

H. ROUANET

D. LÉPINE

Problèmes de méthodologie statistique. I. Introduction à l'étude de la robustesse des méthodes usuelles d'inférence statistique

Mathématiques et sciences humaines, tome 46 (1974), p. 21-33

http://www.numdam.org/item?id=MSH_1974__46__21_0

© Centre d'analyse et de mathématiques sociales de l'EHESS, 1974, tous droits réservés.

L'accès aux archives de la revue « Mathématiques et sciences humaines » (<http://msh.revues.org/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

PROBLÈMES DE MÉTHODOLOGIE STATISTIQUE

I. - INTRODUCTION A L'ÉTUDE DE LA ROBUSTESSE DES MÉTHODES USUELLES D'INFÉRENCE STATISTIQUE

par

H. ROUANET ET D. LÉPINE

RÉSUMÉ

La connaissance de la robustesse des méthodes usuelles d'inférence statistique vis-à-vis de divers types d'écarts par rapport au modèle de base est essentielle pour leur bonne utilisation. Dans cet article sont exposés un certain nombre de résultats (pour la plupart classiques, mais parfois mal connus) concernant la robustesse de ces méthodes vis-à-vis de la non-normalité (pour les comparaisons de moyennes, puis pour les comparaisons de variances), vis-à-vis de la non-équidistribution et de la non-indépendance, enfin vis-à-vis des valeurs aberrantes.

SUMMARY

If we want to apply correctly the usual methods of statistical inference to various types of deviation in relation to the basic models, we have to be aware of their robustness. A series of results have been presented in this article (most of which are classical but sometimes hardly known) showing the robustness of these methods in relation to non-normality (for mean comparisons, then for variance comparisons), in relation to non-equidistribution and non-independence, and finally in relation to aberrant values.

1. INTRODUCTION

Les méthodes courantes d'inférence statistique (tests de signification et intervalles de confiance) sont construites de la manière suivante : on pose un modèle probabiliste que nous appellerons *modèle de base*, caractérisé par certaines conditions, et sous ce modèle on dérive une méthode présentant des propriétés d'optimalité. Fondamentalement, une « méthode » est caractérisée par la donnée d'une variable aléatoire privilégiée (sous le modèle de base), dont la distribution est entièrement déterminée par le modèle, et à partir de laquelle on construit, soit la statistique de test, soit l'intervalle de confiance. Par exemple, pour l'inférence sur une moyenne, on pose le modèle selon lequel les observations constituent un échantillon au hasard d'une distribution parente normale et sous ce modèle on dérive la méthode du *t de Student*.

La variable aléatoire (v.a.) privilégiée est ici la différence entre la moyenne observée M et la moyenne parente μ , rapportée à l'estimation de son écart type $\frac{s}{\sqrt{n}}$ (procédure dite de « studentisation »); sous le modèle normal, cette v.a. $\frac{M - \mu}{s/\sqrt{n}}$ est (quels que soient les paramètres μ et σ^2) distribuée comme un t de Student à $n-1$ degrés de liberté, ce qui permet d'une part de tester l'hypothèse nulle $H_0 : \mu = \mu_0$ (où μ_0 est une valeur donnée) au moyen de la statistique de test $\frac{M - \mu_0}{s/\sqrt{n}}$, d'autre part de construire l'intervalle de confiance au seuil α : $(M - t_\alpha \frac{s}{\sqrt{n}}, M + t_\alpha \frac{s}{\sqrt{n}})$ (où t_α est la valeur critique de la distribution de Student au seuil α).

Or il peut arriver que certaines conditions du modèle de base ne soient pas réalistes dans une situation donnée. Si une condition du modèle de base n'est pas remplie, mais qu'un certain *contre-modèle*, plus ou moins éloigné du modèle de base, soit vrai, la méthode statistique conduit à des résultats qui peuvent être *faussés*. Mais ce faussage peut être plus ou moins grand ; on dira qu'une méthode statistique est plus ou moins *robuste* vis-à-vis d'un certain type d'écart au modèle de base selon que ce faussage est plus ou moins grand.

La connaissance de la robustesse des méthodes d'inférences courantes présente donc un intérêt considérable, car c'est finalement cette connaissance qui seule peut permettre d'apprécier dans quelle mesure ces méthodes peuvent être utilisées à bon escient. En effet, le problème essentiel, dans l'utilisation d'une méthode statistique, n'est pas de savoir si le modèle sous lequel elle est fondée (modèle de base) est réaliste, mais de savoir si cette méthode est suffisamment robuste. C'est pourquoi il n'est pas étonnant que de nombreuses recherches aient été consacrées à l'étude de la robustesse. Le but de la présente note est de constituer une introduction à cette étude ; nous présenterons un certain nombre de résultats importants relatifs à deux conditions classiques :

- condition de normalité ;
- condition d'équidistribution et d'indépendance des observations.

Auparavant nous indiquerons quelques notions générales :

— Pour un *intervalle de confiance* : soit un intervalle calculé à un seuil α à partir du modèle de base : sous ce modèle la probabilité que l'intervalle ne contienne pas le paramètre est égale à α . On appellera α le *seuil nominal*. Sous le contre-modèle la probabilité de l'événement précédent prend la valeur α' : on appellera α' le *seuil réel*. Le faussage $\alpha' - \alpha$ est positif si $\alpha' > \alpha$, négatif si $\alpha' < \alpha$. On peut également dire que si le faussage est *positif* l'intervalle de confiance est *trop étroit*, et que si le faussage est *négatif* il est *trop large*.

— Pour un *test de signification*, on distinguera le faussage sous l'hypothèse nulle : faussage du seuil, et le faussage sous une contre-hypothèse : faussage de la puissance. Soit α un seuil donné : la probabilité d'observer un résultat significatif à ce seuil si l'hypothèse nulle est vraie est par définition égale à α sous le modèle de base ; sous un contre-modèle donné cette probabilité devient α' : α est appelé le seuil nominal, α' le seuil réel. Le faussage du seuil, $\alpha' - \alpha$, est *positif* si $\alpha' > \alpha$: si l'hypothèse nulle est vraie on obtient alors *trop de résultats significatifs*. Si $\alpha' < \alpha$ le faussage du seuil est *négatif* : si l'hypothèse nulle est vraie on obtient alors *trop peu de résultats significatifs*. Par ailleurs, la probabilité d'obtenir un résultat significatif au seuil α si une certaine contre-hypothèse est vraie est par définition la puissance du test pour cette contre-hypothèse. Soit π la valeur de la puissance sous le modèle de base, π' sa valeur sous le contre-modèle. Le faussage de la puissance est $\pi' - \pi$ (c'est naturellement une fonction de la contre-hypothèse). Lorsque $\pi' > \pi$ (faussage de la puissance positif) on obtient *trop de résultats significatifs* si l'hypothèse nulle est fautive ; lorsque $\pi' < \pi$ (faussage de la puissance négatif) on obtient *trop peu de résultats significatifs* si l'hypothèse nulle est fautive.

Un faussage négatif du seuil ou un faussage positif de la puissance pourraient apparaître comme des circonstances favorables. Mais lorsque le faussage du seuil est positif, il en est généralement de même du faussage de la puissance (et inversement) au moins pour des contre-hypothèses proches de l'hypothèse nulle (ce qui se conçoit intuitivement par continuité) : nous dirons alors, généralement parlant, que le test est *faussé positivement*, et qu'il conduit à *trop de résultats significatifs*. Dans le cas contraire, nous dirons que le test est *faussé négativement*, et qu'il conduit à *trop peu de résultats significatifs*.

N.B. :

- 1) Un test faussé négativement est parfois appelé « test conservateur » : terminologie appropriée aux contextes où on cherche à favoriser l'hypothèse nulle (par exemple, pour des raisons méthodologiques, dans des cas où l'hypothèse de travail est précisément la négation de cette hypothèse nulle) : dans ces cas-là, un faussage négatif présente moins d'inconvénients qu'un faussage positif.
- 2) Lorsqu'un même problème donne lieu d'une part à une méthode d'intervalle de confiance, d'autre part à un test de signification : à partir des conclusions sur le faussage relatif à l'intervalle de confiance on passe aux conclusions sur le faussage relatif au test et inversement : à un faussage positif pour l'intervalle de confiance correspond un faussage positif pour le test et inversement. Aussi, dans la suite de ce texte, nous bornons-nous aux tests de signification.
- 3) Lorsque dans un problème concret un contre-modèle particulier apparaît plus réaliste que le modèle de base, on pourra parfois estimer un « facteur de faussage » et procéder à un « test corrigé » ; souvent, il sera au moins possible de connaître le sens (positif ou négatif) du faussage : on pourra alors tirer des observations des *conclusions a fortiori*, qui dans certains cas pourront suffire à répondre au problème d'inférence que l'on se pose. Ainsi, si le faussage est positif, une conclusion de compatibilité avec H_0 sous le modèle de base suffit à entraîner une conclusion de compatibilité sous le contre-modèle ; si le faussage est négatif, une conclusion d'incompatibilité avec H_0 sous le modèle de base entraînera *a fortiori* une conclusion d'incompatibilité sous le contre-modèle.
- 4) Les considérations développées ici s'appliquent directement à des situations du type suivant : on dispose de peu d'informations sur le modèle vrai, dont le modèle de base peut cependant représenter une formalisation admissible; on s'apprête à utiliser une méthode d'inférence que l'on sait optimale sous le modèle de base; la question principale qui se pose est alors de savoir dans quelle mesure, s'il se trouvait que le modèle vrai s'écartât du modèle de base, les conclusions en seraient affectées : l'étude de la *robustesse de la méthode* permet de répondre à cette question.

En revanche, dans d'autres situations, on peut avoir de bonnes raisons (soit *a priori*, soit au vu des données si celles-ci sont nombreuses) de penser que le modèle de base est inadmissible. Or, une méthode optimale (par construction) sous un modèle de base peut ne pas l'être sous un contre-modèle; un problème préalable se pose donc : celui du *choix d'une méthode d'inférence appropriée* (sous un modèle différent du modèle de base). Ce problème fait intervenir de nouvelles considérations que nous ne traiterons pas dans cet article¹ (telles que le conflit efficacité-robustesse que nous nous proposons d'aborder ultérieurement²) mais il est clair qu'un élément de solution indispensable de ce problème sera fourni par l'étude de la robustesse des méthodes. C'est ainsi, par exemple, que s'il apparaît que la méthode privilégiée sous le contre-modèle n'apporte pas un gain d'efficacité appréciable (lorsque le contre-modèle est vrai) par rapport à la méthode fondée sur le modèle de base, on pourra choisir d'utiliser encore celle-ci pourvu qu'elle soit robuste, alors même que l'on sait que le modèle de base n'est pas acceptable.

2. ROBUSTESSE VIS-A-VIS DE LA NON-NORMALITÉ : COMPARAISON DE MOYENNES

La condition de normalité des distributions parentes intervient dans les modèles de base de nombreuses méthodes courantes d'inférence. Or cette condition est souvent peu réaliste : contrairement à un préjugé encore assez répandu chez certains mathématiciens, il y a peu de caractères concrets qui soient distribués même approximativement selon une distribution normale. Il est donc essentiel de connaître la robustesse vis-à-vis de la non-normalité des méthodes fondées sur le modèle normal. Or il se trouve que cette robustesse diffère du tout au tout selon les méthodes. Nous présenterons ici, pour les contraster, les méthodes de comparaison de *moyennes* et les méthodes de comparaison de *variances*.

Les méthodes de comparaison de moyennes dont il sera ici question sont les suivantes :

- pour un groupe d'observations : inférence sur une moyenne par la méthode du *t de Student* ;
- pour deux groupes indépendants : comparaison des moyennes par la méthode du *t de Student* ;

1. La distinction entre « robustesse d'une méthode » et « choix d'une méthode optimale » recoupe la distinction proposée par Box et Tiao (1973, p. 152 *sq.*) entre « robustesse du critère » et « robustesse de l'inférence ».

2. H. Rouanet et D. Lépine : « Étude d'un conflit robustesse-efficacité dans le problème de la comparaison de deux moyennes (groupes indépendants) », à paraître dans *Math. Sci. hum.*, n° 47.

— pour plusieurs groupes indépendants : comparaison des moyennes par la méthode du \mathcal{F} de *Snedecor* (en d'autres termes : étude de l'effet d'un facteur systématique en analyse de variance).

Dans toutes ces méthodes, le modèle de base postule que la distribution parente des observations dans chaque groupe est normale. Une première remarque s'impose : lorsque les échantillons sont de taille élevée, on peut montrer (au moyen de théorèmes de convergence classiques ¹) que la condition de normalité devient inutile et qu'on peut la remplacer par une condition beaucoup plus faible (telle que la condition : distributions parentes de variance finie). Il en résulte, par continuité, l'importante première propriété suivante : *la robustesse vis-à-vis de la non-normalité est d'autant plus grande que les effectifs sont plus élevés.*

Pour de petits échantillons, on trouvera dans Scheffé (1959, chapitre 10) un exposé des résultats essentiels que nous résumons ici. Le plus important de ces résultats est le suivant : dès que les effectifs des échantillons ne sont plus très faibles, le faussage dû à la non-normalité est minime (disons moins de 0,01 pour un seuil nominal de 0,05) tant que les distributions ne s'écartent pas considérablement de la distribution normale. On peut considérer cette propriété comme une généralisation des théorèmes d'approximation normale classiques. Considérons par exemple le cas de la moyenne d'un groupe. Si M est la moyenne d'un échantillon d'effectif n extrait d'une distribution parente de moyenne μ et de variance σ^2 , la variable aléatoire

$\frac{M - \mu}{\sigma/\sqrt{n}}$ est distribuée à peu près normalement dès que n n'est plus très faible ; si on remplace σ par son

estimation S à partir de l'échantillon (S^2 étant l'estimateur usuel sans biais de la variance σ^2), la propriété

de robustesse exprime que, dans des conditions comparables, la variable aléatoire $\frac{M - \mu}{S/\sqrt{n}}$ est distribuée

à peu près comme un t de Student (nous préciserons plus loin cet énoncé).

Dans les études du faussage lorsque les effectifs sont très faibles, on prend généralement comme indices d'écart à la normalité le « coefficient d'asymétrie » ($\gamma' = \frac{E(X - \mu)^3}{\sigma^3}$) et le « coefficient d'aplatissement »

($\gamma'' = \frac{E(X - \mu)^4}{\sigma^4} - 3$) de la distribution parente. Rappelons à titre indicatif : le coeffi-

cient d'asymétrie est nul pour toutes les distributions symétriques ; il vaut + 2 pour la distribution exponentielle ; le coefficient d'aplatissement est égal à 0 pour la distribution normale, à + 6 pour la distribution exponentielle, à - 1,2 pour la distribution rectangulaire (distribution uniforme sur un segment). D'une façon générale : $\gamma'' > 0$ correspond à une distribution plus pointue et/ou plus étalée aux extrémités que la normale ; $\gamma'' < 0$ à une distribution plus aplatie et/ou moins étalée que la normale (à noter qu'on a toujours $\gamma'' \geq - 2$).

Sous le modèle normal, M est distribué normalement avec $E(M) = \mu$, $Var(M) = \frac{\sigma^2}{n}$ et S^2

est distribué comme $\sigma^2 \frac{\chi^2_{[n-1]}}{n-1}$, d'où on déduit que $\frac{M - \mu}{S/\sqrt{n}}$ est distribué comme $t_{[n-1]}$.

Le facteur prédominant, pour le faussage de la distribution de $\frac{M - \mu}{S/\sqrt{n}}$, est l'écart à la normalité de la distribution de M . Or le coefficient d'asymétrie $\gamma'(M)$ et le coefficient d'aplatissement $\gamma''(M)$ de la distribution de la moyenne s'expriment de la manière suivante en fonction de γ' et γ'' : $\gamma'(M) = \frac{\gamma'}{\sqrt{n}}$,

1. Voir par exemple : H. Rouanet : « Un théorème de convergence probabiliste et son application à la justification de méthodes statistiques courantes », *Math. Sci. hum.*, N° 42, 1973.

$\gamma''(M) = \frac{\gamma''}{n}$ (ces formules découlent des propriétés classiques des moments et cumulants). On voit que comme fonction de n , $\gamma'(M)$ décroît plus lentement que $\gamma''(M)$. Il en résulte que l'influence de la non-normalité décroît lorsque n augmente, le facteur essentiel étant l'asymétrie. Mais par ailleurs on démontre (résultat bien intuitif) que l'effet de l'asymétrie est généralement négligeable pour le test bilatéral, et ne peut être notable que pour les tests unilatéraux. Plus précisément, dans le cas d'une distribution parente symétrique, on démontre (cf. Box et Tiao, 1973, p. 155) que la v.a. $\frac{(M - \mu)^2}{s^2/n}$ est approximativement distribuée comme un \mathcal{F} de Fisher-Snedecor à δ et $\delta(n-1)$ degrés de liberté, où $\delta \simeq 1 + \frac{\gamma''}{n}$.

En conclusion, la distribution de $\frac{M - \mu}{S/\sqrt{n}}$ peut être approchée par celle de t_{n-1} pourvu que γ' et surtout γ'' ne soient pas trop éloignés de 0.

Les résultats précédents se généralisent à la comparaison de deux ou plusieurs moyennes (t de Student et \mathcal{F} de Fisher-Snedecor en *Analyse de variance*), la situation vis-à-vis de l'asymétrie étant encore plus favorable que dans le cas d'une moyenne lorsque (éventualité fréquente) les coefficients d'asymétrie des distributions parentes sont de même signe : prenons par exemple le cas de deux moyennes, on vérifie que le coefficient d'asymétrie de la différence de deux moyennes tend à être plus faible en valeur absolue que le coefficient d'asymétrie de chaque moyenne si du moins les tailles des groupes et les variances ne sont pas trop différentes (lorsque les deux distributions parentes ont même coefficient d'asymétrie et même variance, on vérifie facilement que si les deux groupes ont même taille, la distribution de la différence des moyennes a un coefficient d'asymétrie égal à 0).

8. ROBUSTESSE VIS-A-VIS DE LA NON-NORMALITÉ : COMPARAISON DE VARIANCES

Toutes les méthodes fondées sur la condition de normalité ne sont pas aussi robustes que les méthodes de comparaison de moyennes. A titre de contraste, nous donnerons l'exemple des tests classiques de comparaison de variances par les méthodes suivantes :

- pour un groupe : comparaison d'une variance à une valeur donnée par la méthode du χ^2 ;
- pour deux groupes indépendants : comparaison des variances par la méthode du \mathcal{F} ;
- pour plusieurs groupes indépendants : comparaison des variances par la méthode du test de Bartlett.

Or, dans toutes ces méthodes la condition de normalité est absolument *cruciale*. On peut le voir simplement sur le problème de l'inférence sur une variance. Cette inférence est fondée sur la distribution de S^2 , laquelle est $\sigma^2 \frac{\chi^2[n-1]}{n-1}$ sous le modèle normal. Or nous avons vu que si le coefficient d'aplatissement de la

distribution parente est égal à γ'' , S^2 est approximativement distribué comme $\sigma^2 \frac{\chi^2[v]}{v}$, avec $v = \frac{n-1}{1 + \frac{\gamma''}{2} - \frac{\gamma''}{2n}}$. Si $\gamma'' > 0$, $Var(S^2) > \frac{2\sigma^4}{n-1}$, valeur de $Var(S^2)$ sous le modèle normal et on a un

faussage positif ; si $\gamma'' < 0$, $Var(S^2) < \frac{2\sigma^4}{n-1}$ et on a un faussage négatif. On remarque que si n

augmente, le rapport de la variance réelle $Var(S^2)$ à sa valeur nominale $\frac{2\sigma^4}{n-1}$ tend vers $1 + \frac{\gamma''}{2}$ et non pas vers 1. Plus précisément, lorsque n est grand, la distribution de S^2 peut être approchée par une distribution normale d'espérance σ^2 et de variance $\frac{2\sigma^4}{n} \mathcal{F}^2$, où $\mathcal{F}^2 = 1 + \frac{1}{2} \gamma''$ (le facteur de faus-

sage \mathcal{F} — ou \mathcal{F}^2 — étant égal à 1 lorsque $\gamma'' = 0$). Tester l'hypothèse $H_0 : \sigma^2 = \sigma_0^2$ (où σ_0^2 est

une valeur donnée) quand n est grand, revient à former la statistique de test $X = \frac{S^2 - \sigma_0^2}{\sigma_0^2 \sqrt{\frac{2}{n}}}$.

Sous H_0 , cette statistique est de la forme $X = \mathcal{F} Z$, où Z est une v.a. normale réduite. L'événement critique au seuil nominal α est $(|X| > z_\alpha)$; sa probabilité nominale est α ; sa probabilité réelle α' est $Prob(|X| > z_\alpha) = Prob(\mathcal{F}|Z| > z_\alpha) = Prob(|Z| > \frac{z_\alpha}{\mathcal{F}})$: le seuil réel α' correspondant au seuil nominal α s'obtient donc

en consultant la table habituelle (bilatérale) de la distribution normale réduite pour la valeur $\frac{z_\alpha}{\mathcal{F}}$.

Exemples : pour $\alpha = 0,05$ et $\gamma'' = +2$ on trouve $\alpha' = 0,16$; pour $\gamma'' = -1$ on trouve $\alpha' = 0,006$: les faussages sont considérables.

La différence avec l'inférence sur des moyennes tient au fait suivant : dans les méthodes de comparaison de moyennes, on procède par « studentisation » c'est-à-dire on rapporte la moyenne à une estimation de son écart-type effectuée à partir de l'échantillon, cette estimation ne dépendant pas d'hypothèses sur la distribution parente. Au contraire, dans la méthode précédente d'inférence sur une variance, tout revient à utiliser une estimation de l'écart-type de la variance fournie par le modèle normal. En d'autres termes, la statistique de test privilégiée sous le modèle normal n'est plus privilégiée sous un modèle plus général.

La discussion serait analogue pour le problème de la comparaison de deux variances par la méthode du \mathcal{F} . Les résultats précédents se généralisent à toutes les méthodes classiques de comparaisons de variances fondées sur l'hypothèse d'une distribution parente normale : selon que les coefficients d'aplatissement γ'' des distributions parentes sont positifs ou négatifs, on a un faussage positif ou négatif. Ce faussage, qui ne tend pas à disparaître pour de grands échantillons, est encore plus élevé lorsqu'on compare plusieurs variances (test de Bartlett) que pour une variance ou deux variances. Par exemple, pour 20 groupes, le faussage asymptotique est tel que pour $\alpha = 0,05$, lorsque $\gamma'' = +2$ le seuil réel α' vaut 0,72 et lorsque $\gamma'' = -1$, $\alpha' = 0,000004$.

En conséquence, les statisticiens avertis présentent aujourd'hui avec réserves les méthodes de comparaisons de variances fondées sur la condition de normalité. Plus particulièrement, ils déconseillent de procéder systématiquement à un tel test de comparaison de variances préalablement à un test de comparaison de moyennes lorsque ce dernier test suppose l'homogénéité des variances parentes. En effet, si γ'' est élevé on risque de rejeter à tort l'hypothèse d'homogénéité des variances, donc de se priver d'effectuer une comparaison de moyennes qui aurait été valide (étant donné la robustesse de ce dernier test vis-à-vis de la non-normalité). Inversement, si γ'' est faible on risque d'accepter à tort l'homogénéité des variances et de procéder au test de comparaison de moyennes dans des conditions défectueuses. (Nous laissons ici de côté le problème de la robustesse du test de comparaison de moyennes vis-à-vis de la non-homogénéité des variances, problème qui sera abordé ailleurs : il est clair en effet que même lorsque le test de comparaison de variances est valide, le problème réel n'est pas de détecter si les variances sont hétérogènes ou non, mais d'apprécier si l'hétérogénéité est assez grande pour être susceptible de fausser le test de comparaison de moyennes fondé sur la condition d'égalité des variances.)

Remarques

1. Test \mathcal{F} d'égalité des variances et test \mathcal{F} d'égalité des moyennes en analyse de variance

Nous avons vu que le test \mathcal{F} d'égalité de deux variances est peu robuste vis-à-vis de la non-normalité, et que le faussage ne tend pas à disparaître lorsque le nombre d'observations dans chacun des deux groupes tend vers l'infini ; au contraire le test \mathcal{F} d'égalité des moyennes en analyse de la variance est robuste et ne donne lieu à aucun faussage asymptotique. Cette différence n'est paradoxale qu'en apparence. En effet, en analyse de variance, le carré moyen entre les k groupes est asymptotiquement sous H_0 distribué comme

$\sigma^2 \frac{\chi^2[k-1]}{k-1}$ (où σ^2 est la variance intragroupe) ; cette propriété généralise la propriété de convergence normale

d'une moyenne. Quant au carré moyen intragroupe, il donne bien lieu à un faussage dû à la non-normalité, mais il converge (en probabilité) vers la variance intragroupe σ^2 . En vertu des théorèmes de convergence classiques rappelés au § 2, le rapport des deux carrés moyens à une distribution de la forme $\frac{\chi^2[k-1]}{k-1}$, d'où absence de faussage asymptotique.

2. Test d'égalité des variances ne dépendant pas de la condition de normalité

Le peu de robustesse des tests usuels d'égalité des variances a conduit certains auteurs à rechercher des tests ne dépendant pas de la condition de normalité. Pour comparer les variances de deux ou plusieurs groupes indépendants, Box et Scheffé (1959, p. 83) ont proposé la méthode suivante : diviser (au hasard) chaque groupe en plusieurs sous-groupes, estimer sur chaque sous-groupe la variance du groupe correspondant, d'où pour chaque groupe plusieurs valeurs estimant la variance du groupe ; procéder ensuite (de préférence après une transformation logarithmique) à une analyse de la variance sur ces valeurs. Cette méthode, moins puissante que le test usuel, est pratiquement indépendante de toute condition de normalité.

4. ROBUSTESSE VIS-A-VIS DE LA NON-ÉQUIDISTRIBUTION ET DE LA NON-INDÉPENDANCE

Dans la plupart des méthodes d'inférence usuelles (méthodes fondées sur le modèle normal aussi bien que méthodes non-paramétriques), le modèle de base stipule que les observations constituent un ou plusieurs échantillons, un échantillon étant la réalisation de variables aléatoires équidistribuées indépendantes (modèle « équiproduit »). Or les conditions d'équidistribution et d'indépendance peuvent ne pas être réalistes, en particulier lorsque les observations sont les réalisations successives d'un processus se déroulant dans le temps. Lorsque les observations ne sont pas équidistribuées, nous disons qu'il y a une *tendance* (nous donnons donc à ce terme une acception très large) ; lorsqu'elles ne sont pas indépendantes, nous dirons qu'il y a des *effets séquentiels*. On étudiera la robustesse vis-à-vis de la non-équidistribution et de la non-indépendance en faisant des hypothèses sur le processus : pour présenter ici ces problèmes de robustesse, nous examinerons le problème de la comparaison d'une moyenne à une valeur donnée. Appelons (X_i) $i = 1, 2, \dots, n$ la famille des n variables

aléatoires dont les observations sont les réalisations ; soit $\mu_i = E(X_i)$, $\sigma_i^2 = Var(X_i)$, soit M la statistique $\frac{\sum X_i}{n}$. Sous le modèle de base, les v.a. X_i sont indépendantes et équidistribuées : si μ est l'espé-

rance commune des X_i , σ^2 leur variance commune, M a pour espérance μ et pour variance $\frac{\sigma^2}{n}$,

et $\frac{M - \mu}{\sigma/\sqrt{n}}$ est une v.a. réduite. Si l'hypothèse nulle est $H_0 : \mu = \mu_0$, où μ_0 est une valeur donnée, on

prendra comme statistique de test $\frac{M - \mu_0}{\sigma/\sqrt{n}}$ si la variance σ^2 est donnée, ou $\frac{M - \mu_0}{S/\sqrt{n}}$ si la variance σ^2

est inconnue (S^2 étant l'estimateur usuel sans biais — sous le modèle de base — de σ^2). La distribution de la statistique de test sous le modèle de base et sous H_0 sera connue lorsque de plus on suppose soit que la distribution parente (distribution commune des X_i) est normale, soit que l'échantillon est de taille élevée. Plus précisément, lorsque la variance σ^2 est donnée et la distribution parente normale, la statistique de test

$\frac{M - \mu_0}{\sigma/\sqrt{n}}$ est sous H_0 une v.a. normale réduite ; lorsque l'échantillon est de taille élevée, la statistique de test

$\frac{M - \mu_0}{S/\sqrt{n}}$ est (approximativement) sous H_0 une v.a. normale réduite. Dans ce qui suit, nous envisagerons

les deux cas.

Nous considérerons d'abord l'écart à l'équidistribution (tendance). Pour simplifier nous supposons que les v.a. X_i ont même variance σ^2 et que seules leurs espérances μ_i diffèrent. Si les v.a. X_i sont toujours

indépendantes, M a pour espérance $\mu = \frac{\sum \mu_i}{n}$ et pour variance $\frac{\sigma^2}{n}$ et $\frac{M - \mu}{\sigma/\sqrt{n}}$ est encore une v.a. réduite.

L'hypothèse nulle qui généralise directement l'hypothèse nulle considérée sous le modèle de base est $\mu = \mu_0$ (μ désignant maintenant la moyenne des μ_i , qui peuvent être distinctes). Examinons le test de cette hypothèse nulle au moyen des statistiques usuelles. Lorsque la variance σ^2 est connue, en prenant $\frac{M - \mu_0}{\sigma/\sqrt{n}}$ comme statistique de test, on n'aura pas de faussage. Mais lorsque la variance σ^2 est inconnue, en prenant $\frac{M - \mu_0}{S/\sqrt{n}}$ comme statistique de test on aura un faussage. En effet, S^2 n'est plus un estimateur sans biais

de σ^2 : comme on le vérifie facilement, $E(S^2) = \sigma^2 + \sigma_\mu^2$ en posant $\sigma_\mu^2 = \frac{\sum (\mu_i - \mu)^2}{n-1}$. Comme S^2

estime une valeur supérieure à σ^2 , en divisant $M - \mu_0$ par S/\sqrt{n} on aura une statistique qui tendra à avoir une valeur trop faible, d'où faussage négatif ¹. Pour corriger ce faussage, on se placera dans un contre-modèle permettant une estimation sans biais de σ^2 : par exemple, on supposera que les v.a. X_i ont la même espérance à l'intérieur de blocs successifs (modèle d'analyse de la variance) ou que la tendance des μ_i a une forme déterminée (modèle de régression).

Nous considérerons maintenant l'écart à l'indépendance (effets séquentiels) ², ce qui suit s'applique aux deux cas : variance donnée et variance inconnue.

Supposons que pour un contre-modèle donné les v.a. (X_1, X_2, \dots, X_n) soient toujours équidistribuées mais ne soient plus indépendantes, la non-indépendance se traduisant par le fait que la variance de M n'est plus égale à $\frac{\sigma^2}{n}$ mais à $\mathcal{F}^2 \frac{\sigma^2}{n}$; on appellera \mathcal{F} le « facteur de faussage » et on a donc :

$$\mathcal{F}^2 = \frac{\text{Var}(M \text{ sous le contre-modèle})}{\text{Var}(M \text{ sous le modèle de base})}$$

(On remarque d'emblée que si les observations sont « corrélées positivement » — nous donnerons des exemples précis plus loin — $\text{Var}(M) > \frac{\sigma^2}{n}$, c'est-à-dire $\mathcal{F}^2 > 1$.)

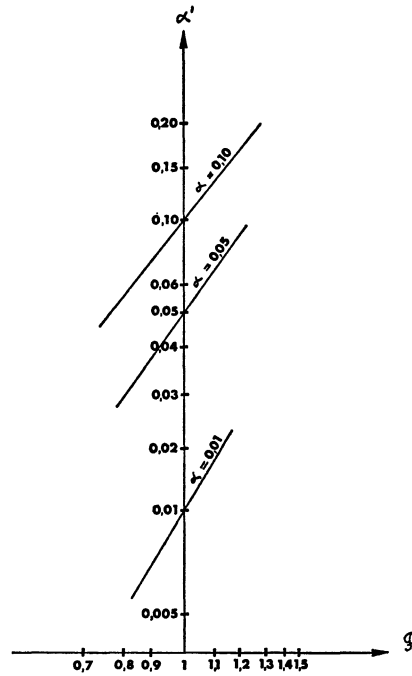
Sous le contre-modèle, M a pour espérance μ et pour variance $\mathcal{F}^2 \frac{\sigma^2}{n}$; $\frac{M - \mu}{\sigma/\sqrt{n}}$ a maintenant pour variance \mathcal{F}^2 ; c'est $\frac{M - \mu}{\mathcal{F} \sigma/\sqrt{n}}$ qui est une v.a. réduite.

On en déduit le faussage du seuil : si la valeur donnée est μ_0 , sous H_0 , la v.a. $X = \frac{M - \mu_0}{\sigma/\sqrt{n}}$ est une v.a. d'espérance 0 et de variance \mathcal{F}^2 ; donc on peut poser $X = \mathcal{F}Z$, où Z est une v.a. normale réduite. Soit α le seuil nominal ; il lui correspond une valeur z_α telle que $\text{Prob}(|Z| \geq z_\alpha) = \alpha$ (z_α est la valeur lue dans la table usuelle de la distribution normale réduite au seuil bilatéral α). La région

1. On aurait un test non faussé en considérant l'hypothèse nulle restrictive : $\forall i, \mu_i = \mu_0$; mais le rejet d'une telle hypothèse nulle sera peu informatif (il existe des μ_i différentes de μ_0 .)

2. On trouvera de nombreux développements sur certains aspects de la robustesse vis-à-vis de la non-indépendance des observations dans le travail de G. Oppenheim : *Faussage et robustesse de tests de comparaisons de moyennes et de fréquences lorsque les observations ne sont pas indépendantes* (Thèse de 3^e cycle, 1972).

critique (bilatérale) au seuil α est ($|X| \geq z_\alpha$), donc le seuil réel est $\alpha' = \text{Prob}_{\mu_0} (|X| \geq z_\alpha) = \text{Prob} (|\mathcal{F}_Z| \geq z_\alpha) = \text{Prob} (|Z| \geq \frac{z_\alpha}{\mathcal{F}})$. Si $\mathcal{F} > 1$, $\frac{z_\alpha}{\mathcal{F}} < z_\alpha$ donc $\alpha' > \alpha$ (faussage positif); si $\mathcal{F} < 1$, $\frac{z_\alpha}{\mathcal{F}} > z_\alpha$ donc $\alpha' < \alpha$ (faussage négatif). L'amplitude du faussage n'est bornée que par les valeurs 0 et 1 : Si $\mathcal{F} \rightarrow \infty$, $\alpha' \rightarrow 1$; si $\mathcal{F} \rightarrow 0$, $\alpha' \rightarrow 0$. On a représenté ci-dessous, en coordonnées bilogarithmiques, pour trois valeurs du seuil nominal α (0,10; 0,05 et 0,01) la variation de α' en fonction du facteur de faussage.



Seuil réel α' en fonction du facteur de faussage \mathcal{F} pour les seuils nominaux $\alpha = 0,10$, $\alpha = 0,05$ et $\alpha = 0,01$.

On étudiera de même le faussage de la puissance. Caractérisons l'écart à l'hypothèse nulle par $\delta = \frac{\mu - \mu_0}{\sigma\sqrt{n}}$; par une méthode analogue à la précédente, on trouve que la puissance au seuil α , définie par

$\text{Prob} \left(\frac{|M - \mu_0|}{\sigma\sqrt{n}} \geq z_\alpha \right)$ est égale à $1 - \Phi \left(\frac{z_\alpha - \delta}{\mathcal{F}} \right) + \Phi \left(\frac{-z_\alpha - \delta}{\mathcal{F}} \right)$, Φ étant la fonction de répartition

de la distribution normale réduite (pour $\delta = 0$, on retrouve naturellement α'). Cette formule permettra de construire des courbes de faussage de la puissance.

Si l'on précise le contre-modèle sous forme de processus, on pourra écrire le facteur de faussage \mathcal{F} comme fonction des paramètres du processus. Par exemple, supposons le processus *markovien*, avec un coefficient d'autocorrélation (coefficient de corrélation entre deux v.a. consécutives) égal à ρ ; le coefficient de corrélation entre deux v.a. X_i et X_j est alors égal à $\rho_{ij} = \rho^{|i-j|}$. On vérifie qu'on a alors (pour n grand)

$$\mathcal{F}^2 \simeq \frac{1 + \rho}{1 - \rho}$$

Un autre processus, qu'on peut considérer comme une approximation du processus markovien, consiste à supposer que $\rho_{ij} = \rho$ pour $|i - j| = 1$ et $\rho_{ij} = 0$ pour $|i - j| > 1$. On vérifie qu'on a alors (pour n grand) $\mathcal{F}^2 \simeq 1 + 2\rho$.

(On vérifie que pour les deux modèles $\mathcal{F}^2 > 1$ (faussage positif) si $\rho > 0$, $\mathcal{F}^2 < 1$ (faussage négatif) si $\rho < 0$.)

Le tableau suivant indique pour le modèle markovien (les résultats sont voisins pour l'autre modèle) les valeurs de ρ entraînant un faussage correspondant à la moitié et au double du seuil α , pour $\alpha = 0,10$; $0,05$ et $0,01$.

ρ	\mathcal{F}	α	α'
- 0,18 0 + 0,26	0,84 1 1,30	0,10 0,10 0,10	0,05 0,10 0,20
- 0,12 0 + 0,17	0,88 1 1,19	0,05 0,05 0,05	0,025 0,05 0,10
- 0,082 0 + 0,098	0,92 1 1,10	0,01 0,01 0,01	0,005 0,01 0,02

Sur ce tableau on voit tout d'abord que *le faussage est considérable même pour de faibles valeurs de $|\rho|$* : pour $|\rho|$ de l'ordre de $0,15$, le seuil varie du simple au double pour $\alpha = 0,05$. Toujours pour $\alpha = 0,05$ on a pour $\rho = - 0,3$, $\alpha' = 0,008$ et pour $\rho = + 0,3$, $\alpha' = 0,15$. On remarque ensuite que le faussage relatif apparaît un peu plus marqué lorsque le seuil est plus faible.

Rappelons que ces considérations s'appliquent au test asymptotique de comparaison d'une moyenne à une valeur donnée : le faussage dû à la non-indépendance ne tend donc pas à disparaître pour de grands échantillons.

Un cas particulier du test précédent est le test asymptotique classique de *comparaison d'une fréquence à une valeur donnée*. Soit F la statistique-fréquence, Φ la fréquence parente, Φ_0 la valeur donnée. La sta-

tistique de test est $\frac{F - \Phi_0}{\sqrt{\frac{\Phi_0 (1 - \Phi_0)}{n}}}$, qui a asymptotiquement une distribution normale réduite. Lorsque

les observations ne sont plus indépendantes, c'est $\frac{F - \Phi_0}{\mathcal{F} \sqrt{\frac{\Phi_0 (1 - \Phi_0)}{n}}}$ qui est asymptotiquement nor-

male réduite. Montrons sur un exemple comment en supposant le processus markovien on peut alors procéder à un test corrigé.

Supposons qu'on ait observé une suite de 100 observations avec 60 observations égales à 1 et 40 observations égales à 0, et qu'on veuille tester $H_0 : \Phi = \frac{1}{2}$. Selon le test classique la statistique

$$\frac{F - \Phi_0}{\sqrt{\frac{\Phi_0 (1 - \Phi_0)}{n}}} \text{ a une valeur égale à } \frac{\frac{60}{100} - \frac{1}{2}}{\sqrt{\frac{1}{2} \times \frac{1}{2} \times \frac{1}{100}}} = 2,00 : \text{ la différence est significative au seuil}$$

de 0,05. Prenons comme contre-modèle un modèle markovien : on est conduit à constituer le « tableau des effectifs de transition », qu'on supposerait être le suivant :

$n-1 \nearrow$	1	0	
1	43	17	60
0	17	23	40

(43 est le nombre de transitions de 1 vers 1, c'est-à-dire le nombre de fois qu'on a observé 1 immédiatement après 1, etc.)

L'estimation du coefficient de corrélation Φ du processus peut être obtenue à partir du tableau; c'est le « coefficient de contingence » r_Φ classique associé au tableau ; on trouve ici $r_\Phi = +0,292$. Ce coefficient est positif, ce qui exprime que la fréquence des transitions de 1 vers 1 ($\frac{43}{60}$) est supérieure à la

fréquence des transitions de 0 vers 1 ($\frac{17}{40}$), ou, de manière équivalente, que la fréquence des transitions

de 0 vers 0 ($\frac{23}{40}$) est supérieure à la fréquence des transitions de 1 vers 0 ($\frac{17}{60}$) : nous dirons que les

effets séquentiels sont positifs, ou encore que la contagion entre les essais successifs est positive (si r_Φ était négatif, nous parlerions de même d'effets séquentiels négatifs). On peut dès lors conclure que le faussage du test est positif, donc que le seuil observé réel est supérieur à 0,05.

On peut en fait aller plus loin en estimant, à partir de la valeur estimée de ρ , le facteur de faussage :

\mathcal{F}^2 est en effet estimé par $\frac{1 + r_\Phi}{1 - r_\Phi} = \frac{1 + 0,292}{1 - 0,292} = 1,83$ d'où l'estimation de \mathcal{F} : 1,4. La statistique

de test corrigée est alors $\frac{2,00}{1,4} = 1,4$, ce qui correspond à un seuil réel observé $P \simeq 0,15$ (au lieu de 0,05) :

on voit que le faussage positif est considérable.

N.B. Dans le cas d'une variable binaire (ou plus généralement discrète), l'hypothèse d'indépendance des observations successives peut elle-même être testée à partir du tableau des transitions : le test est identique au test Khi-deux que l'on effectuerait pour tester l'indépendance entre deux variables si le tableau de transition était un « tableau de contingence ». Dans le cas présent, la valeur de Khi-deux (égale à $nr^2\Phi$) est 8,5 significative au seuil de 0,01. Anderson et Goodman ont montré que ce test est approprié pour tester l'hypothèse nulle d'indépendance (chaîne de Markov d'ordre 0) sous le modèle plus général d'une chaîne de Markov d'ordre 1.

Les considérations sur la robustesse vis-à-vis de la non-indépendance s'étendent au cas de plusieurs groupes d'observations. Par exemple, pour comparer les moyennes de deux groupes indépendants, pour lesquels les facteurs de faussage sont respectivement \mathcal{F}_1 et \mathcal{F}_2 , on montre que le facteur de faussage \mathcal{F}

relatif à la comparaison des deux moyennes est donné par $\mathcal{F}^2 = \frac{\mathcal{F}_1^2 \frac{\sigma_1'}{n_1} + \mathcal{F}_2^2 \frac{\sigma_2'}{n_2}}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ (on remarque que

dans ce cas, on peut avoir annulation du faussage pour des tailles d'échantillons convenables, lorsque les faussages \mathcal{F}_1 et \mathcal{F}_2 sont de sens contraire, éventualité d'ailleurs assez académique).

5. AUTRES PROBLÈMES DE ROBUSTESSE LIÉS A DES MÉTHODES D'INFÉRENCE USUELLES

L'utilisation des méthodes classiques d'inférence pour l'analyse des protocoles usuels non élémentaires soulève de nouveaux problèmes de robustesse. En particulier, dans les problèmes de comparaison de moyennes et plus généralement en analyse de la variance, les méthodes classiques sont fondées sur des conditions d'homogénéité de variances. Dans un prochain article, nous nous proposons de présenter en détail le problème de la robustesse de la comparaison de moyennes vis-à-vis de la non-homogénéité des variances dans le cas de deux groupes indépendants ¹. D'autres problèmes, qu'on peut rattacher à des problèmes de robustesse, sont liés au choix d'un modèle et d'une méthode d'analyse appropriée : nous envisagerons ces problèmes dans un article ultérieur.

6. ROBUSTESSE VIS-A-VIS DES VALEURS ABERRANTES

Un type de robustesse qui a été beaucoup étudié récemment est la robustesse vis-à-vis des valeurs aberrantes (« outliers »). On peut poser le problème des valeurs aberrantes de la façon suivante : on considère que la majeure partie des observations constitue un échantillon d'une distribution parente (que nous appellerons distribution de base), mais que certaines observations (les valeurs aberrantes) peuvent ne pas appartenir à cette distribution, en fournissant des valeurs anormalement basses ou élevées (exemple : erreurs grossières de mesure, etc.). L'inférence porte sur les paramètres de la distribution de base : on recherchera des méthodes d'inférence robustes (c'est-à-dire peu sensibles au « parasitage » occasionné par les valeurs aberrantes) et en même temps aussi efficaces que possible sous le modèle de base.

En fait, dans les travaux actuels, on s'est généralement placé dans le cadre d'un modèle de base normal et on a surtout proposé des méthodes d'inférence robustes pour la moyenne de cette distribution de base. Parallèlement on a proposé des méthodes pour détecter les valeurs aberrantes. Donnons à titre d'exemple le principe de deux de ces méthodes : sous le modèle normal, on peut tabuler la distribution de l'étendue w de l'échantillon en fonction de l'écart-type σ de la distribution de base : si on connaît σ , une valeur significativement trop grande de w est une indication que l'une ou l'autre des valeurs extrêmes de l'échantillon sont des valeurs aberrantes. Une autre méthode (qui n'exige pas la connaissance de σ) consiste à considérer une valeur extrême et à examiner le rapport de la distance de cette valeur extrême à la valeur adjacente à l'étendue de l'échantillon : la distribution, sous le modèle normal, de ce rapport a été également tabulée : si ce rapport est trop élevé on conclut que la valeur extrême est aberrante, etc. Si l'on a une idée de la probabilité *a priori* des valeurs aberrantes, on peut, en appliquant des règles approchées qui ont été établies, en déduire le seuil auquel on doit travailler pour la détection des valeurs aberrantes ². Ces méthodes d'élimination de valeurs aberrantes ne doivent pas être appliquées de façon trop aveugle (dans la mesure en particulier où elles reposent sur l'hypothèse d'une distribution de base normale) ; mais elles marquent un progrès décisif par rapport aux pratiques courantes mais inadmissibles consistant à éliminer des observations d'une manière purement empirique *a posteriori*, sans disposer d'un modèle spécifiant le processus d'obtention des observations.

Une fois éliminées les valeurs présumées aberrantes, on procédera à l'estimation de la moyenne μ de la distribution de base à partir de la moyenne calculée sur les valeurs restantes, ce qui revient à calculer une « moyenne tronquée » à partir de l'échantillon initial. On peut également, au lieu de tester les valeurs extrêmes pour élimination, envisager une procédure de « troncature *a priori* ». On montre que les moyennes tronquées constituent sous le modèle de base des estimateurs relativement très efficaces de la moyenne μ par rapport à la moyenne habituelle (non tronquée), optimale sous le modèle de base mais peu robuste vis-à-vis des valeurs aberrantes. L'efficacité diminue naturellement à mesure que l'on augmente l'ordre de troncature (on dira qu'une moyenne est tronquée d'ordre k si on a éliminé les k valeurs les plus élevées et les k valeurs

les plus faibles) ; si $k = \frac{n-1}{2}$ avec n impair, on retrouve la médiane, estimateur très robuste mais peu

1. H. Rouanet et D. Lépine : « Etude d'un conflit robustesse-efficacité dans le problème de la comparaison de deux moyennes (groupes indépendants) » (à paraître).

2. On trouvera les tables des distributions dans la nouvelle édition de l'ouvrage de W. Dixon et F. Massey : *Introduction to statistical analysis*, New York, McGraw-Hill, 1969.

efficace de μ (rappelons que sous le modèle normal l'efficacité relative de la médiane par rapport à la moyenne, évaluée à partir du rapport des variances d'échantillonnage des deux statistiques, est de l'ordre de 64 %).

A côté des moyennes tronquées, on a proposé des moyennes « winsorisées » : une moyenne winsorisée d'ordre 1 est obtenue en remplaçant chaque valeur extrême par la valeur adjacente, etc. L'efficacité de la winsorisation est légèrement supérieure à celle de la troncature.

On a développé des méthodes de test valables pour les données tronquées ou winsorisées. Par exemple, si on calcule un rapport T_w de Student à partir d'une winsorisation d'ordre k , on démontre que

$T^* = \frac{n - 2k - 1}{n - 1} T_w$ est approximativement sous H_0 , distribué comme t avec $n - 2k - 1$ d.l.

On ne doit pas confondre le problème de la robustesse vis-à-vis des valeurs aberrantes et celui de la robustesse vis-à-vis de la non-normalité : dans ce dernier cas, on suppose que les observations constituent un échantillon d'une distribution non-normale dont on cherche à estimer les paramètres (et non pas un échantillon d'une distribution normale « parasitée » dont on cherche à estimer les paramètres). Cependant, certaines méthodes d'estimation sont communes aux deux problèmes. Par exemple, on sait que pour des distributions très étalées aux extrémités (disons de variance infinie) on prendra comme paramètre de centralité non plus la moyenne, mais la médiane, et la médiane des observations sera l'estimateur de la médiane parente ; dans le cas du modèle normal parasité, la médiane est un estimateur robuste du paramètre de centralité (moyenne ou médiane) de la distribution normale de base. De même, pour des distributions parentes assez éloignées de la normale mais pas trop extrêmes, les moyennes tronquées ou winsorisées fournissent des estimateurs raisonnables des paramètres de centralité. Dans le même ordre d'idées, signalons l'estimateur d'Hodges et Lehman qui consiste à prendre la médiane des moyennes des paires des observations.

BIBLIOGRAPHIE

- [1] SCHEFFÉ H., *The analysis of variance*, New York, Wiley, 1959.
- [2] BOX G.E. and TIAO G.C., *Bayesian inference in statistical analysis*, Londres/Reading, Mass., Addison Wesley, 1973.