Chapitre 5 : Tests d'hypothèses

- Le problème de décision
- Formalisation du problème de test paramétrique sur un échantillon
- Tests sur la moyenne d'une loi normale
- Lien entre tests d'hypothèses et intervalles de confiance
- Procédure pour construire un test d'hypothèses
- Tests sur la variance d'une loi normale
- Tests sur une proportion
- Le test du χ^2

Grenoble INP Ensimag

Le problème de décision

But : prendre des décisions sur une activité risquée au vu de résultats d'expériences ou d'observation de phénomènes dans un contexte incertain.

- *élections* : au vu du sondage, se prononcer sur le vainqueur de l'élection.
- *essais thérapeutiques* : décider si un nouveau traitement médical est meilleur qu'un ancien au vu du résultat de son expérimentation sur des malades.
- *santé* : trancher sur la nocivité ou non des OGM ou des antennes de téléphonie mobile.
- *finance* : au vu du marché, décider si on doit ou pas se lancer dans une opération financière donnée.
- *justice* : décider si l'accusé est innocent ou coupable à partir des informations acquises pendant le procès.

Hypothèses et erreurs

Dans chaque cas, le **problème de décision** consiste à trancher, au vu d'observations, entre une hypothèse appelée **hypothèse nulle**, notée H_0, et une autre hypothèse dite **hypothèse alternative**, notée H_1.

Un **test d'hypothèses** est une procédure qui permet de choisir entre ces deux hypothèses.

Hypothèses et erreurs

Dans chaque cas, le **problème de décision** consiste à trancher, au vu d'observations, entre une hypothèse appelée **hypothèse nulle**, notée H_0, et une autre hypothèse dite **hypothèse alternative**, notée H_1.

Un **test d'hypothèses** est une procédure qui permet de choisir entre ces deux hypothèses.

Dans un problème de décision, deux types d'erreurs sont possibles :

- **erreur de première espèce** : décider que H_1 est vraie alors que H_0 est vraie.
- **erreur de seconde espèce** : décider que H_0 est vraie alors que H_1 est vraie.

Les conséquences de ces deux erreurs peuvent être d'importances diverses. En général, une des erreurs est plus grave que l'autre.

Probabilités de bonne et mauvaise décision

A toute décision correspond une probabilité de décider juste et une probabilité de se tromper :

Vérité / Décision	H_0	H_1
H_0		
H_1		

Probabilités de bonne et mauvaise décision

A toute décision correspond une probabilité de décider juste et une probabilité de se tromper :

Vérité / Décision	H_0	H_1
H_0		
H_1	α	

- la probabilité de l'erreur de première espèce, qui est la probabilité de rejeter à tort H_0, est notée α et est appelée **seuil** ou **niveau de signification** du test.

Probabilités de bonne et mauvaise décision

A toute décision correspond une probabilité de décider juste et une probabilité de se tromper :

Vérité / Décision	H_0	H_1
H_0		$1-\beta$
H_1	α	

- la probabilité de l'erreur de première espèce, qui est la probabilité de rejeter à tort H_0, est notée α et est appelée **seuil** ou **niveau de signification** du test.
- la probabilité de l'erreur de deuxième espèce est notée $1-\beta$.

Probabilités de bonne et mauvaise décision

A toute décision correspond une probabilité de décider juste et une probabilité de se tromper :

Vérité / Décision	H_0	H_1
H_0		$1-\beta$
H_1	α	β

- la probabilité de l'erreur de première espèce, qui est la probabilité de rejeter à tort H_0, est notée α et est appelée **seuil** ou **niveau de signification** du test.
- la probabilité de l'erreur de deuxième espèce est notée $1-\beta$.
- β est la probabilité de décider H_1 ou de rejeter H_0 à raison. Elle est appelée **puissance** du test.

Probabilités de bonne et mauvaise décision

A toute décision correspond une probabilité de décider juste et une probabilité de se tromper :

Vérité / Décision	H_0	H_1
H_0	$1-\alpha$	$1-\beta$
H_1	α	β

- la probabilité de l'erreur de première espèce, qui est la probabilité de rejeter à tort H_0, est notée α et est appelée **seuil** ou **niveau de signification** du test.
- la probabilité de l'erreur de deuxième espèce est notée $1-\beta$.
- β est la probabilité de décider H_1 ou de rejeter H_0 à raison. Elle est appelée **puissance** du test.
- $1-\alpha$ est parfois appelée **niveau de confiance** du test.

Choix de H_0 et H_1

Il est impossible de minimiser les deux risques d'erreur en même temps.

Dans la pratique, on va donc considérer que l'une des deux erreurs est plus importante que l'autre, et tâcher d'éviter que cette erreur se produise.

Par exemple, dans le cas du procès, on fait en général tout pour éviter de condamner un innocent, quitte à prendre le risque d'acquitter un coupable.

Choix de H_0 et H_1

Il est impossible de minimiser les deux risques d'erreur en même temps.

Dans la pratique, on va donc considérer que l'une des deux erreurs est plus importante que l'autre, et tâcher d'éviter que cette erreur se produise.

Par exemple, dans le cas du procès, on fait en général tout pour éviter de condamner un innocent, quitte à prendre le risque d'acquitter un coupable.

$\Rightarrow$ On va choisir H_0 et H_1 de sorte que l'erreur que l'on cherche à éviter soit l'erreur de première espèce.

Mathématiquement cela revient à se fixer la valeur du seuil du test α. Plus la conséquence de l'erreur est grave, plus α sera choisi petit.

Le **principe de précaution** consiste à limiter au maximum la probabilité de se tromper, donc à prendre α très petit.

Règle de décision

Règle de décision : règle qui permet de choisir entre H_0 et H_1 au vu des observations $x_1, \ldots, x_n$, sous la contrainte que la probabilité de rejeter à tort H_0 est égale à α fixé.

Une idée naturelle est de conclure que H_0 est fausse si il est très peu probable d'observer $x_1, \ldots, x_n$ quand H_0 est vraie.

Exemple : on doit décider si une pièce est truquée ou pas au vu de 100 lancers de cette pièce. Si on observe 90 piles, il est logique de conclure que la pièce est truquée et on pense avoir une faible probabilité de se tromper en concluant cela. Mais si on observe 65 piles, que conclure ?

Région critique

Région critique du test : W est l'ensemble des valeurs des observations $x_1, \ldots, x_n$ pour lesquelles on rejettera H_0.

W dépend du seuil α et est déterminée a priori, indépendamment de la valeur des observations.

Si les observations appartiennent à W, on rejette H_0, sinon on ne la rejette pas.

Région critique

Région critique du test : W est l'ensemble des valeurs des observations $x_1, \ldots, x_n$ pour lesquelles on rejettera H_0.

W dépend du seuil α et est déterminée a priori, indépendamment de la valeur des observations.

Si les observations appartiennent à W, on rejette H_0, sinon on ne la rejette pas.

Remarque : il vaut mieux dire "*ne pas rejeter H_0*" que "*accepter H_0*".

- Si on rejette H_0, c'est que les observations sont telles qu'il est très improbable que H_0 soit vraie.
- Si on ne rejette pas H_0, c'est qu'on ne dispose pas de critères suffisants pour pouvoir dire que H_0 est fausse. Mais cela ne veut pas dire que H_0 est vraie.

Récapitulatif

1. Choisir H_0 et H_1 de sorte que ce qui importe, c'est de ne pas se tromper en rejetant H_0.
2. Se fixer α selon la gravité des conséquences de l'erreur de première espèce.
3. Déterminer la région critique W.
4. Regarder si les observations se trouvent ou pas dans W.
5. Prendre la décision, c'est-à-dire conclure au rejet ou au non-rejet de H_0.

Grenoble INP Ensimag

Pour le même problème de décision, plusieurs tests de même seuil sont souvent possibles. Le meilleur de ces tests est celui qui minimisera la probabilité de l'erreur de seconde espèce.

Tests paramétriques : tests dont les hypothèses peuvent se traduire sur la valeur d'un paramètre d'une loi de probabilité.

Exemple de l'élection : le problème est de trancher entre les deux hypothèses "$p \leq 1/2$" et "$p > 1/2$".

Tests non paramétriques : tous les autres tests.

Suite du chapitre :

- tests d'hypothèses paramétriques quand l'observation est un échantillon d'une loi de probabilité.
- le plus célèbre des tests d'hypothèses : le test du χ^2.

Formalisation du problème de test paramétrique sur un échantillon

Observations : $x_1, \ldots, x_n$ réalisations de variables aléatoires $X_1, \ldots, X_n$ indépendantes et de même loi, dépendant d'un paramètre inconnu θ.

On supposera $\theta \in \mathbf{R}$. Si θ est un paramètre vectoriel, on fera des tests sur chacune de ses composantes.

Par exemple, on fera des tests sur la moyenne de la loi normale, puis des tests sur la variance, mais pas sur les deux en même temps.

Hypothèse simple = hypothèse du type "$\theta = \theta_0$", où θ_0 est un réel fixé.

Hypothèse composite ou **multiple** : hypothèse du type "$\theta \in A$" où A est une partie de $\mathbf{R}$ non réduite à un élément.

Tests d'hypothèses simples

Les hypothèses nulle et alternative sont simples toutes les deux :

$$\text{Test de } H_0 : \text{“}\theta = \theta_0\text{”} \text{ contre } H_1 : \text{“}\theta = \theta_1\text{”}.$$

Un tel test est un cas d'école : il permet de dire laquelle des deux valeurs θ_0 et θ_1 est la plus vraisemblable au vu des observations. Mais il ne prend pas en compte la possibilité que θ ne soit égal ni à θ_0 ni à θ_1. Pour cela, il faudra faire un test d'hypothèses composites.

Tests d'hypothèses simples

Les hypothèses nulle et alternative sont simples toutes les deux :

$$\text{Test de } H_0 : \text{“}\theta = \theta_0\text{” contre } H_1 : \text{“}\theta = \theta_1\text{”}.$$

Un tel test est un cas d'école : il permet de dire laquelle des deux valeurs θ_0 et θ_1 est la plus vraisemblable au vu des observations. Mais il ne prend pas en compte la possibilité que θ ne soit égal ni à θ_0 ni à θ_1. Pour cela, il faudra faire un test d'hypothèses composites.

Seuil du test = probabilité de rejeter à tort H_0 :

$$\alpha = P\left((X_1, \ldots, X_n) \in W; \theta_0\right)$$

Tests d'hypothèses simples

Les hypothèses nulle et alternative sont simples toutes les deux :

$$\text{Test de } H_0 : \text{“}\theta = \theta_0\text{”} \text{ contre } H_1 : \text{“}\theta = \theta_1\text{”}.$$

Un tel test est un cas d'école : il permet de dire laquelle des deux valeurs θ_0 et θ_1 est la plus vraisemblable au vu des observations. Mais il ne prend pas en compte la possibilité que θ ne soit égal ni à θ_0 ni à θ_1. Pour cela, il faudra faire un test d'hypothèses composites.

Seuil du test = probabilité de rejeter à tort H_0 :

$$\alpha = P\left((X_1, \ldots, X_n) \in W; \theta_0\right)$$

Puissance du test = probabilité de rejeter à raison H_0 :

$$\beta = P\left((X_1, \ldots, X_n) \in W; \theta_1\right)$$

Tests d'hypothèses composites

L'une au moins des deux hypothèses est composite.

Les tests les plus usuels sont du type :

- **Test bilatéral** :

$$\text{test de } H_0 : \text{“}\theta = \theta_0\text{”} \text{ contre } H_1 : \text{“}\theta \neq \theta_0\text{”}$$

- **Tests unilatéraux** :

$$\text{test de } H_0 : \text{“}\theta \leq \theta_0\text{”} \text{ contre } H_1 : \text{“}\theta > \theta_0\text{”}$$
$$\text{test de } H_0 : \text{“}\theta \geq \theta_0\text{”} \text{ contre } H_1 : \text{“}\theta < \theta_0\text{”}$$

Tests d'hypothèses composites

La puissance d'un test a été définie comme la probabilité de rejeter H_0 à raison, c'est à dire de rejeter H_0 quand H_1 est vraie.

Or, dans les exemples ci-dessus, il y a une infinité de valeurs de θ pour lesquelles H_1 est vraie.

Donc la puissance du test doit dépendre de la vraie valeur (inconnue) de θ.

Puissance du test : $\beta(\theta) = P\left((X_1, \ldots, X_n) \in W; \theta\right) =$ probabilité de rejeter H_0 quand la vraie valeur du paramètre est θ.

Seuil du test : $\alpha = \underset{H_0}{Sup}\, \beta(\theta) =$ probabilité maximale de rejeter H_0 à tort.

- test bilatéral : $\alpha = \beta(\theta_0)$.
- premier test unilatéral : $\alpha = \underset{\theta \leq \theta_0}{Sup}\, \beta(\theta)$.

Exemple : essais thérapeutiques

Pour apaiser un certain type de maux de tête, on a l'habitude de traiter les malades avec un médicament A.

Une étude statistique a montré que la durée de disparition de la douleur chez les malades traités avec A était une variable aléatoire de loi normale $\mathcal{N}\left(m_0, \sigma_0^2\right)$, avec $m_0 = 30$ mn et $\sigma_0 = 5$ mn.

Un laboratoire pharmaceutique a conçu un nouveau médicament B et désire tester son efficacité.

Pour cela, le nouveau médicament a été administré à $n = 12$ malades cobayes, et on a mesuré la durée de disparition de la douleur pour chacun d'entre eux : $x_1, \ldots, x_n$.

25	28	20	32	17	24	41	28	25	30	27	24

Une étude de statistique descriptive a amené à admettre que cette durée est une variable aléatoire de loi normale $\mathcal{N}\left(m, \sigma^2\right)$.

Exemple : essais thérapeutiques

L'effet du nouveau médicament se traduit sur la valeur de la durée moyenne de disparition de la douleur :

- "$m = m_0$" : B a en moyenne le même effet que A.
- "$m < m_0$" : B est en moyenne plus efficace que A.
- "$m > m_0$" : B est en moyenne moins efficace que A.

Pour savoir s'il faut commercialiser B, il faut trancher entre ces 3 hypothèses.

Exemple : essais thérapeutiques

L'effet du nouveau médicament se traduit sur la valeur de la durée moyenne de disparition de la douleur :

- "$m = m_0$" : B a en moyenne le même effet que A.
- "$m < m_0$" : B est en moyenne plus efficace que A.
- "$m > m_0$" : B est en moyenne moins efficace que A.

Pour savoir s'il faut commercialiser B, il faut trancher entre ces 3 hypothèses.
L'important est de ne pas se tromper si on décide de changer de médicament : il est préférable de conserver un médicament moins performant que le nouveau que d'adopter un médicament moins performant que l'ancien.

$\Rightarrow$ il faut tester H_0 : "$m \geq m_0$" contre H_1 : "$m < m_0$".

Test sur la moyenne d'une loi normale : première idée

Puisque $\bar{X}_n$ est l'ESBVM de m, une première idée est de conclure que $m < m_0$ si et seulement si $\bar{x}_n < m_0$.

$\Rightarrow$ on conclut que B est meilleur que A si la durée moyenne de disparition de la douleur sur les malades traités avec B est plus petite que ce qu'elle est sur les malades traités avec A.

Cela revient à proposer comme région critique du test :

$$W = \{(x_1, \ldots, x_n); \bar{x}_n < m_0\}$$

Test sur la moyenne d'une loi normale : première idée

Puisque $\bar{X}_n$ est l'ESBVM de m, une première idée est de conclure que $m < m_0$ si et seulement si $\bar{x}_n < m_0$.

$\Rightarrow$ on conclut que B est meilleur que A si la durée moyenne de disparition de la douleur sur les malades traités avec B est plus petite que ce qu'elle est sur les malades traités avec A.

Cela revient à proposer comme région critique du test :

$$W = \{(x_1, \dots, x_n); \bar{x}_n < m_0\}$$

Seuil du test :

$$\alpha = \underset{H_0}{Sup}\, \beta(m) = \underset{m \geq m_0}{Sup}\ P\left(\bar{X}_n < m_0; m\right) = \dots = 1/2.$$

Il y a donc une chance sur deux de se tromper si on décide que B est plus efficace que A quand $\bar{x}_n < m_0$. C'est beaucoup trop.

Test sur la moyenne d'une loi normale : deuxième idée

Il faut rejeter H_0 quand $\bar{x}_n$ est **significativement plus petit** que m_0.

Cela revient à prendre une région critique de la forme :

$$W = \{(x_1, \ldots, x_n); \bar{x}_n < l_\alpha\}, \text{ où } l_\alpha < m_0.$$

Test sur la moyenne d'une loi normale : deuxième idée

Il faut rejeter H_0 quand $\bar{x}_n$ est **significativement plus petit** que m_0.

Cela revient à prendre une région critique de la forme :

$$W = \{(x_1, \dots, x_n); \bar{x}_n < l_\alpha\}, \text{ où } l_\alpha < m_0.$$

Seuil du test :

$$\alpha = \underset{m \geq m_0}{Sup}\ P\left(\bar{X}_n < l_\alpha; m\right) = \underset{m \geq m_0}{Sup}\ \phi\left(\sqrt{n}\,\frac{l_\alpha - m}{\sigma}\right) = \phi\left(\sqrt{n}\,\frac{l_\alpha - m_0}{\sigma}\right)$$

$$\Rightarrow l_\alpha = m_0 + \frac{\sigma}{\sqrt{n}}\phi^{-1}(\alpha) = m_0 - \frac{\sigma}{\sqrt{n}}u_{2\alpha}$$

Un test de seuil α de H_0 : "$m \geq m_0$" contre H_1 : "$m < m_0$" est déterminé par la région critique :

$$W = \left\{(x_1, \dots, x_n); \bar{x}_n < m_0 - \frac{\sigma}{\sqrt{n}}u_{2\alpha}\right\}$$

Test sur la moyenne d'une loi normale : troisième idée

Problème : test inutilisable car σ est inconnu. $\Rightarrow$ on remplace σ par S'_n.

Seuil du test :

$$\alpha = \underset{m \geq m_0}{Sup}\ P\left(\bar{X}_n < l_\alpha; m\right) = \cdots = F_{St(n-1)}\left(\sqrt{n}\,\frac{l_\alpha - m_0}{S'_n}\right)$$

$$\Rightarrow l_\alpha = m_0 - \frac{S'_n}{\sqrt{n}}\, t_{n-1,2\alpha}.$$

Un test de seuil α de H_0 : "$m \geq m_0$" contre H_1 : "$m < m_0$" est déterminé par la région critique :

$$\begin{aligned} W &= \left\{(x_1, \ldots, x_n); \bar{x}_n < m_0 - \frac{s'_n}{\sqrt{n}}\, t_{n-1,2\alpha}\right\} \\ &= \left\{(x_1, \ldots, x_n); \sqrt{n}\,\frac{\bar{x}_n - m_0}{s'_n} < -t_{n-1,2\alpha}\right\} \end{aligned}$$

Exemple de l'essai thérapeutique

Avec le médicament A, la durée moyenne de disparition de la douleur était $m_0 = 30$ mn. On a administré le médicament B à $n = 12$ malades et relevé les durées de disparition de la douleur suivants :

25	28	20	32	17	24	41	28	25	30	27	24

$\bar{x}_n = 26.75$. $s'_n = 6.08$. $\alpha = 5\%$.

$\Rightarrow \sqrt{n}\,\dfrac{\bar{x}_n - m_0}{s'_n} = -1.853$ et $t_{n-1,2\alpha} = 1.796$.

$-1.853 < -1.796$, donc les observations sont dans la région critique.

$\Rightarrow$ on rejette H_0.

$\Rightarrow$ on conclut que B est plus efficace que A, avec moins de 5% de chances de se tromper.

$\Rightarrow$ on peut lancer la commercialisation du médicament B.

La p-valeur - 1

Si on avait pris $\alpha = 1\%$, on aurait eu $t_{11,0.02} = 2.718$.

Comme $-1.853 > -2.718$, on n'aurait pas rejeté H_0, donc on n'aurait pas adopté le médicament B.

$\Rightarrow$ nuance entre"ne pas rejeter H_0" et "accepter H_0" : on conclut que rien ne prouve que B est plus efficace que A, mais on ne conclut évidemment pas que A est plus efficace que B.

La p-valeur - 1

Si on avait pris $\alpha = 1\%$, on aurait eu $t_{11,0.02} = 2.718$.

Comme $-1.853 > -2.718$, on n'aurait pas rejeté H_0, donc on n'aurait pas adopté le médicament B.

$\Rightarrow$ nuance entre"ne pas rejeter H_0" et "accepter H_0" : on conclut que rien ne prouve que B est plus efficace que A, mais on ne conclut évidemment pas que A est plus efficace que B.

Il existe un seuil critique α_c tel que :

- $\forall \alpha > \alpha_c$, on rejettera H_0.
- $\forall \alpha \leq \alpha_c$, on ne rejettera pas H_0.

α_c est appelée la **p-valeur** du test.

La p-valeur - 2

$$\sqrt{n}\,\frac{\bar{x}_n - m_0}{s'_n} = -t_{n-1,2\alpha_c} \Rightarrow \alpha_c = F_{St(n-1)}\left(\sqrt{n}\,\frac{\bar{x}_n - m_0}{s'_n}\right).$$

Essai thérapeutique : $\alpha_c = 4.5\%$.

- $\forall \alpha > 4.5\%$, on rejettera H_0 (cas de 5%).
- $\forall \alpha \leq 4.5\%$, on ne rejettera pas H_0 (cas de 1%).

La p-valeur peut être comprise comme étant la probabilité, sous l'hypothèse nulle, que la statistique de test soit encore plus extrême que ce qui a été observé.

Si cette probabilité est forte, il n'y a pas de raison de douter de la véracité de H_0.

Si elle est faible, on peut en douter.

$\Rightarrow$ plus la p-valeur est petite, moins on prend de risque en rejettant H_0.

```
>  medic<-c(25,28,20,32,17,24,41,28,25,30,27,24)
>  t.test(medic,alternative="less",mu=30)

        One Sample t-test

data: medic
t = -1.8526, df = 11, p-value = 0.04547
alternative hypothesis: true mean is less than 30
95 percent confidence interval:
      -Inf 29.90056
sample estimates:
mean of x
    26.75
```

Les tests de Student

Tests de Student sur la moyenne d'une loi normale.

- Test de "$m \leq m_0$" contre "$m > m_0$" :

$$W = \left\{ \sqrt{n}\,\frac{\bar{x}_n - m_0}{s'_n} > t_{n-1,2\alpha} \right\}$$

- Test de "$m \geq m_0$" contre "$m < m_0$" :

$$W = \left\{ \sqrt{n}\,\frac{\bar{x}_n - m_0}{s'_n} < -t_{n-1,2\alpha} \right\}$$

- Test de "$m = m_0$" contre "$m \neq m_0$" :

$$W = \left\{ \left| \sqrt{n}\,\frac{\bar{x}_n - m_0}{s'_n} \right| > t_{n-1,\alpha} \right\}$$

Lien entre tests d'hypothèses et intervalles de confiance

Dans le test bilatéral, on rejette l'hypothèse "$m = m_0$" à condition que

$$\left| \sqrt{n}\, \frac{\bar{x}_n - m_0}{s'_n} \right| > t_{n-1,\alpha}$$

Cela revient à rejeter "$m = m_0$" si

$$m_0 \notin \left[\bar{x}_n - \frac{s'_n}{\sqrt{n}}\, t_{n-1,\alpha}, \bar{x}_n + \frac{s'_n}{\sqrt{n}}\, t_{n-1,\alpha} \right]$$

$\Rightarrow$ on rejette "$m = m_0$" au seuil α si m_0 n'appartient pas à l'intervalle de confiance bilatéral de seuil α pour m.

Grenoble INP Ensimag

t.test

En R, la commande `t.test` permet à la fois d'effectuer un test et d'obtenir un intervalle de confiance sur la moyenne de la loi normale.

Exemple des niveaux de bruit :

```
>  t.test(bruit, conf.level=0.95)

        One Sample t-test

data: bruit
t = 55.7889, df = 19, p-value <  2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 61.82992 66.65008
sample estimates:
mean of x
    64.24
```

Procédure pour construire un test d'hypothèses - 1

Si on connaît un estimateur $\hat{\theta}_n$ de θ, on procèdera de la façon suivante :

- Test de "$\theta \leq \theta_0$" contre "$\theta > \theta_0$" : on rejette H_0 si $\hat{\theta}_n$ est "trop grand".

$$W = \left\{\hat{\theta}_n > l_\alpha\right\}$$

- Test de "$\theta \geq \theta_0$" contre "$\theta < \theta_0$" : on rejette H_0 si $\hat{\theta}_n$ est "trop petit".

$$W = \left\{\hat{\theta}_n < l_\alpha\right\}$$

- Test de "$\theta = \theta_0$" contre "$\theta \neq \theta_0$" : on rejette H_0 si $\left|\hat{\theta}_n - \theta_0\right|$ est "trop grand" ou bien si $\hat{\theta}_n$ est "soit trop grand, soit trop petit".

$$W = \left\{\hat{\theta}_n < l_{1,\alpha} \text{ ou } \hat{\theta}_n > l_{2,\alpha}\right\}, \text{ avec } l_{1,\alpha} < l_{2,\alpha}$$

Procédure pour construire un test d'hypothèses - 2

Pour déterminer $l_\alpha, l_{1,\alpha}, l_{2,\alpha}$, on écrit que le seuil du test est

$$\alpha = \underset{H_0}{Sup}\, P\left((X_1, \ldots, X_n) \in W; \theta\right)$$

Par exemple, dans le premier cas, $\alpha = \underset{\theta \leq \theta_0}{Sup}\, P\left(\hat{\theta}_n > l_\alpha\right)$.

Pour pouvoir calculer $P\left(\hat{\theta}_n > l_\alpha\right)$, il faut utiliser une fonction pivotale.

Si cette procédure de bon sens ne marche pas, il faut utiliser des résultats mathématiques plus sophistiqués.

Tests sur la variance d'une loi normale

Observations : $x_1, \ldots, x_n$ réalisations de v.a. $X_1, \ldots, X_n$ i.i.d. de loi $\mathcal{N}(m, \sigma^2)$. On souhaite tester par exemple

$$H_0 : \text{“}\sigma^2 \leq \sigma_0^2\text{”} \text{ contre } H_1 : \text{“}\sigma^2 > \sigma_0^2\text{”}.$$

L'ESBVM de σ^2 est $S_n'^2 \Rightarrow$ on propose

$$W = \left\{ s_n'^2 > l_\alpha \right\}$$

Fonction pivotale : $\dfrac{(n-1)S_n'^2}{\sigma^2}$ est de loi χ^2_{n-1}.

Seuil du test :

$$\alpha = \underset{\sigma^2 \leq \sigma_0^2}{Sup} P(S_n'^2 > l_\alpha) = \ldots = 1 - F_{\chi^2_{n-1}} \left(\frac{(n-1) l_\alpha}{\sigma_0^2} \right)$$

$$\Rightarrow l_\alpha = \frac{\sigma_0^2}{n-1} F^{-1}_{\chi^2_{n-1}}(1-\alpha) = \frac{\sigma_0^2}{n-1} z_{n-1,\alpha}$$

Grenoble INP Ensimag

Tests sur la variance d'une loi normale

- Test de "$\sigma^2 \leq \sigma_0^2$" contre "$\sigma^2 > \sigma_0^2$" :

$$W = \left\{ \frac{(n-1){s'_n}^2}{\sigma_0^2} > z_{n-1,\alpha} \right\}$$

- Test de "$\sigma^2 \geq \sigma_0^2$" contre "$\sigma^2 < \sigma_0^2$" :

$$W = \left\{ \frac{(n-1){s'_n}^2}{\sigma_0^2} < z_{n-1,1-\alpha} \right\}$$

- Test de "$\sigma^2 = \sigma_0^2$" contre "$\sigma^2 \neq \sigma_0^2$" :

$$W = \left\{ \frac{(n-1){s'_n}^2}{\sigma_0^2} < z_{n-1,1-\alpha/2} \text{ ou } \frac{(n-1){s'_n}^2}{\sigma_0^2} > z_{n-1,\alpha/2} \right\}$$

Exemple des essais thérapeutiques

Avec le médicament A, l'écart-type était $\sigma_0 = 5$ mn.

Avec le médicament B, on estime σ par $s'_n = 6.08$ mn.

La variabilité du second médicament est-elle significativement supérieure à celle du premier ?

$$\Rightarrow \text{test de } H_0 : \text{“}\sigma^2 \leq 25\text{”} \text{ contre } H_1 : \text{“}\sigma^2 > 25\text{”}$$

Au seuil $\alpha = 5\%$, $z_{11,5\%} = 19.68$. $\dfrac{(n-1){s'_n}^2}{\sigma_0^2} = 16.25$.

$16.25 < 19.68 \Rightarrow$ on n'est pas dans la région critique.

$\Rightarrow$ on ne rejette pas H_0 : on n'a pas de preuves suffisantes pour conclure que la variabilité de l'effet de B est supérieure à celle de A.

p-valeur : $z_{n-1,\alpha_c} = F^{-1}_{\chi^2_{n-1}}(1-\alpha_c) = 16.25 \ \Rightarrow\ \alpha_c = 1-F_{\chi^2_{11}}(16.25) = 13.2\%$

Donc même au seuil 10%, on ne rejettera pas H_0.

Grenoble INP Ensimag

Tests sur une proportion-1

Observations : $x_1, \ldots, x_n$ réalisations de v.a. $X_1, \ldots, X_n$ i.i.d. de loi de Bernoulli $\mathcal{B}(p)$. On souhaite faire des tests sur la valeur de p.

On sait que $T = \sum_{i=1}^{n} X_i$ est de loi binomiale $\mathcal{B}(n, p)$.

Pas de fonction pivotale simple.

Fonction pivotale asymptotique : pour n grand, $\dfrac{T - np}{\sqrt{np(1-p)}}$ est approximativement de loi $\mathcal{N}(0, 1)$.

Tests sur une proportion-2

- Test de "$p \leq p_0$" contre "$p > p_0$" :

$$W = \left\{ \frac{t - np_0}{\sqrt{np_0(1 - p_0)}} > u_{2\alpha} \right\}$$

- Test de "$p \geq p_0$" contre "$p < p_0$" :

$$W = \left\{ \frac{t - np_0}{\sqrt{np_0(1 - p_0)}} < -u_{2\alpha} \right\}$$

- Test de "$p = p_0$" contre "$p \neq p_0$" :

$$W = \left\{ \left| \frac{t - np_0}{\sqrt{np_0(1 - p_0)}} \right| > u_{\alpha} \right\}$$

Grenoble INP Ensimag

Exemple du sondage

$n = 800$ personnes interrogées, $t = 420$ déclarent voter pour A.

$\Rightarrow$ le pourcentage p de voix qu'obtiendra le candidat A est estimé par $\hat{p}_n = 420/800 = 52.5\%$.

Un intervalle de confiance de seuil 5% pour ce pourcentage est [49%, 56%], dont une partie est située sous les 50%.

La seule chose qui intéresse A, c'est de savoir s'il va être élu ou pas.

$$\Rightarrow \text{Test de } H_0 : \text{“}p \leq 1/2\text{”} \text{ contre } H_1 : \text{“}p > 1/2\text{”}$$

$\dfrac{t - np_0}{\sqrt{np_0(1-p_0)}} = \dfrac{420 - 800/2}{\sqrt{800/4}} = 1.414$. Au seuil 5%, $u_{0.1} = 1.645$.

$1.414 < 1.645$, donc on n'est pas dans la région critique, donc on ne rejette pas H_0 : on ne peut pas affirmer que A sera élu avec moins de 5% de chances de se tromper.

```
>  binom.test(420, 800, p=0.5, alternative="greater")

	Exact binomial test

data: 420 and 800
number of successes = 420, number of trials = 800, p-value =
0.08395
alternative hypothesis: true probability of success is
greater than 0.5
95 percent confidence interval:
 0.4953009 1.0000000
sample estimates:
probability of success
           0.525
```

Le test du χ^2

Exemple introductif. On jette un dé 300 fois.

face obtenue	1	2	3	4	5	6
nombre de lancers	42	43	56	55	43	61

Peut-on en conclure que le dé est équilibré ?

Idée naturelle : si le dé est équilibré, on devrait avoir à peu près $300/6 = 50$ fois chaque face.

Si le résultat s'éloigne trop de 50 sur quelques unes des faces, on peut douter du fait que le dé est équilibré.

Or on observe 61 fois la face 6 et 42 fois la face 1 : est-ce trop ou trop peu pour un dé équilibré ?

Idée naturelle : si le dé est équilibré, on devrait avoir à peu près $300/6 = 50$ fois chaque face.

Si le résultat s'éloigne trop de 50 sur quelques unes des faces, on peut douter du fait que le dé est équilibré.

Or on observe 61 fois la face 6 et 42 fois la face 1 : est-ce trop ou trop peu pour un dé équilibré ?

face obtenue	1	2	3	4	5	6
nombre de lancers	42	43	56	55	43	61
résultat "attendu"	50	50	50	50	50	50

$\Rightarrow$ on va rejeter l'hypothèse que le dé est équilibré si la "distance" entre les vecteurs $(42, 43, 56, 55, 43, 61)$ et $(50, 50, 50, 50, 50, 50)$ est "trop grande".

Il reste à choisir une distance appropriée.

Cas général

On s'intéresse à une expérience qui a k issues possibles.

Sous une certaine hypothèse H_0, les probabilités d'apparition de ces k issues sont respectivement $p_1, \ldots, p_k$ (avec $\sum_{j=1}^{k} p_j = 1$).

On fait n expériences identiques et indépendantes et on compte les nombres n_j de fois où l'issue j s'est produite (avec $\sum_{j=1}^{k} n_j = n$).

Le problème est de décider si l'observation de $n_1, \ldots, n_k$ est compatible avec l'hypothèse H_0 que les probabilités des issues sont $p_1, \ldots, p_k$.

Exemple du dé : $k = 6$, $\forall j \in \{1, \ldots, 6\}, p_j = \dfrac{1}{6}$ et $n = 300$.

Sous H_0, on s'attend à observer en moyenne np_j fois l'issue j (ex : 50 fois chaque face).

Il s'agit donc de déterminer si les n_j sont significativement proches ou éloignés des np_j.

$\Rightarrow$ Idée de région critique :

$$W = \left\{ \sum_{j=1}^{k} (n_j - np_j)^2 > l_\alpha \right\}$$

Pour déterminer l_α, il faut connaitre la loi de probabilité sous H_0 de $\sum_{j=1}^{k} (N_j - np_j)^2$, ou d'une variable aléatoire analogue.

Théorème de Pearson

Pour tout k-uplet d'entiers $(n_1, \ldots, n_k)$ tels que $\sum_{j=1}^{k} n_j = n$, on a :

$$P(N_1 = n_1, \ldots, N_k = n_k) = \frac{n!}{n_1! \ldots n_k!} p_1^{n_1} \ldots p_k^{n_k}$$

$\Rightarrow$ le vecteur $(N_1, \ldots, N_k)$ est de loi **multinomiale** $\mathcal{M}(n; p_1, \ldots, p_k)$.

Théorème de Pearson

Si $(N_1, \ldots, N_k)$ est de loi $\mathcal{M}(n; p_1, \ldots, p_k)$, alors :

$$\Delta_n^2 = \sum_{j=1}^{k} \frac{(N_j - np_j)^2}{np_j} \xrightarrow{\mathcal{L}} \chi^2_{k-1}$$

Test du χ^2

On appelle **test du khi-deux** le test de H_0 : "les probabilités des k issues sont $p_1, \ldots, p_k$" contre $H_1 = \bar{H}_0$ défini par la région critique :

$$W = \left\{ \sum_{j=1}^{k} \frac{(n_j - np_j)^2}{np_j} > z_{k-1,\alpha} \right\}$$

Exemple du dé : l'hypothèse que le dé est équilibré s'écrit H_0 : "$\forall j, p_j = \frac{1}{6}$".

La statistique de test vaut $\delta_n^2 = \frac{(42-50)^2}{50} + \ldots + \frac{(61-50)^2}{50} = 6.88$.

Au seuil $\alpha = 5\%$, $z_{5,0.05} = 11.07$.

$6.88 << 11.07$, donc on ne rejette pas H_0 : rien n'indique que le dé n'est pas équilibré.

```
>   de<-c(42,43,56,55,43,61)
>   chisq.test(de,p=rep(1/6,6))

        Chi-squared test for given probabilities

data:  de
X-squared = 6.88, df = 5, p-value = 0.2297
```

p-valeur $= 23\% \Rightarrow$ même en s'autorisant 20% de chances de se tromper, on ne rejetterait pas H_0.

$\Rightarrow$ pour un dé équilibré, il n'est pas du tout improbable d'observer 61 fois la face 6, même si on s'attend à ne l'observer en moyenne que 50 fois.

Le résultat du théorème de Pearson n'est qu'asymptotique. En pratique, on considère que l'on peut effectuer le test si pour tout j, $n_j \geq 5$..