DÉCISION STATISTIQUE (G3)

La **théorie de la décision** statistique a pour objet l'étude des **problèmes de décision** statistique. Il s'agit donc de définir des modes de prises de décision les plus appropriés en environnement **aléatoire** (on peut toujours admettre un tel contexte).

(i) Un **problème (de décision) statistique** comporte :

(a) un **modèle statistique** $(\Omega, \mathcal{T}, \mathcal{P})$, appelé **modèle de base**, ou **modèle initial**, constitué d'un **ensemble** fondamental Ω dont les éléments ω sont des **événements élémentaires** (**expériences aléatoires**, épreuves ou **unités statistiques**), d'une **tribu** d'évènements $\mathcal{T}$ définie sur Ω (et dont les éléments sont parfois appelés **évènements « complexes »**) et d'une **famille** de probabilités $P_\theta \in \mathcal{P}$, définies sur $\mathcal{T}$, où $\theta \in \Theta$ (ensemble des valeurs du « **paramètre** » θ). L'une de ces probabilités, en général inconnue (sauf lorsque le modèle décrit une **simulation**, auquel cas cette probabilité est fixée a priori), est réputée gouverner l'apparition des événements décrits par le **phénomène** considéré. L'ensemble Θ est souvent doté d'une **tribu de parties** $\mathcal{B}(\Theta)$, aussi notée $\mathcal{B}_\Theta$, qui en fait un **espace mesurable** : ainsi, dans le cadre bayésien, on suppose donnée une (mesure de) **probabilité a priori** Π, définie sur $\mathcal{B}_\Theta$.

(b) un **espace d'observation** mesurable $(\mathcal{X}, \mathcal{B})$ dont la signification est la suivante. A tout **événement** élémentaire ω, une fonction $X : \Omega \mapsto \mathcal{X}$ associe une grandeur (en principe observable) $X(\omega)$, appelée **observation**.

Mathématiquement, X est une **variable aléatoire** ou une **statistique** (eg un **échantillon**) et l'espace $\mathcal{X}$ est aussi appelé **espace des résultats**, ou **espace d'échantillonnage**. Les **mesures** $X(\omega) = x$ effectuées peuvent être quantitatives ou qualitatives : résultats numériques d'une épreuve ou d'une **expérience**, caractères descriptifs associés à des unités (cf **variable qualitative**, **variable quantitative**). A chaque probabilité P_θ sur $\mathcal{T}$, X associe une probabilité P_θ^X sur $\mathcal{B}$, appelée **loi de probabilité** de X. L'image de la famille $(P_\theta)_{\theta \in \Theta}$ par X est donc la famille de lois $(P_\theta^X)_{\theta \in \Theta}$, dont l'une d'entre elles est ainsi supposée régir l'apparition des résultats X.

En général, on ne distingue pas entre le modèle de base $(\Omega, \mathcal{T}, (P_\theta)_{\theta \in \Theta})$ et le **modèle image** $(\mathcal{X}, \mathcal{B}, (P_\theta^X)_{\theta \in \Theta})$ ainsi défini : en effet, on peut faire abstraction du modèle initial en identifiant $\mathcal{X}$ avec Ω, $\mathcal{B}$ avec $\mathcal{T}$ et en posant $X = id_{\mathcal{X}}$. On étudie alors directement le **modèle image**.

Les données précédentes définissent simplement deux modèles statistiques, l'un de base, l'autre image. Ces modèles décrivent la génération d'une observation à l'aide d'un **schéma probabiliste** dépendant d'un paramètre inconnu. On en déduit un **problème statistique** en les complétant à l'aide :

(c) d'un **espace de décision(s)**, ou **espace d'action(s)**, ou **espace décisionnel**, $(D, \mathcal{B}_D)$, dont les éléments d sont appelés **décisions**, ou **actions**.

(d) d'une **fonction de perte**, ou **fonction de coût**, ou encore **fonction de paiement**, $L : \Theta \times D \mapsto \mathbf{R}_+$, supposée $(\mathcal{B}_\Theta, \mathcal{B}_D)$-mesurable, qui indique la perte subie par le décideur lorsque la décision d est prise au vu du résultat $X(\omega) = x$, alors que la **vraie valeur** du paramètre est θ.

(e) d'une **règle de décision**. Si l'on décide $d = \delta(x)$ au vu du résultat x, l'application (supposée) mesurable, $\delta : \mathcal{X} \mapsto D$, tq $x \mapsto \delta(x)$, est appelée **règle de décision pure**, ou **fonction de décision pure**, ou encore **stratégie de décision pure**, ou simplement **décision**. On note souvent Δ l'espace des règles de décision (pures) δ.

Plus généralement, on définit une **règle de décision mixte**, ou **fonction de décision mixte**, ou encore **stratégie de décision mixte** (ou parfois **aléatoire**) m comme **probabilité de transition** de $\mathcal{X} \times \mathcal{B}_D$ vers $[0, 1]$ selon :

(1) $(x, B) \in \mathcal{X} \times \mathcal{B}_D \mapsto m_x(B) = m(x, B) \in [0, 1]$.

Cette probabilité, définie sur $\mathcal{B}_D$, possède la signification suivante : au vu de l'**observation** $x \in \mathcal{X}$, le **statisticien** décide d avec la probabilité $m(x, d)$. On note Δ_m ou Δ_M l'ensemble des règles de décision mixtes m ainsi définies.

Une règle de décision est donc pure ssi $m = \delta_x$ (**loi de DIRAC** au point x). La fonction de perte L peut être considérée comme une **variable aléatoire**, $L(\theta, \delta(x)) = L(\theta, \delta \circ X)$, qui s'interprète, $\forall d \in D$, comme une **vars** $\Omega \mapsto \mathbf{R}_+$.

Lorsqu'il décide selon la règle mixte m, le **statisticien** encourt une **perte moyenne** :

(2) $\bar{L}(\theta, x) = \int L(\theta, a)\, m(x, da)$

égale à $L(\theta, \delta(x))$ dans le cas pur, donc un **risque**, défini comme l'**espérance mathématique** de cette perte moyenne :

(3) $R(\theta, m) = E_\theta \bar{L}(\theta, m) = \int_{\mathcal{X}} L(\theta, x)\, dP_\theta^X(x)$,

égal à $E_\theta L(\theta, \delta \circ X)$ dans le cas pur, auquel cas on le note aussi :

(4) $$R(\theta, \delta) = E_\theta L(\theta, \delta \circ X) = \begin{array}{l} \int_\Omega L\{\theta, \delta(X(\omega))\}\, dP_\theta(\omega) \\ \int_{\mathcal{X}} L\{\theta, \delta(x)\}\, dP_\theta^X(x) \end{array}$$

(cf **théorème de transfert des mesures**).

On appelle alors **fonction de risque** l'**application mesurable** : $R_m : \Theta \mapsto \mathbf{R}_+$ (ou $R_\delta : \Theta \mapsto \mathbf{R}_+$ dans le cas pur) définie par la première application partielle de R :

(5) $\theta \in \Theta \mapsto R_m(\theta) \in \mathbf{R}_+$ (ou $R_\delta(\theta) \in \mathbf{R}_+$ dans le cas pur).

La fonction de risque permet de définir, sur l'ensemble Δ_M des règles de décision mixtes m, une **relation de préférence**, ie un **préordre** $\prec$ défini par :

(6) $m' \prec m'' \quad \Leftrightarrow \quad R_{m'} \leq R_{m''}$ sur Θ,

avec une écriture analogue dans le cas pur.

Si Π est une **probabilité a priori** définie sur $\mathcal{B}_\Theta$, la relation de préférence précédente, appelée **relation de Π-préférence** sur Δ_M, s'écrit :

(7) $m' \prec m'' \quad \Leftrightarrow \quad R_{m'} \leq R_{m''}$ sur Θ (Π-p.s.)

où R_m est ici le **risque de T. BAYES**, défini par (cf **risque bayésien**) :

(8) $R_m(\Pi) = \int R_m(\theta)\, d\Pi(\theta)$.

(ii) La relation de préférence précédente permet ainsi de choisir une **règle de décision optimale** : l'**optimalité** s'apprécie donc ici pr à la fonction de risque. Est optimale, dans ce sens, toute règle de décision qui minimise le risque (ce qui nécessite, le plus souvent, la résolution d'un **problème d'optimisation**).

Cependant, il n'existe pas, en général, de solution unique à ce problème, la classe des règles de décision optimales étant encore trop vaste. Aussi, selon le **problème de décision**, on doit s'imposer divers **principes de « réduction »** de cette classe : **principe d'invariance**, principe « minimax » (cf **règle minimax**), **principe de parcimonie** (réduction de l'espace des paramètres), etc.

(iii) On considère souvent le triplet (θ, x, d) comme un jeu à deux joueurs, ie : la **Nature** (qui choisit un point $\theta \in \Theta$ et gouverne ainsi le **phénomène** aléatoire considéré) et le **décideur** (**statisticien** ou **homme de l'art**), qui choisit un point $d \in D$ en guise de réponse (raisonnée) à une « **information** » $x \in \mathcal{X}$.

On dit que $\theta \in \Theta$ est un **« état de la Nature »**. Ce paramètre, supposé « choisi » par la Nature, n'est pas connu du statisticien (paramètre **inobservable**), mais présente un intérêt pour lui (**paramètre d'intérêt**) dans la mesure où sa connaissance détermine celle de la loi P_θ^X qui engendre les observations (cf **identification**). Ceci suppose que le modèle est, lui-même, « exact » et que le problème statistique est bien posé.

On peut, dans les données du problème, considérer P_θ^X comme une **probabilité de transition** $\Theta \times \mathcal{B} \mapsto [0 , 1]$, que l'on peut aussi noter P, tq l'application $(\theta , B) \mapsto P(\theta , B)$ signifie que le statisticien observe le résultat d'un tirage aléatoire de x en fonction des valeurs du paramètre $\theta \in \Theta$.

(iv) La signification du paramètre $\theta \in \Theta$ est, en pratique, variable selon la nature du problème. Ce paramètre peut être :

(a) une grandeur composée d'un nombre fini de scalaires (paramètre vectoriel, ie à valeurs dans un **espace vectoriel** réel de dimension finie), dont la connaissance détermine entièrement la (**forme** de la) **loi** P_θ^X ;

(b) ou encore un élément d'un ensemble plus complexe (eg un ensemble **caractéristique** : espace de **fonctionnelles**, de **densités**, de **fr**, de **fc**, etc) associé à la famille des lois P_θ^X.

(v) Un **problème de décision statistique** peut donc se noter par l'un des symboles :

(9) $\{(\Omega, \mathcal{T}, (P_\theta)_{\theta \in \Theta}), (D, \mathcal{B}_D), L\}$ ou $\{(\mathcal{X}, \mathcal{B}, (P_\theta^X)_{\theta \in \Theta}), (D, \mathcal{B}_D), L\}$,

selon qu'on se réfère au modèle de base ou au modèle image (modèle transformé).

(ii) Un problème de décision statistique peut aussi se formaliser dans un cadre « non paramétré », notamment en **Statistique non paramétrique**. Dans ce cadre :

(a) le modèle initial se note $(\Omega, \mathcal{T}, \mathcal{P})$ et le modèle image $(\mathcal{X}, \mathcal{B}, \mathcal{P}^X)$ représente l'image par la va $X : \Omega \mapsto \mathcal{X}$ du modèle initial ;

(b) une décision est définie comme précédemment ;

(c) la fonction de perte est de la forme $L(P , \delta(X))$ ou, selon le cas, $L(\gamma , \delta(X))$ en notant Γ l'ensemble des caractéristiques γ considérées sur $\mathcal{P}^X$;

(d) la fonction de risque s'en déduit directement.

(iii) De nombreuses **procédures statistiques** se ramènent à un problème de décision statistique, eg (cf eg **plan d'expérience**, **sondage**, **estimateur**, **test**, **classification**, **prévision**, etc) :

(a) dans un **problème de test** (d'hypothèses) on pose eg $D = \{d_0 , d_1\}$ ou encore $D = \{0 , 1\}$, d'où Card D = 2. Si l'on partitionne Θ selon $\{\Theta_0 , \Theta_1\}$, avec $\Theta_0 \neq \emptyset$, et $\Theta_1 \neq \emptyset$, la fonction de perte L est définie selon :

(12) $L(\theta , d) = \alpha\, \mathbf{1}_{\Theta 0}(\theta) . \mathbf{1}_{\{d0\}}(d) + \beta\, \mathbf{1}_{\Theta 1}(\theta) . \mathbf{1}_{\{d1\}}(d)$,

avec $\alpha \geq 0$ et $\beta \geq 0$ (en notant d0 pour d_0 , $\Theta 0$ pour Θ_0 , etc). Elle signifie que le statisticien décide d_0 si $\theta \in \Theta_1$ et d_1 si $\theta \in \Theta_0$. La fonction de risque associée à L s'écrit :

(13) $R(q , d) = \mathbf{1}_{\Theta 0} . P_\theta([\delta(X) = d_0]) + \mathbf{1}_{\Theta 1} . P_\theta([\delta(X) = d_1])$.

(b) dans un **problème d'estimation ponctuelle** (cf **estimateur ponctuel**), on pose $D = \Theta$ (ie on décide une valeur unique pour le paramètre θ) ; dans un problème d'estimation ensembliste (cf **estimateur ensembliste**), on pose $D \subset \mathcal{P}(\Omega)$ (ie on décide d'une « zone » de valeurs vraisemblables pour θ).

(c) dans un **problèmes de décision multiple**, ou **problème de classification**, on pose Card D = M (eg $D = \{d_l ,..., d_M\}$). Des exemples en sont fournis par (α) le **problème de test à 3 modalités** (**test avec doute**) $\{d_1 , d_2 , d_3\}$, où Θ est partitionné selon $\{\Theta_0 , \Theta_1 , \Theta_2\}$, avec $\Theta_0 \neq \emptyset$, et $\Theta_1 \neq \emptyset$, et $\Theta_2 \neq \emptyset$. Par suite, Θ_2 est tq le statisticien décide d_3 (ie de ne pas conclure) si $\theta \in \Theta_2$, (β) le **problème de décision monotone**, (γ) le **problème de décision séquentielle** (eg **estimation séquentielle**, **test séquentiel**) ou encore (δ) le problème de **classification statistique**.