Reconstruire le cadre bayésien de raisonnement statistique

De la subjectivité personnelle au projet d'ingéniérie

Éric Parent, Ing. Gen. Ponts Eaux & Forêts

IHP, 23 Septembre 2022

Équipe Solstis, MIA INRAe/AgroParisTech Paris-Saclay

Mac Grayne (2012): *The theory that would not die*

- Jacques Bernoulli (1654-1705) : Première démonstration de la loi des grands nombres ($\bar{X}$ converge presque sûrement vers $\mathbb{E}(x)$).
- Abraham de Moivre (1667-1754) : Montre la convergence de la loi binomiale vers la loi normale

Causes connues $\rightarrow$ Effets attendus (approche probabiliste)

- Thomas Bayes (1701-1761) et Pierre-Simon de Laplace (1749-1827) : Probabilités inverses; Causes inconnues à partir d'événements observés (approche statistique); on connaît la fréquence des boules tirées d'une urne et on voudrait connaître la composition inconnue de cette urne.
- Turing : le *machine learning* = l'inférence bayésienne
- Wald : les seuls estimateurs *intéressants* (non dominés) sont les estimateurs bayésiens (et leurs limites)
- Definetti, puis Savage : interprétation subjective et existence mathématique du prior et de la vraisemblance.
- puis le *Bayesian computation* : a statistical revolution (Brooks, 2003)...

Présentation pratico-pratique

Aujourd'hui : la nature de la probabilité, *épistémique*, de *ressemblance* ou de variabilité *expérimentale* est une question philosophique sans intérêt.

- Approche classique $\hat{\theta} = ArgMax[y|\theta]$
- Approche bayésienne

$$[\theta|y] \propto [y|\theta] \times [\theta] \tag{1}$$

$$[Y_{new}|y_{obs}] = \int_{\theta} [Y_{new}|\theta] \times [\theta|y_{obs}] d\theta \tag{2}$$

Un enseignement que je ne fais jamais : prendre le temps de réfléchir moins loin que l'arsenal technique et remettre à plus tard les joies du MCMC...

Je vais me lancer,... Good Luck!

Programme

- Comment toute incertitude, en tant qu'anticipation personnelle rationelle, se mesure par des probabilités.
- Prior et Vraisemblance, sous-concepts utiles *engendrés* par la considération de symétrie minimale (variables échangeables).
- L'aide à la décision, souvent prônée , rarement mise en oeuvre.
- La prévision probabiliste, interprétation et challenges.

Quantifier l'incertitude par des probabilités *personnelles*

Première approche de la subjectivité contrôlée

La crédibilité, p, d'un événenement A, est le *prix* p euros que *vous* considérez équitable pour le pari qui *vous* donne 1 euro si A se produit, et 0 euro sinon.

Sous des hypothèses:

- d'un marché intra-personnel de paris où prix de vente$=$ prix d'achat
- d'un comportement d'évitement des pertes sûres.
- sans déformation de préférence par aversion au risque.

Principles of uncertainty. J. B. Kadane. Chapman and Hall/CRC(2020).

Understanding uncertainty. D. V. Lindley. John Wiley & Sons (2013).

Principe d'évitement des pertes sûres = paris non arbitraires

Considérons une partition E_i, $i = 1, \dots, n$, avec mises S_i Le système des bilans de gains si E_j se réalise $S_j - \sum_i^{1:n} p_i S_i$ doit être un système linéaire *indéterminé.*

Idée de démonstration:

$$\begin{vmatrix} 1-p_1 & -p_2 & \dots & -p_r \\ -p_1 & 1-p_2 & \dots & -p_r \\ \dots & \dots & \dots & \dots \\ -p_1 & -p_2 & \dots & 1-p_r \end{vmatrix} = 1 - \sum p_i = 0$$

La notation p, qui signifait *prix*, refère maintenant à *probabilité*.

Poursuite d'élaboration de la subjectivité contrôlée

La crédibilité conditionnelle, p, d'un événenement A subordonnée à l'événement B, est le *prix* p euros que *vous* considérez équitable pour un pari tel que:

- si B ne se produit pas, le pari est annulé (et vous récupérez votre mise)
- si B se produit: *vous* recevez 1 euro si A se produit aussi (AB se produit), et 0 euro sinon (si $\bar{A}B$)

Evitement des pertes sûres = paris non arbitraires

Idée de démo: Bilan d'une mise S sur le pari événement A subordonné à B

- Si AB: $(1-p_{AB})S_{AB} + (1-p_B)S_B + (1-p)S$
- Si $\bar{A}B$: $-p_{AB}S_{AB} + (1-p_B)S_B - pS$
- Si $A\bar{B}$: $(1-p_{AB})S_{AB} - p_B S_B - 0$

La cohérence implique l'indétermination de ce système linéaire, soit:

$$\begin{vmatrix} 1-p_{AB} & 1-p_B & 1-p \\ -p_{AB} & 1-p_B & -p \\ -p_{AB} & -p_B & 0 \end{vmatrix} = p_{AB} - p \times p_B = 0$$

Conséquences : les crédibilités ont toutes les propriétés mathématiques d'une probabilité

- La construction peut être plus formelle, mais essentiellement, on a besoin d'un ensemble d'événements $\Omega \subset$ ensemble des parties de E contenant E lui-même, stable par complémentation, $\cup$ et $\cap$. Et on prend... *n'importe* quelle application $\mathbb{P}$ de l'ensemble des événements dans $[0, 1]$ $t.q.$

$$\mathbb{P}(A) \geq 0,$$
$$\mathbb{P}(E) = 1$$
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \; si \; A \cap B = \emptyset$$

- La crédibilité bayésienne donne une interprétation de la probabilité en terme de système de paris personnels cohérents
- La fréquence limite des occurrences d'un événement observable se comporte aussi selon ces exigences mathematiques: cf Jacques Bernoulli (1654-1705)

Moralité

- Les paris ont l'avantage de pouvoir être pris sur des événements *singuliers* .
- *Toutes* les incertitudes se mesurent à l'aulne du calcul de probabilités.
- Probabilité *personnelle*: expression d'une relation entre vous et l'état du monde. Proximité entre préférences, pari et décision. Comment fait-on pour arriver à un consensus d'experts?
- Concept plutôt intuitif en contexte prédictif
- Assez loin de la notion de modèle; une difficulté est que la tribu, plus que l'événement fait sens (cf paradoxe de Bertrand et Lindley, 2013)
- Unités de mesures probabilistes :
  - la cote(French): L'argent que vous recevez sur un pari gagnant en pariant 1 euro, i.e. $\frac{1}{p}$ euros.
  - *Odds against*(English): Le revenu net sur un pari gagnant en pariant 1 euro: $\frac{1}{p} - 1 = \frac{1-p}{p}$

Approches de même orthodoxie

Extension de la logique binaire en situation d'incertitude par construction d'une machine à logique inductive, ce robot calculateur devant obéir à cinq exigences.

- Cohérence : si la logique interne de cette machine permet de résoudre un problème de plusieurs façons, toutes les réponses doivent être identiques.
- Continuité : Le fonctionnement ne doit pas dépendre des chiffres utilisés dans le problème et répondre de façon continue à un changement marginal.
- Universalité : le robot calculateur doit pouvoir s'adapter à n'importe quelle catégorie de problème.
- Utiliser des énoncés sans équivoque.
- Ne pas refuser d'information.

The algebra of probable inference R. T. Cox and E. Jaynes. American Journal of Physics 31 66–67 (1963)

Rational Descriptions, Decisions and Designs. M. Tribus; Pergamon Unified Engineering Series(1972 puis 2016)

The Foundations of Decision under Uncertainty: An Elementary Exposition. J. W. Pratt, H. Raiffa, and R. Schlaifer; J.A.S.A. 59 (1964)

Hétérodoxie

- Ces jeux d'intellectuels mondains seraient-ils bien sérieux?
  Imagine-t-on évaluer un pari sur l'événement: *survenue d'une guerre nucléaire l'année prochaine*?
- Un nombre unique p serait insuffisant pour décrire notre incertitude:
  Prix de vente $\neq$ prix d'achat

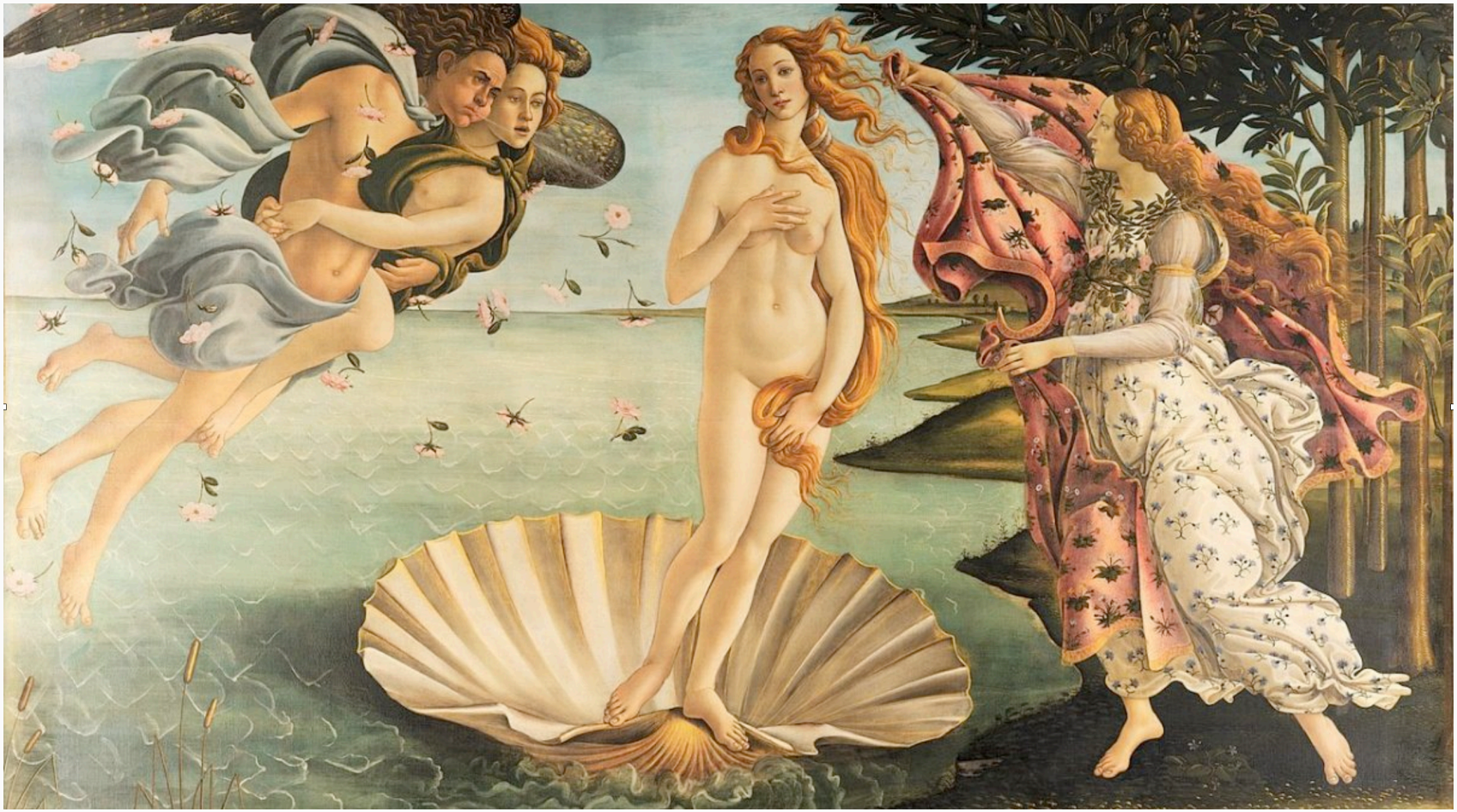
Statistical reasoning with imprecise probabilities. P. Walley (1991)

Decision making with imprecise probabilities: Dempster-Shafer theory and application. W. F. Caselton & W. Luo Water Resources Research 28 3071–3083.

Estimation de la valeur de l'information hydrologique à l'aide de probabilités imprécises.V. Fortin. These de l'INRS-Eau (1997)

Fuzzy rule-based modeling with applications to geophysical, biological, and engineering systems L. Duckstein & A Bardossy & J. Bogardi. CRC press (1995)

Echangeabilité



Symetrie minimale : Variables aléatoires échangeables

- Une séquence infinie $Y_1, Y_2, ...Y_n, ...$ de v.a. est dite *échangeable* si pour tout $n = 2, 3, ...,$ pour toute permutation τ de $\{1, ..., n\}$

$$[Y_1, Y_2, ...Y_n] = [Y_{\tau(1)}, Y_{\tau(2)}, ...Y_{\tau(n)}]$$

- Si $Y_1, Y_2, ...Y_n$ sont indépendantes et identiquement distribuées, elles sont échangeables, mais pas réciproquement.
- Processus de Polya : Tirer une boule depuis une urne. La remettre et en ajouter une de la même couleur. Continuer selon ce mode opératoire...
- Exemple continu: $Y_1, Y_2, ...Y_n$ normal centré avec matrice de covariance t.q.$\mathbb{E}(Y_i) = 0, \quad \mathbb{E}(Y_i^2) = 1$ et $\mathbb{E}(Y_i Y_j) = \rho, i \neq j$
- Application pratique: Thèse M Courbariaux *Contributions statistiques aux prévisions hydrométéorologiques par méthodes d'ensemble* avec P. Barbillon, A.C Favre et L. Perreault

Urne de Polya

Loi de probabilité :

$$[Y_1 = 1, Y_2 = 1, Y_3 = 0] = \frac{a}{a+b}\frac{a+1}{a+b+1}\frac{b}{a+b+2}$$

$$\frac{a}{a+b}\frac{b}{a+b+1}\frac{a+1}{a+b+2} = [Y_1 = 1, Y_2 = 0, Y_3 = 1]$$

$$\frac{b}{a+b}\frac{a}{a+b+1}\frac{a+1}{a+b+2} = [Y_1 = 0, Y_2 = 1, Y_3 = 1]$$

Urne de Polya, modèle échangeable

$$[Y_1 = 1, Y_2 = 1, Y_3 = 0] = \frac{a}{a+b}\frac{a+1}{a+b+1}\frac{b}{a+b+2}$$
$$\frac{a}{a+b}\frac{b}{a+b+1}\frac{a+1}{a+b+2} = [Y_1 = 1, Y_2 = 0, Y_3 = 1]$$

Une expression plus synthétique :

$$[Y_1, Y_2, ...Y_n] = \frac{\Gamma(a+b)\Gamma(a+\sum y_k)\Gamma(b+n-\sum y_k)\Gamma(n+1)}{\Gamma(a+b+n)\Gamma(n+1-\sum y_k)\Gamma(a)\Gamma(b)\Gamma(\sum y_k+1)}$$

Une propriété connue depuis longtemps (exercice clasique de calcul des probabilités) :

$$[Y_1, Y_2, ...Y_n] = \int_{\theta=0}^{1} \left(\prod_{i=1}^{n} dbinom(Y_i, 1, \theta)\right) dbeta(\theta, a, b)d\theta$$

Normal échangeable

$Y_1, Y_2, ...Y_n$ centrés t.q.

$$\Sigma = \begin{pmatrix} 1 & \rho & \rho & ... & \rho \\ \rho & 1 & \rho & ... & \rho \\ \rho & \rho & ... & ... & ... \\ ... & ... & ... & ... & \rho \\ \rho & \rho & \rho & ... & 1 \end{pmatrix}$$

$\rho > 0$ car $\rho > \frac{1}{n-1}, \forall n$

Une écriture plus élégante : $Y_i = \theta + \varepsilon_i \ , \ \ \varepsilon_i \perp \varepsilon_j, i \neq j$

$$[Y_1, Y_2, ...Y_n] = \int_{\theta=-\infty}^{+\infty} \left(\prod_{i=1}^{n} dnorm(Y_i, \theta, \sqrt{1-\rho}^2) \right) dnorm(\theta, 0, \rho) d\theta$$

NB: Existe aussi un modèle gamma échangeable, intéressant pour la représentation des membres de précipitations.

Mise en situation: Contrôle de la qualité par attribut

Vous êtes responsable qualité dans une usine de yaourts

n items échantillonnés,

$Y_k = 1$ si item k est défectueux;

Que proposer pour le tirage suivant de la collection

$$1110101110111111?$$

$Y_{16} = ?$

$\mathbb{P}(Y_{16} = 1) \geq \mathbb{P}(Y_{16} = 0)$ ou bien $\mathbb{P}(Y_{16} = 1) \leq \mathbb{P}(Y_{16} = 0)$?

Contrôle de la qualité par attribut

n items échantillonnés,

$Y_k = 1$ si item k est défectueux;

Que proposer pour le tirage suivant de la collection

$$000101000100000?$$

Contrôle de la qualité par attribut

n items échantillonnés,

$Y_k = 1$ si item k est défectueux;

la loi $[Y_1, Y_2, ...Y_n]$ *correspond* à des jugements probabilistes dépendants

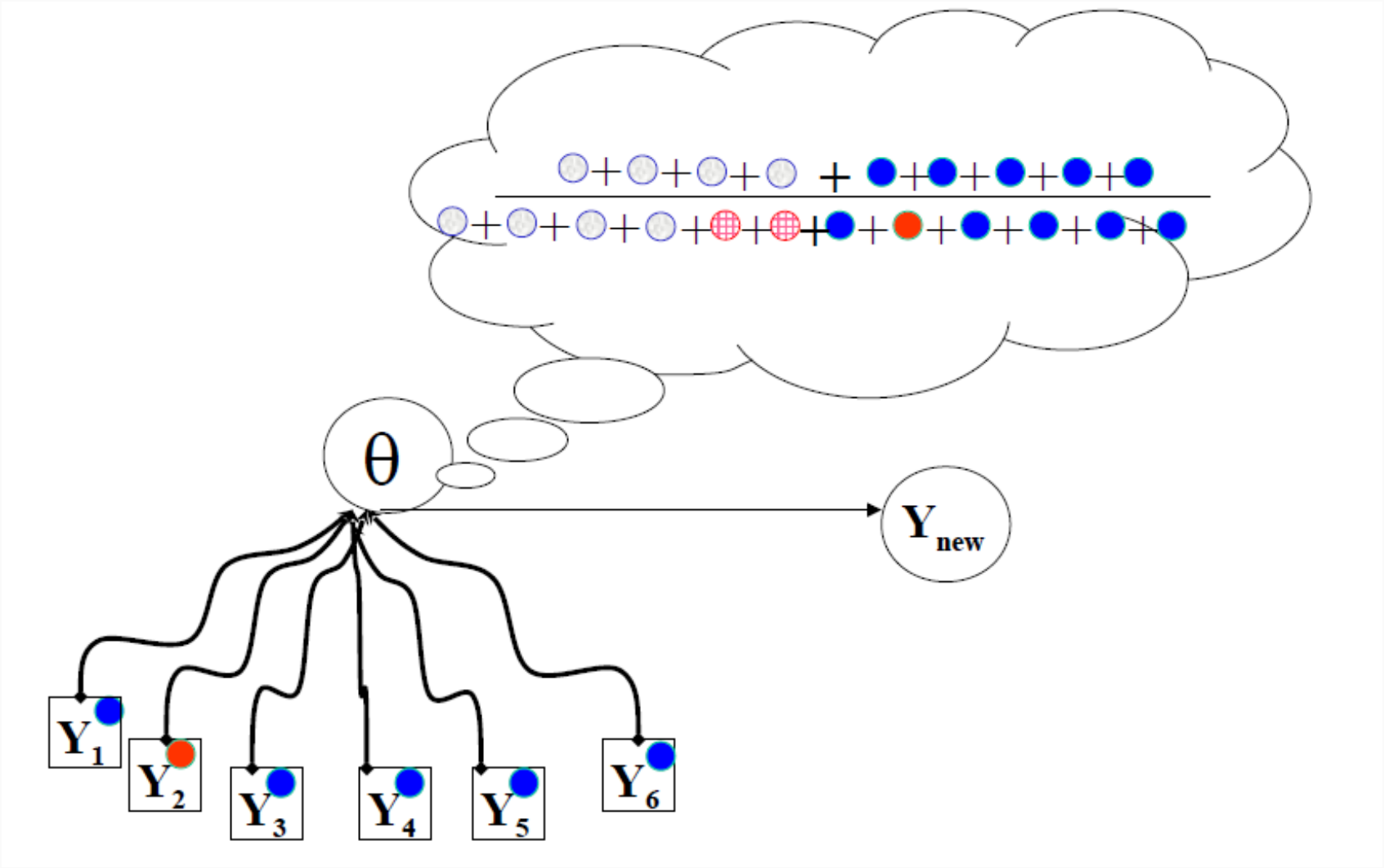
...

Puisque $\mathbb{P}(Y_{16} = 1) \geq \mathbb{P}(Y_{16} = 0)$ ou $\mathbb{P}(Y_{16} = 1) \leq \mathbb{P}(Y_{16} = 0)$ selon les valeurs de $Y_{1:15}$

Que cache notre raisonnement?

Mise en situation de probabilité personnelle prédictive

$Y_1 = y_1, Y_2 = y2, ...Y_n = y_n, Y_{n+1} = ?$



Théorème de représentation de DeFinetti (déconvolution)

$Y_1, Y_2, ...Y_n$ binaires et échangeables sont des mélanges de Bernoulli's: DeFinetti (1931)

$\exists G$, t.q.

$$[Y_1, Y_2, ...Y_n] = \int_{\theta=0}^{1} \left(\prod_{i=1}^{n} \theta^{y_i} (1-\theta)^{1-y_i} \right) dG(\theta)$$

$$\theta = \lim_{n \to +\infty} \frac{\sum_{k=1}^{n} Y_k}{n}, \quad [\theta < y] = G(y)$$

Pour une démonstration, voir:

- Bernardo & Smith, Bayesian Theory, Wiley, 1994, Chap 4 p 172-173, via permutations et théoreme de Helly
- Feller , Chap 7, p 219-230 via Bernstein polynômes de Bernstein, fonctions caractéristiques et martingales

Extensions du théorème de représentation de De Finetti

Hewitt et Savage (1955) : $Y_1, Y_2, ...Y_n$ échangeables si et seulement si, il existe une mesure de probabilité μ sur l'ensemble des mesures de probabilité pour les Y_i , t.q.

$$[Y_1 \in A_i, Y_2 \in A_2, ...Y_n \in A_n] = \int_{\mu} \left(\prod_{i=1}^{n} P(A_i) \right) \mu(dP)$$

$$F(A) = \lim_{n \to +\infty} \hat{F}_n(A) = \lim_{n \to +\infty} \frac{\sum_{k=1}^{n} 1_A(Y_k)}{n}, F \sim \mu()$$

$$[Y_1 \in A_i, Y_2 \in A_2, ...Y_n \in A_n | F = P] = \prod_{i=1}^{n} P(A_i)$$

L'échangeabilité engendre prior et vraisemblance

Le théorème de représentation en paramétrique

$Y_1, Y_2, ...Y_n$ échangeable et supposons que toutes les distributions aient des densités.

Il vient: $\exists G$, t.q. $\theta \sim G()$

$$[Y_1 = y_i, Y_2 = y_2, ...Y_n = y_n] = \int_\theta \left(\prod_{i=1}^{n} f_\theta(y_i) \right) g(\theta) d\theta$$

une valeur de θ correspond à la séquence infinie (et inconnue) $y_1, y_2, ..y_n, ...$ $i.e.$ la situation en *information parfaite*.

$$F_\theta(A) = \lim_{n \to +\infty} \hat{F}_n(A) = \lim_{n \to +\infty} \frac{\sum_{k=1}^{n} 1_A(y_k)}{n}$$

$$[Y_1 \in A_i, Y_2 \in A_2, ...Y_n \in A_n | \theta] = \prod_{i=1}^{n} F_\theta(A_i)$$

Le cadre bayésien se construit mathématiquement à partir du point de vue prédictif

$$[Y_1 = y_i, Y_2 = y_2, ...Y_n = y_n] = \int_\theta \left(\prod_{i=1}^{n} f_\theta(y_i) \right) g(\theta) d\theta$$

- *Vraisemblance* ($f_\theta(y_i)$) et loi *a priori* ($g(\theta)$), co-existent dans un raisonnement prédictif probabiliste. Les deux concepts trouvent des existences mathématiquement prouvées.
- Le travail du statisticien (bayésien) est de fournir des structures pour modéliser la *vraisemblance* ($f_\theta(y_i)$) et le *prior* ($g(\theta)$), les deux constituant le modèle bayésien. Cf. exposés de P. Gloaguen et M Authier

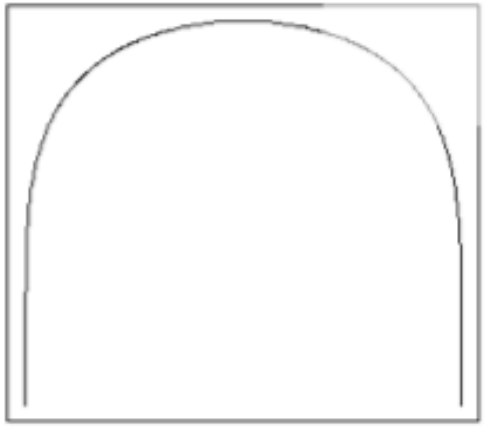
La pratique clame l'existence du prior

Les données ne suffisent pas à elles seules,
le prior *existe*

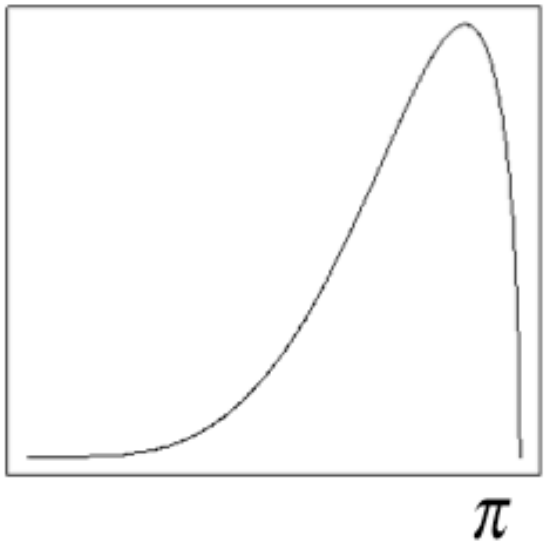
- Une buveuse de thé au lait, affirme qu'elle est capable de dire si c'est le lait ou le thé qui ont été versé en premier dans la tasse. Sur 6 essais, elle détermine correctement le premier ingrédient
- Un amateur de musique annonce qu'il est capable de distinguer Haydn de Mozart. Sur 6 morceaux écoutés, il trouve le bon auteur.
- Une nuit fort arrosée, votre collègue se prétendait devin. Lançant 6 fois une pièce en l'air, il prédit correctement les occurrences de pile ou face ?

Le prior existe selon la théorie...mais la pratique le confirme aussi

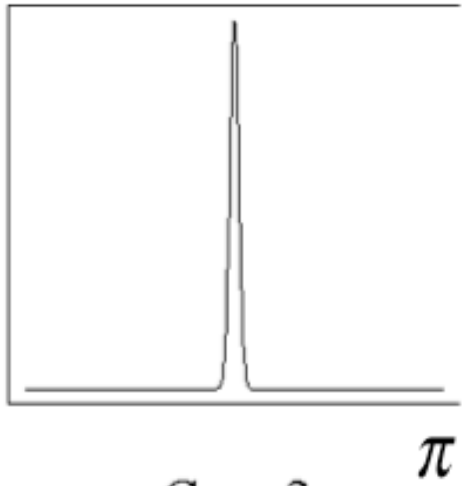
6 succès sur 6 essais mais...



Case 1 :
Lady drinking tea



Case 2 :
Music lover



Case 3 :
Flipping a coin

Le cadre bayésien se justifie mathématiquement si on formalise le point de vue prédictif

$$[Y_1 = y_i, Y_2 = y_2, ...Y_n = y_n] = \int_\theta \left(\prod_{i=1}^{n} f_\theta(y_i) \right) g(\theta) d\theta$$

- Les prédictions conditionnelles aux observations *doivent* utiliser la règle de Bayes pour mettre à jour la loi *a priori*:
  $[\theta] \rightarrow [\theta | y_1, y_2, ...y_n]$

$$[Y_{new} | y_1, y_2, ...y_n] = \frac{[Y_{new}, Y_1, Y_2, ...Y_n]}{[Y_1, Y_2, ...Y_n]}$$

$$= \int_\theta f_\theta(y_{new}) \left(\frac{g(\theta) \left(\prod_{i=1}^{n} f_\theta(y_i) \right)}{\int_\theta \left(\prod_{i=1}^{n} f_\theta(y_i) \right) g(\theta) d\theta} \right) d\theta$$

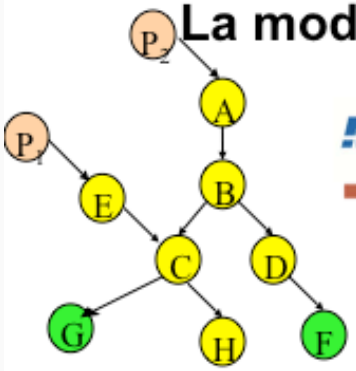
L'approche bayésienne d'un point de vue pratique

$$[Y_1 = y_i, Y_2 = y_2, ...Y_n = y_n] = \int_\theta \int_z \left(\prod_{i=1}^{n} [y_i|z, \theta][z|\theta] \right) [\theta] d\theta dz$$

- Conditionner sur des variables latentes z pour retrouver l'échangeabilité = modéliser
- La famille exponentielle offre nombre de propriétés avantageuses, ses éléments peuvent assemblés grâce à des modèles hiérarchiques.
- Les outils probabilistes (voir la révolution MCMC et autres avancées MC) sont maintenant disponibles pour le praticien.
- Stat. fréquentiste $\in$ Bayes+ information parfaite sur θ dispo. La perspective décisionelle bayésienne redonne d'ailleurs tous les estimateurs admissibles.
- Search analysis for the underwater wreckage of Air France Flight 447. L. D. Stone, C. M. Keller, T. M. Kratzke, and J. P. Strumpfer. 14th International Conference on Information Fusion 1–8 (2011)

La recherche de l'airbus Rio-Paris AF447 : Bayes pour assimiler la donnée

La modélisation hierarchique bayésienne en action



METRON
SCIENTIFIC SOLUTIONS
Mathematics Physics Computer Science

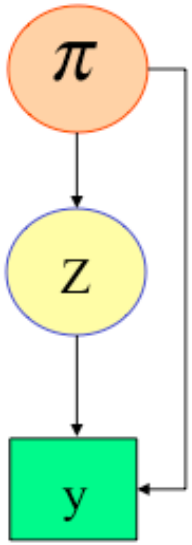
Modéliser : Conditionner par morceaux pour traduire la complexité

Paramètres, Etats du processus et mécanismes d'Observation sont modélisés indépendemment

Probabilités de détection et d'occurrence

Où se trouve vraiment l'avion

Les indices que l'on a observés



$[\pi]$ Paramètres

$[Z|\pi]$ Z: Process Latent

Ressemblance

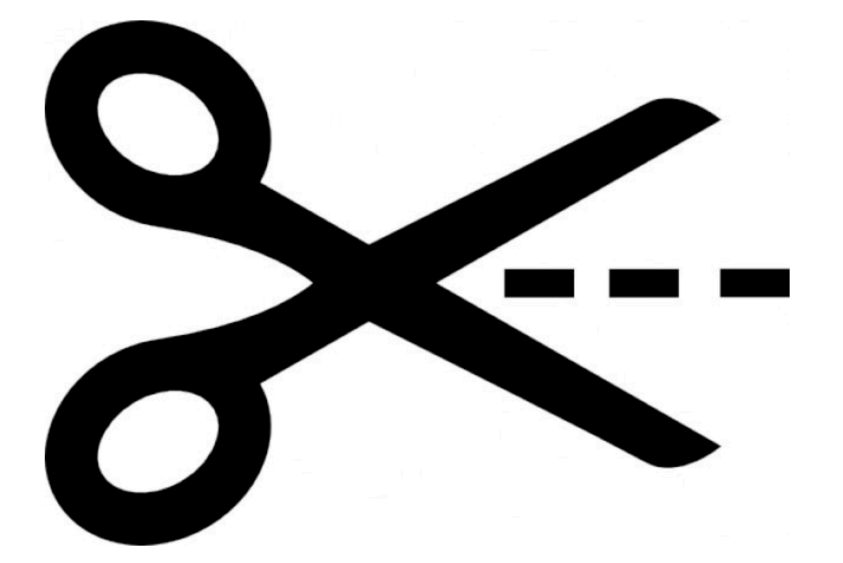
Individus, populations, temps, espace

$[Y|\pi, Z]$ Y: Observations

Rôle des variables cachées…

Modélisation [effets|causes]

Décider en situation d'incertitude



Décision en avenir incertain

- L'ensemble des décisions $D = \{d\}$
- L'ensemble des états de la Nature $\Theta = \{\theta\}$
- Une fonction de coût $W(d, \theta) : D \times \Theta \rightarrow R$
- Une variable aléatoire Y et un échantillon d' observations $\mathbf{y} = \{y_1, y_2, y_3, ..y_n\}$
- Une famille paramétrique de lois P_θ indicée par les états de la Nature Θ, $dP_\theta(Y = y) \equiv [y|\theta]dy$
- Une règle de décision $\delta : \mathbf{y} \longmapsto d = \delta(\mathbf{y})$
- Le risque statistique (Wald, 1950) $R(\delta, \theta) = \int_y W(\delta(y), \theta)[y|\theta]dy = E_{Y|\theta}\left(W(\delta(Y), \theta)\right)$
- Le risque bayésien (Ulmo&Bernier,1973) $\widehat{R(\delta)} = \int_y W(\delta(y), \theta)[y|\theta][\theta]dyd\theta = E_\theta E_{Y|\theta}\left(W(\delta(Y), \theta)\right)$

Décision statistique conventionnelle : les δ?

$$R(\delta, \theta) = \int_y W(\delta(y), \theta)[y|\theta]dy = E_{Y|\theta}\left(W(\delta(Y), \theta)\right)$$

- Test classique d' hypothèse si $D = \{d_{accept}, d_{reject}\}$; region critique $W = \delta^{-1}(d_{reject})$

Fonction de coût conventionelle

$W(d, \theta)$	$\theta \in H_0$	$\theta \notin H_0$
d_{accept}	0	1
d_{reject}	1	0

$$R(\delta, \theta \in H_0) = [y \in W|\theta \in H_0] = \alpha(\theta)$$
$$R(\delta, \theta \notin H_0) = [y \notin W|\theta \notin H_0] = \beta(\theta)$$

Les problèmes statistiques conventionnels$\sim$: Estimation et choix de modèles

- Choix de modèle $D = \{d_1, d_2, ...d_k\}$ avec $\Theta = \biguplus_{j=1}^{k} M_j$

Fonction de coût conventionelle

$W(d,\theta)$	$\theta \in M_1$	$\theta \in M_2$	...	$\theta \in M_k$
d_1	0	1	...	1
d_2	1	0	...	1
...	1	...	...	...
d_k	1	1	...	0

- Estimation ponctuelle

$D = \Theta = \{\hat{\theta}\}$; $\delta : \mathbf{y} \longmapsto d = \delta(\mathbf{y}) \in \Theta$;

$W(\hat{\theta}, \theta) : \Theta \times \Theta \to R$ le plus souvent $\left\|\hat{\theta} - \theta\right\|^2$

- Estimation par intervalle $D = \{B \subset \Theta ...\}...$

Exemple non conventionnel 1: Contrôle de qualité

- les actions possibles $\{d_1 = "livrer", d_2 = "rebut\}$
- θ inconnu est la proportion de pièces mauvaises dans le lot (N pièces ds la fabrication)
- Coûts : $W(d, \theta)$ Coût C ($= 1$ si unitaire pour la fabrication), D indemnité de dommage pour fabrication non conforme

$$W(d_1, \theta) = D \times N\theta$$
$$W(d_2, \theta) = C \times N$$

- Echantillon de taille $n << N$,$i = 1..n$

$$Y_i \sim Bernoulli(\theta)$$
$$[y_i = 1|\theta] = \theta$$
$$[y_i = 0|\theta] = 1 - \theta$$

Exemple Contrôle de qualité suite

- Règle *de bon sens* δ_k : $d_1 =$ "livrer" si $y < k$
- $R(\delta_k, \theta) =$
  $N(D \times \theta \sum_{j=0}^{k} C_n^j \theta^j (1-\theta)^{n-j} + 1 \times \sum_{j=k+1}^{n} C_n^j \theta^j (1-\theta)^{n-j})$
- Risque bayésien avec prior beta(a,b) pour $[\theta]$; $[\theta|y]$ par conjugaison beta(a+y,b+n-y)

$$\hat{R}(\delta) = \mathbb{E}_\theta \mathbb{E}_{Y|\theta} W(\delta(Y), \theta) = \mathbb{E}_Y \mathbb{E}_{\theta|Y} W(\delta(Y), \theta)$$

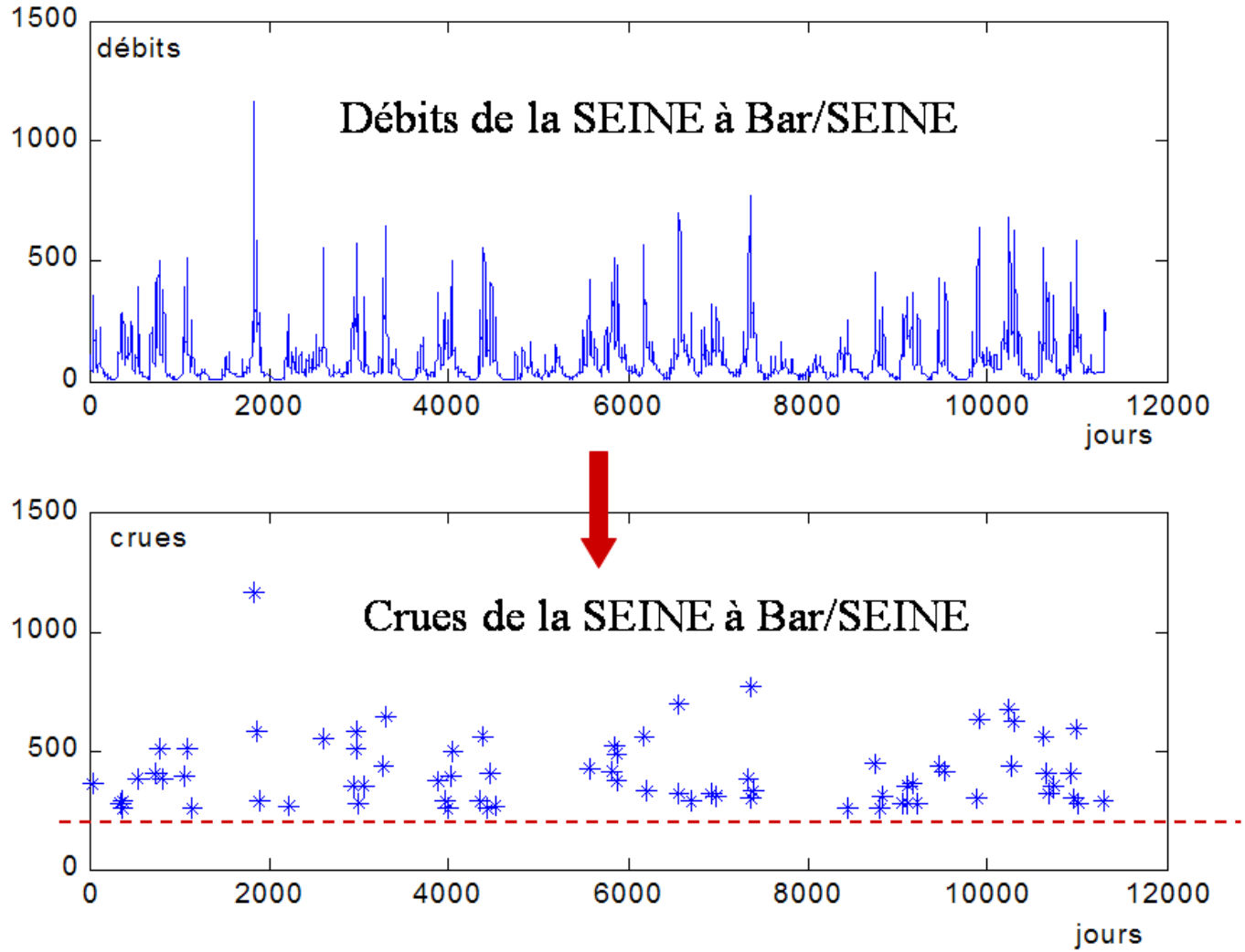
$$\hat{R}(\delta) = N \mathbb{E}_Y (D \times \mathbb{E}_{\theta|Y}(\theta) \times 1_{\delta(Y)=d_1} + 1 \times 1_{\delta(Y)=d_2})$$

d'où $\delta(y) = d_1$ si $D\frac{a+y}{a+b+n} < 1$

avec $k = \frac{a+b+n}{D} - a$ donnant la règle *bayésienne*

- Cas plus complexes : cf. recherche de N. Commeau

Exemple non conventionnel 2: POT



Exemple non conventionnel 2: Digue

- Choisir une hauteur de digue$\{h \in \mathbb{R}^+\}$
- Coûts :
  - construction $C \times h$,
  - dommages vis à vis de Y hauteur de crue annuelle %

$$0 \text{ si } Y < h$$

$$D(Y-h) \text{ si } Y > h$$

Coût moyen pour l'aménageur $W = C \times h + \int_h^{+\infty} D(x-h) f(x) dx$

- Loi exponentielle pour la crue

$$f(x) = \theta^{-1} \exp(-\frac{x}{\theta}) = [x|\theta]$$

alors

$$W(h, \theta) = Ch + D\theta \exp(-\frac{h}{\theta})$$

- Information $y^*_{40} = 104 m^3/s$ le plus fort débit (janvier 1996) sur la période 1968-2007 $(N = 40 ans)$
- Règle de décision $h = k_1 \times y^*_{40}$...Trouver k_1?

Règle de décision

- Loi du maximum

$$\Pr(Y_N^* < y) = (1 - e^{-\frac{y}{\theta}})^N$$

- Densité associée

$$[y^*|\theta] = \frac{N}{\theta} e^{-\frac{y}{\theta}} (1 - e^{-\frac{y}{\theta}})^{N-1}$$

- Règle : $y_N^* \mapsto \delta^1(y_N^*) = k_1 \times y_N^*$.

$$W(\delta^1(Y_N^*), \theta) = k_1 \times C \times Y_N^* + D\theta \exp(-\frac{k_1 \times Y_N^*}{\theta})$$

Risque avec décision = k_1 fois crue maximale sur 40 ans

Le risque $R(\delta, \theta)$ est l' espérance mathématique des coûts entraînés par les décisions prises dans le cadre de la règle δ.

$$E\left(W(\delta^1(Y_N^*), \theta)\right) = \int k_1 \times C \times y \times \frac{N}{\theta} e^{-\frac{y}{\theta}} (1 - e^{-\frac{y}{\theta}})^{N-1} dy$$

$$+ \int D \times \theta \times \exp(-\frac{k_1 \times y}{\theta}) \times \frac{N}{\theta} e^{-\frac{y}{\theta}} (1 - e^{-\frac{y}{\theta}})^{N-1} dy$$

$$R\left(\delta^1, \theta\right) = \theta \times Z(N, k_1)$$

$$Z_1(N, k) = Ck \int_0^\infty Nue^{-u}(1 - e^{-u})^{N-1} du$$

$$+ D \int_0^\infty Ne^{-ku}e^{-u}(1 - e^{-u})^{N-1} du$$

$$k_1^{opt} = 0.9819 \text{ avec } \frac{C}{D} = \frac{1}{50} \text{ et } Z_1(40, k_1^{opt}) = 5.497$$

Exemple : Digue avec information sur le régime moyen

- On observe la moyenne

$$\bar{Y}_N = \frac{\sum_{i=1}^{N} Y_i}{N}$$

- Densité Gamma associée

$$[\bar{y}|\theta] = \frac{N^N \theta^{-N}}{\Gamma(N)} \bar{y}^{N-1} e^{-\frac{N\bar{y}}{\theta}}$$

- Règle : $\bar{y}_N \mapsto \delta^2(\ \bar{y}_N\%$

$$W(\delta^2(\bar{Y}_N), \theta) = k_2 \times C \times \bar{Y}_N + D\theta \exp(-\frac{k_2 \times \bar{Y}_N}{\theta})$$

Exemple : Digue avec information sur le régime moyen

Risque

$$E\left(W(\delta^2(\bar{Y}_N),\theta)\right) = k_2 \times C \times \theta + \int D \times \theta \times \exp(-\frac{(k_2+N)\times\bar{y}}{\theta}) \times \frac{N^N\theta^{-N}}{\Gamma(N)}\bar{y}^{N-1}d\bar{y}$$

$$R\left(\delta^2,\theta\right) = \theta \times Z_2(N,k_2)$$

$$Z_2(N,k) = Ck + D\left(\frac{N}{k+N}\right)^N$$

$$k_2^{opt} = 4.0044 \text{ avec } \frac{C}{D} = \frac{1}{50} \text{ et } Z_2(40,k_2^{opt}) = 5.1047$$

Bilan

Une règle de décision δ est *admissible* dans Δ s'il n'existe pas de règle de Δ qui la domine. On note $\delta_2 \succeq \delta_1$. La relation $\succeq$ n'est qu'un *pré-ordre partiel* sur Δ.

Exemple :$\bar{y} = 30m^3/s$,$y^*_{40ans} = 104m^3/s$

$\delta_2 : \bar{y} \mapsto h = 4.0044 * \bar{y}$

$\delta_1 : y^*_{40ans} \mapsto h = 0.9819 * y^*_{40ans}$

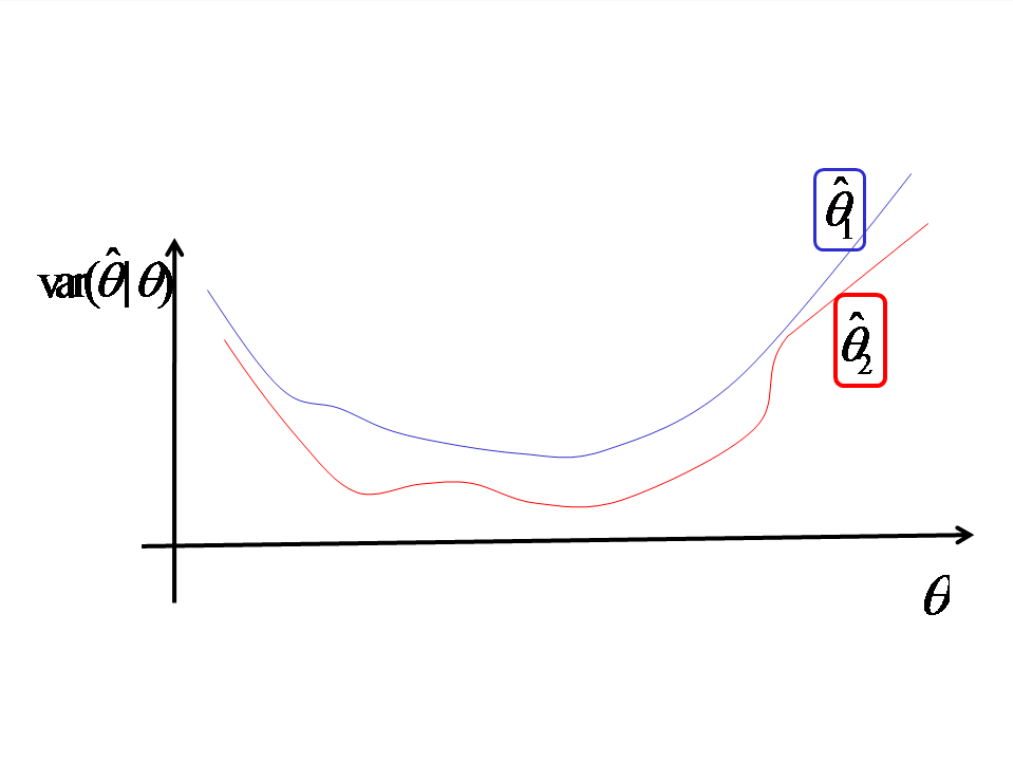
$R(\delta_2, \theta) = \theta \times MinZ_2(k) = \theta \times 5.1047$

$R(\delta_1, \theta) = \theta \times MinZ_1(k) = \theta \times 5.4969$

Problème identique pour l'estimation $D = \Theta = \{\hat{\theta}\}$

Que faire avec $R(\hat{\theta}, \theta)$? souvent $R(\hat{\theta}, \theta) = \int_y (\hat{\theta}(y) - \theta)^2 [y|\theta] dy$

- Admissibilité $\hat{\theta}_2 \succ \hat{\theta}_1$



- non biaisé $E(\hat{\theta}|\theta) = \theta$
- Et étude du comportement asymptotique : $\lim_{n \to \infty} Var(\hat{\theta}|\theta, n) \to 0$

Le bayésien pour soulager les questionnements fréquentistes

- Théorème 1 de Wald: Tout estimateur de Bayes (tel que $\pi(\theta) > 0, \forall \theta \in \Theta$) est admissible.
- On étend l'estimateur de Bayes par une séquence de priors $\pi_1, ...\pi_m, ...$ pour le théorème 2 de Wald (dit de laclasse complète)$\}$ Sous des conditions de régularité, les estimateurs admissibles sont soit des estimateurs de Bayes, soit une limite d'estimateurs de Bayes
- Théorème de Rao-Blackwell Si la fonction de coût est convexe et s'il existe une statistique exhaustive, alors pour toute règle $\delta(Y)$ il existe une règle de décision $\breve{\delta}(T(Y))$ (construite à partir de cette statistique) qui la domine.

$$\breve{\delta}(T(Y)) = E\left(\delta(Y)|T(Y)\right)$$

Application au pb de la digue

$$\breve{\delta}(T(Y)) = E\left(\delta(Y)|T(Y)\right)$$

$\delta_1(\vec{y}) = 0.9819 \times \max(y_1, y_2, ...y_{40})$

$R(\delta_1, \theta) = \theta \times 5.4968$

$\breve{\delta}_1(\vec{y}) = 0.9819 \times E(\max(y_1, y_2, ...y_{40})|\bar{y})$

$\breve{\delta}_1(\vec{y}) = 0.9819 \times k \times \bar{y}$

...Calcul de $\breve{\delta}_1$ par simulation...

$R(\breve{\delta}_1, \theta) = \theta \times 5.1220$!

Modéliser pour apprendre et décider

- Bayes = théorie de l'apprentissage + théorie de la décision
- Le rôle central du calcul des probabilités conditionnelles: il garantit la cohérence et sa mise à jour en fonction de l'information disponible. La probabilité change!
- L'incertitude se quantifie par des probabilités (bayésiennes): un pari décisionnel personnel. Mesure une relation individu/ état inconnu du monde. Quantifie de façon cohérente *l'opinion personnelle (d'un scientifique)*.
- Les acquis de la science naissent du partage argumenté des points de vue et de la reconnaissance collective de la convergence de leur pertinence.
- H. Sapiens utilise essentiellement des raccourcis cognitifs pour réagir face à l'incertitude.

Système 1

Système 2

$$17 \times 24$$

- Kahneman, D. (2011). Thinking, fast and slow.
- Piattelli-Palmarini, M. (1995). La réforme du jugement : ou comment ne plus se tromper.

Héterodoxie

- Les économistes (Micro-économie des risques)
- Les psycho-cognitivistes (Psychologues comportementalistes et neuro-sciences)
- Normatif, descriptif ou prescriptif?

Le développement récent des sciences de la décision (un regard critique sur la statistique décisionnelle Bayesienne) B. Munier and E. Parent UNESCO Doc. Techn. Hydro no. 20, UNESCO, Paris.(1998).

Risk, decision and rationality. B. Munier Springer Science & Business Media(2012)

Paradigme Bayésien et neurosciences

Le Cerveau Bayésien

Apprentissage
&
Sensation-Perception-Action

STIMULUS

- Connaissances a priori
- Information

Mise à jour bayésienne

Perte

- {Actions possibles}
- Etats prévisibles
- Information

- Risque
- Décision optimale

Règle de décision

ACTION

Le cerveau bayésien?

Le Cerveau Bayésien
Le « code » neuronal

- L'algorithme neuronal (entrées)
  - Quelle représentation de nos perceptions ?
  - Comment les manipule-t-il?
- L'algorithme neuronal (implémentation)
  - Une structure biologique pour l'apprentissage? Pour la décision?
  - Des mécanismes biologiques au service d'un algorithme bayésien?

Les grands principes de l'apprentissage, 2017, Stanislas Dehaene
https://www.youtube.com/watch?v=4NYAuRjvMNQ

Challenge : la validation

Comparer une loi de probabilité et une observation

- Tout modèle bayésien produit une distribution de probabilité pour exprimer une connaissance, la loi prédictive a posteriori, à interpréter comme un jugement probabiliste conditionnel (un pari!)

$$[Y_{new}|Y_{old}] = \int_{\theta} [Y_{new}|\theta] \times [\theta|Y_{old}]\, d\theta$$

- L' approche classique oublie un niveau d'incertitude (l'estimation $\hat{\theta}$ est généralement *pointwise*)

$$\hat{Y}_{new} \sim \left[Y|\hat{\theta}(Y_{old})\right]$$

- Face à des données de validation y sous l'input d'un nouvel environnement x, se pose le problème de comparer la distribution prédictive $[Y|Y_{old}, X = x]$ et les nouvelles observations (y, x)
- La vérification s'appuie sur une vision *fréquentielle* alors que la prédiction doit s'interpréter sous forme d'un *pari*.

Un score est un coût décisionnel moyen de prévision imparfaite

Le prévisionniste annonce la loi P pour le Y à venir...

- Avec la fonction de coût $W(d, Y)$, le décideur, qui fait confiance au prévisionniste, adopte l'action optimale $d_p^* = ArgMin_a \mathbb{E}_{Y\sim P}(W(a, Y))$. La fonction de score: $S(P, y) = W(d_p^*, y)$
- ...Mais la Nature se manifeste avec une loi F de Y différente de la prédictive P du prévisionniste, avec pour conséquence un coût moyen engendré par l'imperfection de P vis à vis de F, i.e le *score* :

$$S(P, F) = \mathbb{E}_{Y\sim F}(W(d_P^*, Y))$$

- Par construction, $S(P, F) \leq S(F, F)$. S est dit *propre* si $S(P, F) = S(F, F) \Rightarrow P = F$ (cette propriété dépend de la famille des lois où l'on place F et P)

Un score est un coût décisionnel moyen de prévision imparfaite

- Dawid appelle *entropie* l'optimum optimorum:

$$H(F) = S(F, F) = \mathbb{E}_{Y \sim F}(W(d_F^*, Y))$$

- Et il appelle *divergence* le manque à gagner par rapport au prévisionniste parfait:
$div(P, F) = S(P, F) - H(F) = \mathbb{E}_{Y \sim F}(W(d_F^*, Y) - W(d_P^*, Y))$

The geometry of proper scoring rules. A. P. Dawid. Annals of the Institute of Statistical Mathematics, 59:77–93 (2007)

Application au pb de la digue

Choix du coût de débordement:

$$W_\omega(d(.), Y) = \int_{s=0}^{1} \omega(s) \left(d(s) + D(s) \left| Y - d(s) \right|^+ \right) ds$$

dont la décision optimale est $d^\star(s) = P^{-1}(s)$.

- Ré-écriture avec $s = \alpha = 1 - 1/D$, $\int \omega(\alpha) d\alpha = 1$ et $H(z) = 1_{z>0}$

$$W_\omega(d_P^*(.), y) = \int_0^1 \frac{\omega(\alpha)}{1-\alpha} \left((H(P^{-1}(\alpha) > y) - \alpha) \times (P^{-1}(\alpha) - y) \right) d\alpha + y$$

- Cas particulier $\omega(\alpha) = 2(1-\alpha)$ c-a-d une beta(1,2), on reconnaît le CRPS.

$$W_{\beta(1,2)}(d_P^*(.), y) = \int_0^1 2 \left((H(P^{-1}(\alpha) > y) - \alpha) \times (P^{-1}(\alpha) - y) \right) d\alpha + y$$

D'autres choix des coefficients de la beta mènent à une famille de Quantile-Weighted Scoring Rule

Ouf!

Merci pour votre attention!

N'imprime pas

Que disait Tom?

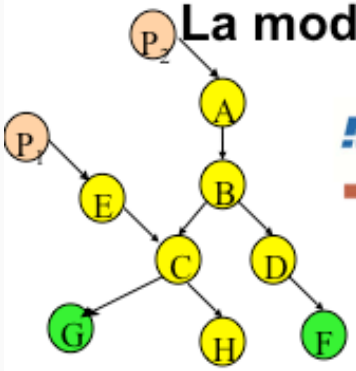
Bayes,1763: The probability of any event is the ratio between the value at which an expectation depending on the happening of the event ought to be computed, and the chance of the thing expected upon its happening.

La probabilité d'un événement est le rapport entre la valeur à laquelle une espérance dépendant de la réalisation de l'événement devrait être calculée, et la chance de la chose attendue lors de sa réalisation

????

La recherche de l'airbus Rio-Paris AF447 : Bayes pour assimiler la donnée

La modélisation hierarchique bayésienne en action



METRON
SCIENTIFIC SOLUTIONS
Mathematics Physics Computer Science

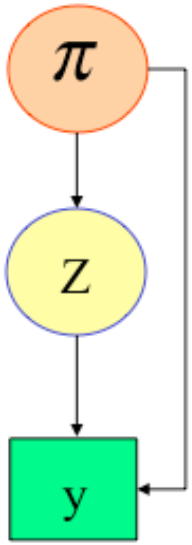
Modéliser : Conditionner par morceaux pour traduire la complexité

Paramètres, Etats du processus et mécanismes d'Observation sont modélisés indépendemment

Probabilités de détection et d'occurrence

Où se trouve vraiment l'avion

Les indices que l'on a observés



$[\pi]$ Paramètres

$[Z|\pi]$ Z: Process Latent

Ressemblance

Individus, populations, temps, espace

$[Y|\pi, Z]$ Y: Observations

Rôle des variables cachées…

Modélisation [effets|causes]

La recherche de l'airbus Rio-Paris AF447 : Bayes pour assimiler la donnée

Recherche de l'Airbus AF447 (2009/06/01)

$$[Y_i = 1 | Z_i = 1] = p_i$$ Probabilité de détection dans la case i

$$\pi_i = [Z_i = 1]$$ Probabilité d'occurrence dans la case i

$$[Z_i = 1 | Y_i = 0] = \frac{[Y_i = 0 | Z_i = 1][Z_i = 1]}{[Y_i = 0 | Z_i = 0][Z_i = 0] + [Y_i = 0 | Z_i = 1][Z_i = 1]}$$

$$= \frac{(1 - p_i)\pi_i}{1 - p_i \pi_i} < \pi_i$$

$$[Z_{j \neq i} = 1 | Y_i = 0] = \frac{[Y_i = 0 | Z_j = 1][Z_j = 1]}{[Y_i = 0]}$$

$$= \frac{\pi_j}{1 - p_i \pi_i} > \pi_j$$

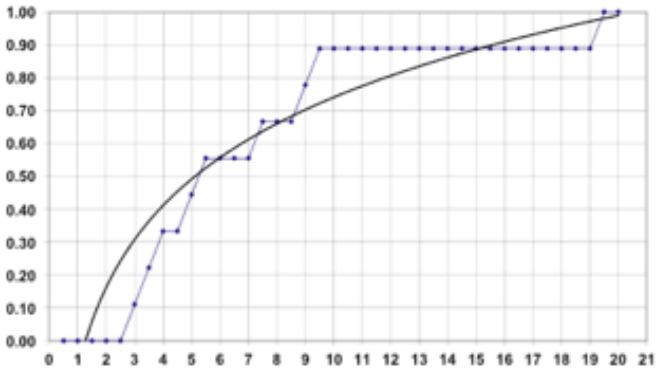
METRON
SCIENTIFIC SOLUTIONS
1818 Library Street, Suite 600
Reston, Virginia 20190
703 787 8700 www.metsci.com

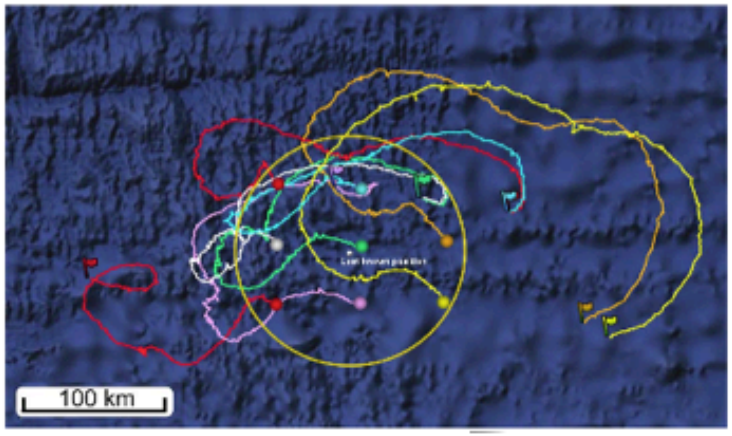
Mathematics Physics Computer Science

Search Analysis for the Location of the AF447 Underwater Wreckage. Report to **Bureau d'Enquêtes et d'Analyses** pour la sécurité de l'aviation civile. By L. D. Stone, C. Keller, T. L. Kratzke., J.Strumpfer. 20 January 2011

La recherche de l'airbus Rio-Paris AF447 : Bayes pour assimiler la donnée

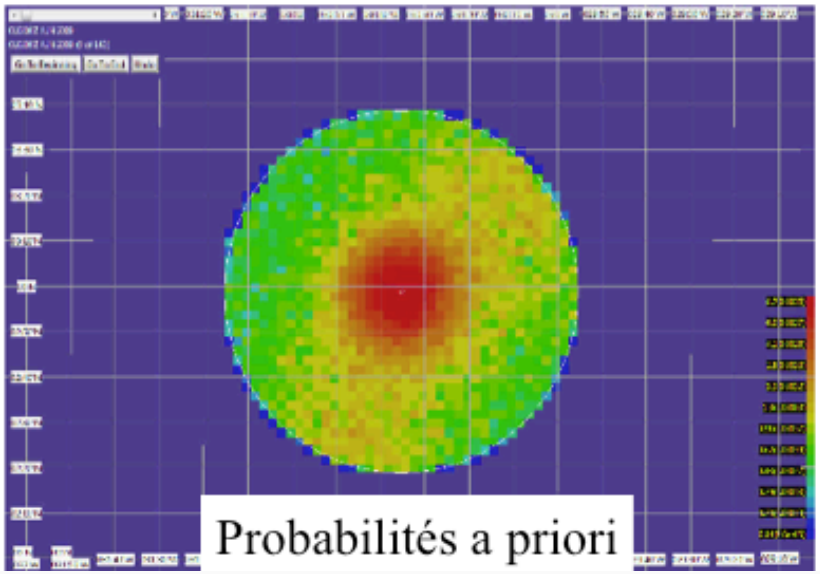
Prior pour l'Airbus AF447 disparu 1er juin





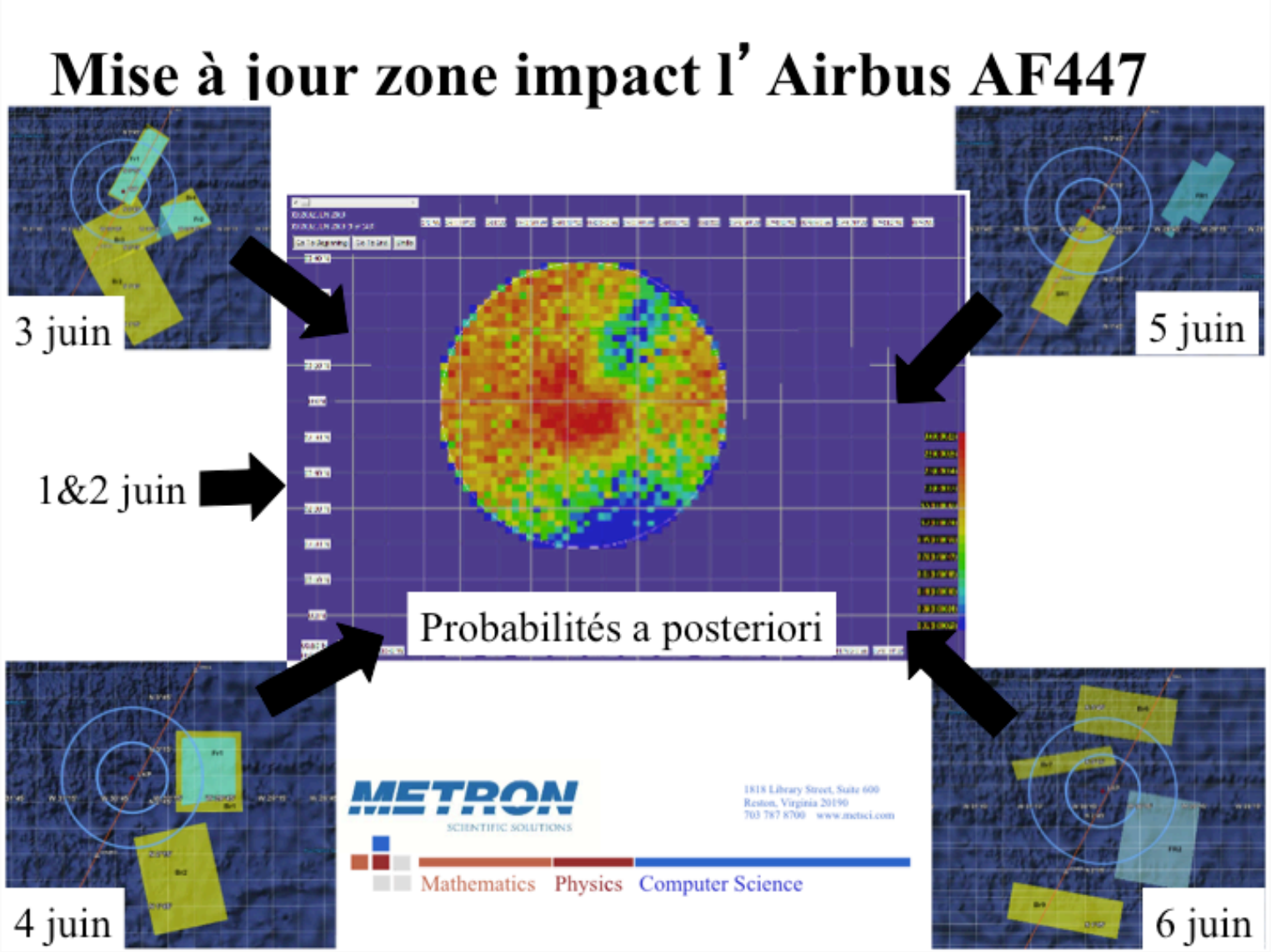
cdf empirique (9 accidents)

Reverse drift



Probabilités a priori

La recherche de l'airbus Rio-Paris AF447 : Bayes pour assimiler la donnée



Exemple non conventionnel de la digue

- $d \in \mathbb{R}^+$ est la hauteur de la digue (ou du paravalanche)
- D le montant des dommages en unités de construction
- la fonction de coût arbitre entre investissement et dommages

$$W(d, Y) = \mathbb{E}_Y(d + D\,|Y - d|^+)$$

Solution explicite pour $D > 1, 1 - \alpha = 1/D$: $d_P^* = P^{-1}(\alpha)$

Application hydro-météorologique : Peak Over Threshold

Quand Y obéit à une loi de Pareto Généralisée

$$P_{\rho,\beta}(y) = \begin{cases} 1-(1-\beta(y))^{\frac{\rho}{\beta}} \text{ si } \beta \neq 0 \\ 1-\exp(-\rho(y) \quad) \text{ si } \beta = 0 \end{cases}$$

$$W(d,P) = d + D\frac{(1-\beta d)^{\frac{\rho}{\beta}}}{\beta+\rho} \text{si } \beta \neq 0$$
$$= d + D\frac{\exp(-\rho \times d)}{\rho} \text{si } \beta = 0$$

L'action optimale est ici définie par: $d^*_{\rho,\beta} = \frac{1-\left(\frac{(\beta+\rho)}{\rho D}\right)^{\frac{\beta}{\rho-\beta}}}{\beta}$ si $\beta \neq 0$
$d^*_\rho = \frac{\log(D)}{\rho}$ si loi exponentielle

Valeur de l'Information parfaite

$$W(h,\theta) = Ch + D\int_{x=h}^{+\infty}(x-h)\,\theta e^{-\theta x}dx$$

Sous connaissance parfaite de θ, que vaudrait h?

$$\begin{aligned}\left(\frac{\partial W(h,\theta)}{\partial h}\right|_{h=h^*} &= 0\\ C + D\exp(-\frac{h^*}{\theta}) &= 0\end{aligned}$$

Estimation statistique de θ et les conséquences

$$\frac{1}{e^{-\frac{h^*}{\theta}}} = T(h^*) = \frac{D}{C} = T(D,C)$$

$$h^*(\theta) = \theta \times Log(T(D,C))$$

Regret

$$\begin{aligned}\tilde{W}(h,\theta) &= W(h,\theta) - W(h^*(\theta),\theta) \\ &= C(h - \theta \times Log(T(D,C))) \\ &\quad + D\theta\left(\exp(-\frac{h}{\theta}) - \frac{1}{T(C,D)}\right)\end{aligned}$$

Le regret mesure le manque à gagner résultant de l'ignorance de l'état réel de la nature.

Souvent approximation quadratique $\tilde{W}(h,\theta) = K(\theta)(h - h^*(\theta))^2$, ici on approximerait le regret pour la digue par $\frac{C}{\theta}(h - \theta \times Log(T(D,C)))^2$ ou linéaire par morceaux

$$\tilde{W}(h,\theta) = \begin{cases} A(h^*(\theta) - h) \text{ si } h < h^*(\theta) \\ B(h - h^*(\theta)) \text{ si } h > h^*(\theta) \end{cases}$$

Risque bayésien

- Considérons une famille de densité de probabilité π sur l'état de la Nature Θ : $d\pi(\theta = \theta) \equiv [\theta|\pi]d\theta$
- Risque bayésien

$$\bar{R}_\pi(\delta) = \int_\theta R(\delta, \theta)[\theta|\pi]d\theta = E_\pi E_\theta \left(W(\delta(Y), \theta)\right)$$

$$\bar{R}_\pi(\delta) = \iint_{y,\theta} W(\delta(y), \theta)[y|\theta][\theta|\pi]d\theta dy$$

- Règle de Bayes (idem avec regret $\tilde{W}(\delta(Y), \theta)$)%

$$\delta^*_\pi = ArgMin\left(\bar{R}_\pi(\delta)\right)$$

%

$$\bar{R}_\pi(\delta) = \iint_{\theta,y} W(\delta(y), \theta)[y, \theta]d\theta dy$$

$$\bar{R}_\pi(\delta) = \int_y \left(\int_\theta W(\delta(y), \theta)[\theta|y]d\theta\right)[y]dy$$

La formule de Bayes sur l'exemple de la digue

Si l' information est la moyenne $\bar{y} = 30$

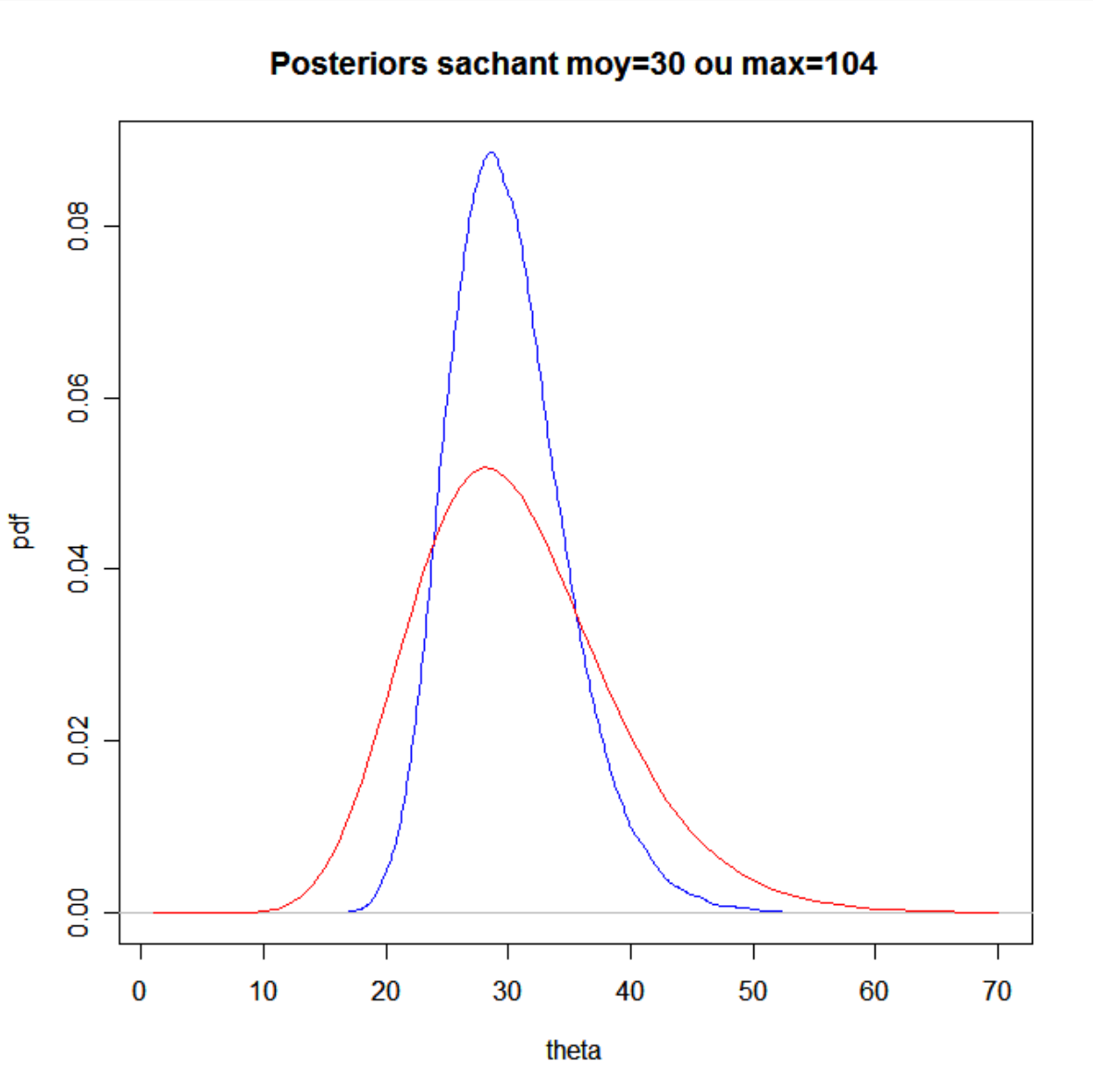
$$\theta^{-1} \sim Gamma(a, b) \Leftrightarrow \pi(\theta) = [\theta]$$
$$\bar{y}|\theta \sim Gamma(N, \frac{N}{\theta})$$
$$[\theta^{-1}|\bar{y}] \sim Gamma(a + N, b + N\bar{y})$$

Si l' information est le max sur 40 ans $y^*_{40} = 30$

$$\theta^{-1} \sim unif(A, B)$$
$$[y^*_N|\theta] = \frac{N}{\theta} \exp(-\frac{y^*_N}{\theta}) \left(1 - \exp(-\frac{y^*_{40}}{\theta})\right)^{N-1}$$
$$[\theta^{-1}|y^*_N] \propto \frac{1}{\theta H(y^*_N)} \exp(-\frac{y^*_N}{\theta}) \left(1 - \exp(-\frac{y^*_{40}}{\theta})\right)^{N-1}$$

avec $H(x) = \int_A^B t \exp(-tx) \left(1 - \exp(-tx)\right)^{N-1}$

Posteriors sur l'exemple de la digue



Coût a posteriori sur l'exemple de la digue

$$t = \theta^{-1}$$

$$V_\pi(h, \bar{y}) = Ch + D \int_t t^{(a+N-1)-1} \exp(-t(h + b + N\bar{y})) \frac{(b + N\bar{y})^{(a+N)}}{\Gamma(a+N)} dt$$

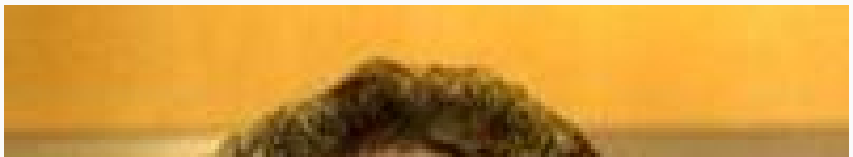
$$= Ch + \frac{D}{(a+N-1)} \frac{(b + N\bar{y})^{(a+N)}}{(h + b + N\bar{y})^{(a+N-1)}}$$

$$h^*(\bar{y}) = (b + N\bar{y}) \left(\left(\frac{D(a+N)}{C(a+N-1)} \right)^{\frac{1}{a+N}} - 1 \right)$$

$$h^*(\bar{y}) \approx \log(D/C) \left(\frac{b}{a} \left(\frac{a}{a+N} \right) + \bar{y} \left(\frac{N}{a+N} \right) \right) \left(1 + \frac{\log(D/C)}{2(a+N)} \right)$$

$$\frac{\log(50/1)}{2(0+40)} = 0.0489$$

Point de vue de Phil Dawid



Un score est un coût décisionnel moyen de prévision imparfaite

Le prévisionniste annonce la loi P pour le Y à venir...

- Avec la fonction de coût $W(d, Y)$, le décideur, qui fait confiance au prévisionniste, adopte l'action optimale $d_p^* = ArgMin_a \mathbb{E}_{Y\sim P}(W(a, Y))$ On appelle fonction de score $S(P, y) = W(d_p^*, y)$
- ...Mais la Nature se manifeste avec une loi F de Y différente de la prédictive P du prévisionniste, avec pour conséquence un coût moyen engendré par l'imperfection de P vis à vis de F, i.e le *score* :

$$S(P, F) = \mathbb{E}_{Y\sim F}(W(d_P^*, Y))$$

- Par construction, $S(P, F) \leq S(F, F)$. S est dit *propre* si $S(P, F) = S(F, F) \Rightarrow P = F$ (cette propriété dépend de la famille des lois où l'on place F et P)
- Dawid appelle *entropie* l'optimum optimorum:

$$H(F) = S(F, F) = \mathbb{E}_{Y\sim F}(W(d_F^*, Y))$$

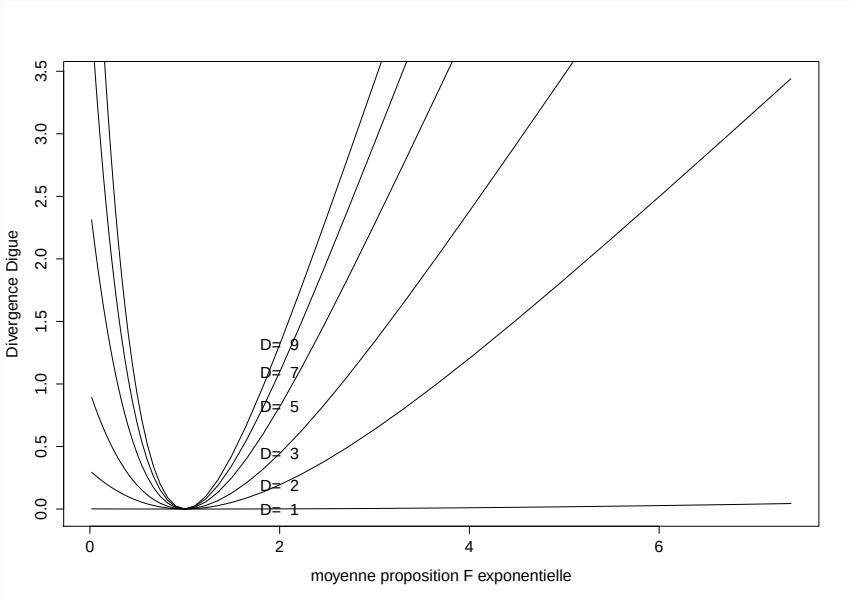
- Et il appelle *divergence* le manque à gagner par rapport au

Exemple monoparamétrique: Score propre de la digue pour la famille de la loi exponentielle

- Loi $F = Exp(\rho_F)$, Prévision $P = Exp(\rho_P)$, on obtient le score:

$$S(\rho_P, \rho_F) = \frac{Log(D)}{\rho_P} + D \times \frac{\exp(-\rho_F \times \frac{Log(D)}{\rho_P})}{\rho_F} = \frac{Log(D)}{\rho_P} + \frac{D^{1-\frac{\rho_F}{\rho_P}}}{\rho_F}$$

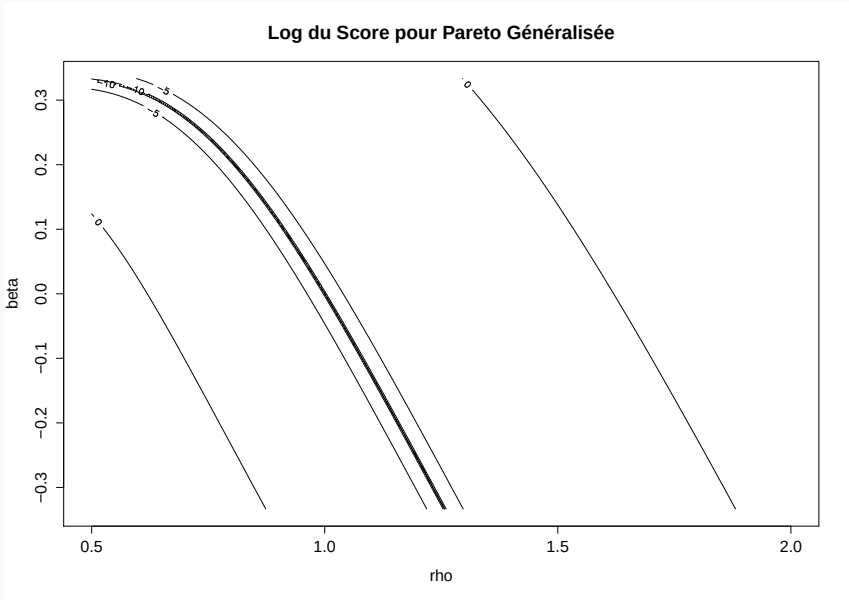
- Application $\rho_F = 1$. En augmentant D , on rend plus rapidement grave (et de façon dissymétrique) la divergence de P par rapport à F



Exemple à deux paramètres: Score NON propre de la digue pour la famille de la loi Pareto Généralisée

Prévision $P = LPG(\beta, \rho)$, Loi $F = Exp(\rho_F)$, on obtient le score:%

$$S((\rho, \beta), \rho_F) = \frac{1 - \left(\frac{(\beta+\rho)}{\rho D}\right)^{\frac{\beta}{\rho-\beta}}}{\beta} + D \times \frac{\exp(-\rho_F \times \frac{1-\left(\frac{(\beta+\rho)}{\rho D}\right)^{\frac{\beta}{\rho-\beta}}}{\beta})}{\rho_F}$$



Pondérer pour parcourir tout le support de la fonction

- Idée: Trouver $P \in \mathcal{F}$ grace à l'action de Bayes, mais si celle-ci est mono-dimensionnelle, on n'en trouvera qu'une statistique! (OK si la famille $\mathcal{F}$ est monoparamétrique) ... Passer au multidimensionnel (OK si la famille $\mathcal{F}$ est paramétrique)), voire au fonctionnel (le cas plus général).
- En dimension 2 équipondérée : Fonction de coût de la digue pour dim 2

$$W_2\left(\begin{pmatrix} d_1 \\ d_2 \end{pmatrix}, Y\right) = d_1 + D_1 \left|Y - d_1\right|^+ + d_2 + D_2 \left|Y - d_2\right|^+$$

- Solution de la digue pour dim 2 , $\alpha = 1 - 1/D$

$$\begin{pmatrix} d_1^* \\ d_2^* \end{pmatrix} = \begin{pmatrix} F^{-1}(\alpha_1) \\ F^{-1}(\alpha_2) \end{pmatrix}$$

- Fonctionnel avec pondérations $\omega(s) > 0$:

$$W_\omega(d(.), Y) = \int_{\mathcal{S}} \omega(s) \left(d(s) + D(s) \left|Y - d(s)\right|^+\right) ds$$

Quantile-Weighted Scoring Rule

- Ré-écriture avec $s = \alpha = 1 - 1/D$, $\int \omega(\alpha) d\alpha = 1$ et $H(z) = 1_{z>0}$

$$W_\omega(d_P^*(.), y) = \int_{\alpha=0}^{1} \omega(\alpha) \left(P^{-1}(\alpha) + \frac{1}{1-\alpha} \left| y - P^{-1}(\alpha) \right|^+ \right) d\alpha$$

$$W_\omega(d_P^*(.), y) = \int_{0}^{1} \frac{\omega(\alpha)}{1-\alpha} \left((1-\alpha) P^{-1}(\alpha) + \left| y - P^{-1}(\alpha) \right|^+ \right) d\alpha$$

$$W_\omega(d_P^*(.), y) = \int_{0}^{1} \frac{\omega(\alpha)}{1-\alpha} \left((H(P^{-1}(\alpha) > y) - \alpha) \times (P^{-1}(\alpha) - y) \right) d\alpha + y$$

- Cas particulier $\omega(\alpha) = 2(1-\alpha)$ c-a-d une beta(1,2), on pondère d'autant plus que la probabilité de dépasser la digue est forte ($1-\alpha$ est la p-value).

$$W_{\beta(1,2)}(d_P^*(.), y) = \int_{0}^{1} 2 \left((H(P^{-1}(\alpha) > y) - \alpha) \times (P^{-1}(\alpha) - y) \right) d\alpha + y$$

- Changement de variable $P(t) = \alpha$ puis intégration par parties donne le Brier intégré uniformément c-a-d le CRPS

Le célèbre Continuous Ranked Probability Score

- Le cumul quadratique entre P (le cdf proposé) et la marche d'escalier (cdf empirique sur l'observation : $H(z) = 1_{z>0}$):

$$S(P, y) = \int_t (P(t) - H(t - y))^2 dt$$

- Biais et variabilité

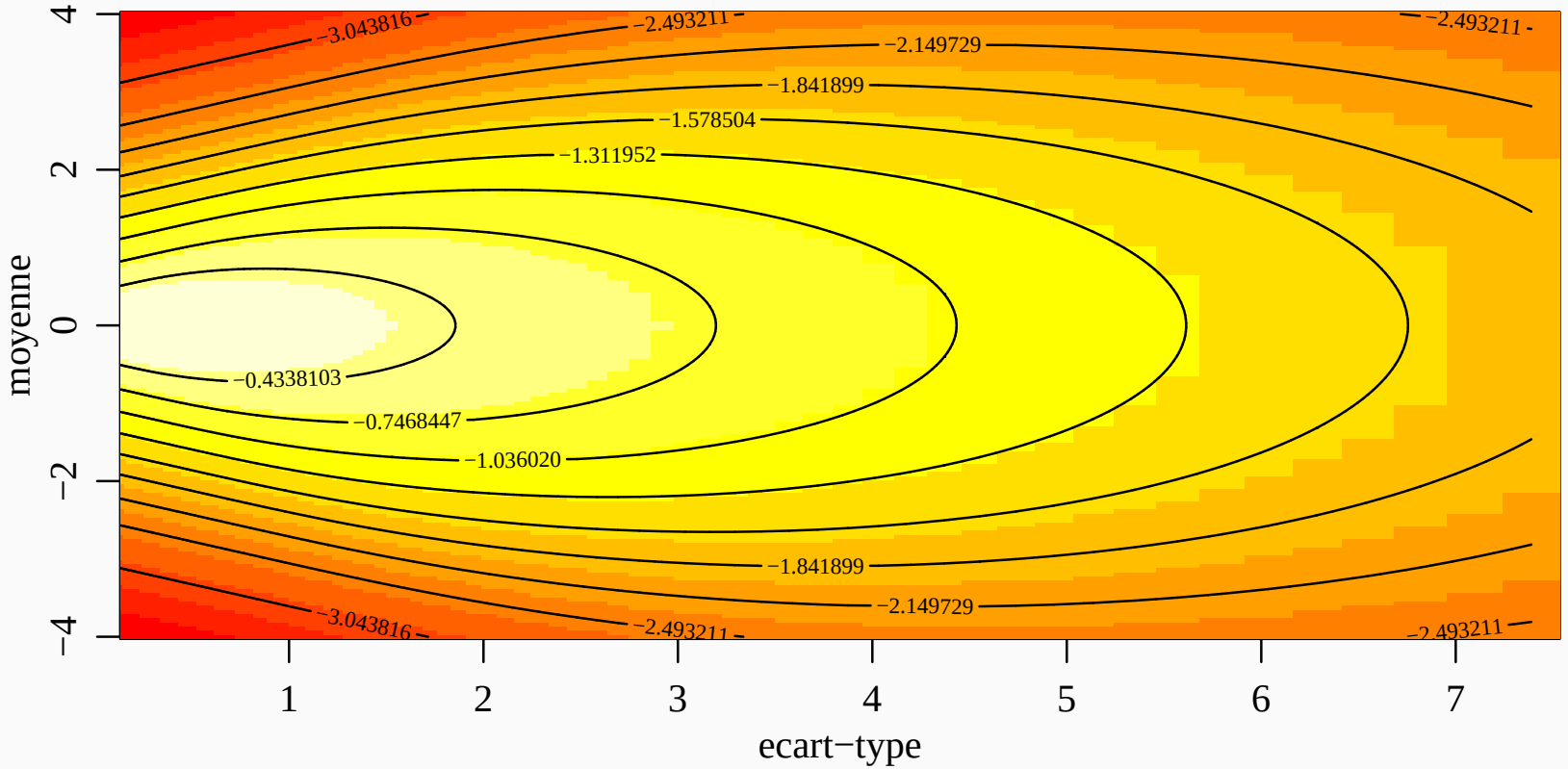
$$S(P, y) = -\mathbb{E}_P(|Z - y|) + \frac{1}{2}\mathbb{E}_P(|Z - Z'|)$$

$$Z, Z' \underset{iid}{\sim} P$$

- même dimension que y à prédire
- Calcul explicite si P normale de paramètres μ et σ :

$$S(P_{\mu,\sigma^2}, y) = -\sigma \left\{ \frac{1}{\sqrt{\pi}} - 2dnorm(\frac{y-\mu}{\sigma}) - \frac{y-\mu}{\sigma}(2pnorm(\frac{y-\mu}{\sigma}) - 1 \right.$$

CRPS Normal


−3.043816
−2.493211
−2.149729
−1.841899
−1.578504
−1.311952
−0.4338103
−0.7468447
−1.036020
−1.841899
−2.149729
−2.493211
−3.043816
−2.493211
−2.493211
moyenne
4
2
0
−2
−4
1
2
3
4
5
6
7
ecart−type

Extending the Quantile-Weighted Scoring Rule

- En pratique, on ne connait pas la loi F d'échantillonnage Y, $S(P,F) = \mathbb{E}_{Y\sim F}(W_\omega(d_P^*(.),Y))$ est déterminé empiriquement, il s'agit souvent plutôt de comparer P_1 vs P_2, les calculs se réalisant en discrétisant les integrations sur Y et sur α
- Ré-écriture du score de la digue avec $s = \alpha = 1 - 1/D$, $\int \omega(\alpha)d\alpha = 1$ et $H(z) = 1_{z>0}$

$$W_\omega(d_P^*(.),y) = \int_0^1 \frac{\omega(\alpha)}{1-\alpha}\left((H(P^{-1}(\alpha) > y) - \alpha) \times (P^{-1}(\alpha) - y)\right) d\alpha + y$$

- CRPS pour $\omega(\alpha) = 2(1-\alpha)$ (une beta(1,2)) on pondère d'autant moins qu'on va vers des fortes valeurs au dépassement.
- Suggestions de pondérations diverses $\omega(\alpha)$ selon une beta(a,b) pour le score quantile de la digue

Effet de la pondération	$\omega(\alpha)$	$\frac{\omega(\alpha)}{1-\alpha}$
Neutre (=CRPS)	beta(1,2)	1