Chapitre 3

Tests d'hypothèses

> «Un statisticien est une personne dont l'ambition principale est d'avoir tort dans 5% des cas.»
>
> Anonyme

Les *tests d'hypothèses* sont utiles pour vérifier si une affirmation sur un modèle théorique correspondant à une expérience aléatoire est cohérente avec les mesures effectuées.

Dans une étude statistique, on peut se demander si les mesures observées peuvent correspondre à une certaine réalité. Par exemple, est-ce que la série de jets d'une pièce de monnaie $(P; P; F; P; F; P)$ peut correspondre à une pièce bien équilibrée ?

Le but d'un test d'hypothèses est de confronter deux hypothèses entre elles : l'*hypothèse nulle* H_0 et l'*hypohèse alternative* H_1. Les *hypothèses* sont des énoncés qui concernent un paramètre d'une population. La confrontation s'effectue à l'aide d'un estimateur du paramètre en question, appelé *statistique de test.*

Dans l'exemple des jets de pièce de monnaie (voir page 38), le paramètre en question est la probabilité p que la pièce tombe sur pile. Son estimateur est donné par $\widehat{P} = \overline{X}$. L'hypothèse nulle sera "la pièce est bien équilibrée ($p = \frac{1}{2}$)" et l'hypothèse alternative sera "la pièce n'est pas bien équilibrée ($p \neq \frac{1}{2}$)".

On va traiter plusieurs cas :

1. Tests d'hypothèses sur une moyenne ou une proportion.
   (a) Variance connue ou variance inconnue.
   Selon le fait que la variance est connue ou inconnue, on utilisera une loi différente pour effectuer les calculs.
   (b) Tests symétriques ou asymétriques.
   Dans le cours, on présente les tests symétriques, les tests asymétriques seront vus en exercices.

   Les tests d'hypothèses sur une proportion sont des tests où la variance est connue, parce que des lois de Bernoulli et binomiale apparaissent.
2. Test d'hypothèses de comparaison de moyenne : données appariées ou non.
3. Les tests du chi-carré
   a) pour l'adéquation à une loi ; b) pour la comparaison d'échantillon ;
   c) pour l'indépendance.

Il existe d'autres tests d'hypothèses. Par exemple, les tests permettant d'inférer la variance σ^2 d'une population. Ces tests utilisent d'autres lois de probabilité, comme, par exemple, la distribution de Fisher à deux paramètres.

3.1 Tests d'hypothèses symétriques sur une moyenne

3.1.1 Tests d'hypothèses symétriques sur une moyenne, variance connue

Considérons des variables aléatoires X_1, ..., X_n indépendantes qui suivent une même loi d'espérance μ (c'est la moyenne théorique qu'on cherche à tester) et de variance σ^2 (supposée connue).

Dans ce modèle de tests d'hypothèses, on teste l'hypothèse «la moyenne théorique vaut μ_0» à partir des n mesures effectuées.

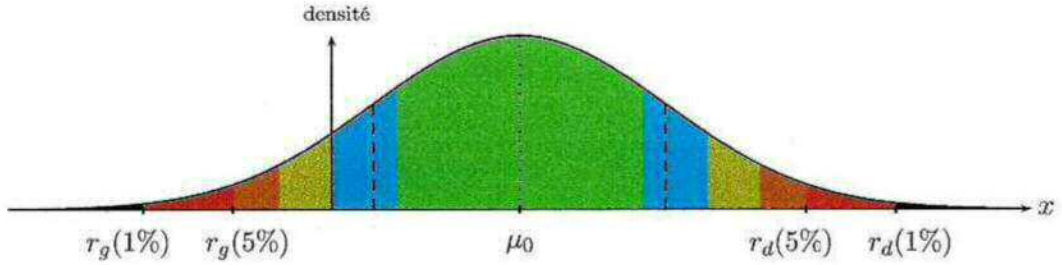
On se trouve face à deux alternatives, appelées *hypothèses*, qui sont

$$\begin{array}{ccc} \textit{hypothèse nulle} & & \textit{hypothèse alternative} \\ H_0 : \mu = \mu_0 & \text{et} & H_1 : \mu \neq \mu_0 \end{array}$$

Comme dans un raisonnement par l'absurde, on suppose qu'on se trouve sous l'hypothèse H_0 et on regarde si les données mesurées permettent d'en tirer une contradiction.

Sous l'hypothèse H_0, les variables aléatoires X_i sont d'espérance μ_0.

Par le théorème de la limite centrale, $\overline{X} = \frac{X_1+\cdots+X_n}{n}$ suit approximativement une loi normale $\mathcal{N}(\mu_0, \frac{\sigma^2}{n})$. En traitillés, on voit les bornes de l'intervalle $\left[\mu_0 - \frac{\sigma}{\sqrt{n}}, \mu_0 + \frac{\sigma}{\sqrt{n}}\right]$; on remarque ainsi que lorsque n grandit la courbe se resserre.



On estime $\overline{X}$ par la moyenne des mesures effectuées $\overline{x} = \frac{x_1+\cdots+x_n}{n}$ (attention à bien faire la différence entre les majuscules et les minuscules). L'estimation $\overline{x}$ donnée par les mesures va donc se trouver quelque part sous la loi.

Il y a une probabilité de 60% que $\overline{x}$ tombe dans la zone ■, il y a une probabilité de 20% que $\overline{x}$ tombe dans la zone ■, il y a une probabilité de 10% que $\overline{x}$ tombe dans la zone ■, il y a une probabilité de 5% que $\overline{x}$ tombe dans la zone ■, il y a une probabilité de 4% que $\overline{x}$ tombe dans la zone ■, il y a une probabilité de 1% que $\overline{x}$ tombe dans la zone ■.

Plus on se trouve dans une zone éloignée de μ_0 donc dans l'ordre : ■, ■, ■, ■, ■, ■; plus les mesures sont en contradiction avec l'hypothèse H_0. On décide ainsi du critère suivant.

On *rejette l'hypothèse H_0 au seuil de signification* 5% si $\overline{x}$ se trouve dans la zone ■ ou ■.
On *rejette l'hypothèse H_0 au seuil de signification* 1% si $\overline{x}$ se trouve dans la zone ■.

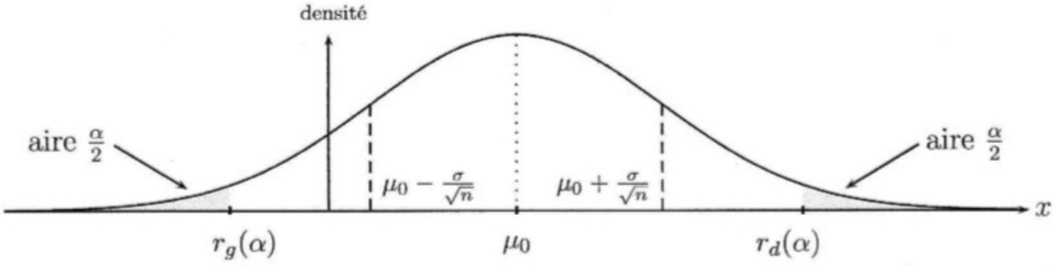
Le *seuil de signification* du test est la probabilité α que l'on a de rejeter l'hypothèse H_0 sachant que H_0 est vraie.

$$\mathbb{P}(\text{ rejeter } H_0 \mid H_0 \text{ est vraie }) = \alpha$$

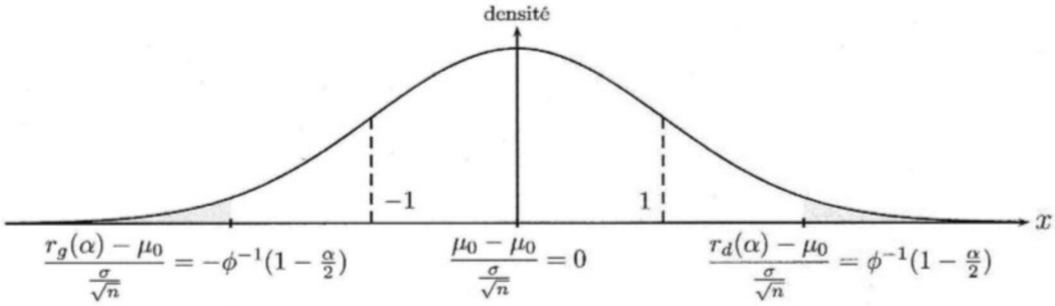
La distribution ci-dessous est la distribution sous l'hypothèse H_0, donc α se voit sur le dessin : c'est l'aire sous la distribution dans la zone de rejet de H_0 (en grisé).

On montre dans le cours OS que si on a une variable N qui suit une loi normale $\mathcal{N}(\mu_0, \frac{\sigma^2}{n})$, alors la variable $\frac{N-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ suit la loi normale $\mathcal{N}(0,1)$. On dit que N a été *centrée-réduite*.

Si la densité de $\overline{X}$ est



Alors la densité de $\dfrac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ est (l'échelle a changé)



où $\phi(x) = \mathbb{P}(Z \leqslant x)$ est la fonction de répartition de la variable aléatoire Z qui suit la loi normale $\mathcal{N}(0,1)$. Les valeurs de ϕ et ϕ^{-1} se trouvent dans la table de la page 30.

On rejette H_0 au seuil de signification α si

pour la variable aléatoire $\overline{X}$	pour la variable aléatoire $\dfrac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$
$\overline{x} < r_g(\alpha)$ ou $\overline{x} > r_d(\alpha)$	$\left\lvert \dfrac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right\rvert > \phi^{-1}\left(1 - \frac{\alpha}{2}\right)$

En particulier, si $\alpha = 5\%$, le critère est $\left\lvert \dfrac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right\rvert > \phi^{-1}(0.975) \overset{\text{table}}{\cong} 1.96.$

De même, si $\alpha = 1\%$, le critère est $\left\lvert \dfrac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right\rvert > \phi^{-1}(0.995) \overset{\text{table}}{\cong} 2.58.$

Exemple

Le test "Gauche-Droite" de Jean Piaget a pour but de vérifier l'acquisition par l'enfant des notions gauche-droite à différents points de vue et d'évaluer ainsi son niveau de socialisation et subjectivisation. Les enfants de 7 ans obtiennent en moyenne 12 comme résultat avec un écart type de 3.4. On applique ces tests à 25 enfants gauchers de 7 ans choisis au hasard et on obtient un résultat moyen de 13.4.

À partir de ces tests, peut-on affirmer qu'il y a une différence significative entre les gauchers et les droitiers ?

Réponse

Ici, on met en doute le fait que la moyenne théorique des gauchers est égale à celle des droitiers. On suppose par ailleurs que l'écart type théorique est le même pour les deux populations [1]. L'approximation par le théorème de la limite centrale est de bonne qualité car $n = 25$. On peut donc effectuer le test avec les alternatives suivantes.

$$\begin{array}{ccc} \textit{hypothèse nulle} & & \textit{hypothèse alternative} \\ H_0 : \mu = 12 & \text{et} & H_1 : \mu \neq 12 \end{array}$$

Prétendre que $\mu = 12$ revient à prétendre que les gauchers suivent le même modèle que celui de l'ensemble de la population (donc a fortiori celui des droitiers).

Les données livrent une moyenne $\overline{x} = 13.4$. On a ainsi

$$\left| \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| = \left| \frac{13.4 - 12}{\frac{3.4}{\sqrt{25}}} \right| \cong 2.059$$

On voit qu'on rejette H_0 au seuil 5%, mais qu'on ne peut pas rejeter H_0 au seuil 1% (en fait $\overline{x}$ est tombé dans la zone ■, mais pas dans la zone ■).

Remarques

1. En supposant que H_0 soit vraie, et qu'un grand nombre d'examinateurs aient fait passer ces tests à d'autres groupes de 25 gauchers, alors 1% des examinateurs auraient eu une moyenne inférieur à 10.25 ou supérieur à 13.75 et 5% des examinateurs auraient eu une moyenne inférieur à 10.7 ou supérieur à 13.3.
2. On ne fixe pas le seuil de signification après avoir fait le test, mais avant. Ceci afin d'éviter de régler le seuil pour que le test donne le résultat escompté après avoir effectué les mesures.
3. Il est habituel de prendre 1% pour confirmer H_0 et 5% pour infirmer H_0, mais ces seuils sont arbitraires. En cas de doute sur la conclusion du test, il est important d'aller regarder de plus près comment les mesures ont été effectuées. On peut aussi refaire l'expérience sur un autre échantillon. Si lors des mesures, une erreur a été effectuée, le test ne sert plus à rien.
4. Pour qu'un test d'hypothèses soit utile, il faut que les mesures aient été effectuées sur un échantillon représentatif de la population.

---

1. Si le lecteur n'est pas d'accord avec cette hypothèse, il peut utiliser le test de Student mis au point par William Gosset (voir section suivante).

Les différents types d'erreurs

Il y a deux probabilités α et β qui décrivent deux types de risque qu'on rencontre en effectuant un test d'hypothèses. On a déjà parlé de α qui est aussi appelé le *seuil de signification d'un test d'hypothèse*, mais il y a aussi β qui est défini ci-dessous.

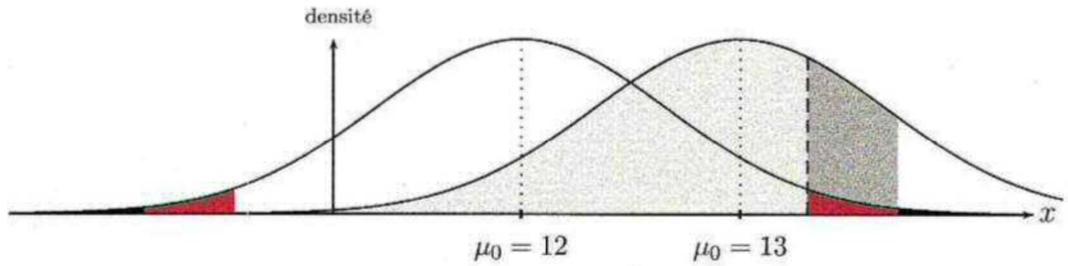
$$\begin{aligned} \alpha &= \mathbb{P}(\text{ rejeter } H_0 \mid H_0 \text{ est vraie }) \\ \beta &= \mathbb{P}(\text{ ne pas rejeter } H_0 \mid H_0 \text{ est fausse }) \end{aligned}$$

Comme il est difficile de calculer β, dans la pratique on calcule plutôt des β_{μ_i} comme définis ci-dessous.

	H_0 est vraie	H_0 est fausse		
	$\mu = \mu_0$	$\mu = \mu_1$	$\dots$	$\mu = \mu_k$
On rejette H_0	Mauvaise décision Probabilité α	Bonne décision Probabilité $1 - \beta_{\mu_1}$	$\dots$	Bonne décision Probabilité $1 - \beta_{\mu_k}$
On ne rejette pas H_0	Bonne décision Probabilité $1 - \alpha$	Mauvaise décision Probabilité β_{μ_1}	$\dots$	Mauvaise décision Probabilité β_{μ_k}

Le seuil α est une probabilité appelée *risque de première espèce* et les nombres β_{μ_i}, tout comme β, sont appelées *risques de deuxième espèce*.

Calcul d'un risque de deuxième espèce (suite de l'exemple précédent)



Pour un seuil $\alpha = 5\%$, le risque de deuxième espèce avec $H_1 : \mu = 13$ donne $\beta_{13} \cong 69\%$ (c'est la proportion de la zone en gris clair déterminée par les zones ■ et ■). Pour un seuil $\alpha = 1\%$, le risque de deuxième espèce avec $H_1 : \mu = 13$ donne $\beta_{13} \cong 87\%$ (c'est la proportion de la zone en gris clair et foncé déterminée par la zone ■).

Autrement dit, au seuil $\alpha = 5\%$. Si H_0 est vrai, la bonne décision est de ne pas rejeter l'hypothèse ; sa probabilité est de 95%. La mauvaise décision a une probabilité de 5%.

Par contre (au même seuil α), si au lieu de H_0, c'est $H_1 : \mu = 13$ qui est vrai, alors la bonne décision est de rejeter H_0, la probabilité est maintenant de 31%. La mauvaise décision a une probabilité de 69%.

On le voit sur cet exemple, si le risque de premier espèce diminue, alors le risque de deuxième espèce augmente.

Pour diminuer le risque de deuxième espèce, il faudrait augmenter le seuil α et par conséquent le risque de première espèce.

C'est un compromis qu'il faut savoir accepter.

3.1.2 Test d'hypothèses symétriques sur une moyenne, variance inconnue

On suppose que les variables aléatoires $X_1, \ldots, X_n$ qui correspondent aux n valeurs observées sont indépendantes et suivent une loi normale $\mathcal{N}(\mu, \sigma^2)$ où les paramètres μ et σ sont inconnus.

Sous ces hypothèses, l'estimateur $\overline{X}$ suit une loi normale de paramètres $\mathcal{N}(\mu, \frac{\sigma^2}{n})$.

On teste l'hypothèse nulle $H_0 : \mu = \mu_0$ contre l'hypothèse alternative $H_1 : \mu \neq \mu_0$.

Comme avant, on se place sous l'hypothèse nulle et on contemple. Malheureusement, on ne peut pas utiliser la loi $\mathcal{N}(\mu_0, \frac{\sigma^2}{n})$ pour calculer les probabilités puisque σ est inconnu. Heureusement, William Gosset, brasseur et mathématicien britannique trouva une loi qui permit de travailler sous l'hypothèse H_0. Il publia son résultat sous le pseudonyme de Student.

Théorème (sans preuve)

Si $X_1, \ldots, X_n$ sont des variables aléatoires indépendantes qui suivent une loi normale $\mathcal{N}(\mu, \sigma^2)$, alors la variable aléatoire[2] $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ suit une distribution de Student avec $n-1$ degrés de liberté.

Théorème

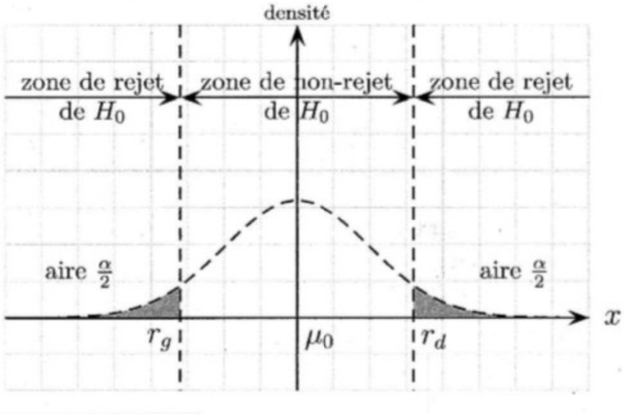
On rejette l'hypothèse H_0 au seuil de signification α si[3]

$$\overline{x} \notin \left[\mu_0 - \phi_{n-1}^{-1}\left(1-\tfrac{\alpha}{2}\right) \cdot \tfrac{s}{\sqrt{n}} \; , \; \mu_0 + \phi_{n-1}^{-1}\left(1-\tfrac{\alpha}{2}\right) \cdot \tfrac{s}{\sqrt{n}}\right] \quad \text{ou} \quad \left|\tfrac{\overline{x}-\mu_0}{s/\sqrt{n}}\right| > \phi_{n-1}^{-1}\left(1-\tfrac{\alpha}{2}\right)$$

où $\phi_{n-1}(x)$ est la fonction de répartition de la loi de Student à $n-1$ degrés de liberté.

Preuve

On se fixe le seuil de signification α. Sous l'hypothèse H_0, on sait que $\overline{X}$ suit une distribution normale $\mathcal{N}(\mu_0; \frac{\sigma^2}{n})$ de la forme suivante :

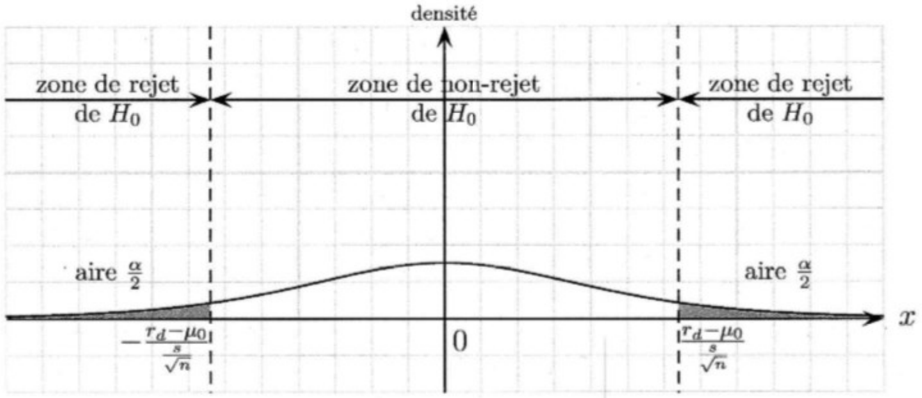


2. Ce n'est pas la variable centrée réduite de $\overline{X}$, car on a remplacé σ par son estimateur sans biais S.
3. Dans l'énoncé de ce théorème, s est l'estimation de l'estimateur S, tout comme $\overline{x}$ est l'estimation de l'estimateur $\overline{X}$.

Malheureusement, comme σ est inconnu, on ne peut pas utiliser cette loi pour calculer la valeur exacte de r_d sous l'hypothèse H_0. Mais grâce à Gosset, on sait que la variable aléatoire $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ suit une distribution de Student avec $n-1$ degrés de liberté.

$$1-\frac{\alpha}{2} = \mathbb{P}\left(\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \leqslant \frac{r_d-\mu}{\frac{s}{\sqrt{n}}}\right) \stackrel{\substack{\text{on se place sous}\\ \text{l'hypothèse } H_0\\ \downarrow}}{=} \mathbb{P}\left(\frac{\overline{X}-\mu_0}{\frac{S}{\sqrt{n}}} \leqslant \frac{r_d-\mu_0}{\frac{s}{\sqrt{n}}}\right) = \phi_{n-1}\left(\frac{r_d-\mu_0}{\frac{s}{\sqrt{n}}}\right)$$

Voici la représentation de cette loi de Student :



Donc, on a :

$$\boxed{\frac{r_d-\mu_0}{\frac{s}{\sqrt{n}}} = \phi_{n-1}^{-1}\left(1-\frac{\alpha}{2}\right)} \iff r_d-\mu_0 = \phi_{n-1}^{-1}\left(1-\frac{\alpha}{2}\right)\cdot\frac{s}{\sqrt{n}} \iff \boxed{r_d = \mu_0 + \phi_{n-1}^{-1}\left(1-\frac{\alpha}{2}\right)\cdot\frac{s}{\sqrt{n}}}$$

Donc, on rejette l'hypothèse H_0 au seuil de signification α si l'estimation $\overline{x}$ est plus grande que r_d ou plus petite que $r_g = \mu_0 - \phi_{n-1}^{-1}\left(1-\frac{\alpha}{2}\right)\cdot\frac{s}{\sqrt{n}}$. □

3.1.3 Résumé et autres statistiques de tests sur les moyennes

Dans les tableaux ci-dessous, on note $z_\alpha = \phi^{-1}(\alpha)$ où ϕ est la fonction de répartition de la loi normale centrée réduite et $t_{\nu,\alpha} = \phi_\nu^{-1}(\alpha)$ où ϕ_ν est la fonction de répartition de la loi de Student à ν degrés de liberté.

Tests d'inférence d'une espérance (rappel)

On considère n variables aléatoires indépendantes X_1, X_2, ..., X_n qui suivent toute la même loi normale $\mathcal{N}(\mu, \sigma^2)$.

H_0	Statistique de test	H_1	Région de rejet
$\mu = \mu_0$	$Z = \dfrac{\overline{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$ σ connu	$\mu \neq \mu_0$ $\mu < \mu_0$ $\mu > \mu_0$	$Z \leqslant z_{\frac{\alpha}{2}}$ ou $Z \geqslant z_{1-\frac{\alpha}{2}}$ $Z \leqslant z_\alpha$ $Z \geqslant z_{1-\alpha}$
$\mu = \mu_0$	$T = \dfrac{\overline{X} - \mu_0}{\sqrt{\frac{S^2}{n}}}$ σ inconnu	$\mu \neq \mu_0$ $\mu < \mu_0$ $\mu > \mu_0$	$T \leqslant t_{n-1,\frac{\alpha}{2}}$ ou $T \geqslant t_{n-1,1-\frac{\alpha}{2}}$ $T \leqslant t_{n-1,\alpha}$ $T \geqslant t_{n-1,1-\alpha}$

Tests de comparaison de deux espérances

On considère m variables aléatoires indépendantes X_1, X_2, ..., X_m qui suivent toute la même loi normale $\mathcal{N}(\mu_X, \sigma_X^2)$ et n variables aléatoires indépendantes Y_1, Y_2, ..., Y_n qui suivent toute la même loi normale $\mathcal{N}(\mu_Y, \sigma_Y^2)$. On suppose en outre que ces familles de variables sont indépendantes.

H_0	Statistique de test	H_1	Région de rejet
$\mu_X - \mu_Y = d_0$	$Z = \dfrac{(\overline{X} - \overline{Y}) - d_0}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}}$ σ_X et σ_Y connus	$\mu_X - \mu_Y \neq d_0$ $\mu_X - \mu_Y < d_0$ $\mu_X - \mu_Y > d_0$	$Z \leqslant z_{\frac{\alpha}{2}}$ ou $Z \geqslant z_{1-\frac{\alpha}{2}}$ $Z \leqslant z_\alpha$ $Z \geqslant z_{1-\alpha}$
$\mu_X - \mu_Y = d_0$	$T = \dfrac{(\overline{X} - \overline{Y}) - d_0}{\sqrt{\frac{S_p^2}{m} + \frac{S_p^2}{n}}}$ $\sigma_X = \sigma_Y$ inconnus $S_p^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}$	$\mu_X - \mu_Y \neq d_0$ $\mu_X - \mu_Y < d_0$ $\mu_X - \mu_Y > d_0$	$T \leqslant t_{m+n-2,\frac{\alpha}{2}}$ ou $T \geqslant t_{m+n-2,1-\frac{\alpha}{2}}$ $T \leqslant t_{m+n-2,\alpha}$ $T \geqslant t_{m+n-2,1-\alpha}$
$\mu_X - \mu_Y = d_0$	$T = \dfrac{(\overline{X} - \overline{Y}) - d_0}{\sqrt{\frac{S_X^2}{m} + \frac{S_Y^2}{n}}}$ $\sigma_X \neq \sigma_Y$ inconnus $\hat{\nu} = \dfrac{\left(\frac{S_X^2}{m} + \frac{S_Y^2}{n}\right)^2}{\frac{\left(\frac{S_X^2}{m}\right)^2}{m-1} + \frac{\left(\frac{S_Y^2}{n}\right)^2}{n-1}}$	$\mu_X - \mu_Y \neq d_0$ $\mu_X - \mu_Y < d_0$ $\mu_X - \mu_Y > d_0$	$T \leqslant t_{\hat{\nu},\frac{\alpha}{2}}$ ou $T \geqslant t_{\hat{\nu},1-\frac{\alpha}{2}}$ $T \leqslant t_{\hat{\nu},\alpha}$ $T \geqslant t_{\hat{\nu},1-\alpha}$

Tests de comparaison de deux espérances d'échantillons appariés

On considère n variables aléatoires indépendantes X_1, X_2, $\ldots$, X_n qui suivent toute la même loi normale $\mathcal{N}(\mu_X, \sigma_X^2)$ et n variables aléatoires indépendantes Y_1, Y_2, $\ldots$, Y_n qui suivent toute la même loi normale $\mathcal{N}(\mu_Y, \sigma_Y^2)$. On suppose de plus que pour chaque i, il y a une dépendance entre X_i et Y_i : on dit que X_i et Y_i sont appariés.

On sait (c'est un théorème) que les variables aléatoires $D_i = Y_i - X_i$ suivent aussi une loi normale $\mathcal{N}(\mu_D, \sigma_D^2)$ où $\mu_D = \mu_Y - \mu_X$. On est ainsi ramené au test d'inférence de l'espérance de D_i qui se trouve sur la page précédente (celui où la variance n'est pas supposée connue).

H_0	Statistique de test	H_1	Région de rejet
$\underbrace{\mu_Y - \mu_X}_{=\mu_D} = d_0$	$Z = \dfrac{\overline{D} - d_0}{\sqrt{\frac{\sigma_D^2}{n}}}$ σ_D connu	$\mu_Y - \mu_X \neq d_0$ $\mu_Y - \mu_X < d_0$ $\mu_Y - \mu_X > d_0$	$Z \leqslant z_{\frac{\alpha}{2}}$ ou $Z \geqslant z_{1-\frac{\alpha}{2}}$ $Z \leqslant z_\alpha$ $Z \geqslant z_{1-\alpha}$
$\underbrace{\mu_Y - \mu_X}_{=\mu_D} = d_0$	$T = \dfrac{\overline{D} - d_0}{\sqrt{\frac{S_D^2}{n}}}$ σ_D inconnu	$\mu_Y - \mu_X \neq d_0$ $\mu_Y - \mu_X < d_0$ $\mu_Y - \mu_X > d_0$	$T \leqslant t_{n-1,\frac{\alpha}{2}}$ ou $T \geqslant t_{n-1,1-\frac{\alpha}{2}}$ $T \leqslant t_{n-1,\alpha}$ $T \geqslant t_{n-1,1-\alpha}$

3.2 Test du chi-carré : adéquation à une loi

Considérons des variables aléatoires $X_1, \ldots, X_n$ indépendantes qui suivent une même loi qu'une variable aléatoire X. On cherche à tester la densité de probabilité (ou distribution) de X.

On découpe les valeurs possibles pour X en k classes $A_1, \ldots, A_k$. On note N_i les variables aléatoires qui comptent le nombre de mesures qui tombent dans la classe A_i.

tableau des probabilités théoriques

A_1	$\cdots$	A_k	total
$\mathbb{P}(X \in A_1)$	$\cdots$	$\mathbb{P}(X \in A_k)$	1

Dans ce modèle de test d'hypothèses, on teste l'hypothèse «la variable X a une certaine densité de probabilité» à partir des n mesures effectuées. À l'aide du découpage en classes, cette hypothèse est reformulée par «les probabilités théoriques $\mathbb{P}(X \in A_i)$ valent p_i».

On se trouve face à deux alternatives qui sont

hypothèse nulle | *hypothèse alternative*

$H_0 : \mathbb{P}(X \in A_i) = p_i$ pour tout i et $H_1 :$ il existe i tel que $\mathbb{P}(X \in A_i) \neq p_i$

Comme dans un raisonnement par l'absurde, on suppose qu'on se trouve sous l'hypothèse H_0 et on regarde si les données mesurées permettent d'en tirer une contradiction.

Sous l'hypothèse H_0, on peut établir le tableau des effectifs théoriques.

effectifs mesurés

A_1	$\cdots$	A_k	total
n_1	$\cdots$	n_k	n

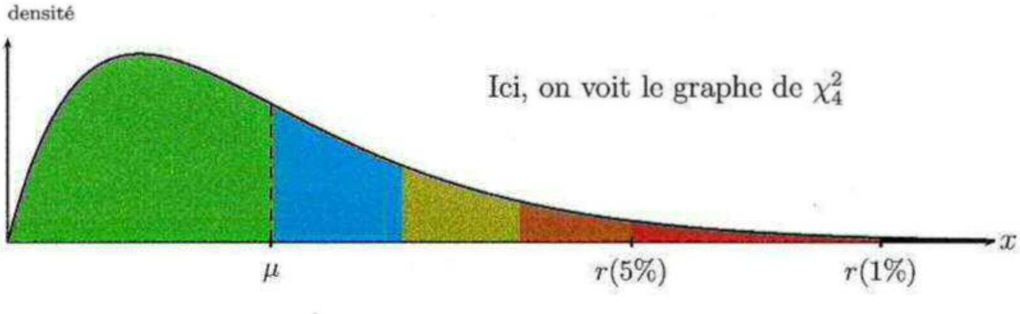
effectifs théoriques

A_1	$\cdots$	A_k	total
np_1	$\cdots$	np_k	n

Les mathématiciens ont montré que si les effectifs théoriques np_i sont de taille au moins 5, alors la variable aléatoire D^2 suivante suit approximativement une loi du chi-carré à $k-1$ degrés de liberté.

$$D^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i}$$

On a $k-1$ degrés de liberté, car les variables aléatoires N_1 à N_k sont liées entre-elles par la condition : $\sum_i N_i = n$.



On estime D^2 à l'aide des mesures effectuées $d^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$ (attention à bien faire la différence entre les majuscules et les minuscules). L'estimation d^2 donnée par les mesures va donc se trouver quelque part sous la loi.

Il y a une probabilité de 60% que d^2 tombe dans la zone ■, il y a une probabilité de 20% que d^2 tombe dans la zone ■, il y a une probabilité de 10% que d^2 tombe dans la zone ■, il y a une probabilité de 5% que d^2 tombe dans la zone ■, il y a une probabilité de 4% que d^2 tombe dans la zone ■, il y a une probabilité de 1% que d^2 tombe dans la zone ■.

Plus on se trouve dans une zone éloignée de 0 donc dans l'ordre : ■, ■, ■, ■, ■, ■; plus les mesures sont en contradiction avec l'hypothèse H_0. On décide ainsi du critère suivant.

On *rejette l'hypothèse* H_0 *au seuil de signification* 5% si d^2 se trouve dans la zone ■ ou ■.

On *rejette l'hypothèse* H_0 *au seuil de signification* 1% si d^2 se trouve dans la zone ■.

On rejette H_0 au seuil de signification α si

$$d^2 > \phi_{k-1}^{-1}(1-\alpha)$$

où $\phi_{k-1}(x) = \mathbb{P}(D^2 \leqslant x)$ est la fonction de répartition de la variable aléatoire D^2 qui suit la loi du chi-carré χ^2_{k-1} avec $k-1$ degrés de liberté. Les valeurs de cette fonction ϕ_{k-1} sont données par la table de la page 32.

En particulier, si $\alpha = 5\%$, le critère est $d^2 > \phi_{k-1}^{-1}(0.95) \overset{\text{table}}{=} 11.07$ si $k = 6$.

De même, si $\alpha = 1\%$, le critère est $d^2 > \phi_{k-1}^{-1}(0.99) \overset{\text{table}}{=} 15.09$ si $k = 6$.

Remarque sur les degrés de liberté

Si lors du calcul des probabilités p_i, on doit utiliser k estimateurs, il faut encore enlever k degrés de liberté.

Ce peut être le cas lorsque, par exemple, on calcule les p_i avec une loi normale où il faut d'abord estimer l'espérance et la variance.

Remarque sur le risque de deuxième espèce

Dans le cas d'un test du χ^2, le risque de deuxième espèce ne se calcule pas aussi facilement que dans le cas d'un test sur un moyenne. Cela dépasse le cadre de ce cours.

Exemples

Reprenons l'exemple du début du cours où on a lancé un dé 60 fois. Les classes choisies sont $A_i = \{i\}$ de sorte que A_i corresponde à l'apparition de la face i.

1. Testons l'alternative suivante.

$$\begin{array}{lcl} H_0 & : & \text{le dé est bien équilibré, c'est-à-dire } p_i = \frac{1}{6}. \\ H_1 & : & \text{le dé n'est pas bien équilibré.} \end{array}$$

   On peut donc construire les tableaux suivants.

   effectifs mesurés

face	1	2	3	4	5	6	total
n_i	13	13	11	8	8	7	60

   effectifs théoriques

face	1	2	3	4	5	6	total
np_i	10	10	10	10	10	10	60

   Pour que l'approximation soit bonne, il faut que $np_i \geqslant 5$ pour chaque i. C'est le cas ici. On a

$$d^2 = \sum_{i=1}^{6} \frac{(n_i - np_i)^2}{np_i} = \underbrace{\frac{(13-10)^2}{10}}_{\text{pour la face 1}} + \underbrace{\frac{(13-10)^2}{10}}_{\text{pour la face 2}} + \cdots + \underbrace{\frac{(7-10)^2}{10}}_{\text{pour la face 6}} = 3.6$$

   Comme $d^2 < \phi_5^{-1}(0.95) \cong 11.07$, on ne peut pas rejeter l'hypothèse H_0 au seuil 5%. Il n'y a pas assez de preuves pour dire que le dé n'est pas bien équilibré.

2. Imaginons qu'on pense qu'il y a une bille de plomb vers le sommet adjacent aux faces 4, 5 et 6 qui fait que le dé montre pour le 75% des lancés les faces 1, 2 ou 3.

   On en déduit que $p_1 = p_2 = p_3 = \frac{3}{12}$ et que $p_4 = p_5 = p_6 = \frac{1}{12}$ (le total des p_i vaut bien 1, et $p_1 + p_2 + p_3 = 3(p_4 + p_5 + p_6)$ tout comme 75% vaut $3 \cdot 25\%$).

   Testons l'alternative suivante.

$$\begin{array}{lcl} H_0 & : & \text{le dé est truqué comme ci-dessus.} \\ H_1 & : & \text{le dé n'est pas truqué comme ci-dessus.} \end{array}$$

   On peut donc construire les tableaux suivants.

   effectifs mesurés

face	1	2	3	4	5	6	total
n_i	13	13	11	8	8	7	60

   effectifs théoriques

face	1	2	3	4	5	6	total
np_i	15	15	15	5	5	5	60

   Pour que l'approximation soit bonne, il faut que $np_i \geqslant 5$ pour chaque i. C'est le cas ici. On a

$$d^2 = \sum_{i=1}^{6} \frac{(n_i - np_i)^2}{np_i} = \underbrace{\frac{(13-15)^2}{15}}_{\text{pour la face 1}} + \underbrace{\frac{(13-15)^2}{15}}_{\text{pour la face 2}} + \cdots + \underbrace{\frac{(7-5)^2}{5}}_{\text{pour la face 6}} = 6$$

   Comme $d^2 < \phi_5^{-1}(0.95) \cong 11.07$, on ne peut pas rejeter l'hypothèse H_0 au seuil 5%. Il n'y a pas assez de preuves pour dire que le dé n'est pas truqué de cette façon.

Les résultats des 60 jets ne sont pas assez différents de ce qu'on aurait pu obtenir avec un dé bien équilibré ou un dé truqué de la façon ci-dessus. La valeur mesurée d^2 indiquerait que la situation la plus probable est que le dé est bien équilibré (la p-valeur est meilleure sur le premier exemple ($\cong 60.8\%$) que sur le deuxième ($\cong 30.6\%$)).

Néanmoins, le mieux serait de lancer ce dé encore 60 fois et de refaire ces tests sur les 120 lancers obtenus.

3.3 Test du chi-carré : comparaison d'échantillons

Pour chaque échantillon i, $1 \leqslant i \leqslant k$, considérons des variables aléatoires $X_1^{(i)}, \ldots, X_{n^{(i)}}^{(i)}$ indépendantes qui suivent une même loi qu'une variable aléatoire $X^{(i)}$. On cherche à tester si les lois $X^{(i)}$ suivent toutes une même loi qu'une variable aléatoire X.

On découpe les valeurs possibles pour les échantillons en l classes $A_1, \ldots, A_l$ (les mêmes pour chaque échantillon). On note $N_j^{(i)}$ les variables aléatoires qui comptent le nombre de mesures de l'échantillon i qui tombent dans la classe A_j. On note aussi $N_j = \sum_i N_j^{(i)}$.

tableau des probabilités théoriques

	A_1	$\cdots$	A_l	totaux
échantillon 1	$\mathbf{P}(X^{(1)} \in A_1)$	$\cdots$	$\mathbf{P}(X^{(1)} \in A_l)$	1
$\vdots$	$\vdots$		$\vdots$	$\vdots$
échantillon k	$\mathbf{P}(X^{(k)} \in A_1)$	$\cdots$	$\mathbf{P}(X^{(k)} \in A_l)$	1
moyenne	$\frac{1}{k}\sum_i \mathbf{P}(X^{(i)} \in A_1)$	$\cdots$	$\frac{1}{k}\sum_i \mathbf{P}(X^{(i)} \in A_l)$	1

Dans ce modèle de tests d'hypothèses, on teste l'hypothèse «les échantillons suivent tous la même loi» à partir des mesures effectuées sur chaque échantillon. Cette hypothèse est reformulée par «les $X^{(i)}$ suivent une même loi X pour tout i».

On se trouve face à deux alternatives qui sont

$$H_0 : \mathbf{P}(X^{(i)} \in A_j) = \mathbf{P}(X \in A_j) \text{ pour tout } i$$

$$H_1 : \text{ il existe } i \text{ tel que } \mathbf{P}(X^{(i)} \in A_j) \neq \mathbf{P}(X \in A_j)$$

Comme dans un raisonnement par l'absurde, on suppose qu'on se trouve sous l'hypothèse H_0 et on regarde si les données mesurées permettent d'en tirer une contradiction.

Sous l'hypothèse H_0, les moyennes ci-dessus deviennent

$$\tfrac{1}{k}\textstyle\sum_i \mathbf{P}(X^{(i)} \in A_j) = \tfrac{1}{k}\sum_{i=1}^k \mathbf{P}(X \in A_j) = \mathbf{P}(X \in A_j)$$

et on peut établir le tableau des effectifs théoriques en notant pour simplifier n pour le nombre total de mesures $n = \sum_i n^{(i)}$ et $p_j = \mathbf{P}(X \in A_j)$.

effectifs mesurés

	A_1	$\cdots$	A_l	totaux
éch. 1	$n_1^{(1)}$	$\cdots$	$n_l^{(1)}$	$n^{(1)}$
$\vdots$	$\vdots$	$n_j^{(i)}$	$\vdots$	$\vdots$
éch. k	$n_1^{(k)}$	$\cdots$	$n_l^{(k)}$	$n^{(k)}$
totaux	n_1	$\cdots$	n_l	n

effectifs théoriques

	A_1	$\cdots$	A_l	totaux
éch. 1	$n^{(1)}p_1$	$\cdots$	$n^{(1)}p_l$	$n^{(1)}$
$\vdots$	$\vdots$	$n^{(i)}p_j$	$\vdots$	$\vdots$
éch. k	$n^{(k)}p_1$	$\cdots$	$n^{(k)}p_l$	$n^{(k)}$
totaux	np_1	$\cdots$	np_l	n

Les mathématiciens ont montré que si les effectifs théoriques $n^{(i)}p_j$ sont de taille au moins 5, alors la variable aléatoire D^2 suivante suit approximativement une loi du χ^2 à $kl - k$ degrés de liberté (on enlève à kl un degré par échantillon, car les variables aléatoires $N_1^{(i)}$ à $N_l^{(i)}$ sont liées entre-elles par la condition : $\sum_j N_j^{(i)} = n^{(i)}$).

$$D^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(N_j^{(i)} - n^{(i)}p_j)^2}{n^{(i)}p_j}$$

Lorsque que les paramètres p_j sont inconnus : on les estime[4] par $p_j = \frac{n_j}{n}$. Ces estimateurs sont naturels, ils font en sorte que les totaux des deux tableaux soient identiques.

Dans ce cas, il faut recalculer les degrés de liberté, il faut encore enlever $(l-1)$ degrés de liberté (les estimateurs p_1 à p_l sont liés entre-eux par la condition $\sum_j p_j = 1$). Ainsi le nombre de degré de liberté vaut $kl - k - (l-1) = kl - k - l + 1 = k(l-1) - (l-1) = (k-1)(l-1)$.

On estime D^2 à l'aide des mesures effectuées $d^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(n_j^{(i)} - n^{(i)} p_j)^2}{n^{(i)} p_j}$ (attention à bien faire la différence entre les majuscules et les minuscules). L'estimation d^2 donnée par les mesures va donc se trouver quelque part sous la loi.

L'allure de la loi de probabilité et le critère de rejet sont les mêmes que pour le test d'adéquation (attention aux degrés de liberté qui changent).

Exemple

Cet exemple est basé sur les relevés concernant les déchets urbains du Jura en 2010, 2000 et 1994. Les déchets sont subdivisés en cinq catégories : les déchets urbains combustibles ; les déchets compostables ; le papier et le carton ; le verre ; l'aluminium, le fer blanc, la ferraille.

Ici les échantillons correspondent aux trois années susmentionnées, et on va tester si la façon dont les types de déchets sont répartis évolue avec les années (si c'est le cas, on peut éventuellement affirmer que les habitants ont pris conscience de l'importance de trier les déchets).

H_0 : la répartition des déchets dans les cinq catégories est la même pour les trois ans.
H_1 : il y a deux ans au moins pour lesquels les déchets sont répartis différemment.

Voici les tableaux des effectifs mesurés et celui des effectifs théoriques (calculés à partir de ces mesures sous l'hypothèse H_0). Les unités sont en kilogrammes par habitant et par année. Les erreurs d'arrondis ont été lissées (pour que les totaux jouent).

effectifs mesurés

	comb.	comp.	papier	verre	fer	totaux
1994	284	19	28	33	6	370
2000	265	75	52	47	23	462
2010	252	117	75	48	13	505
totaux	801	211	155	128	42	1337

effectifs théoriques

	comb.	comp.	papier	verre	fer	totaux
1994	222	58	43	35	12	370
2000	277	73	54	44	14	462
2010	302	80	58	49	16	505
totaux	801	211	155	128	42	1337

Pour que l'approximation par la loi du chi-carré soit bonne, il faut que les effectifs théoriques soient plus grands ou égaux à 5. C'est le cas ici, et on a

$$d^2 = \sum_{i=1}^{3} \sum_{j=1}^{5} \frac{(n_{i;j} - np_i q_j)^2}{np_i q_j} = \frac{(284-222)^2}{222} + \frac{(19-58)^2}{58} + \cdots + \frac{(13-16)^2}{16} \cong 89$$

(sans arrondir les effectifs théoriques, on trouve environ 88.952)

Les paramètres p_j ont été estimés, on se retrouve avec $(3-1)(5-1) = 8$ degrés de liberté. Comme $d^2 > \phi_2^{-1}(0.99) \cong 20.090$, on rejette l'hypothèse H_0 au seuil de signification de 1% (c'est-à-dire avec une probabilité de 1% de chance de rejeter à tort).

Les données laissent à penser que les citoyens trient de mieux en mieux leurs déchets.

4. Il s'agit des estimateurs du maximum de vraisemblance , notés $\widehat{p_j}_{\mathrm{MLE}}$.

3.4 Test du chi-carré : indépendance

Considérons des couples de variables aléatoires $(X_1; Y_1)$, ..., $(X_n; Y_n)$ indépendantes qui suivent les mêmes lois que le couple de variables aléatoires $(X; Y)$. On cherche à tester l'indépendance entre X et Y.

On découpe les valeurs possibles pour X en k classes A_1, ..., A_k, et celles pour Y en l classes B_1, ..., B_l. On note $N_{i;j}$ les variables aléatoires qui comptent le nombre de couples de mesures qui tombent dans $(A_i; B_j)$. On note aussi $N_{i;\heartsuit} = \sum_j N_{i;j}$ et $N_{\heartsuit;j} = \sum_i N_{i;j}$.

tableau des probabilités théoriques

	B_1	$\cdots$	B_l	totaux
A_1	$\mathbf{P}\big((X \in A_1) \cap (Y \in B_1)\big)$	$\cdots$	$\mathbf{P}\big((X \in A_1) \cap (Y \in B_l)\big)$	$\mathbf{P}(X \in A_1)$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
A_k	$\mathbf{P}\big((X \in A_k) \cap (Y \in B_1)\big)$	$\cdots$	$\mathbf{P}\big((X \in A_k) \cap (Y \in B_l)\big)$	$\mathbf{P}(X \in A_k)$
totaux	$\mathbf{P}(Y \in B_1)$	$\cdots$	$\mathbf{P}(Y \in B_l)$	1

Dans ce modèle de tests d'hypothèses, on teste l'hypothèse «les variables X et Y sont indépendantes» à partir des n mesures effectuées. En utilisant le découpage en classes et la théorie des probabilités, cette hypothèse est reformulée par «les probabilités théoriques satisfont $\mathbf{P}\big((X \in A_i) \cap (Y \in B_j)\big) = \mathbf{P}(X \in A_i) \cdot \mathbf{P}(Y \in B_j)$».

On se trouve face à deux alternatives qui sont

$$H_0 : \mathbf{P}\big((X \in A_i) \cap (Y \in B_j)\big) = \mathbf{P}(X \in A_i) \cdot \mathbf{P}(Y \in B_j) \text{ pour tout } i \text{ et } j$$

$$H_1 : \text{ il existe } i \text{ et } j \text{ tels que } \mathbf{P}\big((X \in A_i) \cap (Y \in B_j)\big) \neq \mathbf{P}(X \in A_i) \cdot \mathbf{P}(Y \in B_j)$$

Comme dans un raisonnement par l'absurde, on suppose qu'on se trouve sous l'hypothèse H_0 et on regarde si les données mesurées permettent d'en tirer une contradiction.

Sous l'hypothèse H_0, on peut établir le tableau des effectifs théoriques en notant pour simplifier $p_i = \mathbf{P}(X \in A_i)$ et $q_j = \mathbf{P}(X \in B_j)$.

effectifs mesurés

	B_1	$\cdots$	B_l	totaux
A_1	$n_{1;1}$	$\cdots$	$n_{1;l}$	$n_{1;\heartsuit}$
$\vdots$	$\vdots$	$n_{i;j}$	$\vdots$	$\vdots$
A_k	$n_{k;1}$	$\cdots$	$n_{k;l}$	$n_{k;\heartsuit}$
totaux	$n_{\heartsuit;1}$	$\cdots$	$n_{\heartsuit;l}$	n

effectifs théoriques

	B_1	$\cdots$	B_l	totaux
A_1	np_1q_1	$\cdots$	np_1q_l	np_1
$\vdots$	$\vdots$	np_iq_j	$\vdots$	$\vdots$
A_k	np_kq_1	$\cdots$	np_kq_l	np_k
totaux	nq_1	$\cdots$	nq_l	n

Les mathématiciens ont montré que si les effectifs théoriques np_iq_j sont de taille au moins 5, alors la variable aléatoire D^2 suivante suit approximativement une loi du χ^2 à $kl - 1$ degrés de liberté (on enlève à kl un degré de liberté, car les variables aléatoires $N_{i;j}$ sont liées entre-elles par la condition : $\sum_{i,j} N_{i;j} = n$).

$$D^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(N_{i;j} - np_iq_j)^2}{np_iq_j}$$

Lorsque que les paramètres p_i et q_j sont inconnus : on les estime[5] par $p_i = \frac{n_{i;\heartsuit}}{n}$ et $q_j = \frac{n_{\heartsuit;j}}{n}$. Ces estimateurs sont naturels, ils font en sorte que les totaux des deux tableaux soient identiques.

Dans ce cas, il faut recalculer les degrés de liberté, il faut encore enlever $(k-1)+(l-1)$ degrés de liberté (les estimateurs p_i et q_j sont liés entre-eux par les conditions $\sum_i p_i = 1$ et $\sum_j q_j = 1$). Ainsi le nombre de degré de liberté vaut $kl - 1 - (l-1) - (k-1) = kl - k - l + 1 = k(l-1) - (l-1) = (k-1)(l-1)$.

On estime D^2 à l'aide des mesures effectuées $d^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(n_{i;j} - np_iq_j)^2}{np_iq_j}$ (attention à bien faire la différence entre les majuscules et les minuscules). L'estimation d^2 donnée par les mesures va donc se trouver quelque part sous la loi.

L'allure de la loi de probabilité et le critère de rejet sont les mêmes que pour le test d'adéquation (attention aux degrés de liberté qui changent).

Exemple

Cet exemple est tiré du livre «Statistics» (2e édition) écrit par Freeman, Pisani, Purves et Adhikari, aux éditions Norton, international student edition.

Une étude basée sur 2237 américains âgés de 25 à 34 a permis de montrer que les femmes sont plus souvent droitières que les hommes.

L'étude consistait à faire un test d'hypothèse sur l'indépendance entre le sexe d'une personne et le fait qu'elle soit droitière ou gauchère (ou ambidextre).

On teste l'alternative suivante.

H_0 : le sexe d'une personne est indépendant du fait qu'elle soit gauchère ou droitière.
H_1 : il n'y a pas indépendance.

Voici les données observées (effectifs mesurés) et les effectifs théoriques (calculés à partir de ces mesures sous l'hypothèse H_0).

effectifs mesurés

	femmes	hommes	totaux
droitiers	1070	934	2004
gauchers	92	113	205
ambidextres	8	20	28
totaux	1170	1067	2237

effectifs théoriques

	femmes	hommes	totaux
droitiers	1048	956	2004
gauchers	107	98	205
ambidextres	15	13	28
totaux	1170	1067	2237

Pour que l'approximation par la loi du chi-carré soit bonne, il faut que les effectifs théoriques soient plus grands ou égaux à 5. C'est le cas ici, et on a

$$d^2 = \sum_{i=1}^{3} \sum_{j=1}^{2} \frac{(n_{i;j} - np_iq_j)^2}{np_iq_j} = \frac{(1070-1048)^2}{1048} + \frac{(934-956)^2}{956} + \cdots + \frac{(20-13)^2}{13} \cong 12$$

(sans arrondir les effectifs théoriques, on trouve environ 11.806)

Les paramètres p_i et q_j ont été estimés, on se retrouve avec $(3-1)(2-1) = 2$ degrés de liberté.

Comme $d^2 > \phi_2^{-1}(0.99) \cong 9.210$, on rejette l'hypothèse H_0 au seuil de signification de 1% (c'est-à-dire avec une probabilité de 1% de chance de rejeter à tort).

Les données laissent penser que les femmes sont plus souvent droitières que les hommes.

5. Il s'agit des estimateurs du maximum de vraisemblance, notés $\widehat{p_i}_{\mathrm{MLE}}$ et $\widehat{q_j}_{\mathrm{MLE}}$.

3.5 La p-valeur associée à un test d'hypothèse

La *p-valeur* est la probabilité sous l'hypothèse nulle H_0 que la statistique de test[6] soit au moins aussi "extrême" que la valeur observée à partir des données.
La signification du mot "extrême" dépend de la façon dont la probabilité indiquée par le seuil de signification α est définie dans le test d'hypothèse en question.

Exemple d'un test bilatéral avec une distribution symétrique

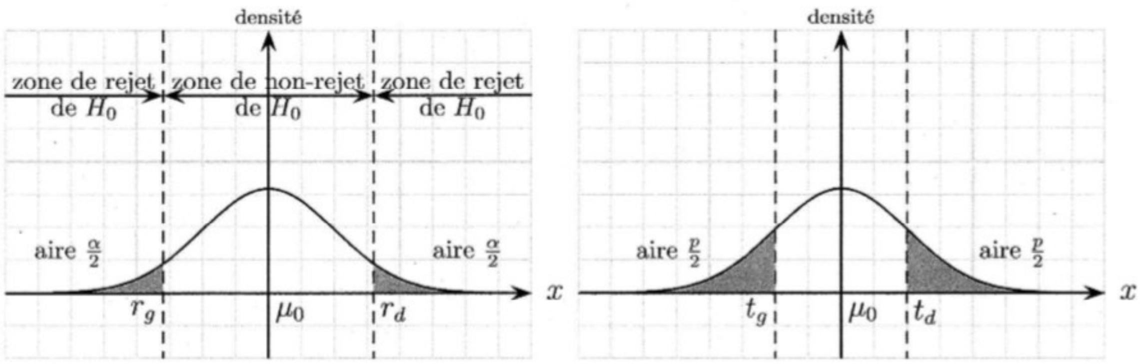
Rappelons que les distributions des lois normales, de Student et binomiales sont des distributions symétriques.
Dans le cas des tests bilatéraux où la statistique de test T possède une distribution symétrique, le seuil de signification α satisfait la condition suivante :

$$\alpha = \underbrace{\mathbb{P}(T < r_g)}_{\frac{\alpha}{2}} + \underbrace{\mathbb{P}(T > r_d)}_{\frac{\alpha}{2}} \quad \text{avec} \quad r_g = F^{-1}(\tfrac{\alpha}{2}) \quad \text{et} \quad r_d = F^{-1}(1 - \tfrac{\alpha}{2})$$

où $F(x)$ est la fonction de répartition associée à la statistique de test T. Les nombres r_g et r_d sont les limites de la zone de rejet de H_0.
De façon similaire, on définit la p-valeur par $\quad p = \underbrace{\mathbb{P}(T < t_g)}_{\frac{p}{2}} + \underbrace{\mathbb{P}(T > t_d)}_{\frac{p}{2}}$

où t_g est la valeur à gauche qui est obtenue par symétrie de t (observation de la statistique de test T) par rapport à l'axe de symétrie de la distribution de T et où t_d est la valeur symétrique de t à droite (l'une des deux est égale à t !).

Description graphique d'un tel test d'hypothèse (pour cette situation, H_0 ne peut pas être rejetée au seuil de signification α).



Utilité de la p-valeur

Par construction de la p-valeur, on rejette H_0 au seuil de signification α lorsque $p < \alpha$.
En pratique, on peut aussi utiliser le tableau suivant :

p-valeur	évidence contre H_0	seuil de signification associé
p-valeur $\geqslant 0.10$	négligeable	$10\% < \alpha$
$0.10 >$ p-valeur $\geqslant 0.05$	faible	$5\% < \alpha \leqslant 10\%$
$0.05 >$ p-valeur $\geqslant 0.01$	modérée	$1\% < \alpha \leqslant 5\%$
$0.01 >$ p-valeur $\geqslant 0.001$	forte	$0.1\% < \alpha \leqslant 1\%$
$0.001 >$ p-valeur	très forte	$\alpha \leqslant 0.1\%$

6. Les *statistiques de test* sont les variables aléatoires utilisées pour les tests d'hypothèses.