Université de Nice
Département de Mathématiques

Année 2005-2006
Licence MI/MP 2e année

Statistiques : notes du cours 4
Tests d'hypothèse : fréquence d'un caractère dans une population

La méthode d'estimation d'un paramètre p par intervalle de confiance, décrite précédemment, fait courir à celui qui l'utilise un risque : celui d'affirmer $p \in I$ alors qu'en réalité $p \notin I$. Il est impossible en effet, en observant seulement un échantillon, d'exclure totalement un tel risque. Cependant la méthode mise en oeuvre assure également que la probabilité de se tromper ne peut dépasser un seuil α, auquel on peut donner à l'avance la valeur que l'on souhaite. On contrôle donc ce risque.

Dans le cas d'un test d'hypothèse présenté ici, la méthode est très proche de celle mise en oeuvre pour une estimation par intervalle de confiance (les formules sont pratiquement les mêmes) mais on court cette fois deux risques différents, comme nous allons l'expliquer. C'est une difficulté nouvelle.

1 Test d'une fréquence

La méthode des tests d'hypothèse appartient comme l'estimation à la statistique inférentielle. Elle est largement utilisée aujourd'hui dans tous les domaines des sciences apliquées et elle est incorporée à la plupart des logiciels de calculs statistiques. Elle remonte pour l'essentiel au mathématicien anglais Karl Pearson (1857-1936).

Il existe de nombreux types de test satistiques (test du χ^2, test de Student, test de normalité, etc...). Dans ce cours nous étudions l'un d'entre eux, le test d'une fréquence, à titre d'exemple.

Le contexte type d'un test de fréquence (qui relève du *contrôle de qualité*) est le suivant : une entreprise reçoit de son fournisseur un lot de pièces qui comporte normalement une proportion au plus égale à p_0 de pièces défectueuses. Elle veut, par examen d'un échantillon, décider entre les deux hypothèses

- H_0 (hypothèse nulle) : la *vraie* proportion p dans le lot vérifie $p \leq p_0$
- $\overline{H}_0$ (hypothèse alternative) : la *vraie* proportion p dans le lot vérifie $p > p_0$

sachant que dans le premier cas elle acceptera le lot et dans le second elle le rejettera.

De nombreuses autres situations relèvent de cette problématique, comme par exemple savoir, avant de mettre un nouveau produit sur le marché, si la proportion de consommateurs prets à l'acheter dépasse un certain seuil, connaître le taux de réussite d'un médicament avant d'autoriser son usage, déterminer l'étendue réelle d'une épidemie ou d'une pollution pour décider d'un traitement de masse, etc...

Le modèle mathématique est le même que celui de l'estimation par intervalle de confiance : une population dans laquelle chaque individu présente ou non un certain caractère, la proportion d'individus présentant le caractère étant notée p, et un échantillon aléatoire de taille n extrait de cette population. La proportion f calculée à partir de l'échantillon est considérée comme une réalisation d'une v.a. $\hat{p}$ de loi binomiale $\mathcal{B}(n,p)$ qu'on peut assimiler, si n est assez grand, à une loi normale $\mathcal{N}(p, \frac{\sqrt{p(1-p)}}{\sqrt{n}})$. Dans un tel modèle, on peut affirmer que :

Proposition 1 *Si l'on suppose la proportion p d'un caractère dans une population égale à p_0, alors tout échantillon aléatoire de taille n extrait de cette population présente ce caractère avec une proportion (variable) f qui appartient, avec une probabilité égale à $1-\alpha$, à l'intervalle*

$$I(\alpha, p_0, n) = \left[p_0 - t_\alpha \frac{\sqrt{p_0(1-p_0)}}{\sqrt{n}} \,; p_0 + t_\alpha \frac{\sqrt{p_0(1-p_0)}}{\sqrt{n}}\right] \tag{1}$$

où t_α est défini par $P(-t_\alpha \leq X \leq t_\alpha) = 1-\alpha$, pour une v.a. X de loi $\mathcal{N}(0,1)$.

On utilise ce résultat pour mettre en place un test en procédant par étapes de la façon suivante :

1. On choisit un seuil α (par exemple $\alpha = 0,05$ ou $\alpha = 0,01$)
2. On fait une hypothèse H_0, par exemple $p = p_0$
3. On choisit la taille n de l'échantillon à prélever.
   Les choix de α, p_0 et n ainsi fait conduisent à un intervalle $I(\alpha, p_0, n)$ de la forme (1), appelé parfois dans ce contexte *intervalle d'acceptation.*
4. On prélève un échantillon pour lequel on calcule la proportion f du caractère.
5. On décide : si $f \notin I(\alpha, p_0, n)$, on rejette l'hypothèse posée selon laquelle on aurait $p = p_0$ et si $f \in I(\alpha, p_0, n)$, on conclut que le test effectué ne permet pas de douter que $p = p_0$, en d'autres termes on accepte l'hypothèse faite.

Remarque : Attention à ne pas conclure, si la proportion f trouvée sur l'échantillon appartient à $I(\alpha, p_0, n)$ sans être égale à p_0 que la *vraie* proportion p de la population est en réalité égale à f. Ne pas oublier qu'un autre échantillon aurait probablement donné une autre valeur de f.

A noter aussi que lorsqu'on rejette l'hypothèse, le test ne fournit pas d'indication fiable sur la *vraie* valeur de p ; une seule affirmation est faite : la proportion trouvée f ne peut pas être due seulement aux fluctuations d'échantillonnage , elle est trop différente de p_0 pour être compatible avec l'hypothèse $p = p_0$. On dit que l'écart entre la proportion trouvée dans l'échantillon et celle supposée de la population dont il est issu est *statistiquement significative.*

2 Risques de première et deuxième espèce

Le seuil α dont dépend la taille de l'intervalle $I(\alpha, p_0, n)$ et donc la décision prise, mesure ici le *risque de rejet à tort.* On l'appelle le risque de première espèce. Si on veut réduire ce risque, il suffit de choisir une valeur de α plus petite. Il est égal à

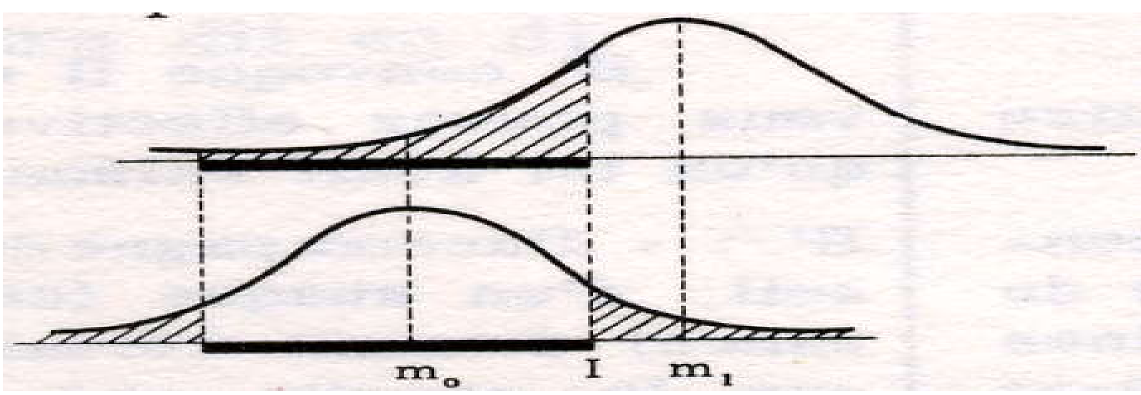
$$P(f \notin I(\alpha, p_0, n)/p = p_0).$$

Mais pour celui qui effectue un test statistique il existe un second risque de se tromper : c'est le *risque d'accepter tort.* C'est ce qui se produit lorsque dans la population la proportion p n'est pas egale à p_0 mais que dans l'échantillon prélevé la proportion soit assez proche de p_0 pour appartenir à l'intervalle $I(\alpha, p_0, n)$ et conduire à accepter l'hypothèse. On ne peut pas réduire totalement ce risque (sauf à prendre comme échantillon la population entière!). On appelle ce risque le risque de deuxième espèce. Si la *vraie* proportion est p_1, il est égal à

$$P(f \in I(\alpha, p_0, n)/p = p_1).$$

Il y a beaucoup de situations *de la vie courante* où ces deux types de risques sont présents. Par exemple le client qui choisit ou non d'acheter un produit qui lui est proposé par un vendeur (voiture d'occasion par exemple) court le risque d'y renoncer à tort (il ne retrouvera pas de meilleure affaire) et aussi celui d'accepter à tort (il achète mais il s'agit d'une mauvaise affaire). De même lors d'une embauche, l'entreprise peut refuser un candidat en sous-évaluant ses qualités mais aussi l'accepter en sous-évaluant ses défauts. Et c'est aussi la situation dans laquelle se trouve un jury d'examen.

Dans le modèle choisi, les deux graphes suivants indiquent les liens entre ces deux type de risques : on peut y voir que l'on ne peut diminuer α sans augmenter β. Mais on note aussi que, puisque σ diminue lorsque n augmente, il est possible de diminuer β à α fixé, à condition d'augmenter la taille n de l'échantillon.



3 Tests unilatéral ou bilatéral

Lorsque l'intervalle d'acceptation $I(\alpha, p_0, n)$ est symétrique comme indiqué ci-dessus on dit qu'il s'agit d'un test bilatéral. C'est le cas lorsque l'hypothèse H_0 est de la forme $p = p_0$ (contre $p \neq p_0$). Par contre lorsque l'hypothèse H_0 est de la forme $p < p_0$ (contre $p \geq p_0$) (on parle alors de test unilatéral), il convient de choisir un intervalle $I(\alpha, p_0, n)$ de la forme $I(\alpha, p_0, n) =]-\infty\,; p_0 + \tilde{t}_\alpha \frac{\sqrt{p_0(1-p_0)}}{\sqrt{n}}]$, où $\tilde{t}_\alpha$ est défini par $P(-\infty \leq X \leq \tilde{t}_\alpha) = 1 - \alpha$ pour X de loi $\mathcal{N}(0, 1)$. Par exemple $\tilde{t}_\alpha \simeq 1,64$ pour $\alpha = 0,95$. Le cas typique d'un test de qualité d'un lot de pièces est un exemple de test unilatéral.