$ is almost always verified.
- Entropy reduction is obtained as opposed to reduction variance.

Conclusion and perspectives :

- Entropy captures uncertainty and allows to oppose « Entropy against variance ».
- To motivate the use of entropy, we explore and compare its role in ranking distributions from univariate to multivariate cases via copulas. We present three typical examples.
- Entropy exists for the Cauchy distribution, contrarily to the variance. In terms of reduction of measure, the concept of entropy reduction is more meaningful than variance reduction in the multivariate normal distribution when we reduce the dimension. Results on variance and entropy for the Farlie-Gumbel-Morgenstern copula do not permit the same conclusion.
- Extension and generalization to the multivariate case is a challenging research topic that needs to be explored to promote entropy in sensitivity analysis.