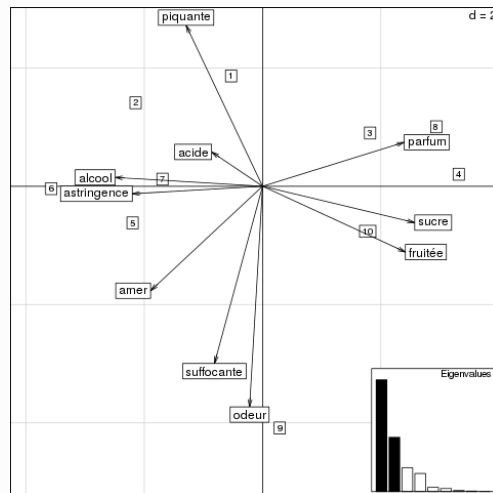


### Exemple : biplot



Après avoir vérifié que les individus étaient bien représentés sur le plan factoriel en examinant les valeurs des cosinus carrés des angles entre les individus et les axes factoriels, il semble se dégager 4 groupes de cidres :

- ▶ groupe 1 : les cidres 3, 8, 4 et 10 qui sont des cidres doux,
- ▶ groupe 2 : les cidres 2, 5, 6 et 7 qui sont des cidres bruts,
- ▶ groupe 3 : le cidre 9 qui est un cidre particulièrement odorant et suffocant,
- ▶ groupe 4 : le cidre 1 qui est particulièrement piquant.

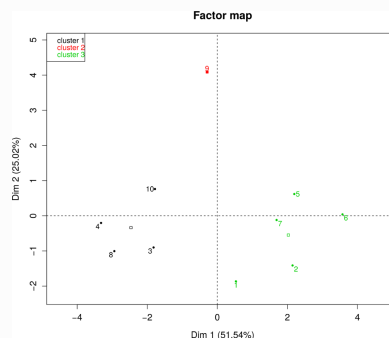
56/120

## 2.5. Classification ascendante hierarchique : Généralité

### Principe

**Contexte** : On dispose de  $n$  produits  $M_1, \dots, M_n$  décrits par  $p$  grandeurs. Comme précédemment, on représente les données sous la forme d'un nuage  $\Gamma$  de  $n$  points dans  $\mathbb{R}^p$ .

Exemple : retour sur les cidres. On a le nuage de points suivant sur le premier plan factoriel :



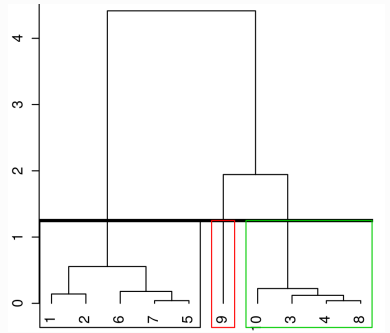
**Classification** : On souhaite partitionner le nuage. En général, on cherche à créer des classes les plus homogènes possibles.

57/120

### Dendrogramme

**Dendrogramme** : C'est une représentation en arbre. On crée facilement des sous-groupes en "coupant" le dendrogramme à un niveau donné : tous les individus solidaires d'une même branche constituent un groupe.

Exemple : le dendrogramme associé au nuage de points précédent



**CAH** : C'est une méthode de classement "automatique". On crée tout d'abord un dendrogramme puis on fixe le nombre de groupe *a posteriori* uniquement.

### Distances

**Distance entre produits** : On la note  $d$  et on définit la matrice de distance  $D = [D_{ij}]_{i,j=1}^n$  qui satisfait

$$D_{ij} = d(M_i, M_j), \text{ où } i, j = 1, \dots, n.$$

C'est une matrice  $n \times n$ , symétrique, de diagonale nulle : on a donc seulement  $\frac{n(n-1)}{2}$  entrées à calculer.

Exemple : distance euclidienne, distance  $\ell^1$ ...

**Écart entre les classes** : À l'aide de la distance  $d$  on définit une distance *dist* entre deux sous-groupes  $\Gamma_1$  et  $\Gamma_2$  de  $\Gamma$ .

Exemple : La distance moyenne. Distance du saut minimum (*i.e* distance des points les plus proches). Distance du diamètre (*i.e* distance entre les points les plus éloignés). Distance de Ward...

## Espace Euclidien

**Distance Euclidienne** : C'est la distance "habituelle". Soit  $M_1$  et  $M_2$  deux produits on a

$$d_2(M_1, M_2) = \sqrt{(M_1^1 - M_2^1)^2 + \dots + (M_1^p - M_2^p)^2}.$$

**Distance de Ward** : Soit  $\Gamma_1$  et  $\Gamma_2$  deux sous groupes de points de  $\Gamma$ . L'effectif de  $\Gamma_k$  est noté  $|\Gamma_k|$  et son centre de gravité est  $G_k = \frac{1}{|\Gamma_k|} \sum_{M_i \in \Gamma_k} M_i$ . On définit la distance de Ward par

$$dist_2(\Gamma_1, \Gamma_2) = \sqrt{\frac{p_1 p_2}{p_1 + p_2} d_2(G_1, G_2)}$$

où  $p_k = \frac{|\Gamma_k|}{n}$ ,  $k = 1, 2$ .

**Inertie** : On appelle inertie d'un sous-nuage  $\Gamma_0 = \{M_1, \dots, M_{n_0}\}$  de  $\Gamma$  la somme pondérée des carrés des distances de ses points au centre de gravité du nuage. Si  $G_0$  désigne le centre de gravité de  $\Gamma_0$  on a

$$\mathcal{I}(\Gamma_0) = \frac{1}{n} \sum_{M_i \in \Gamma_0} (d(M_i, G_0))^2$$

Attention les poids  $\frac{1}{n}$  sont fixés : on a  $\mathcal{I}(\Gamma_0) = \frac{n_0}{n} Var(\Gamma_0) \dots$

60/120

## Décomposition de l'inertie

**Theorem de Huygens** : Soit  $\Gamma$  un nuage de points partitionné en classes  $\Gamma_1, \dots, \Gamma_k$  (i.e disjointes deux à deux et recouvrant tout  $\Gamma$ ). On a

$$\mathcal{I}(\Gamma) = \mathcal{I}_{intra} + \mathcal{I}_{inter}$$

où l'inertie intra est la somme des inerties :

$$\mathcal{I}_{intra} = \mathcal{I}(\Gamma_1) + \dots + \mathcal{I}(\Gamma_k)$$

et l'inertie inter mesure la dispersion des classes autour du centre de gravité  $G$  de  $\Gamma$  :

$$\mathcal{I}_{inter} = p_1(d(G_1, G))^2 + \dots + p_k(d(G_k, G))^2,$$

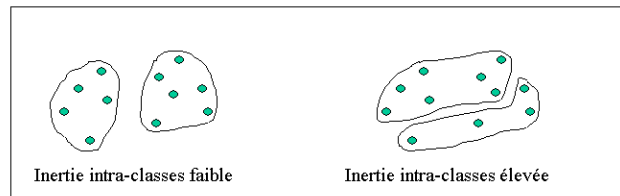
où  $p_\ell = \frac{|\Gamma_\ell|}{n}$ ,  $\ell = 1, \dots, k$ .



61/120

### Principe de la méthode

**Partition :** La qualité globale d'une partition est liée à l'homogénéité interne des sous-nuages et donc également à l'éloignement des sous-nuages.



**Principe :** A chaque étape de l'algorithme on regroupe les deux classes dont l'agrégation minimise le gain d'inertie intraclasse (i.e minimise la perte d'inertie interclasse). Étant données deux sous nuages  $\Gamma_1$  et  $\Gamma_2$  d'un nuage  $\Gamma$ , la perte d'inertie inter est d'après le théorème de Huygens,

$$p_1 d_2(G_1, G) + p_2 d_2(G_2, G) = \frac{p_1 p_2}{p_1 + p_2} d_2(G_1, G_2)$$

Autrement dit, c'est la distance de Ward (au carré) entre les deux classes  $\Gamma_1$  et  $\Gamma_2$ .

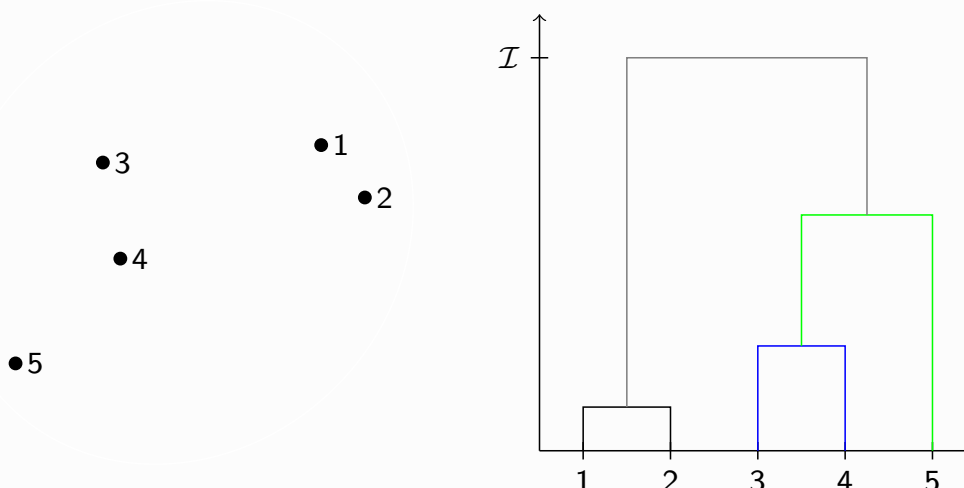
62/120

## 2.5. Classification ascendante hierarchique : Méthode de Ward

### Algorithme

**Dendrogramme :** on construit itérativement :

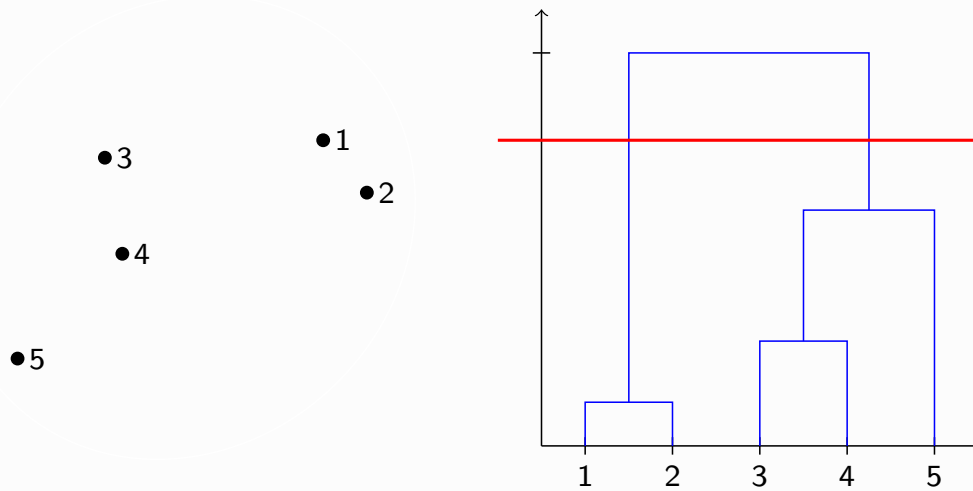
- ▶ Initialisation : chaque individu est seul dans une classe (i.e  $\mathcal{I} = \mathcal{I}_{inter}$ ).
- ▶ Règle de progression : on rassemble les deux classes les plus proches au sens de la distance de Ward.
- ▶ Règle d'arrêt : quand tous les individus sont dans la même classe (i.e  $\mathcal{I} = \mathcal{I}_{intra}$ ).



La hauteur de chaque nœud du dendrogramme est choisie ici proportionnelle à l'inertie intra. Mais attention, dans diverses implémentations de la méthode de Ward pour la CAH, la hauteur du noeud reliant  $\Gamma_i$  et  $\Gamma_j$  est  $\sqrt{2} \text{dist}_2(\Gamma_i, \Gamma_j) \dots$

63/120

### Classes



**Classes :** Le choix du nombre de classe est déterminé *a posteriori* en coupant le dendrogramme niveau donné et tous les individus solidaires d'une même branche constituent un groupe.

## Sommaire

### 3. Inférence et décision

#### 3.1 Rappels sur les tests statistiques

- Les hypothèses
- Niveau et puissance du test
- Décision

#### 3.2 Tester une différence entre produits

- Test triangulaire
- Épreuve à choix forcés (k-AFC)
- Méthode Séquentielle de Wald
- Épreuve d'appariement
- Test d'indiscernabilité globale
- Test de confusion de deux des produits

#### 3.3 Classer des produits

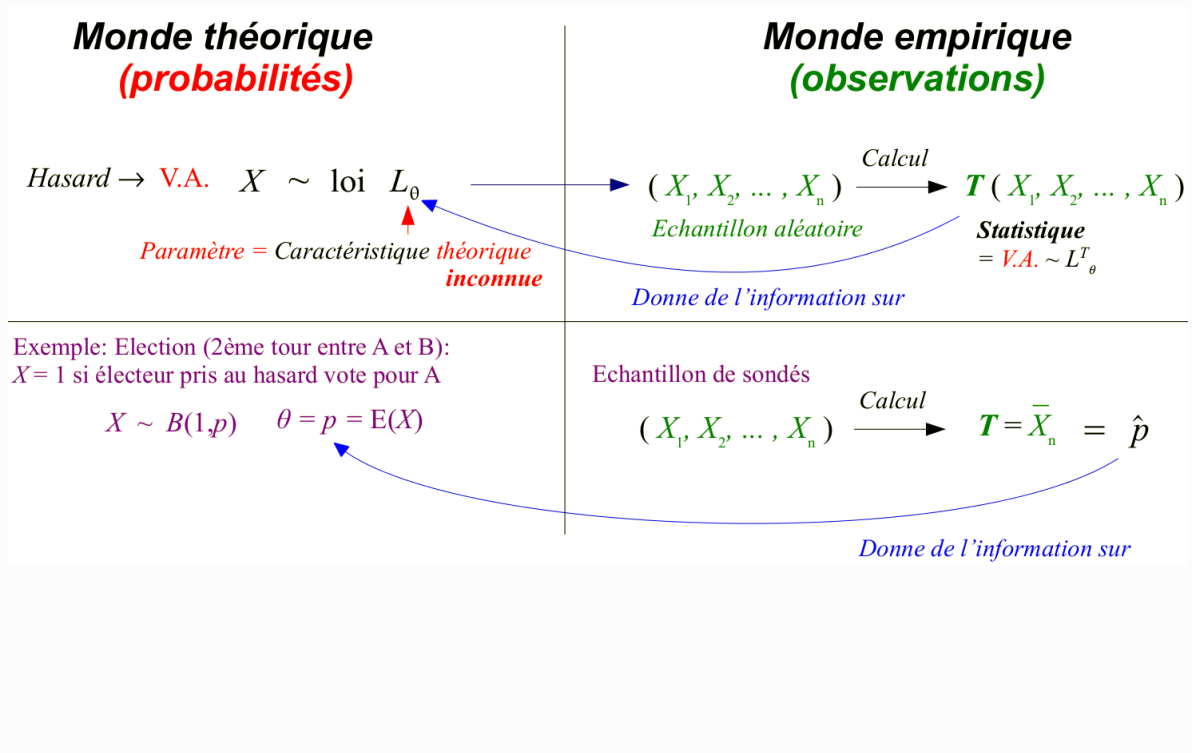
- Test de Friedman
- Test de Page
- Tests de comparaison multiples

#### 3.4 Comparer deux produits sur une échelle continue

- Le cas gaussien
- Le cas non gaussien

## 3.1. Rappels sur les tests statistiques :

### Principe



65/120

## 3.1. Rappels sur les tests statistiques :

### Test d'une hypothèse concernant le modèle

**Objectif :** Au vu des observations, peut-on considérer comme "valide" une certaine hypothèse  $H_0$  concernant le modèle, ou ces observations la "contredisent"-elles trop manifestement, en faveur d'une hypothèse alternative  $H_1$  ?

Exemple : on dispose des notes d'amertume de 36 juges concernant deux liqueurs de gentiane A et B. La gentiane A est en moyenne notée plus amère que B par les juges. Mais nous savons que les fluctuations d'échantillonnage existent. Au vu de ces observations particulières, peut-on considérer qu'en général, la gentiane A sera trouvée plus amère en moyenne que la B ?

On a le modèle suivant :  $X_i = A_i - B_i = a + \varepsilon_i$  avec  $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ ,  $a = \mathbb{E}(X)$ . On a donc  $X \sim \mathcal{N}(a, \sigma^2)$  et

$$H_0 : \quad ; \quad H_1 :$$

**Principe :** On calcule à partir des observations une quantité adéquate  $T = T(X_1, \dots, X_n)$ . Typiquement, un estimateur du paramètre à tester qui, sous  $H_0$ , a un certain comportement connu, et sous  $H_1$ , un autre comportement. On regarde alors si la valeur qu'on a calculée pour  $T$  est suffisamment "peu probable" sous  $H_0$  par rapport à ce qu'elle l'est sous  $H_1$ . Si tel est le cas, on rejette  $H_0$  en faveur de  $H_1$ .

66/120

### Choix des hypothèses

**Les hypothèses :** Nous sommes dans un modèle paramétrique : les hypothèses concernent donc un paramètre  $\theta$  qui appartient à un ensemble  $\Theta \subset \mathbb{R}^n$  décrivant la loi de la v.a  $X$ . On les notes

- ▶  $H_0(\theta)$  : hypothèse nulle (on cherche à la *réfuter*)
- ▶  $H_1(\theta)$  : hypothèse alternative (on cherche à la *démontrer*)

Le choix des hypothèses est primordial et souvent délicat car il oriente la décision en modifiant les priorités.

Remarque : il est indispensable de connaître la loi de la statistique de test sous  $H_0$ .

**En pratique :** on choisit  $\mathcal{V}_0$  et  $\mathcal{V}_1$  deux sous-ensembles de  $\Theta$  avec  $\mathcal{V}_0 \cap \mathcal{V}_1 = \emptyset$ . Les hypothèses sont de la forme  $H_0 : \theta \in \mathcal{V}_0$  et  $H_1 : \theta \in \mathcal{V}_1$ . On parle d'hypothèse :

- ▶ "simple" :  $\mathcal{V}_i$  est un singleton (i.e  $H_0 : \theta = \theta_0$ ).
- ▶ "composite" :  $\mathcal{V}_i$  est un ensemble.

Exemple : Si  $\Theta \subset \mathbb{R}$  et  $H_0 : \theta = \theta_0$ , on dit que le test est

- ▶ unilatéral si  $H_1 : \theta > \theta_0$  ou  $H_1 : \theta < \theta_0$  (on test si "plus ou moins", si "améliore ou détériore").

$$\Theta$$


---

- ▶ bilatéral si  $H_1 : \theta \neq \theta_0 = \theta \in ] - \infty, \theta_0[ \cup ] \theta_0, +\infty[$  (i.e on test si "différent de", si "à un effet").

$$\Theta$$


---

### Allégorie du tribunal

Un tribunal doit statuer sur l'innocence ou la culpabilité d'un prévenu (décision) au vu d'un dossier (observations). Il y a deux erreurs possibles :

1. condamner un innocent
2. gracier un coupable

Tant qu'il reste une part d'incertitude, la seule façon d'être sûr de :

- ▶ ne jamais condamner un innocent est de relaxer systématiquement le prévenu.  
⇒ on maximise le risque de relaxer un coupable.
- ▶ ne jamais relaxer un coupable est de condamner systématiquement le prévenu.  
⇒ on maximise le risque de condamner un innocent.

**Hypothèse "nulle" :** On choisit une hypothèse "par défaut", qui sera maintenue sauf signes "suffisants" du contraire. Dans un tribunal pénal, le prévenu est présumé innocent sauf indications suffisantes du contraire (présomption d'innocence).

**Hypothèse alternative :** Le "contraire" de l'hypothèse nulle. Dans un tribunal pénal, c'est la culpabilité du prévenu, qu'il s'agit d'établir avec un dossier suffisamment étayé.

*On souhaite en priorité limiter le risque de condamner un innocent. On ne condamnera le prévenu que si le dossier est significativement incompatible avec l'innocence.*

#### Exemple : contamination alimentaire

Une norme alimentaire fixe la quantité moyenne théorique  $\tau$  (en grammes) d'une aflatoxine dans des pistaches ensachées. Il doit être  $< 10^{-3}$ g. On prélève 100 sachets dans une production de pistaches et on dose l'aflatoxine. Le taux empirique moyen observé est de  $7 \cdot 10^{-4}$ g et les mesures ont une variance de  $4 \cdot 10^{-8}$ .

On a observé un taux moyen empirique  $< 10^{-3}$ g mais il faut compter avec la variabilité de l'échantillonnage : soit l'échantillon est "conforme" à la réalité (alors  $\tau < 10^{-3}$ ), soit "on n'a pas eu de chance" et l'échantillon trahit la réalité (alors  $\tau \geq 10^{-3}$ ).

Les hypothèses ne sont pas symétriques! On peut limiter le risque sanitaire de manière prioritaire :

$$\begin{cases} H_0(\tau) : \tau \geq 10^{-3} & \text{les pistaches sont considérées toxiques par défaut...} \\ H_1(\tau) : \tau < 10^{-3} & \text{sauf si les mesures inclinent suffisamment dans l'autre sens.} \end{cases}$$

ou limiter le risque économique pour l'entreprise de manière prioritaire :

$$\begin{cases} H_0(\tau) : \tau \leq 10^{-3} & \text{les pistaches sont considérées non toxiques par défaut...} \\ H_1(\tau) : \tau > 10^{-3} & \text{sauf si les mesures inclinent suffisamment dans l'autre sens.} \end{cases}$$

### 3.1. Rappels sur les tests statistiques : Niveau et puissance du test

#### Risque de 1ère espèce

Chaque décision (accepter  $H_0$  ou rejeter  $H_0$  en faveur de  $H_1$ ) comporte un risque :

décision \ réal.	$H_0$ vraie	$H_1$ vraie
	accepte $H_0$	$1 - \alpha$
accepte $H_1$	$\alpha$	$1 - \beta$

**Risque de 1er espèce** : noté  $\alpha$  c'est la probabilité de rejeter  $H_0$  à tort (i.e choisir  $H_1$  alors que  $H_0$  est vraie). Si  $H_0$  est simple (i.e de la forme  $H_0 : \theta = \theta_0$ ) on a :

$$\alpha = \mathbb{P}_{\theta_0}(\text{choisir } H_1) = \text{niveau du test}$$

Il est fixé *a priori* car c'est le risque que l'on veut contrôler en priorité.

En pratique, pour les hypothèse  $H_0$  composite, on choisit l'hypothèse nulle  $H_0 : \theta \in \mathcal{V}_0$  avec  $\mathcal{V}_0$  un intervalle fermé. Cela permet de contrôler le niveau du test, car on a alors :

$$\alpha = \mathbb{P}_{H_0}(\text{choisir } H_1) \leq \sup_{\theta_0 \in \mathcal{V}_0} \mathbb{P}_{\theta_0}(\text{choisir } H_1)$$



#### Choix de la statistique

**Statistique de test :** Une fois choisie  $H_0$  et  $H_1$ , on cherche une statistique de test  $T = T(X_1, \dots, X_n)$  qui jouera le rôle de variable de décision. On doit connaître sa loi sous  $H_0$  (son comportement doit être différent sous  $H_1$ ). Le choix de  $T$  n'est pas unique : en général on prend le meilleur estimateur qu'on ait du paramètre  $\theta$  (il donne l'information la plus précise sur  $\theta$ ).

Exemple de la contamination alimentaire avec le principe de précaution sanitaire :

$$\begin{cases} H_0(\tau) : \tau \geq 10^{-3} & \text{les pistaches sont considérées toxiques par défaut. . .} \\ H_1(\tau) : \tau < 10^{-3} & \text{sauf si les mesures inclinent suffisamment dans l'autre sens.} \end{cases}$$

Le meilleur estimateur sans biais de  $\tau$  est  $\bar{X}$ . On le normalise et on pose  $T = \sqrt{n} \frac{(\bar{X} - \tau)}{\sigma_0}$

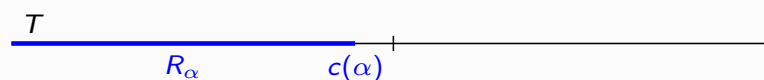
71/120

#### Région de rejet

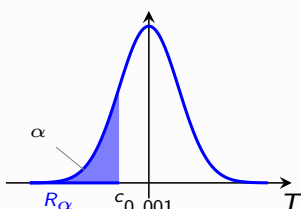
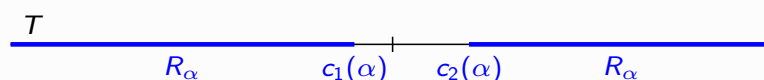
**Région de rejet :** Étant donné un niveau  $\alpha \in [0, 1]$ , on appelle région de rejet  $R_\alpha$  un ensemble de valeurs de la statistique de test  $T$  pour lesquelles  $H_0$  est rejeté. On a par définition

$$\mathbb{P}_{H_0}(\text{rejeter } H_0) = \mathbb{P}_{H_0}(T \in R_\alpha) = \alpha$$

L'allure de la région de rejet est dictée par la forme de  $H_1$ . Par exemple pour un test unilatéral on a  $R_\alpha = [c(\alpha), +\infty[$  ou  $R_\alpha = ]-\infty, c(\alpha)[$



et pour un test bilatérale on a  $R_\alpha = ]-\infty, c_1(\alpha)[ \cup ]c_2(\alpha), +\infty[$ .



... dans notre exemple on souhaite montrer que  $\tau < 10^{-3}$  : on étudie le comportement de  $T$  sous les lois dans  $H_0$ . On prend  $R_\alpha = ]-\infty, c_\theta(\alpha)[$  et on choisit la loi (i.e le  $\theta$ ) pour laquelle il est le plus difficile de rejeter  $H_0$  (i.e "qui est la plus proche de  $H_1$ "). C'est  $\theta = 10^{-3}$ .

72/120

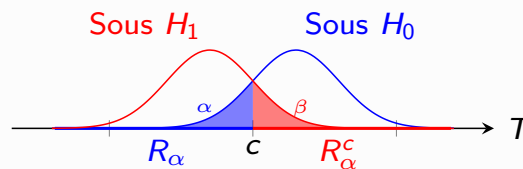
#### Risque de 2ème espèce

**Risque de 2ème espèce** : c'est la probabilité d'accepter  $H_0$  à tort (i.e choisir  $H_0$  alors que  $H_1$  est vraie). Si  $H_1 : \theta \in \mathcal{V}_1$ , on a pour tout  $\theta_1 \in \mathcal{V}_1$

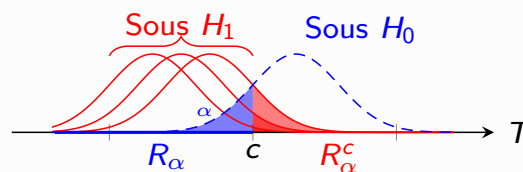
$$\beta(\theta_1) = \mathbb{P}_{\theta_1}(\text{choisir } H_0)$$

Le risque  $\alpha$  étant fixé, le risque  $\beta(\theta_1)$  sera déterminé comme le résultat d'un calcul (c'est donc une fonction de  $\alpha$ ). Le risque  $\beta$  varie en sens contraire de  $\alpha$  : si l'on veut diminuer  $\alpha$  on est conduit à ne rejeter  $H_0$  que dans des rares cas. On conserve donc  $H_0$  (bien souvent) à tort, donc on augmente  $\beta$ .

Comportement de  $T$  :



$\beta$  est calculable si on connaît exactement la loi de  $T$  sous  $H_1$ .



73/120

#### Puissance d'un test

**Définition** : La puissance est la capacité d'un test de détecter la véracité de  $H_1 : \theta \in \mathcal{V}_1$ . Lorsque  $H_1$  est composite, la puissance est définie pour tout  $\theta \in \mathcal{V}_1$  par :

$$\pi(\theta) = 1 - \beta(\theta) = 1 - \mathbb{P}_{\theta \in \mathcal{V}_1}(\text{rejeter } H_0) \in [0, 1]$$

Abaisser le niveau  $\alpha$  d'un test abaisse sa puissance. De manière générale,  $\alpha$  est à maîtriser en priorité... mais il est important de vérifier que cette maîtrise n'est pas trop cher payée par le risque de 2ème espèce  $\beta$ .

**Objectif en théorie des tests** : à niveau  $\alpha$  fixé on veut trouver les tests les plus puissants possible.

**En pratique** : Imaginons une procédure de test qui ne rejette pratiquement jamais  $H_0$  : son niveau  $\alpha$  est très bas... mais la probabilité  $\beta$  d'accepter  $H_0$  à tort devient grande. Autrement dit, on ne choisit jamais  $H_1$  et la puissance du test est alors très faible ce qui le rend inutile pour la décision. Si c'est le cas on peut :

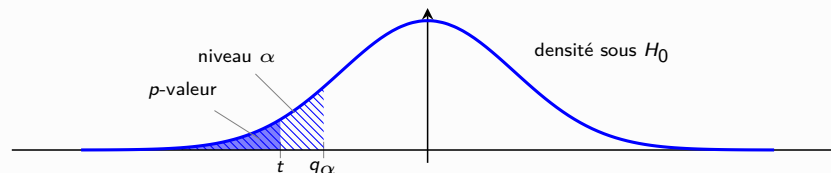
- ▶ baisser ses prétentions sur le niveau  $\alpha$ .
- ▶ augmenter la taille  $n$  de l'échantillon pour avoir une information plus précise.
- ▶ changer la procédure de test pour améliorer sa puissance.

74/120

#### P-valeur et Décision

**Décision :** La décision finale (rejet de  $H_0$  en faveur de  $H_1$  ou non) dépend du niveau  $\alpha$  du test. Ce niveau est (trop !) souvent arbitrairement fixé à 5%. Mais suivant le problème, c'est un risque acceptable ou non. . .

**La  $p$ -valeur :** c'est une v.a parfois appelée probabilité critique ou seuil de signification ou encore  $p$ -value. Étant donné des observations, c'est le plus petit niveau du test pour lequel on rejette  $H_0$ . En général, si la  $p$ -valeur est très petite devant  $\alpha$  (la valeur  $t$  de  $T$  n'est que très peu vraisemblable sous  $H_0$ ) alors on rejette  $H_0$  "très largement".



Exemple : Pour un test unilatéral, avec  $H_0 : \theta = \theta_0$  simple,  $T$  statistique de décision continue et une région de rejet de type  $R_\alpha = ]-\infty, c(\alpha)]$ , la  $p$ -valeur est le plus petit  $P \in [0, 1]$  qui satisfait

$$P = \mathbb{P}_{\theta_0}(T \leq c(P))$$

Remarque :  $P$  est une v.a qui, sous  $H_0$ , suit une loi uniforme sur  $[0, 1]$ .

**Avantages :** il n'y a pas plus de travail que pour un test "standard" (on détermine un quantile dans tous les cas) et elle apporte plus d'information (on sait si la décision se fait "largement" ou pas).

#### Exemple (fin)

Récapitulons le déroulement d'un test :

1. **Modélisation :** mesure d'aflatoxine  $X : X = \tau + \varepsilon$  avec  $\varepsilon \sim \mathcal{N}(0, \sigma_0^2)$  où  $\sigma_0$  est connu. On a donc  $\tau = \mathbb{E}(X)$  et  $X = \mathcal{N}(\tau, \sigma_0^2)$
2. **Choix des hypothèses :** principe de précaution sanitaire :  $\begin{cases} H_0(\tau) : \tau \geq 10^{-3} \\ H_1(\tau) : \tau < 10^{-3} \end{cases}$
3. **Choix du risque de 1ère espèce :** e.g  $\alpha = 0.05$ .
4. **Choix d'une statistique de test :**  $T = \sqrt{n} \frac{\bar{X} - \tau}{\sigma_0} \sim \mathcal{N}(0, 1)$  car  $\bar{X} \sim \mathcal{N}(\tau, \frac{\sigma_0^2}{n})$ .
5. **Allure de la région de rejet :**  $R_\alpha = ]-\infty, c(\alpha)]$ .
6. **Détermination de la région de rejet par la limitation du risque de 1ère espèce :**  $H_0$  est composite, on limite à  $\alpha$  la probabilité de rejet à tort de  $H_0$ , pour la "pire" des valeurs de  $\tau$  dans  $H_0$  : c'est  $\tau = 10^{-3}$ .

$$\mathbb{P}_{\tau \geq 10^{-3}}(\{\text{rejette } H_0\}) \leq \mathbb{P}_{\tau = 10^{-3}}(\{\text{rejette } H_0\}) = \mathbb{P}_{\tau = 10^{-3}}(T < u_{0.05})$$

Avec une table de quantile on détermine  $u_{0.05} = -1.65$  et  $R_\alpha = ]-\infty, -1.65]$ .

7. **Décision :** la valeur observée pour  $T_n$  est  $t_n = \sqrt{100} \frac{7 \cdot 10^{-4} - 10^{-3}}{2 \cdot 10^{-4}} = -15$ . On a donc  $t_n \in R_\alpha$  et on rejette  $H_0$  en faveur de  $H_1$  avec un risque (très) inférieur à 5% de se tromper. . .

### Introduction

**Objectifs :** Tester au choix la :

- ▶ Discriminabilité de 2 produits  $A$  et  $B$  : une épreuve par juge pour plusieurs juges.
- ▶ Capacité d'un juge à discerner une nuance entre deux produits : plusieurs épreuves pour un même juge.

**Contexte :** la différence sensorielle entre les produits testés est souvent très faible (seuil de l'infra-liminaire et du liminaire). Elle n'est pas connue des dégustateurs (test à choix libre).

**Intérêt :** facile à mettre en œuvre, à interpréter. Test universellement connu.

**Inconvénients :** Test peu puissant (faible probabilité de rejeter  $H_0$  alors que  $H_1$  est vrai).

77/120

## 3.2. Tester une différence entre produits : Test triangulaire

### Déroulement de l'épreuve

**Une épreuve :** On présente 3 échantillons anonymes  $X, Y, Z$ , constitués des produits  $A$  et  $B$ , l'un des deux étant présenté en double, et l'autre, en singulier. La question est alors posée : "lequel des échantillons est le produit singulier ?"

**Ordre de présentation des échantillons  $X, Y$  et  $Z$  :** il peut engendrer des biais d'appréciation. Pour éviter cela, on peut

1. présenter le même nombre de fois chaque produit dans la même position.
2. doubler le même nombre de fois  $A$  et  $B$  dans les présentations.

Le nombre de présentation est alors un multiple de

Un bon effectif est

**Remarque :** Épreuve généralisable directement à la présentation de  $J$  échantillons dont un seul est singulier (plus  $J$  est grand, plus le test est puissant, mais plus l'épreuve est fastidieuse pour les testeurs).

78/120

## 3.2. Tester une différence entre produits : Test triangulaire

### Un exemple

**Exemple :** test d'indiscernabilité de deux chocolats

Echantillons																														
X	A	A	B	B	B	A	A	A	B	B	B	A	A	A	B	B	B	A	A	A	B	B	B	A	A	A	B	B	B	A
Y	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
Z	B	A	A	A	B	B	B	A	A	A	B	B	B	A	A	A	B	B	B	A	A	A	B	B	B	A	A	A	B	B
Réponse	Y	Y	Y	Z	X	Y	X	Y	X	X	Y	X	Z	Y	Y	Z	X	X	Y	Y	X	Y	Z	Y	Z	Y	Z	Y		

↑ ↑

On note  $N$  le nombre de bonnes réponses (ici  $N =$  ) et  $n$  le nombre d'épreuves (ici  $n =$  ).

**Problème :** Mais certaines bonnes réponses pourrait être dues au hasard. . .

Question : le nombre de bonnes réponses est-il significativement supérieur au nombre de bonnes réponses générées en choisissant un des trois produit au hasard ?

**Objectifs :** Intuitivement, il y a une relation entre la proportion de bonnes réponses  $p$  (inconnue) dans ce type de tests et la proportion  $\pi$  (inconnue) de juges percevant la différence entre  $A$  et  $B$ . On va donc chercher à estimer  $\pi$  à l'aide de l'estimateur  $\hat{P} = \frac{N}{n}$  de  $p$  qui prend la valeur  $\hat{p} =$  dans notre exemple.

79/120

## 3.2. Tester une différence entre produits : Test triangulaire

### Estimer $\pi$ (ponctuellement)

**Schéma probabiliste :** La population dont sont tirés les testeurs comprend une proportion  $\pi$  (inconnue) de personnes sentant une différence entre  $A$  et  $B$  (on veut savoir si  $\pi = 0$  ou si  $\pi > 0$ ).

On note  $p$  la proportion théorique (inconnue) de bonnes réponses lors d'un test triangulaire quand on choisit les juges au hasard au sein de la population. On introduit les évènements

$PD = \{\text{le juge perçoit la différence}\}$  et  $BR = \{\text{le juge donne une bonne réponse}\}$

On a

- ▶  $p = \mathbb{P}(BR)$  et  $\pi = \mathbb{P}(PD)$  (par définition)
- ▶  $\mathbb{P}(BR|PD) = 1$  et  $\mathbb{P}(BR|PD^c) = \frac{1}{3}$  (hypothèses raisonnables. . .).

On a alors

$$\begin{aligned}
 p &= \mathbb{P}(BR) = \mathbb{P}(BR \cap PD) + \mathbb{P}(BR \cap PD^c) && \text{principe de partition} \\
 &= \mathbb{P}(BR|PD)\mathbb{P}(PD) + \mathbb{P}(BR|PD^c)\mathbb{P}(PD^c) && \text{probabilités conditionnelles} \\
 &= 1 \cdot \pi + \frac{1}{3}(1 - \pi)
 \end{aligned}$$

**Relation entre  $p$  et  $\pi$  :** On a donc  $\pi = \frac{3p-1}{2}$ .

Exemple : on a une estimation de  $p$  qui est donnée par  $\hat{p} =$  . On en déduit une estimation  $\hat{\pi} =$  de  $\pi$ .

80/120

### Estimer $\pi$ (par IC)

On fait une estimation de  $p$  par IC avec l'estimateur  $\hat{P} = \frac{N}{n}$ ... et on en déduit un IC pour  $\pi$  à l'aide de la relation entre  $p$  et  $\pi$ .

**Estimation de  $p$  :** On pose  $BR_i = \{\text{le } i\text{ème juge donne une bonne réponse}\}$  et  $N = \sum_{i=1}^n \mathbb{1}_{BR_i}$ . Ainsi (par indépendance) :

$$N \sim \text{Bin}(n, p)$$

Avec  $n$  connu et  $p$  que l'on estime avec  $\hat{P} = \frac{N}{n}$ . D'après le TCL on a :

$$\frac{\hat{P} - \mathbb{E}(\hat{P})}{\sqrt{\text{Var}(\hat{P})}} = \sqrt{n} \frac{\hat{P} - p}{\sqrt{p(1-p)}} \approx \sqrt{n} \frac{\hat{P} - p}{\sqrt{\hat{P}(1-\hat{P})}} \approx \mathcal{N}(0, 1).$$

Et on a approximativement  $IC_{1-\alpha}(p) = \left[ \hat{P} \pm u_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \right]$ .

Exemple : avec  $n = 24$ ,  $\hat{p} =$  et  $\alpha = 0.05$  : l'intervalle de confiance est  $ic_{0.95}(p) =$

**Estimation de  $\pi$  :** on estime  $\pi$  avec l'estimateur  $\hat{\pi} = \frac{3\hat{P}-1}{2}$ . On a  $\mathbb{P}(z_1 < ax + b < z_2) = 1 - \alpha \Leftrightarrow \mathbb{P}\left(\frac{z_1-b}{a} < x < \frac{z_2-b}{a}\right) = 1 - \alpha$  car  $a > 0$ . On en déduit que

$$IC_{1-\alpha}(\pi) = \left[ \hat{\pi} \pm \frac{3}{2} u_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \right]$$

Exemple : avec  $n = 24$ ,  $\hat{p} =$  (et donc  $\hat{\pi} =$ ) et  $\alpha = 0.05$  :  $ic_{0.95}(\pi) =$

**Remarque :** Pour une précision double (largeur IC/2), il faut  $n = 24 \times 4 = 96$  juges. Précision d'autant meilleure que  $p(1-p)$  proche de 0. On peut rapprocher  $p$  de 1 en augmentant  $\pi$  (i.e en entraînant les juges).

81/120

### Test exact : les hypothèses

**Hypothèses :**  $\begin{cases} H_0 : A \text{ et } B \text{ sont indiscernables} & \Leftrightarrow H_0 : \pi = 0 \\ H_1 : A \text{ et } B \text{ sont discernables} & \Leftrightarrow H_1 : \pi > 0 \end{cases}$

**Statistique de test :** Si personne ne perçoit de différences ( $\pi = 0$ ), la proportion de bonnes réponses observées  $\hat{p} \simeq \frac{1}{3}$ . Au contraire, si  $\pi > 0$ , la proportion sera plus élevée (et d'autant plus que  $\pi$  est grand).

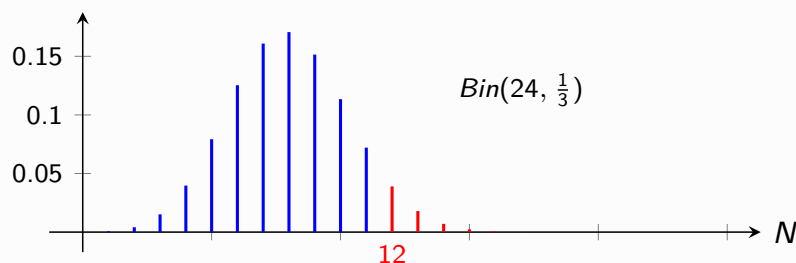
Le nombre de bonnes réponses  $N$  est la statistique de test et la région de rejet aura la forme

$$R_\alpha = [c(\alpha), +\infty[$$

où  $c(\alpha)$  est un réel qui dépend de  $\alpha$ . Remarque : si  $\alpha_1 < \alpha_2$  alors  $R_{\alpha_1} \subset R_{\alpha_2}$ .

**Test exact :** Sous  $H_0$ , on a  $\pi = 0 \Leftrightarrow p = \frac{1}{3} : N \sim \text{Bin}(n, \frac{1}{3})$

Exemple : on a observé  $N = 12$ . De plus, sous  $H_0$ ,  $\mathbb{P}(N \geq 12) = \mathbb{P}\left(\text{Bin}(24, \frac{1}{3}) \geq 12\right) = .07$  (c'est la  $p$ -valeur).  $H_0$  n'est pas rejeter pour tout test de niveau  $\alpha \leq 7\%$ ...



82/120

### Test exact : niveau et puissance

**Niveau** : Si on veut s'assurer que les données indiquent bien une *discernabilité significative* : il faut réduire la région de rejet en prenant  $\alpha$  petit.

**Puissance** : Mais si on veut s'assurer que les données indiquent une *indiscernabilité significative* : il faut contrôler la probabilité du risque de deuxième espèce  $\beta$ . Cela peut être délicat en pratique : on peut changer d'approche en s'assurant que  $H_0$  continue à ne pas être rejeté pour des valeurs de  $\alpha$  assez grande (e.g 75%).

**Risque de deuxième espèce du test exact** : probabilité de ne pas détecter  $H_1$ . En l'état on ne peut pas calculer explicitement le risque du test : il faut fixer un niveau, disons  $\alpha = 0.07$  correspondant à  $R_{0.07} = [12, +\infty[$  et on peut alors calculer  $\beta$  pour des valeurs de  $\pi > 0$  données :

$$\pi = .25 \Leftrightarrow p = \frac{1}{2} : \beta = \mathbb{P}_{p=\frac{1}{2}}(N \leq 11) = \mathbb{P}(Bin(24, \frac{1}{2}) \leq 11) = 0.42$$

$$\pi = .5 \Leftrightarrow p = \frac{2}{3} : \beta = \mathbb{P}_{p=\frac{2}{3}}(N \leq 11) = \mathbb{P}(Bin(24, \frac{2}{3}) \leq 11) = 0.03$$

Il est donc bien plus facile de détecter une différence lorsque les juges sont bien entraînés (si différence il y a, le nombre de bonnes réponses est important avec grande probabilité).

On peut aussi jouer sur la taille de l'échantillon  $n$  (voir ci-dessous).

83/120

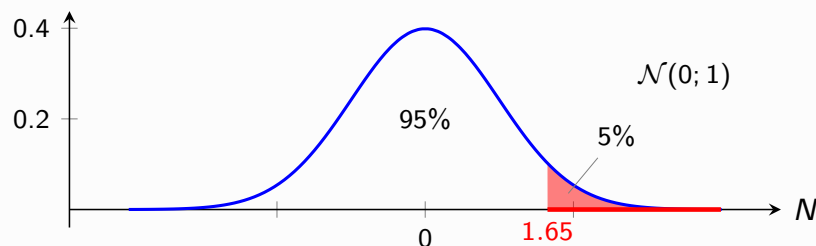
## 3.2. Tester une différence entre produits : Test triangulaire

### Test approché : principe

**Test approché** : Pour calculer la taille minimum de l'échantillon qui garantit une puissance suffisante on utilise (pour faciliter les calculs) l'approximation de la loi binomiale par une loi Normale :

$$N \sim Bin(n, p) \Rightarrow \tilde{N} = \sqrt{n} \frac{\hat{P} - p}{\sqrt{p(1-p)}} \approx \mathcal{N}(0, 1)$$

Le test peut alors porter sur la statistique  $\tilde{N}$  sans (trop) changer la décision (vérifier les conditions sur  $n$ ,  $np$  et  $n(1-p)$  pour garantir une bonne approximation).



**Rejet** : Sous  $H_0$ , on a  $p = p_0$ . On rejette si,

$$\{N > c(\alpha)\} = \{\hat{P} > d(\alpha)\} = \left\{ \sqrt{n} \frac{\hat{P} - p_0}{\sqrt{p_0(1-p_0)}} > e(\alpha) \right\} \approx \left\{ \sqrt{n} \frac{\hat{P} - p_0}{\sqrt{p_0(1-p_0)}} > u_{1-\alpha} \right\}$$

Exemple : on a  $n = 24$  et  $p_0 = \frac{1}{3}$ . Cela donne  $\sqrt{24} \frac{3\hat{p}-1}{\sqrt{2}} = 1.73 > 1.65 = u_{0.95}$  et...  $H_0$  est rejeté de justesse au niveau 5% (c'est une approximation du test)

84/120

### Test approché : Calcul de la taille de l'échantillon

**Principe :** on fixe un risque de deuxième espèce  $\beta_0 = 0.05$  et on calcule le plus petit nombre de répétitions  $n$  nécessaire pour avoir un test de puissance plus grande que  $1 - \beta_0 = 0.95$ .

**En pratique :** si  $\pi = \pi_1$  est donné (i.e  $p = p_1$  est donné aussi) on cherche  $n$  tel que

$$\beta(p_1) = \mathbb{P}_{p=p_1}(\tilde{N} < u_{1-\alpha}) < \beta_0$$

(i.e proba d'accepter  $H_0$  à tort soit petite). Autrement dit :

$$\begin{aligned} & \mathbb{P}_{p=p_1} \left( \sqrt{n} \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}} < u_{1-\alpha} \right) < \beta_0 \\ \Leftrightarrow & \mathbb{P}_{p=p_1} \left( \hat{p} < p_0 + u_{1-\alpha} \sqrt{\frac{p_0(1-p_0)}{n}} \right) < \beta_0 \\ \Leftrightarrow & \mathbb{P}_{p=p_1} \left( \sqrt{n} \frac{\hat{p} - p_1}{\sqrt{p_1(1-p_1)}} < \sqrt{n} \frac{p_0 + u_{1-\alpha} \sqrt{\frac{p_0(1-p_0)}{n}} - p_1}{\sqrt{p_1(1-p_1)}} \right) < \beta_0 \\ \Leftrightarrow & \mathbb{P} \left( \mathcal{N}(0, 1) < \frac{\sqrt{n}(p_0 - p_1) + u_{1-\alpha} \sqrt{p_0(1-p_0)}}{\sqrt{p_1(1-p_1)}} \right) < \beta_0 \end{aligned}$$

Remarque : Pour toute v.a. réelle  $Z$  : on a si  $t_1 < t_2$ ,  $\{Z < t_1\} \subset \{Z < t_2\}$  et donc  $\mathbb{P}(Z < t_1) \leq \mathbb{P}(Z < t_2)$ .

85/120

### Test approché : Calcul de la taille de l'échantillon

**Condition sur  $n$  :** On note  $u_{\beta_0}$  le quantile d'ordre  $\beta_0$  de la loi Normale. Cela nous donne la condition sur  $n$  suivante :

$$\frac{\sqrt{n}(p_0 - p_1) + u_{1-\alpha} \sqrt{p_0(1-p_0)}}{\sqrt{p_1(1-p_1)}} < u_{\beta_0} \Leftrightarrow n > \left( \frac{(u_{1-\alpha} \sqrt{p_0(1-p_0)} - u_{\beta_0} \sqrt{p_1(1-p_1)})}{p_1 - p_0} \right)^2$$

car  $p_1 > p_0$ .

Exemple : si  $\pi_1 = \frac{1}{4}$  (i.e  $p_1 = \frac{1}{2}$ ) et  $\beta_0 = 0.05$ , on a  $n > (6u_{0.95} (\frac{\sqrt{2}}{3} + \frac{1}{2}))^2 = 9.62^2 = 92.48$  beaucoup trop grand en pratique !

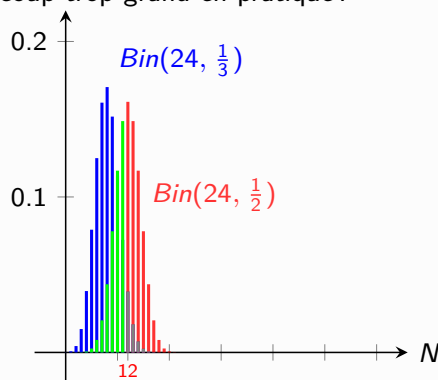


FIGURE -  $n = 24$  et  $\beta = 0.42$

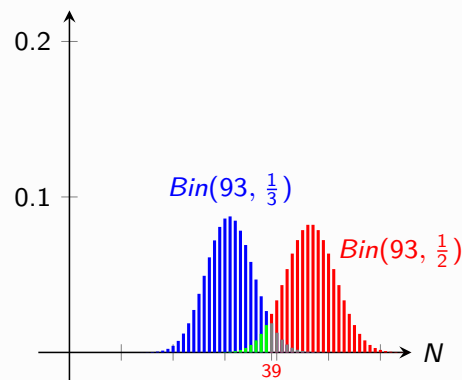


FIGURE -  $n = 93$  et  $\beta < 0.05$

En gris : risque  $\alpha$ . En vert : risque  $\beta$ . Lorsque  $n$  augmente, les lois se séparent

86/120



### Extensions : Capacité d'un juge à discerner une nuance

**Procédure :** Pour tester la capacité d'un juge à discerner une nuance entre deux stimuli proches  $A$  et  $B$ , on peut procéder à  $K$  épreuves successives pour un seul juge  $j$ . Le modèle de jugement change : on note  $\pi_j$  la probabilité de discerner la différence entre  $A$  et  $B$  pour ce juge.

**Extension :** Plus généralement, on peut procéder à  $K$  épreuves pour chacun des  $J$  juges, ce qui produit  $n = KJ$  observations : on teste alors l'existence d'une discernabilité des produits pour l'un au moins des juges :

$$\begin{cases} H_0 : \text{personne ne discerne} & \Leftrightarrow \forall j, \pi_j = 0 \Leftrightarrow \forall j, p_j = \frac{1}{3} \\ H_1 : \text{certains juges discernent} & \Leftrightarrow \exists j, \pi_j > 0 \Leftrightarrow \exists j, p_j > \frac{1}{3} \end{cases}$$

Que se passe-t-il alors ?

- ▶ Sous  $H_0$  : toutes les réponses sont des résultats d'expériences aléatoires indépendantes et de même loi. Ceci permet le calcul correct de la région de rejet de niveau  $\alpha$  (et donc... de contrôler l'erreur faite si on rejette  $H_0$ )
- ▶ Sous  $H_1$  : on n'impose pas que tous les juges aient la même probabilité de discerner (*i.e* les  $\pi_j$  peuvent être différents). Les réponses ne pourront plus être vues comme des expériences i.i.d. et on ne peut plus exprimer simplement la loi de la statistique de test. En résumé : le calcul du risque de deuxième espèce n'est plus valide !

87/120

## 3.2. Tester une différence entre produits : Épreuve à choix forcés (k-AFC)

### Principe

**Procédure :** On ne change pas la structure probabiliste du test triangulaire à choix libre (même hypothèse, même statistique). On pose une question *différente* aux juges en vue de changer le schéma cognitif du testeur : on espère alors augmenter le nombre de bonnes réponses sous  $H_1$  (*i.e* diminuer le risque  $\beta$  : test plus puissant).

**Déroulement de l'épreuve :** par exemple, pour évaluer la discernabilité de deux produits  $A$  et  $B$  du point de vue de l'acidité, et sachant qu'objectivement  $A < B$ , on posera les questions suivantes si on présente

- ▶  $\{X, Y, Z\} = \{A, A, B\}$  : "quel est l'échantillon le plus acide ?"
- ▶  $\{X, Y, Z\} = \{A, B, B\}$  : "quel est l'échantillon le moins acide ?"

... au lieu de demander "quel est l'échantillon singulier ?"

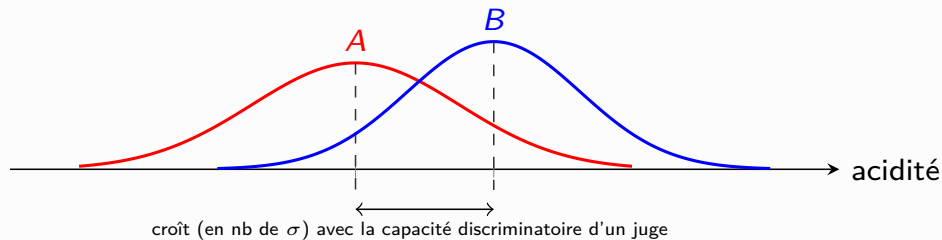
**Que se passe-t-il alors ?** Sous  $H_0$ , rien : les juges ne discernent rien et répondent "au hasard". Sous  $H_1$ , on suppose que les personnes qui discernent une différence le font plus ou moins sûrement et non pas totalement sûrement (le modèle est plus fin que le précédent).

88/120

### 3.2. Tester une différence entre produits : Épreuve à choix forcés (k-AFC)

#### Heuristique : le modèle cognitif de Thurstone

Aléas perceptifs d'un juge sur des produits A et B :



- ▶ La courbe rouge représente la distribution des perceptions du juge pour A.
- ▶ La courbe bleue représente la distribution des perceptions du juge pour B.

89/120

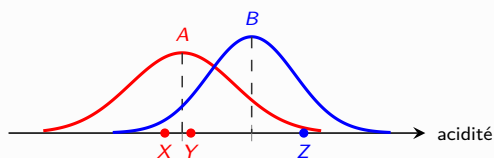
### 3.2. Tester une différence entre produits : Épreuve à choix forcés (k-AFC)

#### Heuristique : Le modèle cognitif de Thurstone

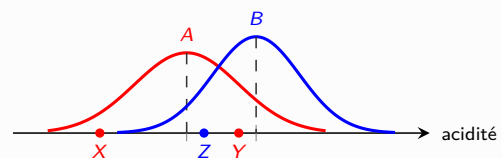
À la présentation d'un échantillon du type "deux fois A, une fois B", 4 cas sont à envisager :

**Cas 1 : Les réponses concordent :**

- a) singulier : Z (Vrai)  
le plus acide : Z (Vrai)

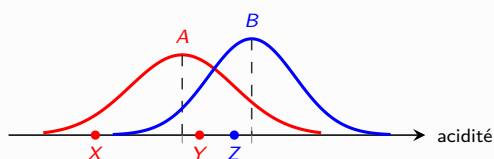


- b) singulier : X (Faux)  
le plus acide : Y (Faux)

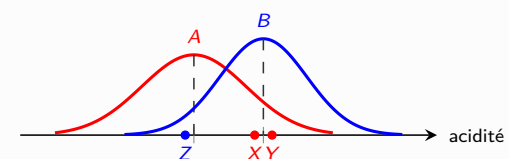


**Cas 2 : Les réponses ne concordent pas :**

- a) singulier : X (Faux)  
le plus acide : Z (Vrai)



- b) singulier : Z (Vrai)  
le plus acide : Y (Faux)



... mais le cas 2a est beaucoup plus fréquent que le cas 2b.

En pratique, on augmente la capacité d'un juge à percevoir une différence lorsqu'il y en a une (le test est alors plus puissant).

90/120

#### Principe

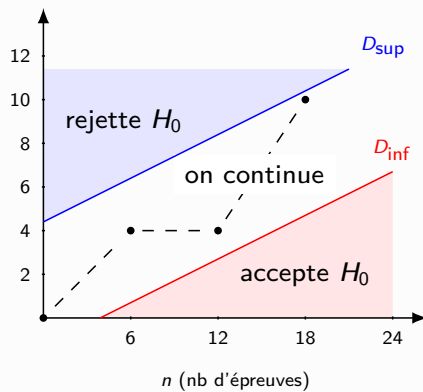
**Objectif** : définir une procédure de test pour contrôler simultanément les risques  $\alpha$  et  $\beta$  tout en minimisant le nombre d'essais à faire.

**Validité** : bon pour les tests triangulaires et 3-AFC.

**Principe** : On ne prend pas la décision à la fin de toutes les épreuves, mais au cours des essais successifs (l'information s'ajoute au cours du temps, on arrête dès que la proportion de réponses penche suffisamment dans un sens ou dans l'autre).

**En pratique** : On fixe  $\alpha$  (e.g 5%) et  $\beta$  (e.g 10%) et  $p_1$  (e.g  $\frac{1}{2}$ ). On trace ensuite deux droites parallèles :

$N$  (nb de bonnes réponses)



$$D_{inf} : N = \frac{\ln\left(\frac{\beta}{1-\alpha}\right)}{\ln\left(\frac{p_1}{p_0}\right) - \ln\left(\frac{1-p_1}{1-p_0}\right)} + n \frac{\ln\left(\frac{1-p_0}{1-p_1}\right)}{\ln\left(\frac{p_1}{p_0}\right) - \ln\left(\frac{1-p_1}{1-p_0}\right)}$$

$$D_{sup} : N = \frac{\ln\left(\frac{1-\beta}{\alpha}\right)}{\ln\left(\frac{p_1}{p_0}\right) - \ln\left(\frac{1-p_1}{1-p_0}\right)} + n \frac{\ln\left(\frac{1-p_0}{1-p_1}\right)}{\ln\left(\frac{p_1}{p_0}\right) - \ln\left(\frac{1-p_1}{1-p_0}\right)}$$

Cela définit trois zones : tant que le chemin reste dans la bande centrale on continue les observations, dès qu'on rentre dans une des deux zones décisionnelle (rejet ou acceptation de  $H_0$ ), on arrête le test.

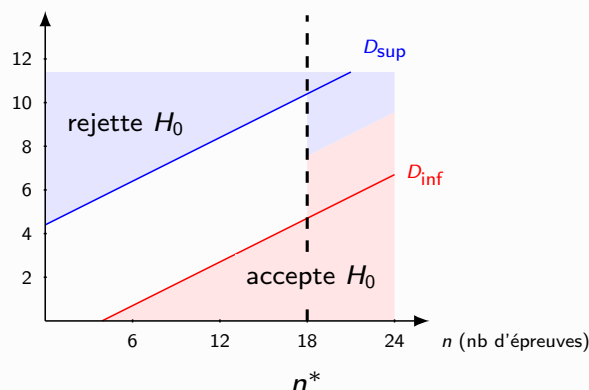
91/120

#### En pratique

On fixe un nombre maximum d'épreuves  $n^*$ . Si on atteint  $n^*$  sans avoir atteint une des deux zones de décision (acceptation ou rejet), on juge par rapport au milieu de la bande verticale délimitée par  $n^*$ ,  $D_{sup}$  et  $D_{inf}$  :

- ▶ si le  $N$  obtenu est supérieur au milieu, on obtient un rejet de  $H_0$  au niveau  $\alpha$ .
- ▶ s'il est inférieur, une acceptation de  $H_0$  au niveau  $\alpha$  ... sans garantie sur  $\beta$ .

$N$  (nb de bonnes réponses)



Remarque : Nombre effectif d'épreuves non précisément maîtrisé... On ne peut pas respecter a priori parfaitement l'équilibre entre les ordres de présentation (biais possible). on avance par groupe de six épreuves, ou présentations aléatoires.

92/120

### Contexte et objectifs

**Objectifs :** On dispose de  $K > 2$  produits. On souhaite tester l'indiscernabilité globale entre eux, sans passer par des comparaisons 2 à 2 (*via* les méthodes précédentes)

**Situations :** Ce test s'applique généralement pour :

- ▶ la reconnaissance de *stimuli* qualitativement différents (*e.g* saveurs) à la frontière de l'infraliminaire. Contrôle de la non-équivalence ou de l'équivalence de produits.

Exemple : plusieurs sirops d'orange utilisant différents édulcorants

- ▶ le test de  $K$  produits par un jury de  $n$  juges (1 épreuve par juge)
- ▶ le test d'un juge sur  $K$  produits ( $n$  épreuves pour le même juge)

93/120

### Déroulement

**L'épreuve :** On présente  $K$  témoins identifiés puis  $J$  échantillons (certains en double, triple, etc. . . ) extraits des témoins. On demande alors, pour chaque échantillon, de l'associer à un témoin.

**Présentation des produits :** Si possible, présenter les produits le même nombre de fois et dans la même position. Ceci peut impliquer des ordonnancements au hasard, si on n'a pas assez d'épreuves pour parcourir toutes les possibilités.

Exemple :  $K = 3$  pastis  $A, B, C$  distincts par les proportions de badiane, d'anis vert et de réglisse. On présente  $J = 4$  échantillons à chacun de  $n = 18$  juges, l'un des échantillons étant un produit doublé. Il y a 36 séquences d'échantillons présentées :

Problème d'effectif : seulement 18 juges. Pour éviter les biais : on tire au sort les produits non doublés dans la moitié des échantillons ( $\frac{n}{2} = 9$ ) et on inverse leur ordre dans l'autre moitié.

94/120

## 3.2. Tester une différence entre produits : Épreuve d'appariement

### Exemple : les 3 pastis

vrai	R	A	B	C
	A	13	10	1
	B	7	17	0
	C	1	1	22
		21	28	23



judge	$X_1$	$X_2$	$X_3$	$X_4$	$R_1$	$R_2$	$R_3$	$R_4$
1	A	A	B	C	A	B	B	C
2	A	C	A	B	A	C	B	B
3	A	B	C	A	B	B	C	A
4	C	A	A	B	C	B	A	A
5	B	A	C	A	B	A	C	B
6	C	B	A	A	C	B	A	B
7	B	B	A	C	A	B	A	C
8	B	C	B	A	A	C	B	A
9	B	A	C	B	B	A	C	B
10	C	B	B	A	C	B	B	B
11	A	B	C	B	B	B	C	B
12	C	A	B	B	C	B	B	A
13	C	C	B	A	C	C	A	B
14	C	A	C	B	C	C	A	B
15	C	B	A	C	C	B	A	C
16	A	C	C	B	B	C	C	A
17	B	C	A	C	A	C	A	B
18	A	B	C	C	A	B	C	C

$K$  produits  $\rightarrow$  tableau croisé  $K \times K$ ,  
dont la case  $(j,k)$  contient  $n_{jk}$ .

95/120

## 3.2. Tester une différence entre produits : Test d'indiscernabilité globale

### Hypothèses

**Hypothèse nulle** : Les appariements se font totalement "au hasard"

$\Leftrightarrow$  Chaque produit est distribué dans sa ligne selon la distribution uniforme  $(\frac{1}{K}, \dots, \frac{1}{K})$

**Hypothèse alternative** : Les appariements non totalement "au hasard"

$\Leftrightarrow$  L'une au moins des distribution théorique (probabilité) en ligne n'est pas uniforme

**Attention** : si deux produits sont systématiquement confondus (e.g  $A$  est pris pour  $B$ ) alors on peut rejeter  $H_0$  même si les produits ne sont pas correctement identifiés. Ce test doit donc davantage être utilisé pour s'assurer de la non-discriminabilité (prendre  $\alpha$  plutôt grand).

96/120

### La statistique de test

**Test :** test d'adéquation des  $K$  lignes à la distribution uniforme.

**Statistique du  $\chi^2$  :** Pour chaque ligne  $k = 1, \dots, K$ , on a sous  $H_0$

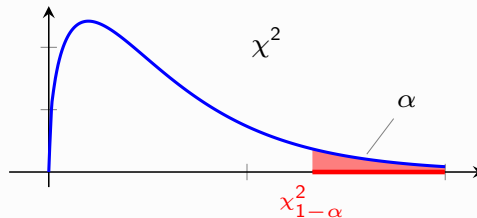
$$D_k = n_k \sum_{\ell=1}^K \frac{\left(\frac{n_{k\ell}}{n_k} - \frac{1}{K}\right)^2}{\frac{1}{K}} = \underbrace{\sum_{\ell=1}^K \frac{\left(n_{k\ell} - \frac{n_k}{K}\right)^2}{\frac{n_k}{K}}}_{\frac{(\text{effectif observé} - \text{effectif attendu})^2}{\text{effectif attendu}}} \approx \chi^2(K-1)$$

Les affectations des  $K$  produits étant indépendantes, les degrés de libertés des  $D_k$  "s'additionnent". On a sous  $H_0$  toujours :

$$D = \sum_{k=1}^K D_k \approx \chi^2(K(K-1))$$

**Région de rejet :** Sous  $H_1$  l'écart entre les effectifs observés et attendus doit être grand.

La région de rejet de niveau  $\alpha$  est de la forme :  $R_\alpha = \{D > \chi_{1-\alpha}^2(K(K-1))\}$



97/120

### Exemple : Résultats du test sur les 3 pastis

Tableaux des  $\sum_{\ell=1}^K \frac{\left(n_{k\ell} - \frac{n_k}{K}\right)^2}{\frac{n_k}{K}}$  :

<i>vrai</i>	R	A	B	C
A		3.13	0.50	6.13
B		0.13	10.13	8.00
C		6.13	6.13	24.50

64.75

La probabilité critique de la loi du  $\chi^2(6)$  associée à cette observation est  $\leq 10^{-4}$ . On rejette donc l'appariement au hasard avec une probabilité très faible de se tromper.

**Attention :** Chaque  $D_k$  permet un test d'appariement concernant le seul produit  $k$ . Mais ce test est soumis à la même réserve que le global : son alternative n'est pas que le produit est reconnu...

### Loi multinomiale

**Principe** : généralisation de la loi binomiale avec une épreuve aléatoire avec  $J$  issus.

**Épreuve aléatoire** : Soit  $X$  une v.a. nominale à  $J$  modalités  $\{E_1, \dots, E_J\}$  et on note  $p_i = \mathbb{P}(E_i) = \mathbb{E}(\mathbb{1}_{E_i})$ . On a alors  $\sum_{i=1}^J p_i = 1$  et la variable aléatoire

$$\begin{pmatrix} \mathbb{1}_{E_1} \\ \vdots \\ \mathbb{1}_{E_J} \end{pmatrix} \sim \text{Mult}(1, \mathbf{p}), \quad \mathbf{p} = \begin{pmatrix} p_1 \\ \vdots \\ p_J \end{pmatrix}$$

Lorsque  $J = 2$  et  $p \in [0, 1]$  :  $X \sim \text{Mult}\left(1, \begin{pmatrix} p \\ 1-p \end{pmatrix}\right) = \text{Bern}(p) = \text{Bin}(1, p)$ .

**Définition** On note  $Y_i$  le nombre de fois que se réalise l'évènement  $E_i$  en  $n$  répétitions indépendantes de  $X$ . On a

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_J \end{pmatrix} = \sum_{i=1}^n \begin{pmatrix} \mathbb{1}_{E_1} \\ \vdots \\ \mathbb{1}_{E_J} \end{pmatrix} \sim \text{Mult}(n, \mathbf{p})$$

On a  $\mathbb{P}(Y_1 = n_1, \dots, Y_J = n_J) = \frac{n!}{n_1! \dots n_J!} p_1^{n_1} \dots p_J^{n_J}$  si  $n_1 + \dots + n_J = n$  et 0 sinon.  
De plus

$$\mathbb{E}(Y) = n\mathbf{p} \quad \text{Var}(Y) = n(\text{Diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}^t)$$

99/120

### Analyse statistique

Pour tester si deux produits  $A$  et  $B$  sont confondus, on note les évènements :

$$\begin{aligned} A_i \rightarrow A &= \{\text{le } i\text{ème juge classe } A \text{ pour } A\} \\ A_i \rightarrow B &= \{\text{le } i\text{ème juge classe } A \text{ pour } B\} \\ A_i \rightarrow C &= (A_i \rightarrow A \cup A_i \rightarrow B)^c \end{aligned}$$

On note  $n_A$  le nombre d'épreuves qui mettent en jeu le produit  $A$ .

**Modélisation** : On pose  $X_i = \begin{pmatrix} \mathbb{1}_{A_i \rightarrow A} \\ \mathbb{1}_{A_i \rightarrow B} \\ \mathbb{1}_{A_i \rightarrow C} \end{pmatrix} \sim \text{Mult}(1; \mathbf{p})$  où  $\mathbf{p} = \begin{pmatrix} p_A \\ p_B \\ p_C \end{pmatrix}$  contient respectivement la proportion de juges qui classent  $A$  vers  $A$ ,  $A$  vers  $B$  et  $A$  vers un autre produit.

**Estimation** : on estime  $\mathbf{p}$  par la moyenne empirique  $\hat{P} = \begin{pmatrix} \hat{P}_A \\ \hat{P}_B \\ \hat{P}_C \end{pmatrix} = \frac{1}{n_A} \sum_{i=1}^{n_A} \begin{pmatrix} \mathbb{1}_{A \rightarrow A} \\ \mathbb{1}_{A \rightarrow B} \\ \mathbb{1}_{A \rightarrow C} \end{pmatrix}$ .

**Approximation** : on a pour  $n_A$  suffisamment grand (TCL multidimensionnelle)

$$\hat{P} \approx \mathcal{N}\left(\begin{pmatrix} p_A \\ p_B \\ p_C \end{pmatrix}, \frac{1}{n_A} \begin{pmatrix} p_A(1-p_A) & -p_A p_B & -p_A p_C \\ -p_B p_A & p_B(1-p_B) & -p_B p_C \\ -p_C p_A & -p_C p_B & p_C(1-p_C) \end{pmatrix}\right)$$

Ce qui implique  $\hat{P}_A - \hat{P}_B \approx \mathcal{N}\left(p_A - p_B, \frac{1}{n_A} (p_A(1-p_A) + p_B(1-p_B) - 2(-p_A p_B))\right)$

### Test

**Question :** lors de sa présentation,  $A$  est-il classé autant vers  $A$  que vers  $B$ , ou plus ?  
(test de comparaison de deux proportions).

**Hypothèse :**  $\begin{cases} H_0 : A \text{ est autant classé vers } A \text{ que vers } B & \Leftrightarrow H_0 : p_A - p_B = 0 \\ H_1 : A \text{ est plus classé vers } A & \Leftrightarrow H_1 : p_A - p_B > 0 \end{cases}$

**Loi de la statistique de test :** Sous  $H_0$  on a  $p_A = p_B = p$  :

$$\hat{P}_A - \hat{P}_B \approx \mathcal{N}\left(0, \frac{2p}{n_A}\right) \Leftrightarrow \sqrt{n_A} \frac{\hat{P}_A - \hat{P}_B}{\sqrt{2p}} \approx \mathcal{N}(0, 1)$$

On peut estimer  $p$  (inconnu) par  $\hat{P} = \frac{1}{2}(\hat{P}_A + \hat{P}_B)$  dans la variance et on a la statistique de test suivante

$$T = \sqrt{n_A} \frac{\hat{P}_A - \hat{P}_B}{\sqrt{\hat{P}_A + \hat{P}_B}} \approx \mathcal{N}(0, 1)$$

**Le test :** Région de rejet est du type  $R_\alpha = [c(\alpha), +\infty[$  (*i.e* on rejette lorsque  $\{T > u_{1-\alpha}\}$ ).

Exemple : On a  $\hat{p}_A - \hat{p}_B = \frac{3}{24} = 0.125$ ;  $\sqrt{\frac{n_A}{\hat{p}_A + \hat{p}_B}} = \sqrt{\frac{24^2}{13+10}} = \sqrt{\frac{24^2}{23}} \simeq 5.004$  et  $t \simeq 0.6255$ .  
Ainsi  $H_0$  n'est pas rejeté au niveau  $\alpha = 5\%$ ...

101/120

## 3.3. Classer des produits :

### Généralités

**Contexte :** Tester le classement de  $K$  produits sur une unique échelle de référence

**Objectifs :** Par exemple :

- ▶ S'assurer qu'une échelle d'intensité est bien reconnue par un ou plusieurs juges
- ▶ S'assurer que  $K$  produits sont correctement classés selon une échelle donnée

**Domaine de validité :** domaine liminaire ou supra-liminaire car il y a une "bonne" relation croissante (*i.e* qui conserve l'ordre) entre intensité du stimulus  $c_i$  et intensité perçue  $i$  (plus vrai dans le cas dans l'infra-liminaire ni dans la saturation).

**Avantage :** La méthode est robuste : pas besoin de connaître la relation exacte entre l'intensité de la perception et le stimulus (*c.f* lois de la perception de Fechner, Stevens, ...), on reconnaît l'ordre quelque soit la modélisation choisie.

102/120



### 3.3. Classer des produits :

#### Épreuve

**Épreuve :**  $K$  produits sont appréciés sur 1 dimension sensorielle quantitative (e.g sucré, amer, acide...) par  $n$  juges. Une seule présentation des  $K$  produits masqués par juge (sauf dans le cas d'un juge unique :  $n$  présentations masquées des produits). Chaque juge classe alors les  $K$  produits dans l'ordre d'intensité croissante et les numérote de 1 à  $K$ .

**Ordre de présentation des échantillons :** Idéalement, les  $K!$  ordres de présentation des  $K$  produits doivent être utilisés le même nombre de fois. Il y a  $K!$  ordres possibles : le nombre de testeurs devrait être un multiple de  $K!$ .

Au delà de  $K = 4$  il y a trop d'ordres possibles. Il faut envisager des variantes faibles de la contrainte comme imposer qu'un produit  $k$  soit présenté autant de fois avant qu'après un autre produit  $\ell$ .

**Résultats :** Le rang  $R_{j,k}$  du produit  $k$  classé par le juge  $j$  :

juge	produit				
	1	...	k	...	K
1	$R_{1,1}$	...	$R_{1,k}$	...	$R_{1,K}$
⋮	⋮		⋮		⋮
$j$	$R_{j,1}$	...	$R_{j,k}$	...	$R_{j,K}$
⋮	⋮		⋮		⋮
$J$	$R_{J,1}$	...	$R_{J,k}$	...	$R_{J,K}$

103/120

### 3.3. Classer des produits :

#### Tests

**Hypothèse nulle :** On part d'une hypothèse très restrictive :

$H_0$  : les échantillons sont classés au hasard par tous les juges.

C'est une hypothèse "plus que défavorable" dans le liminaire ou supra-liminaire, car elle contredit la supposition même qu'on soit dans ce domaine. On cherche à "établir statistiquement" l'hypothèse alternative, qui elle devra être compatible avec le domaine liminaire et supra-liminaire.

**Hypothèse alternative :** Plusieurs alternatives sont envisageables

- ▶ test de Friedman
- ▶ test de Page

104/120

#### Principe et hypothèses

**Hypothèses :**  $H_0$  : échantillons classés au hasard par tous les juges.  
 $H_1$  : échantillons non classés au hasard par tous les juges.

Attention à la formulation : les échantillons peuvent ne pas être classés systématiquement au hasard sans pour autant que l'ordre soit reconnu...

**Principe du test :** On compare la moyenne des rangs de chaque produit.

- ▶ Sous  $H_0$ , tous les produits obtiennent en moyenne le même rang, aucun n'étant classé avant un autre : pour tout  $k = 1, \dots, K$

$$\mathbb{E}(R_k) = \frac{1}{K} (1 + 2 + \dots + K) = \frac{K + 1}{2}$$

- ▶ Sous  $H_1$  il devrait exister des disparités : on le voit sur l'exemple ci contre où les sommes des rangs semblent significativement différentes.

	1	2	3	4
2	1	1	4	3
1	2	2	4	3
1	2	3	4	3
2	3	1	1	4
2	3	1	1	4
1	2	3	4	4
3	1	4	2	2
1	2	4	4	3
4	2	3	3	1
3	1	4	4	2
1	2	3	4	4
2	1	4	4	3
2	3	1	4	4
2	1	4	3	4
1	2	3	4	4
1	3	2	4	4
2	3	1	4	4
3	1	4	2	4
1	3	2	4	4
2	1	4	3	4
2	3	1	4	4
3	1	4	2	4
2	3	1	4	4
1	3	4	2	4
	45	49	70	76

#### Déroulement du test

**Statistique de décision :** La statistique  $F$  de Friedman est définie par

$$F = n \frac{12}{K(K+1)} \sum_{k=1}^K \left( \bar{R}_k - \frac{K+1}{2} \right)^2 = \left[ \frac{12}{nK(K+1)} \sum_{k=1}^K (S_k)^2 \right] - 3n(K+1)$$

où  $S_k = \sum_{j=1}^J R_{j,k}$ .

**Loi de F sous  $H_0$  :** on a  $F \approx \chi^2(K - 1)$ . L'approximation est rapidement satisfaisante même pour des valeurs modérés de  $n$  et  $K$ . Il faut utiliser des lois tabulées si  $K = 3$  et  $n < 10$  ou  $K = 4$  et  $n < 5$ .

**Comportement de F sous  $H_1$  :**  $F \rightarrow \infty$  quand  $n \rightarrow \infty$ . On rejette lorsque l'évènement suivant se produit :

$$\{F > c_{1-\alpha}(K - 1)\}$$

où  $c_{1-\alpha}(K - 1)$  est le quantile de la loi du  $\chi^2(K - 1)$ .

#### Exemple

Dans cet exemple on a  $K = 4$  et  $n = 24$

k	1	2	3	4
$S_k$				
$(S_k)^2$				

On a  $\sum_k (S_k)^2 =$  ,  $F =$  et la  $p$ -valeur est inférieure à .

Conclusion : Dire que le classement n'est pas fait au hasard entraîne un risque d'erreur inférieur à . Mais, l'alternative n'est pas orientée vers la reconnaissance de l'échelle !

Par exemple, si les juges ont systématiquement produit le classement inverse du classement correct, on se trouve dans l'alternative  $H_1$

	1	2	3	4
2	1	4	3	3
1	2	4	3	3
1	2	4	3	3
2	3	1	4	4
2	3	1	4	4
1	2	3	4	4
3	1	4	2	2
1	2	4	3	3
4	2	3	1	1
3	1	4	2	2
1	2	3	4	4
2	1	4	3	3
2	3	1	4	4
2	1	4	3	3
1	2	3	4	4
1	3	2	4	4
2	3	1	4	4
3	1	4	2	2
1	3	2	4	4
2	1	4	3	3
2	3	1	4	4
3	1	4	2	2
2	3	1	4	4
1	3	4	2	2
	45	49	70	76

**Rejeter  $H_0$  en faveur de  $H_1$  ne signifie pas que l'échelle est reconnue correctement !**

### 3.3. Classer des produits : Test de Page

#### Déroutement du test

**Hypothèses :**  $H_0$  : échantillons classés au hasard par tous les juges.  
 $H_1$  : l'ordre perçu est le bon

**Principe du test :** on construit une statistique de test qui suit une loi connue sous  $H_0$ , mais qui ne prend de grandes valeurs que sous la nouvelle  $H_1$  (pour cela on pondère chaque rang différemment).

**Statistique de décision :** Statistique  $L$  de Page :

$$L = S_1 + 2S_2 + \dots + kS_k + \dots + KS_K$$

**Lois de  $L$  sous  $H_0$  :** elles ont été tabulées par Page. On rejette lorsque

$$\{ T > l_{1-\alpha} \}$$

où  $l_{1-\alpha}$  est le quantile d'ordre  $1 - \alpha$  d'une loi tabulée.

**Approximation :** Elle n'est valable que si l'ordre moyen observé ne contredit pas trop le vrai ordre (pas de gros inversion/écart). On peut alors considérer que

$$T = \frac{(12L - 3nK(K + 1))^2}{nK^2(K^2 - 1)(K + 1)} \approx \chi^2(1)$$

La région de rejet approchée est alors  $R_\alpha = [c_{1-\alpha}(1), +\infty[$  avec  $c_{1-\alpha}(1)$  le quantile de loi du  $\chi^2(1)$ .

Exemple

Dans cet exemple on a  $K = 4$  et  $n = 24$

k	1	2	3	4
$S_k$	45	49	70	76
$kS_k$				

On a  $L = \sum_k kS_k = \dots$ ,  $T = \dots$  et la  $p$ -valeur est inférieure à  $\dots$ .

Conclusion : l'hypothèse du classement au hasard est rejetée en faveur du classement correct, avec un risque d'erreur  $< \dots$ .

La probabilité critique est très inférieure à celle du test de Friedman : le test de basé sur  $F$  inclut dans l'événement  $\{F > 17.25\}$  les manières de s'écarter de  $H_0$  autres que le classement est reconnu.

	1	2	3	4
2	1	4	3	
1	2	4	3	
1	2	4	3	
2	3	1	4	
2	3	1	4	
1	2	3	4	
3	1	4	2	
1	2	4	3	
4	2	3	1	
3	1	4	2	
1	2	3	4	
2	1	4	3	
2	3	1	4	
2	1	4	3	
1	2	3	4	
1	3	2	4	
2	3	1	4	
3	1	4	2	
1	3	2	4	
2	1	4	3	
2	3	1	4	
3	1	4	2	
2	3	1	4	
1	3	4	2	
	45	49	70	76

3.3. Classer des produits : Tests de comparaison multiples

Niveau

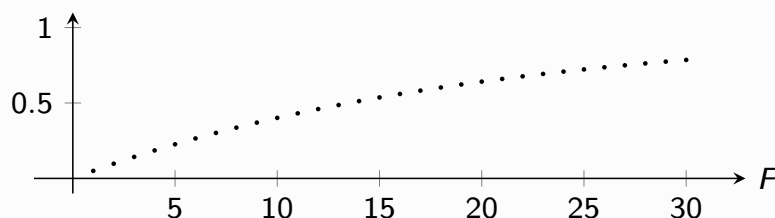
**Objectif** : une fois que l'existence de différences entre  $K$  produits est admise, on souhaite connaître les différences qui sont statistiquement significatives. L'idée est de faire un test pour les  $\frac{K(K-1)}{2}$  paires de produits possibles.

**Niveau du test global** : Attention au risque du test global suivant :

$$\begin{cases} H_0 : \forall k, \ell \mathbb{E}(R_k) = \mathbb{E}(R_\ell) \\ H_1 : \exists k, \ell \mathbb{E}(R_k) \neq \mathbb{E}(R_\ell) \end{cases} \Leftrightarrow \begin{cases} H_0 : \forall k, \mathbb{E}(R_k) = \frac{K+1}{2} \\ H_1 : \mathbb{E}(R_1) \neq \mathbb{E}(R_2) \text{ ou } \mathbb{E}(R_1) \neq \mathbb{E}(R_3) \text{ ou } \dots \end{cases}$$

On réalise  $\frac{K(K-1)}{2}$  tests avec les mêmes données et la même hypothèse  $H_0$  : le niveau du test global peut être très grand !

**Problème** : Lorsqu'un test de niveau  $\alpha$  est répété  $F$  fois de manière indépendante, la probabilité de rejeter  $H_0$  au moins une fois à tort est alors  $1 - (1 - \alpha)^F$ . Cette probabilité peut être très grande :



#### Correction de Bonferoni

**Inégalité de Bonferoni :** On note  $A_i = \{H_0 \text{ est rejeté à tort dans le test } i\}$ . La probabilité  $\alpha_K$  de rejeter  $H_0$  à tort au moins une fois parmi les  $\frac{K(K-1)}{2}$  tests est :

$$\alpha_K = \mathbb{P}\left(\bigcup_{i=1}^{\frac{K(K-1)}{2}} A_i\right) \leq \sum_{i=1}^{\frac{K(K-1)}{2}} \mathbb{P}(A_i) = \frac{K(K-1)}{2} \alpha.$$

**Correction du niveau :** Pour contrôler  $\alpha_K$  il faut alors limiter le risque de première espèce de chaque test à :

$$\gamma = \frac{\alpha}{\frac{K(K-1)}{2}}$$

La correction de Bonferoni est très conservatrice pour les grandes valeurs de  $K$ . Pour chaque paire, on rejette beaucoup plus difficilement  $H_0$  (on augmente la tolérance pour l'hypothèse d'indiscernabilité).

111/120

#### En pratique

**Déroulement du test :** Soit une paire  $(k, \ell)$ . On montre que sous l'hypothèse d'indiscernabilité des deux produits sur l'échelle :

$$\Delta_{k,\ell} = \frac{S_k - S_\ell}{\sqrt{\frac{nK(K+1)}{6}}} \approx \mathcal{N}(0, 1)$$

On compare  $\Delta_{k,\ell}$  au quantile de la loi normale et il est trouvé statistiquement différent de 0 au risque  $\gamma$  si :

$$|\Delta_{k,\ell}| > u_{1-\frac{\gamma}{2}} \Leftrightarrow |S_k - S_\ell| > u_{1-\frac{\gamma}{2}} \sqrt{\frac{nK(K+1)}{6}}$$

**PPDS :** On appelle "plus petite différence significative" (PPDS) la quantité :  $u_{1-\frac{\gamma}{2}} \sqrt{\frac{nK(K+1)}{6}}$ .

112/120

### 3.3. Classifier des produits : Tests de comparaison multiples

#### Exemple

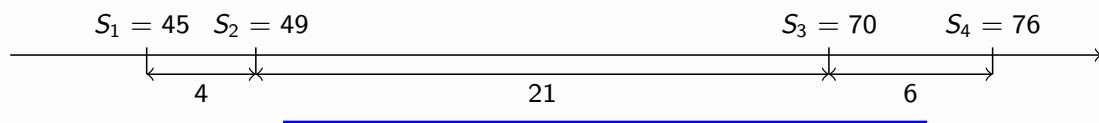
Exemple (la suite) : On dispose toujours de 4 produits : cela fait  $\alpha = 0.05$  paires à tester. On a

$$\gamma =$$

$$u_{1-\frac{\gamma}{2}} =$$

$$PPDS =$$

Dans notre exemple, on a  $S_1 < S_2 < S_3 < S_4$  :



En conclusion : le produit 3 est trouvé différent du produit 1, le 4 du 2, et le 4 du 1.

113/120

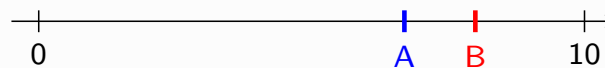
### 3.4. Comparer deux produits sur une échelle continue :

#### Les données

**Épreuve** :  $J$  juges notent deux produits  $A$  et  $B$  sur une échelle continue. Dispositif :

- ▶ Présentation de quelques références (e.g min, max [et milieu] de l'échelle) ;
- ▶ Pointage par le juge sur l'échelle (e.g une bande de 10 cm sans graduation)

**Données** : Les données brutes sont de la forme :



... on les représente sous forme de tableau :

Juge	1	...	j	...	J
A	$X_1^A$	...	$X_j^A$	...	$X_J^A$
B	$X_1^B$	...	$X_j^B$	...	$X_J^B$

**Problème** : Estimer la différence de notes entre  $A$  et  $B$ . Établir la significativité de cette différence.

114/120

## Le modèle gaussien

**Modèle :** Soit  $X_j^i$  la note du juge  $j$  pour le produit  $i$  :

$$X_j^i = c_j + b_i + \varepsilon_j^i$$

où

- ▶  $c_j$  est l'effet juges.
- ▶  $b_i$  est l'effet produit.
- ▶  $\varepsilon_j^i$  erreur Gaussienne erreur d'espérance nulle.

**Le cas de 2 produits :**

$$X_j^A = c_j + b_A + \varepsilon_j^A$$

$$X_j^B = c_j + b_B + \varepsilon_j^B$$

d'où

$$\underbrace{X_j^A - X_j^B}_{Y_j} = \underbrace{(b_A - b_B)}_{=b} + \underbrace{(\varepsilon_j^A - \varepsilon_j^B)}_{=\delta_j}$$

On a alors,  $Y_j \sim \mathcal{N}(b, \sigma^2)$  pour tout  $j$ .

**Estimateur de  $b$  :** Moyenne empirique  $\bar{Y}$ .

115/120

## Intervalle de confiance

Dans le modèle gaussien, on peut avoir un intervalle de confiance exact pour  $b$ .

**Variance connue :** on a  $\sqrt{n} \frac{\bar{Y} - b}{\sigma} \sim \mathcal{N}(0, 1)$  et

$$IC_{1-\alpha}(b) = \left[ \bar{Y} \pm u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right].$$

où  $u_{1-\frac{\alpha}{2}}$  est le quantile de la loi normale centrée réduite.

**Variance inconnue :** on "remplace"  $\sigma$  par son estimateur  $S_Y = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (Y_j - \bar{Y})^2}$ .

$$IC_{1-\alpha}(b) = \left[ \bar{Y} \pm t_{1-\frac{\alpha}{2}}(n-1) \frac{S_Y}{\sqrt{n}} \right]$$

où  $t_{1-\frac{\alpha}{2}}(n-1)$  est le quantile de la loi de Student  $\mathcal{T}(n-1)$ .

116/120

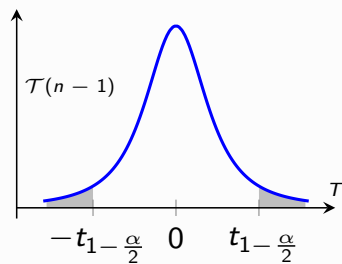
### Test de différence de moyenne de Student

**hypothèse :**  $\begin{cases} H_0 : b = 0 \Leftrightarrow \text{pas de différence entre les produits} \\ H_1 : b > 0 \text{ ou } H'_1 : b < 0 \text{ ou } H''_1 : b \neq 0 \end{cases}$

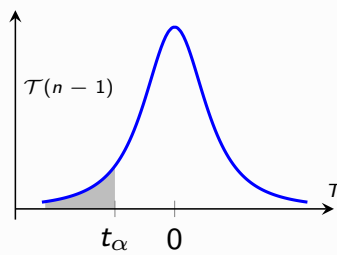
**Statistique de test :**  $\bar{Y}$  (c'est la différence des moyennes empiriques). Sous  $H_0$ , on sait que

$$T = \sqrt{n} \frac{\bar{Y}}{S_Y} \sim \mathcal{T}(n-1).$$

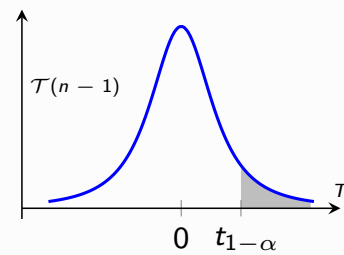
**Rejet de  $H_0$  :** dépend de l'hypothèse alternative :



$$H_1 : \{|T| > t_{1-\frac{\alpha}{2}}(n-1)\}$$



$$H'_1 : \{T < t_{\alpha}(n-1)\}$$



$$H_1 : \{T > t_{1-\alpha}(n-1)\}$$

### Modèle et intervalle de confiance

**Modèle :** Le modèle reste le même que pour le cas gaussien, seule la loi du bruit change. On la suppose centrée (espérance nulle). Par exemple, dans le cas de 2 produits notés sur une échelle continue :

$$X_j^A = b_A + \varepsilon_j^A$$

$$X_j^B = b_B + \varepsilon_j^B$$

d'où

$$\underbrace{X_j^A - X_j^B}_{Y_j} = \underbrace{(b_A - b_B)}_{=b} + \underbrace{(\varepsilon_j^A - \varepsilon_j^B)}_{=\delta_j}$$

On a alors,  $Y_i$  i.i.d pour tout  $i$ .

**Intervalles de confiance pour  $b$  :** Ils peuvent être calculés par diverses méthodes :

- ▶ Loi limite : dans les hypothèses du TCL si  $n$  est grand ( $n \geq 30$ ) on a un intervalle de confiance approché.
- ▶ Loi connue : Au cas par cas. . .
- ▶ Autres méthodes : quantiles empiriques, inégalités de concentration.



## Le test des signes et de rang de Wilcoxon

**Objectif :** Comparer les notes de 2 produits sur une échelle continue dans le cas non gaussien. Lorsque l'aléa est loin de la normalité, il faut un test qui continue d'être valable (robuste). Le test de rang de Wilcoxon est non seulement robuste, mais aussi puissant que celui de Student dans le cas gaussien, et beaucoup plus en dehors.

**Principe :** Plutôt que de tester la nullité de la moyenne  $b$  de  $Y = X^A - X^B$ , on va tester celle de sa médiane  $\mu$ .

**Remarque :** 1) Dans le cas gaussien, la moyenne et la médiane sont égales. 2) La moyenne peut ne pas être définie.

**Hypothèses :** 
$$\begin{cases} H_0 : \mu = 0 \Leftrightarrow \mathbb{P}(Y < 0) = \mathbb{P}(Y > 0) \Leftrightarrow \mathbb{P}(X^A < X^B) = \mathbb{P}(X^A > X^B) \\ H_1 : \mu > 0 \Leftrightarrow \mathbb{P}(Y < 0) < \mathbb{P}(Y > 0) \Leftrightarrow \mathbb{P}(X^A < X^B) < \mathbb{P}(X^A > X^B) \end{cases}$$

On peut bien sûr considérer  $H_1' : \mu < 0$  ou  $H_1'' : \mu \neq 0$  en modifiant la région de rejet ci-dessous.

119/120

## Le test des signes et de rang de Wilcoxon

**Procédure :**

- ▶ On classe les  $Y_i = X_i^A - X_i^B$  dans l'ordre croissant de leur valeur absolue  $|Y_i|$ . (on élimine les valeurs nulles et on diminue  $n$  d'autant).
- ▶ On note  $R_i$  le rang de  $Y_i$  dans ce classement (en cas d'ex æquo, attribution à tous de leur rang moyen).
- ▶ On calcule la somme  $R^+$  des  $R_i$  correspondant aux  $Y_i > 0$ , et celle  $R^-$  des  $R_i$  correspondant aux  $Y_i < 0$ .

**Comportement sous  $H_0$  :**  $R^+$  et  $R^-$  doivent être "statistiquement proches". On a

$$R^+ + R^- = n(n+1)/2$$

Ainsi,  $R^+$  et  $R^-$  vaudront en moyenne théorique :  $n(n+1)/4$ .

**Comportement sous  $H_1$  :** les  $Y_i$  positifs vont être plus nombreux et plus élevés que les négatifs :  $R^+ \gg R^- \Leftrightarrow R^+ \gg \frac{n(n+1)}{2} - R^+ \Leftrightarrow R^+ \gg \frac{n(n+1)}{4}$ .

**Statistique de test :** Sous  $H_0$  on peut montrer que  $S = \frac{R^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \approx \mathcal{N}(0, 1)$

**Rejet :** Si on observe l'évènement  $\{S > u_{1-\alpha}\}$  où  $u_{1-\alpha}$  est le quantile de la loi normale.

120/120