THÉORIE DES TESTS (1) (I)
(09 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

La **théorie des tests** suit, de façon générale, une démarche fondée sur une **table de décision à deux entrées et deux sorties**, parfois appelée **table (entrées-sorties) 2 x 2**, ou simplement **table 2 x 2**. Cette table élémentaire comporte en ligne deux décisions possibles, d_1 et d_2 , et en colonne deux états de la **Nature** (**événements aléatoires** ou contextes) possibles E_1 et E_2 .

Le **statisticien** peut alors raisonner selon deux approches classiques :

(a) **décision → conséquence** (cf tableau ci-après) :

théorie des tests : décisions → conséquences

état / décision	E_1	E_2
d_1	c_{11} (p_{11})	c_{12} (p_{12})
d_2	c_{21} (p_{21})	c_{22} (p_{22})

$(a)_1$ lorsque l'état E_1 est réalisé et que la décision d_1 est prise, une conséquence c_{11} survient (avec une probabilité p_{11} donnée), tandis qu'une conséquence c_{12} survient (avec la probabilité $p_{12} = 1 - p_{11}$) si la décision d_2 est prise ;

$(a)_2$ de même, lorsque l'état E_2 est réalisé : la décision d_1 entraîne une conséquence c_{21} (avec une probabilité p_{21}), tandis que la décision d_2 entraîne une conséquence c_{22} (avec une probabilité $p_{22} = 1 - p_{21}$) ;

(b) **décision → risque** (cf tableau ci-après) :

théorie des tests : décisions → risques

état / décision	E_1	E_2
d_1	$p_{11} = 1 - p_{12}$	p_{12}
d_2	p_{21}	$p_{22} = 1 - p_{21}$

$(b)_1$ si la décision d_1 est prise alors que l'état E_2 est réalisé, le statisticien subit un risque p_{12} , tandis que ce risque est $1 - p_{12}$ si l'état E_1 est réalisé ;

$(b)_2$ de même, si la décision d_2 est prise alors que l'état E_1 est réalisé, le statisticien subit un risque p_{21} , tandis que ce risque est $1 - p_{21}$ si l'état E_2 est réalisé.

C'est cette seconde approche qui est suivie en théorie des tests : d_1 y est notée d_0 et d_2 notée d_1 ou d_a ; p_{12} est notée β et $1 - p_{12}$ est notée $1 - \alpha$; enfin, p_{21} est notée α et $1 - p_{21}$ est notée $1 - \beta$ ou η (puissance du test) (cf tableau ci-après) :

théorie des tests : décisions $\rightarrow$ risques

état / décision	E_1	E_2
d_0	$1 - \beta$	β seconde espèce
d_1 ou d_a	α première espèce	$1 - \alpha$

Un **problème de test** est ainsi un **problème de décision** statistique dans lequel l'ensemble des décisions possède deux éléments (cas des tests à deux décisions usuels).

Il mobilise les concepts de base suivants : **risque de première espèce**, **risque de seconde espèce**, **fonction de test** (pure ou mixte), **fonction puissance**, **statistique de test** (définie à partir de l'échantillon X et dépendant de la propriété sujette à hypothèse), comparaison des fonctions de test à l'aide d'un **préordre** adéquat, et de **région de rejet** (ou **région critique**).

(i) Disposant d'un **modèle statistique** $(\Omega, \mathcal{T}, \mathcal{P})$ et d'un **espace d'observation** $(\mathcal{X}, \mathcal{B})$, le **statisticien**, qui observe un échantillon $X : \Omega \mapsto \mathcal{X}$, doit souvent rechercher si l'**hypothèse** plus « précise » H_0 , selon laquelle le « vrai » modèle statistique gouvernant le **phénomène** étudié est $(\Omega, \mathcal{T}, \mathcal{P}_0)$, avec $\mathcal{P}_0 \subset \mathcal{P}$ et $\mathcal{P}_0 \neq \mathcal{P}$, est vérifiée.

La **théorie des tests d'hypothèses**, ou simplement **théorie des tests**, a pour objet une telle vérification.

Plus généralement, elle peut aussi avoir pour objet de vérifier si le « vrai » modèle statistique satisfait à l'hypothèse :

(1) $\quad H_0 : (\Omega, \mathcal{T}, \mathcal{P}) = (\Omega, \mathcal{T}, \mathcal{P}_0)$, $\quad$ avec $\mathcal{P}_0 \subset \mathcal{P}$ et $\mathcal{P}_0 \neq \mathcal{P}$,

tout en s'assurant qu'il ne satisfait pas une autre hypothèse :

(2) $H_1 : (\Omega, \mathcal{T}, \mathcal{P}) = (\Omega, \mathcal{T}, \mathcal{P}_1)$, avec $\mathcal{P}_1 \subset \mathcal{P}$ et $\mathcal{P}_1 \neq \mathcal{P}$.

Dans un **problème de test**, les sous-familles $\mathcal{P}_0$ et $\mathcal{P}_1$ de $\mathcal{P}$ peuvent être :

(a) disjointes ($\mathcal{P}_0 \cap \mathcal{P}_1 = \emptyset$), auquel cas on parle d'**hypothèses disjointes**, ou non disjointes ($\mathcal{P}_0 \cap \mathcal{P}_1 \neq \emptyset$) (**hypothèses non disjointes**). On parle encore d'**hypothèses « séparées »** ;

(b) emboîtées ($\mathcal{P}_0 \subset \mathcal{P}_1$ ou $\mathcal{P}_1 \subset \mathcal{P}_0$), auquel cas on parle d'**hypothèses emboîtées**, ou non emboîtées (**hypothèses non emboîtées**).

Le cadre général précédent est sous-jacent à la plupart des problèmes de **test d'hypothèses**. On dit que l'on teste l'**hypothèse de base** :

(1)' $H_0 : P \in \mathcal{P}_0$,

aussi appelée **hypothèse privilégiée**, ou (dans certains **contextes statistiques**) **hypothèse nulle**, contre l'**hypothèse alternative** (ou simplement **alternative**) :

(2)' $H_1 : P \in \mathcal{P}_1$,

aussi appelée **hypothèse secondaire** ou **hypothèse concurrente**.

La théorie s'intègre à la **théorie de la décision** statistique. L'ensemble D des **décisions** est alors en **bijection** avec l'ensemble $\{\mathcal{P}_0 , \mathcal{P}_1\}$ et le choix d'une décision $d \in D$ équivaut au choix de l'une des deux parties $\mathcal{P}_i$ (i = 0, 1) de $\mathcal{P}$.

En pratique, la conduite d'un test d'hypothèses s'effectue, en général, directement sur le modèle $(\mathcal{X}, \mathcal{B}, \mathcal{P}^X)$, **modèle image** du modèle initial par X : les hypothèses portent alors sur les familles $\mathcal{P}_0^X$ et $\mathcal{P}_1^X$, images respectives des familles $\mathcal{P}_0$ et $\mathcal{P}_1$ par la **va** X.

Ceci conduit naturellement à formuler un **problème de décision** statistique (cf **théorie de la décision**), appelé **problème de test** (**d'hypothèses**). Ce problème peut être :

(a) un **problème de test « statique »** : dans ce cas, l'espace $(\mathcal{X}, \mathcal{B})$ est un espace produit, et la taille N de l'échantillon $X = (X_1 ,..., X_N)$, est fixée (ou donnée a priori) ;

(b) un **problème de test « dynamique »**, auquel cas on parle de **problème de test séquentiel**, et la taille de l'échantillon X peut augmenter jusqu'au moment où une condition requise a priori est vérifiée par la procédure de test mise en oeuvre (cf eg **test de WALD**, **test séquentiel**). Dans ce dernier, en effet, l'expérience (ou expérimentation) se poursuit jusqu'à une **décision finale**.

Autrement dit, un test « statique » aboutit à une décision sans faire appel à des observations supplémentaires, contrairement à un test séquentiel pour lequel la **décision finale** est prise au bout d'un certain nombre (de répétitions) d'expériences.

On présente deux approches formalisées de la théorie (statique), selon que le **modèle statistique** considéré se présente sous forme paramétrée ou non.

(ii) Un **problème de test paramétrique** consiste en la donnée :

(a) d'un modèle image ($\mathscr{X}$, $\mathscr{B}$, $(P_\theta^X)_{\theta \in \Theta}$), ie d'un modèle statistique obtenu comme image d'un modèle statistique initial (Ω, $\mathscr{T}$, $(P_\theta)_{\theta \in \Theta}$) par une **variable aléatoire** (ou **échantillon**) $X : \Omega \mapsto \mathscr{X}$, où ($\mathscr{X}$, $\mathscr{B}$) désigne un **espace d'observation**. Très généralement, $\mathscr{X} = \mathbf{R}^K$ (avec $K \geq 1$) et $\mathscr{B} = \mathscr{B}(\mathbf{R}^K)$;

(b) d'un **espace de décisions** (D, $\mathscr{B}_D$) tq $D = \{d_0 , d_1\}$ (ensemble à deux éléments), où $\mathscr{B}_D$ est la **tribu discrète** de D. L'ensemble D est défini, de façon biunivoque, à l'aide d'une **partition** $\{\Theta_0 , \Theta_1\}$ donnée de Θ, en sorte que :

(1)
si $\theta \in \Theta_1$, on décide d_0 ,
si $\theta \in \Theta_2$, on décide d_1 .

(c) d'une **fonction de perte** $L : D \times \Theta \mapsto \mathbf{R}_+$.

Concrètement, il s'agit de choisir, au vu d'une **observation** $\omega \in \Omega$, ou d'une **mesure** $X(\omega) = x \in \mathscr{X}$ faite sur elle (eg « résultat d'une **expérience** »), entre deux **hypothèses statistiques** : l'une, généralement notée H_0 , est l'hypothèse selon laquelle $\theta \in \Theta_0$, l'autre, notée H_1 , est l'hypothèse selon laquelle $\theta \in \Theta_1$. On appelle **hypothèse** (**statistique**) le fait pour θ d'appartenir soit à Θ_0 , soit à Θ_1 .

On dit alors que :

(a) l'hypothèse $H_0 : \theta \in \Theta_0$ est l'**hypothèse de base**, ou **hypothèse fondamentale**, ou **hypothèse privilégiée**, ou **hypothèse principale**, ou encore (dans certaines circonstances) l'**hypothèse nulle** du problème de test ;

(b) l'hypothèse $H_1 : \theta \in \Theta_1$ est l'**hypothèse alternative**, ou **hypothèse secondaire**, du problème. On note parfois même, par abus, $\theta \in H_0$ et $\theta \in H_1$.

L'hypothèse alternative est parfois notée H_a , auquel cas d_1 est notée d_a .

Dans le cadre de la **théorie de la décision**, l'ensemble D des **décisions** est tq, $w \in \mathscr{X}$ étant une **partie mesurable** donnée ;

(2)
si $X \in w$, on décide d_0 (acceptation de H_0),
si $X \in w^c$, on décide d_1 (acceptation de H_1).

La partie $w \in \mathscr{X}$ est appelée **région critique** car le fait pour l'échantillon X de tomber dans w conduit à « refuser » l'hypothèse de base.

Par suite, on appelle, dans ce **contexte** :

(a) **test pur** toute **règle de décision pure** (ou règle de décision non aléatoire), ie toute **application mesurable** :

(3) $\delta : \mathscr{X} \mapsto D$, où $D = \{d_0 , d_1\}$ (ensemble à deux éléments).

Un test pur δ est plus souvent noté φ, ou ψ, etc. C'est donc un **test mixte** particulier (cf aussi **fonction de test**).

Cette règle est appelée **test pur**, ou **test non aléatoire**, ou encore **test déterministe** : ainsi, une règle de décision δ équivaut au choix d'une partition de $\mathscr{X}$ selon $\{w, w^c\}$. Un **test pur** $\delta : \mathscr{X} \mapsto D$ est souvent appelé **fonction de test** (ou **fonction-test**), ou encore simplement un **test** ;

(b) **règle de décision aléatoire** t ou **règle de décision mixte** (ou encore **règle de décision « randomisée »**) le résultat d'un tirage aléatoire d'une décision d dans l'ensemble $D = \{d_0 , d_1\}$ au vu de $x = X(\omega)$. Ceci revient à définir une **probabilité de transition** m de $\mathscr{X} \times \mathscr{B}_D$ dans [0, 1] tq, par définition :

(4)
$$m_x(d) = \begin{cases} m(x, \{d_0\}) & \text{ssi } d = d_0 , \\ m(x, \{d_1\}) & \text{ssi } d = d_1 . \end{cases} \qquad \forall x \in \mathscr{X},$$

La fonction $m(x, \{d_0\})$ représente alors la probabilité de rejeter H_0 sachant que x a été observé.

On qualifie souvent la règle mixte $m : \mathscr{X} \times \mathscr{B}_D \mapsto [0,1]$ de **test mixte**, de **fonction de test** (**aléatoire ou mixte**), ou même simplement de **test**.

(iii) Un **problème de test non paramétrique** consiste en la donnée :

(a) d'un **modèle image** ($\mathscr{X}$, $\mathscr{B}$, $\mathscr{P}^X$), ie d'un **modèle statistique** obtenu comme image d'un modèle initial sous forme non paramétrée (Ω, $\mathscr{T}$, $\mathscr{P}$) par une va (ou **échantillon**) $X : \Omega \mapsto \mathscr{X}$;

(b) d'un **espace de décision** (D, $\mathscr{B}_D$) tq D = {d_0 , d_1} (ensemble à deux éléments), où $\mathscr{B}_D$ est la **tribu discrète** de D. L'ensemble D est défini à l'aide d'une partition {$\mathscr{P}_0$, $\mathscr{P}_1$} donnée de la famille $\mathscr{P}$, en sorte que :

si $P^\xi \in \mathscr{P}_0$, on décide d_0 ;

(1)'

si $P^\xi \in \mathscr{P}_1$, on décide d_1 ,

P^ξ désignant l'une quelconque des lois de probabilité qui ont engendré (ou qui sont susceptibles d'être suivies par) ξ ;

(c) d'une **fonction de perte** $L : D \times \; \mapsto \mathbf{R}_+$.

Il s'agit alors de choisir, au vu de $x = X(\omega)$, entre d_0 , qui correspond à l'hypothèse $H_0 : P \in \mathscr{P}_0$, et d_1 , qui correspond à l'hypothèse $H_1 : P \in \mathscr{P}_1$.

La terminologie précédente est encore utilisée ici, et l'on définit les notions de **région critique** (associée à un test d'hypothèse) (cf **région d'acceptation**, **région de confiance**), de **règle de décision pure** et de **règle de décision mixte** (ie les tests purs et les tests aléatoires).

(iv) Remarques :

(a) hormi les généralités du début, on a supposé principalement que H_0 et H_1 (ou H_a) étaient des hypothèses « disjointes », ie que $\Theta_0 \cap \Theta_1 = \varnothing$ (resp $\mathscr{P}_0 \cap \mathscr{P}_1 = \varnothing$). La théorie a été développée au cas d'**hypothèses non disjointes** (ie $\Theta_0 \cap \Theta_1 \neq \varnothing$) ou d'**hypothèses emboîtées** (ie $\Theta_0 \subset \Theta_1$ ou $\Theta_1 \subset \Theta_0$) ;

(b) un problème de test est essentiellement un **« problème de choix »**, ou un **« problème d'arbitrage »**, entre H_0 et H_1 , ie entre deux éventualités qui peuvent être à l'origine des observations X. La loi P_θ^X (resp P^X) qui engendre X appartient soit à une sous-famille non vide $\mathscr{P}_0^X$, soit à la sous-famille (éventuellement complémentaire) (supposée non vide) $\mathscr{P}_1^X = (\mathscr{P}_0^X)^c$, et ceci au vu de la réalisation $X(\omega) = x$;

(c) la **fonction de perte** L associée à un test pur $\varphi : \mathcal{X} \mapsto D$ vérifie naturellement (eg cas paramétré) :

(5) $L(d_i, \theta) = 0, \quad \forall \theta \in \Theta_i, \quad$ où $i \in \{0, 1\}$.

La **fonction de risque** associée est donc :

(6) $R(w, \theta) = \mathbf{1}(\theta \in H_0) \,.\, P_\theta(w) \,.\, L(d_1, \theta) + \mathbf{1}(\theta \in H_1) \,.\, (1 - P_\theta(w)) \,.\, L(d_0, \theta)$,

où $\theta \mapsto \alpha_\theta(w)$ est la **(fonction de) risque de première espèce** et $\theta \mapsto \beta_\theta(w)$ est la **(fonction de) risque de seconde espèce** (cf **risque de première espèce**, **risque de seconde espèce**). La fonction de risque R (w, θ) définie en (6) est aussi notée $R_\theta(w)$.

Le **préordre** ≥ associé au risque R est donc défini selon :

(7) $w' \geq w'' \quad \Leftrightarrow \quad \{\alpha_\theta(w') \leq \alpha_\theta(w'') \text{ sur } \Theta_0\}$ et $\{\beta_\theta(w') \leq \beta_\theta(w'') \text{ sur } \Theta_1\}$.

Ce préordre ne permet cependant pas d'obtenir un **test optimal**. Différents principes statistiques (**principe bayésien**, **principe de NEYMAN**, **principe d'invariance**) permettent généralement de résoudre ce problème.

(d) dans le cadre de la **théorie bayésienne**, si $\mathcal{B}_\Theta$ est une **tribu de parties** de Θ et si Π est une **probabilité a priori** définie sur $\mathcal{B}_\Theta$, le **risque de BAYES** du test δ (ou de sa **région critique** associée w) correspondant à Π s'écrit :

(8) $R_\Pi(w) = \int R(w, \theta)\, d\Pi(\theta)$.

(v) La notion de **statistique de test** est centrale dans cette théorie. Il s'agit d'une statistique tq S = s (X) qui « caractérise » l'hypothèse principale (ou, alternativement, l'hypothèse secondaire) et s'interprète en conséquence. Ainsi, une valeur « élevée » de S peut signifier que l'une des hypothèses est vraisemblable (resp peu plausible), selon l'interprétation.

La **loi de probabilité**, à distance finie ou la **loi asymptotique** (grand échantillon X, avec N >> 0) P^S de S peuvent être définies (ou calculées), entièrement ou partiellement (eg moments), au moins sous l'hypothèse de base H_0 : toute **région critique** de **seuil** donné a priori $\alpha \in]0, 1[$, définie à l'aide de P^S, permet alors de conclure le test.

(vi) En pratique, les **risques d'espèce**, ou **« types » de risques**, jouent aussi un rôle important. L'idéal serait de les minimiser tous les deux.

Cependant, à cause de leur caractère antagoniste, la procédure la plus courante consiste à fixer (ou à majorer) a priori le risque de première espèce, puis à minimiser le risque de seconde espèce, ie à maximiser la **fonction puissance**. Cette procédure fonde la **théorie de J. NEYMAN - E.S. PEARSON**. Compte tenu du modèle et des observations, elle aboutit soit à accepter l'hypothèse H_0 , soit à accepter (ie à ne pas refuser) l'hypothèse H_1 . Il s'agit donc d'un « arbitrage par préférence » et non pas d'un « jugement ».

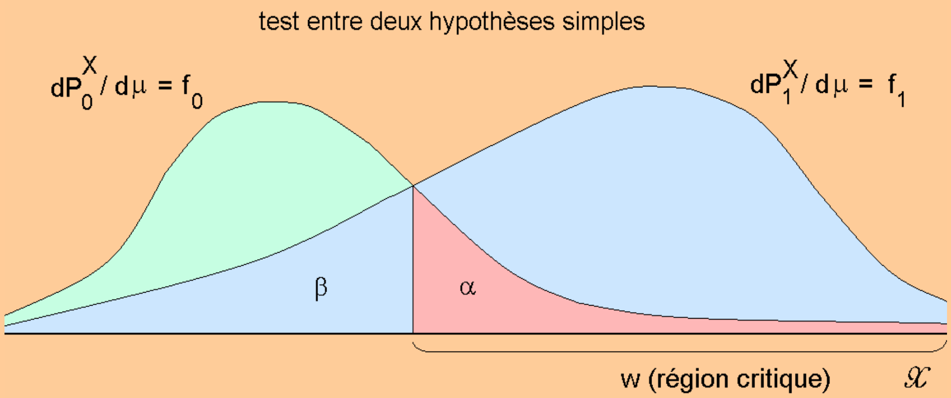
(vi) D'autres procédures de test ont été recherchées, notamment dans le cadre du **problème de décision multiple** :

(a) **tests à plusieurs conclusions**. Les notions de test ci-dessus sont des notions comportant deux « issues » (ou « possibilités »). Il existe des **tests à plusieurs issues**. Ainsi dans un test à trois issues, D = $\{d_a , d_d , d_r\}$, d_a représente l'**acceptation** de H_a ($\theta \in \Theta_a$, dans le cas paramétré), d_r l'acceptation de H_r ($\theta \in \Theta_r$) (ou **refus** de H_a) et d_d le « **doute** », ie l'acceptation d'une troisième hypothèse H_d : $\theta \in \Theta_d$), où $\{\Theta_a , \Theta_d , \Theta_r\}$ est une partition donnée de Θ) à laquelle s'associe naturellement une partition $\{w_a , w_d , w_r\}$ de l'ensemble des observations $\mathscr{X}$;

(b) **problème de classification**. Ce type de **problème statistique** constitue aussi une généralisation à plusieurs issues du test d'hypothèses du type décrit ci-dessus (cf **classification**, **discrimination**, etc).

(vii) Par ailleurs, la théorie des tests entretient des liens étroits avec la **théorie de l'estimation** « ensembliste » (cf **estimateur ensembliste**) : le complémentaire d'une **région de confiance** (dotée d'un **seuil de confiance** 1 - α) est alors une **région de rejet** (dotée d'un **seuil critique** α).

test d'une hypothèse H_0 : $P_0^X \in \mathscr{P}_0^X$ contre une alternative H_1 : $P_1^X \in \mathscr{P}_1^X$



(vii) Le terme **risque** peut désigner :

(a) selon la terminologie de la théorie des tests, une probabilité tq les p_{ij} (cf supra) ;

(b) selon une autre acception, le couple (p_{ij} , c_{ij}) associant une probabilité p_{ij} et une conséquence c_{ij} (valeur ou magnitude). Cette notion, plus riche, permet de « séparer » ce qui relève d'un contexte aléatoire (eg « hasard ») et ce qui relève d'un contexte exogène (eg économique : budget, assurance, etc). En effet, elle permet de définir des arbitrages entre situations associées aux couples du type précédent : eg arbitrage entre une situation avec probabilité élevée ($p >> 0$) et conséquence faible ($c << \infty$), et une situation inverse, avec probabilité faible ($p << 0$) et conséquence élevée ($c >> \infty$).

Les tables de décision indiquées au début, qui peuvent s'appliquer à d'autres contextes, se généralisent à plusieurs entrées / sorties, dont le nombre n'est pas nécessairement égal (cf tableau m x n ci-après).

décisions multiples : décisions → conséquences

état / décision	E_1	...	E_n
d_1	c_{11} (p_{11})	...	c_{1n} (p_{1n})
...	...	c_{ij} (p_{ij})	...
d_m	c_{m1} (p_{m1})	...	c_{mn} (p_{mn})

Par suite, la **fonction d'utilité** υ du statisticien peut s'écrire $C \mapsto u = \upsilon(C)$, avec pour argument la (m,n)- **matrice** $C = (c_{ij})_{(i,\ j)}$ des conséquences. La **décision statistique** (multiple) peut donc se prendre à partir de u (C) et de la matrice $P = (p_{ij})_{(i,\ j)}$ des probabilités, ce qui revient à utiliser la notion d'utilité espérée (espérance de l'utilité calculée avec P).