

REVUE DE STATISTIQUE APPLIQUÉE

JEAN-BAPTISTE DENIS

JEAN VAILLANT

**Analyse statistique de données de piégeage d'insectes pour
mettre en évidence des comportements agrégatifs**

Revue de statistique appliquée, tome 33, n° 3 (1985), p. 29-42

http://www.numdam.org/item?id=RSA_1985__33_3_29_0

© Société française de statistique, 1985, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (<http://www.sfds.asso.fr/publicat/rsa.htm>) 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/>

ANALYSE STATISTIQUE DE DONNÉES DE PIÉGEAGE D'INSECTES POUR METTRE EN ÉVIDENCE DES COMPORTEMENTS AGRÉGATIFS

Jean-Baptiste DENIS et Jean VAILLANT

Laboratoire de Biométrie
INRA, Route de Saint-Cyr
F78000 Versailles

RÉSUMÉ

Pour l'analyse statistique de données de piégeage d'insectes, on utilise le plus souvent des tests de χ^2 qui reposent sur l'hypothèse implicite que les insectes sont piégés indépendamment les uns des autres. Pour ne pas s'astreindre à cette hypothèse, un modèle de comportement agrégatif est proposé. Il est ensuite approché pour le traitement statistique par un modèle plus simple basé sur des distributions de Polya. Un test de l'agrégativité est construit pour des probabilités quelconques d'attraction relative des pièges, un intervalle de confiance du paramètre d'agrégativité est donné si on suppose ces mêmes probabilités connues. Deux exemples illustrent la démarche.

SUMMARY

χ^2 tests are often used to analyse insect trapping data, based on the assumption that insects are caught independently of each other. An aggregative behaviour model is proposed in order to escape from this constraint. For statistical analysis, this model is approached using a simpler model based on Polya distribution.

An aggregativity test is proposed for different probabilities of trap attraction and confidence intervals for the aggregativity parameter are calculated for given probabilities. Two examples are illustrated.

PLAN

I. Introduction

II. Modèles probabilistes

- II.1. Modèle Poissonien
- II.2. Modèle d'urne de Polya
- II.3. Relations entre les 2 modèles agrégatifs
- II.4. Cas de plusieurs pièges
- II.5. Cas de plusieurs répétitions

III. Aspects statistiques

- III.1. Introduction
- III.2. Cas de 2 pièges identiques (ou connus)
- III.3. Cas de plusieurs pièges connus avec répétitions
- III.4. Cas de plusieurs pièges inconnus avec répétitions
 - III.4.a. estimation des p_i
 - III.4.b. test de l'agrégativité ($s=0$)
 - III.4.c. estimation de s

IV. Quelques exemples

Remerciements

Références Bibliographiques

I. INTRODUCTION

Lorsqu'on étudie des effectifs de capture d'insectes par piégeage dans le but de mettre en évidence des variations dans le niveau de population ou encore dans leur évolution par rapport à une variable quantitative (le temps, la distance à un massif forestier...), on se sert généralement de χ^2 pour tester l'hypothèse de variation nulle. Ces tests, qui sont asymptotiques, reposent sur une modélisation probabiliste peu contraignante mais qui suppose en particulier que les insectes soient capturés indépendamment les uns des autres. Prenons le cas le plus simple : celui de la comparaison de 2 pièges. Soient n_1 et n_2 les nombres d'insectes capturés dans l'un et l'autre des 2 pièges. La distribution implicite de n_1 conditionnellement au nombre total d'insectes capturés (n_+) est celle d'une loi binomiale $B(n_+, p_1)$. Le test de χ^2 est alors un test asymptotique. Ce modèle binomial repose sur 2 suppositions :

- Chacun des n_+ insectes capturés a la même probabilité p_i d'être capturé dans le piège i ; $i = 1, 2$; $p_1 + p_2 = 1$ (b1)
- Les insectes sont capturés indépendamment les uns des autres (b2)

Si on peut considérer que les conditions sont identiques sur toute la durée du piégeage, (b1) est remplie. Pour que (b2) le soit aussi, il ne faut pas qu'il y ait de comportements agrégatifs ou répulsifs. L'effet de l'agrégativité sur le test de χ^2 mentionné ci-dessus est de conduire à déclarer trop fréquemment les 2 pièges différents alors qu'ils ne le sont pas : le niveau réel du test est plus faible que celui qu'on croit pratiquer (Chambon et al., 1984). Pour s'en convaincre intuitivement, il suffit de supposer que les insectes se déplacent par groupes de k individus et que ce sont ces groupes de k individus qui se comportent indépendamment les uns des autres : il faudrait alors diviser par k , la statistique classique pour qu'elle suive un χ^2 , car il faut dénombrer les groupes et non les insectes.

L'objet de cette étude est de proposer une approche probabiliste réaliste de comportements agrégatifs puis d'en déduire des procédures statistiques. Le développement principal sera basé sur le cas de 2 pièges, il permet de bien mettre en évidence toutes les idées et les difficultés que comporte le problème. On envisagera ensuite le cas d'un nombre quelconque de pièges. Enfin quelques données seront analysées pour illustrer la démarche. Pour préciser les notions générales d'échantillonnage sur lesquelles repose cette approche, le lecteur est renvoyé à Frontier (1983).

NOTATIONS : lorsqu'un indice est remplacé par le signe +, cela signifie que l'on a fait la somme sur cet indice ($n_+ = n_1 + n_2$),

$B(n, p)$: loi binomiale de paramètre p et de taille n ,

$P(x)$: loi de Poisson de paramètre x ,

$Po(n, p, s)$: loi de Polya de paramètres p et s , de taille n ,

χ^2_d : loi de khi deux à d degrés de liberté,

$N(m, s)$: loi normale de moyenne m et de variance s ,

$n_1/(n_+)$: variable aléatoire n_1 conditionnée par la variable aléatoire n_+ ,

$E(n_1/n_+ = t)$: espérance conditionnelle de n_1 pour $n_+ = t$, si le contexte le permet sera notée simplement $E(n_1)$,

$\binom{n}{p}$: nombre de combinaisons différentes de p éléments pris parmi n ,

$i(j)k$: suite d'entiers de i à k avec un pas de progression de j .

II. MODÈLES PROBABILISTES

II.1. Modèle Poissonnien

Le modèle binomial évoqué dans l'introduction peut se déduire d'un modèle plus général basé sur des distributions de POISSON (FELLER, 1968). Si n_1 et n_2 suivent des lois de POISSON de paramètres λ_1 et λ_2 indépendantes $n_1/(n_+)$ suit une binomiale $B(n_+, \lambda_1/\lambda_+)$. Reprenant cette même perspective, nous avons bâti un modèle dit Poissonnien. Nous supposons que les insectes se déplacent en groupes et qu'un nombre variable d'insectes de chaque groupe représenté est capturé. Soit N_1 et N_2 , les nombres respectifs de groupes représentés dans les pièges 1 et 2; soit F_i le nombre d'insectes capturés dans le groupe i , que celui-ci ait été représenté dans le piège 1 ou le piège 2, alors notre modèle postule :

$$N_1 \sim P(\lambda_1), N_2 \sim P(\lambda_2) \text{ et elles sont indépendantes} \quad (c1)$$

$$F_i = 1 + G_i \text{ où les } G_i \sim P(\mu), \text{ sont indépendantes entre elles et indépendantes de } N_1 \text{ et } N_2 \quad (c2)$$

Ce modèle est une généralisation du modèle binomial, puisqu'il s'y ramène si $\mu = 0$. De fait $\mu + 1$ s'interprète comme le paramètre d'agrégativité relative au type de piège. Remarquons que l'on suppose implicitement que le taux de représentation d'un groupe d'insectes ne dépend pas du piège, $p = \lambda_1/\lambda_+$ s'interprète comme l'attractivité relative du piège 1; enfin λ_+ traduit le potentiel de capture des 2 pièges. Ce dernier paramètre n'intéresse en général pas car il dépend trop des conditions de piégeage, en particulier de sa durée, du climat... qui ne sont pas suffisamment contrôlées. Dans ce modèle, les nombres d'insectes capturés dans les 2 pièges, n_1 et n_2 suivent des lois de POISSON généralisées indépendantes (ROGERS, 1974). Par exemple

$$n_1 = N_1 + \sum_{i=1}^{N_1} G_i$$

et a pour espérance et variance :

$$E(n_1) = (1 + \mu) \lambda_1 \quad V(n_1) = (1 + 3\mu + \mu^2) \lambda_1 \quad (1)$$

Reprenant l'approche binomiale et considérant que λ_+ , qui caractérise le niveau global des captures, n'est pas interprétable, il est logique de s'intéresser à la distribution de n_1 conditionnelle à n_+ .

En effet N_+ est inconnu et n_+ caractérise le niveau global des captures. On arrive alors aux résultats suivants en calculant $P(n_1 = k/n_+ = t, N_+ = \ell)$ pour $1 \leq \ell \leq t$ et en déconditionnant par rapport à N_+ .

$$P(n_1 = k/n_+ = t) = \binom{t}{k} F(k, \lambda_1, \mu) \cdot F(t-k, \lambda_2, \mu) / F(t, \lambda_+, \mu)$$

$$\text{pour } 0 \leq k \leq t \quad (2)$$

$$= 0 \text{ pour les autres valeurs de } k$$

$$\text{où } F(n, \lambda, \mu) = \sum_{i=0}^n \binom{n}{i} (\lambda e^{-\mu})^i \cdot (i\mu)^{n-i}$$

$$E(n_1/n_+ = t) = p \cdot t; \text{ ce résultat ne dépend pas de } (c2) \quad (3)$$

$$\begin{aligned} V(n_1/n_+ = t) &= p(1-p)t + \\ &= p(1-p)t + p(1-p) \sum_{\ell=1}^t [(t-\ell)(t+\ell-1)/\ell] P(N_+ = \ell/n_+ = t) \\ &= p(1-p)t + p(1-p)[1 - E(N_+/n_+ = t) + t(1-t) \cdot E(N_+^{-1}/n_+ = t)] \quad (4) \end{aligned}$$

avec

$$P(N_+ = \ell/n_+ = t) = \frac{H(\ell, \mu, t, \lambda_+)}{\sum_{j=1}^t H(j, \mu, t, \lambda_+)} \quad \text{pour } 0 \leq \ell \leq t$$

et où

$$H(\ell, \mu, t, \lambda) = \frac{e^{-\mu\ell} (\ell\mu)^{t-\ell} \cdot e^{-\lambda_+} \cdot \lambda_+^\ell}{(t-\ell)! \ell!}$$

Si $\mu = 0$, on retrouve bien les formules du cas binomial. Si $\mu > 0$, alors l'espérance est inchangée mais la variance est augmentée du second terme de l'expression. On retrouve le résultat fondamental : l'agrégativité n'a pas d'effet sur l'espérance mais gonfle la variance (d'où l'utilisation classique du rapport variance/espérance). On constate malheureusement que ni au niveau de la distribution, ni au niveau de la variance, le conditionnement par n_+ n'a fait disparaître le paramètre d'échelle λ_+ !

II.2. Modèle d'urne de POLYA

Cette distribution est extensivement développée dans JOHNSON et KOTZ (1969, 1977), rappelons en brièvement le fondement pour fixer nos notations. On dispose d'une urne contenant 2 catégories de boules (qui correspondront à nos 2 pièges) et on effectue une série de n_+ tirages indépendants successifs (qui correspondront aux n_+ captures de nos insectes). Dans l'urne, il y a m_1 (m_2) boules de la 1^{ère} (2^{de}) catégorie. On rajoute après chaque tirage k boules de la catégorie tirée, modifiant ainsi la probabilité du tirage suivant.

Cette procédure conduit à augmenter la probabilité de tirage de la catégorie qui est tirée, d'où un effet agrégatif. Si on pose $p_1 = m_1/(m_1 + m_2)^{-1}$ et $s = k \cdot (m_1 + m_2)^{-1}$ alors n_1 suit une Po (n_+, p_1, s); $0 \leq p_1 \leq 1$ et $0 \leq s \cdot p_1$ garde la même interprétation que précédemment et s caractérise à n_+ fixé, l'agrégativité, s'il est nul on retrouve le modèle binomial.

On établit les relations suivantes :

$$P(n_1 = k/n_+ = t) = \binom{t}{k} \Psi(k, p_1, s) / \Psi(t-k, 1-p_1, s) / \Psi(t, 1, s)$$

pour $0 \leq k \leq t$

(5)

$$\text{où} \quad \Psi(m, q, s) = \prod_{i=1}^m [q + (i-1)s]$$

$$E(n_1) = p_1 \cdot t \quad (6)$$

$$V(n_1) = p_1(1-p_1)t + p_1(1-p_1)t(t-1) \frac{s}{1+s} \quad (7)$$

II.3. Relations entre les 2 modèles agrégatifs

Les formules des 2 premiers moments des 2 modèles présentent des analogies certaines. En particulier, les espérances sont identiques, ce qui confirme la même interprétation pour le paramètre p_1 . Les variances se mettent toutes les 2 sous forme de « variance du modèle binomial + ajout dû à l'agrégativité ».

Contrairement au modèle dit de POLYA, le modèle Poissonnien est fondé sur une traduction directe d'un mécanisme explicatif de la réalité. Malheureusement les formules de vraisemblances sont complexes et il comprend un paramètre supplémentaire peu intéressant : λ_+ . Les 2 modèles ne sont pas emboîtés dans le sens où aucun des 2 n'est un cas particulier de l'autre. Cependant, on peut préjuger que le modèle Poissonnien est le plus complexe puisqu'il dépend de 4 paramètres (n_+ , p , μ , λ_+) alors que le modèle de POLYA ne dépend que de 3 (n_+ , p , s). Aussi avons-nous procédé de la manière suivante : à partir d'un certain nombre de combinaisons des valeurs des paramètres du modèle Poissonnien, nous avons calculé les distributions correspondantes, leurs espérances et leurs variances, utilisant ces valeurs nous en avons déduit la valeur de s telle que la distribution de POLYA ait même variance. Enfin nous avons calculé la distance de Hellinger; cette distance est comprise entre 0 et 1; contrairement à la « distance de χ^2 », elle est symétrique. Si r et q sont des probabilités discrètes de termes génériques r_i et q_i , alors :

$$d^2(r, q) = 1 - \sum_i [r_i \cdot q_i]^{1/2}$$

Elle sera utilisée comme indice pour observer dans quels domaines de l'espace des paramètres, les 2 distributions de probabilité sont proches. A titre d'illustration les figures 1 et 2 présentent un cas où les distributions sont différentes et un autre où elles sont proches.

Dans le premier cas la distribution du modèle Poissonnien est plurimodale, de plus elle comporte des poids importants aux extrémités, sans rapport avec l'allure générale. Il est facile de saisir intuitivement la situation : il s'agit de la distribution de

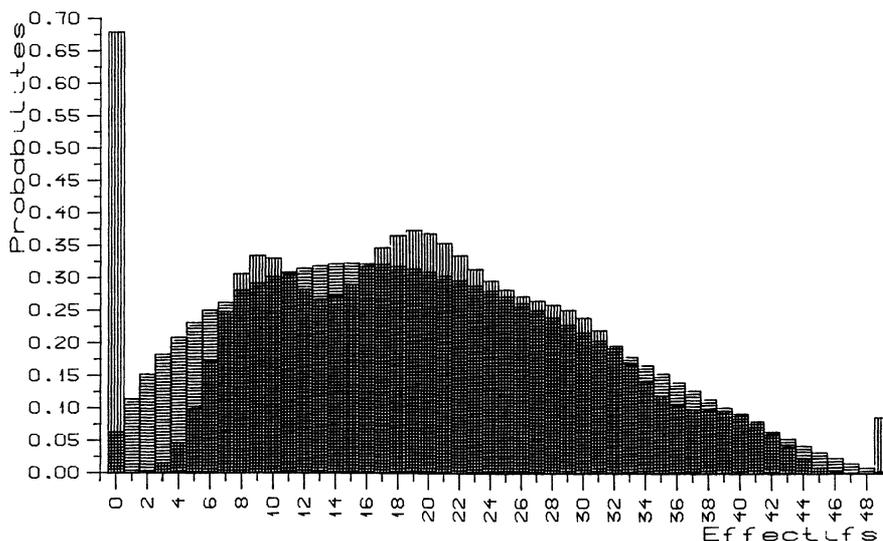


Figure 1. — Première comparaison du modèle Poissonnien (hachurage vertical) et du modèle de Polya (hachurage horizontal)

$$(n_+ = 49; p_1 = 0,4; \mu = 8; \lambda_+ = 4,36 \text{ et } s = 0,23)$$

Les classes des histogrammes sont d'amplitude 1 la distance de Hellinger entre les 2 distributions est de 0,0418.

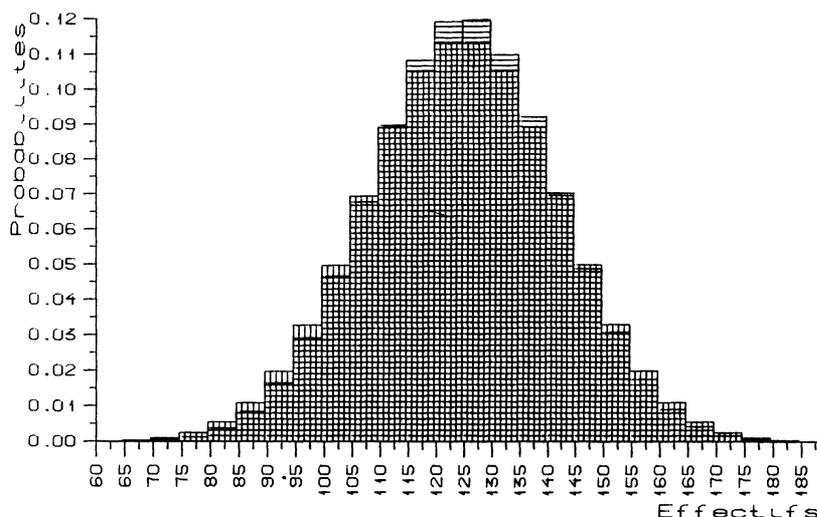


Figure 2. — Seconde comparaison du modèle Poissonnien (hachurage vertical) et du modèle de Polya (hachurage horizontal)

($n_+ = 249$; $p_1 = 0,5$; $\mu = 3$; $\lambda_+ = 62,25$ et $s = 0,0132$)

Les classes des histogrammes sont d'amplitude 5 la distance de Hellinger entre les 2 distributions est de 0,0008.

n_1 conditionnellement à n_+ , or les insectes sont supposés arriver par groupes de taille moyenne $\mu + 1 = 9$. Le nombre de groupes représentés tournera donc autour de $49/9$ soit environ 6, qui est le nombre de modes (pour les valeurs de $n_1 = 0, 9, 19, 28, 40$ et 49). Etant donné que la distribution de POLYA est unimodale, elle ne peut suivre le mouvement et s'écarte donc notablement. Dans le second cas où $\mu = 3$ est petit devant $n_+ = 249$, l'effet d'arrivée en groupe est dilué et les distributions sont très proches. Le tableau 1 confirme cette interprétation.

TABLEAU 1

Distances de Hellinger ($\times 10^6$) en fonction de μ et de n_+ pour $p = 0,5$ et $\lambda_+ = n_+ / (\mu + 1)$. Cette dernière condition est prise pour que n_+ soit égal à son espérance dans le modèle Poissonnien

$\mu =$	3	5	8	10
29	39	284	1 160	1 188
$n_+ = 59$	3	12	88	250
89	1	4	14	35

Le Tableau 2 expose quelques résultats où on constate d'autre part que plus p_1 est proche de 0,5, meilleure est la proximité entre les 2 distributions. La valeur « 29 » en bas à droite fait exception; ceci provient du fait que $n_+ = 249$ est très improbable pour les valeurs considérées des paramètres : c'est l'espérance moins 3,55 fois l'écart-type ! En outre, plus n_+ est grand par rapport à son espérance (dans des limites réalistes) meilleure est la proximité.

TABLEAU 2

Distances de Hellinger ($\times 10^4$) en fonction de p , de n_+ et de λ_+ fixé par la relation $\alpha = E(n_+)/n_+$. Pour $\mu = 8$.

$\alpha =$	0,5	0,5	0,5	1	1	1	2	2	2
$p_1 =$	0,2	0,4	0,5	0,2	0,4	0,5	0,2	0,4	0,5
$n_+ = 49$	1 395	484	355	1 275	388	273	1 149	302	201
$n_+ = 99$	353	29	14	285	22	11	224	17	9
$n_+ = 249$	20	2	1	17	2	1	14	9	29

Sur quels domaines de l'espace des paramètres l'approximation du modèle Poissonien par le modèle de POLYA est-elle valide ? Question difficile mais capitale. Après avoir examiné de multiples histogrammes, nous sommes arrivés à la conclusion que si la distance de HELLINGER était inférieure à $40 \cdot 10^{-4}$ l'approximation était satisfaisante. C'est bien-sûr une limite arbitraire dont le seul mérite est de fixer les idées. Nous avons vérifié qu'elle était largement respectée dès que $\lambda_+ \leq 2\mu$, c'est-à-dire dès que la taille des groupes n'est pas importante vis à vis du nombre de groupes représentés dans le piègeage; pour remplir cette condition, il faudra laisser les pièges en place un temps suffisant.

Un autre point important est l'expression de s en fonction des paramètres du modèle Poissonien. Son calcul pour toutes les combinaisons de $N_+ = 45(5)60$; $\lambda_+ = 10(1)13$ et $\mu = 3, 5, 8$ a montré que si c'était surtout une fonction de $(\mu + 1)/n_+$, λ_+ intervenait aussi. Il sera donc délicat de comparer des estimations de s obtenues dans des conditions de piègeages différentes au sens où elles modifieraient λ_+ .

Par contre $s = 0 \Leftrightarrow \mu = 0$ puisque dans les 2 modèles on retrouve la distribution binomiale, le test de l'agrégativité garde donc le même sens.

II.4. Cas de plusieurs pièges

Dans ce paragraphe, nous discutons l'extension des modèles proposés au cas où le nombre de pièges considéré est $I > 2$, n_i est le nombre d'insectes capturés dans le piège i .

Le modèle binomial (non-agrégatif), se généralise très simplement en modèle multinomial, à chaque piège correspond une probabilité d'attraction p_i (FELLER, 1968).

Le modèle Poissonien se généralise aussi d'une manière naturelle, le nombre de groupes d'insectes représentés dans le piège i est N_i , il suit une $P(\lambda_i)$; les distributions des N_i sont indépendantes. Malheureusement l'expression de la loi conjointe n'est pas très commode à utiliser, aussi nous limiterons-nous à l'examen de l'extension du modèle basé sur des distributions de POLYA; car c'est dans le cadre de ce modèle que nous ferons des développements statistiques.

La distribution de POLYA se généralise par une POLYA multivariable (JOHNSON et KOTZ, 1977) dont les paramètres sont p_1, p_2, \dots, p_I probabilités d'attraction relative des différents pièges (c'est la même généralisation que lorsque l'on passe de la binomiale à la multinomiale) et s paramètre de l'agrégativité. On remarque que n_i suit une $Po(n_i, p_i, s)$; marginalement, c'est donc toujours une approximation du modèle Poissonien; c'est une condition nécessaire mais pas suffisante puisque les marges d'une densité ne définissent pas la densité conjointe. Le volume des calculs nécessaires

nous a dissuadé d'envisager une comparaison numérique. Dans le cadre de ce modèle, on vérifie que :

$$\begin{aligned}
 E(n_i) &= n_+ p_i \\
 V(n_i) &= n_+ p_i (1 - p_i) \frac{1 + s.n_+}{1 + s} \\
 \text{Cov}(n_i, n_{i'}) &= - n_+ p_i p_{i'} \frac{1 + s.n_+}{1 + s}, \quad i \neq i'
 \end{aligned} \tag{8}$$

II.5. Cas de plusieurs répétitions

Il peut arriver que l'on dispose de données relatives à un même groupe de pièges, répétées J fois. Par exemple les pièges sont relevés sur des périodes de temps différentes; autre possibilité, on a mis en place le même ensemble de pièges (distincts entre eux par leur forme, leur couleur ou leur proximité à la lisière...) dans différents lieux. Il convient alors d'introduire pour l'analyse des données ce nouveau facteur que nous avons baptisé répétition mais qu'on pourrait rapprocher du facteur bloc dans les dispositifs expérimentaux classiques.

D'un point de vue probabiliste, la différence essentielle est qu'un insecte capturé se caractérise par le piège dans lequel il est arrivé et la période durant laquelle il a été capturé. Le résultat du piégeage est le fruit d'un processus aléatoire : il aurait pu être capturé par un autre piège; la répétition lui est imposée, il vivait à cette période là ou dans ce lieu là.

Soit n_{ij} le nombre d'insectes capturés dans le piège i durant la période j , avec $i = 1, \dots, I$ et $j = 1, \dots, J$. Pour chacune des répétitions, la distribution des n_{ij} à j et à n_{+j} fixés est celle donnée en II.4, enfin ces distributions sont indépendantes d'une répétition à l'autre. Si on considère le modèle dit de POLYA, on vérifie alors que:

$$\begin{aligned}
 E(n_{ij}) &= n_{+j} p_i \\
 V(n_{ij}) &= n_{+j} p_i (1 - p_i) \frac{1 + s.n_{+j}}{1 + s} \\
 \text{Cov}(n_{ij}, n_{i'j}) &= - n_{+j} p_i p_{i'} \frac{1 + s.n_{+j}}{1 + s}, \quad i \neq i' \\
 n_{ij} \text{ et } n_{i'j} &\text{ indépendantes si } j \neq j'.
 \end{aligned} \tag{9}$$

III. ASPECTS STATISTIQUES

III.1. Introduction

Par rapport au modèle binomial, les 2 modèles envisagés ci-dessus mettent en évidence que l'introduction de l'agrégativité augmente la variance, laissant l'espérance inchangée; on perd donc la liaison fonctionnelle entre moyenne et variance [formules (4) et (7)]. Du point de vue de l'estimation des paramètres, il faudra donc d'une manière ou d'une autre se donner les moyens, par des répétitions, d'estimer la variabilité. Nous ne considérerons les problèmes statistiques que dans le cadre du modèle de POLYA. En effet le modèle Poissonien conduit à des équations difficiles, de plus le modèle de POLYA comporte un paramètre de moins et demandera donc moins de données pour l'estimation.

La recherche d'estimateurs et de tests dans le contexte de ce modèle n'est ni classique ni simple. Nous nous sommes d'abord intéressés à la théorie du maximum de vraisemblance. Malheureusement les difficultés rencontrées dans la recherche d'estimateurs ne nous permettait pas de proposer autre chose qu'un calcul numérique; plus gênant encore : il n'était pas possible d'utiliser les résultats généraux dans le cas de répétitions puisque les effectifs capturés d'une répétition à l'autre sont différents. C'est pourquoi, nous avons préféré nous tourner vers une autre voie, celle de l'approximation Normale à partir des 2 premiers moments. L'avantage est de nous faire utiliser les statistiques classiques du modèle multinomial en leur donnant une interprétation différente. Nous mettons ainsi bien en évidence les dangers que présentent les comportements agrégatifs pour l'interprétation des données s'ils ne sont pas pris en compte.

Enfin pour nous rendre compte de la pertinence du passage du modèle de POLYA à la distribution normale, nous avons fait quelques simulations dans le cas de 2 pièges, 3 répétitions (d'effectifs respectifs 45, 50 et 55) pour $p_1 = 0,4; 0,5$ et $s = 0; 0,15$; pour chaque jeu de paramètres 1 000 tirages ont été effectués à partir desquels ont été calculés les moyennes et les écarts-types de nos statistiques.

III.2. Cas de 2 pièges identiques (ou connus)

Nous supposons dans un premier temps que p_1 est connu. Ce cas n'est pas invraisemblable si les 2 pièges sont identiques et qu'a priori rien ne les distingue, alors $p_1 = 0,5$. Autre possibilité, ce que nous appelons piège est constitué d'un nombre d'éléments différents et on admet que les captures respectives sont proportionnelles au nombre d'éléments.

Dans ce cas extrêmement simple nous n'avons qu'un paramètre inconnu à considérer : s . Si on fait l'approximation Normale

$$n_i / (n_+) \sim N \{ n_+ p_1, n_+ p_1 (1 - p_1) (1 + n_+ \cdot s) (1 + s)^{-1} \}$$

alors

$$S(p_1, s) = \frac{(n_i - n_+ p_1)^2}{n_+ p_1 (1 - p_1) (1 + n_+ s) (1 + s)^{-1}} \quad (10)$$

suit un χ^2_1 .

Utilisant le fait que $S(p_1, s)$ est une fonction monotone décroissante de s ; il est possible de reporter sur s , une estimation faite sur le χ^2 et même un intervalle de confiance. Par exemple si les 2 pièges sont identiques et si X_i et X_s sont les valeurs seuils d'un χ^2_1 telles que :

$$P(X_i < S(p_1, s) < X_s) = 1 - \alpha \quad (11)$$

alors en inversant S , on en déduit un intervalle de confiance de s

$$P(s_i = S^{-1}(p_1, X_i) < s < s_s = S^{-1}(p_1, X_s)) = 1 - \alpha \quad (12)$$

La figure 3 fournit l'interprétation graphique de ce raisonnement. On y constate que 2 petites difficultés peuvent être éprouvées. D'abord $S^{-1}(p_1, X_s)$ peut être négatif, dans ce cas on prendra $s_i = 0$; enfin si l'asymptote horizontale est supérieure à X_i , $S^{-1}(p_1, X_i)$ n'existe pas, on prendra alors $s_s = +\infty$.

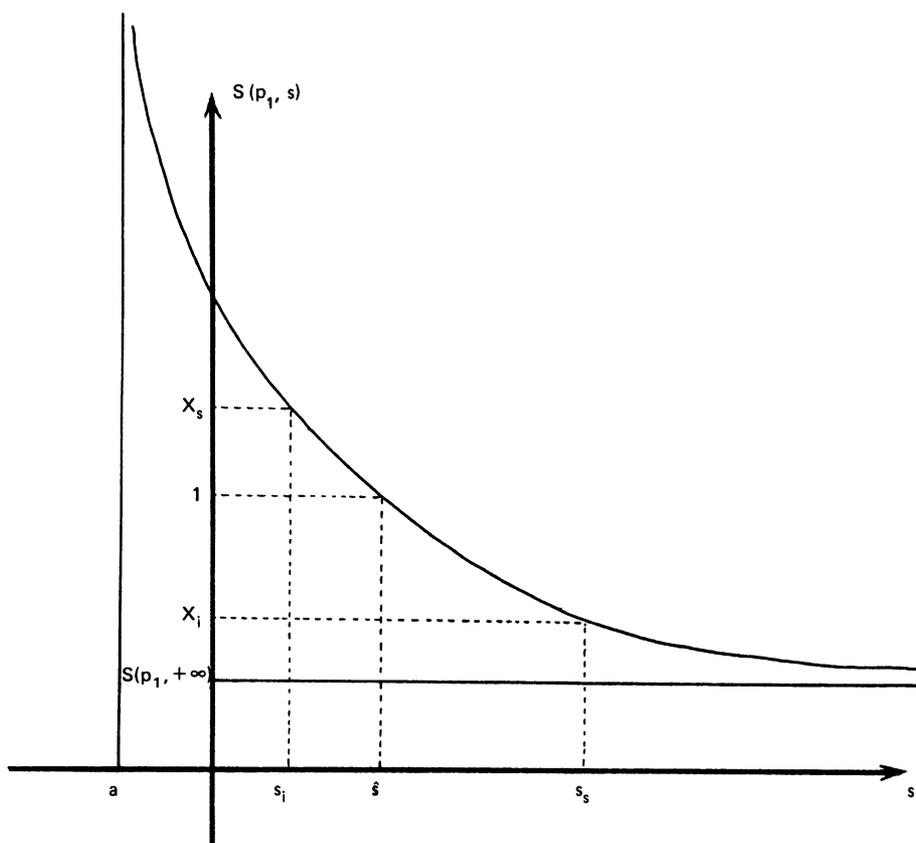


Figure 3. — Représentation graphique de calcul de l'estimation de s et des bornes de l'intervalle de confiance dans le cas où p_1 est supposé connu. Pour des raisons de lisibilité l'échelle des abscisses a été déformée.

$$a = -1/n_+$$

$$\text{si } p_1 = 0,5; S(p_1, +\infty) = \sum_{i=1}^2 4(n_i - n_+/2)^2/n_+^2$$

TABLEAU 3

Résultats de la simulation pour l'estimation, le calcul d'un intervalle de confiance de s , le test de $s = 0$ lorsque p est supposé égal à 0,5. Les statistiques sur X_i ont été calculées pour 2,5%. Celles pour X_s également mais on n'a pas tenu compte des valeurs infinies. Les 2 dernières lignes sont les proportions de cas où l'hypothèse nulle a été rejetée.

pour $s =$ pour $p =$	0 0,5	0 0,4	0,15 0,5	0,15 0,4
moyenne de X_i	10^{-4}	0,004	0,030	0,044
écart-type de X_i	0,001	0,008	0,033	0,043
moyenne de \hat{s}	0,006	0,046	0,161	0,233
écart-type de \hat{s}	0,012	0,040	0,155	0,226
moyenne de X_s	0,530	5,153	5,384	6,653
écart-type de X_s	1,05	22,6	24,9	19,0
proportion de bornes ∞	0,011	0,329	0,730	0,825
test significatif à 2,5 %	0,018	0,420	0,768	0,863
test significatif à 5 %	0,045	0,528	0,807	0,896

De la même manière on prendra comme estimation de s , $S^{-1}(p_1, 1)$ et comme statistique de test de l'agrégativité $S(p_1, 0)$ qui, si $s = 0$, suit un χ^2_1 . On verra au moment de l'interprétation des exemples le danger que comporte une telle pratique, puisque c'est le test habituel de l'hypothèse $p = p_1$ qui sert à tester l'agrégativité. Notons enfin qu'il n'est pas nécessaire de rechercher un intervalle de confiance fini, dans l'optique du test de l'agrégativité, on peut rechercher un intervalle le plus écarté possible de 0, c'est-à-dire de la forme $[s_i, +\infty[$, il suffit pour cela de prendre $X_i = 0$.

Les résultats des simulations présentés dans le tableau 3 confortent tout à fait notre point de vue. En particulier, on observe qu'une erreur d'appréciation de p_1 (0.4 au lieu de 0.5) conduit à déclarer une agrégativité qui n'existe pas; d'autre part la proportion de rejet sous l'hypothèse nulle est proche du niveau attendu.

III.3. Cas de plusieurs pièges connus avec répétitions

Si de la même manière que dans III.2, on utilise l'approximation Normale à partir des espérances et des variances déjà calculées (9), la fonction aléatoire norme au carré du vecteur multinomial centré réduit devient simplement :

$$S(p_1, \dots, p_I, s) = \sum_{j=1}^J \sum_{i=1}^I \frac{(n_{ij} - n_{+j} p_i)^2}{n_{+j} \cdot p_i (1 + n_{+j} \cdot s) / (1 + s)} \quad (13)$$

qui suit un $\chi^2_{(I-1)J}$.

Là encore $S(\dots, p_i, \dots, s)$ est une fonction décroissante de s . Si les p_i sont connus, nous pouvons donc l'utiliser pour estimer s , déterminer un intervalle de confiance, faire un test comme précédemment en prenant $J(I-1)$ comme degrés de liberté pour le χ^2 .

III.4. Cas de plusieurs pièges inconnus avec répétitions

Jusqu'à présent, nous avons supposé que nous connaissons les p_i , probabilités d'attraction relative des pièges. Ce n'est généralement pas le cas, si les pièges sont différents. Nous allons maintenant examiner ce qu'il est possible de faire dans cette situation. La première remarque est que dans le cas de 2 pièges, le nombre de paramètres libres à estimer est 2 (p_1 et s); il faudra donc au minimum 2 résultats pour les estimer, d'où la nécessité de disposer des répétitions. Présenté autrement, il nous faut estimer l'espérance et la variance, la première caractérisant l'attractivité du piège 1, la seconde dépendant de la première et du degré d'agrégativité.

Dans le cas de plusieurs pièges, la problématique est identique. Pour estimer plus que les p_i , il faut donc aussi des répétitions.

III.4.a) Estimation des p_i

Nous proposons de prendre l'estimation que fournit la méthode des moments basée sur les effectifs totaux (n_{i+}) récoltés dans chaque type de pièges :

$$E(n_{i+}) = n_{++} \cdot p_i = n_{i+} \quad (14)$$

Un des avantages de cet estimateur est d'être identique à celui du modèle non-agrégatif (multinomial). On obtient donc :

$$\hat{p}_i = n_{i+} / n_{++} \quad (15)$$

III.4.b) Test de l'agrégativité ($s = 0$)

On utilise la fonction aléatoire définie en (13) en mettant $s = 0$ et en remplaçant les p_i inconnus par leurs estimateurs. Sous l'hypothèse nulle de non-agrégativité, tenant compte du fait qu'on a estimé $(I - 1)$ paramètres indépendants, cette statistique suit un $\chi^2_{(I-1)(J-1)}$. Le test s'effectue donc sans difficulté. Contrairement au cas des p_i connus, il n'est pas possible d'en tirer un intervalle de confiance de s , puisque cette distribution asymptotique de χ^2 n'est valable que pour $s = 0$.

III.4.c) Estimation de s

Cette estimation ne sera pratiquée que si le test précédent est significatif. En cas contraire, on prendra comme estimation de s la valeur nulle. Nous proposons une estimation basée sur une approximation de la méthode des moments.

La statistique tout indiquée est bien sûr celle qui nous a servi pour faire le test de s nul, ou d'estimer s dans le cas où les p_i étaient connus. Nous avons donc calculé $E(S(\hat{p}_i, s))$ malheureusement le calcul ne se simplifie pas car il fait intervenir $E(1/n_{i+})$ que nous ne connaissons pas explicitement; nous avons essayé de considérer l'espérance conditionnelle de $S(\hat{p}_i, s)$ aux n_{i+} , là non plus les simplifications ne sont pas suffisantes et nous retrouvons des p_i . C'est pourquoi, nous proposons une estimation très empirique qui consiste à résoudre l'équation :

$$S(\hat{p}_i, s) = J \cdot (I - 1) \quad (16)$$

L'empirisme consiste à supposer que les p_i sont égaux à leurs estimations (15). Dans ce cas nous nous retrouvons dans la même situation qu'en (13), l'espérance est donc $J(I - 1)$. On note \hat{s} cet estimateur.

Lorsque s est petit, cette estimation tend plutôt à la sous-estimation, puisque si on suppose $s = 0$, on sait que la statistique est un $\chi^2_{(I-1)(J-1)}$ dont l'espérance est $(I - 1)(J - 1) < J(I - 1)$ et comme on a déjà vu que S est décroissante en s , par continuité on peut supposer que s est sous-estimé.

Comme la résolution de l'équation (16) pour estimer s est conditionnée par le test présenté en III.4.b significatif, nous sommes pratiquement sûrs d'obtenir une estimation positive. Nous le serions complètement si nous avions choisi $(I - 1)(J - 1)$ et non pas $J(I - 1)$ puisque l'intervalle de confiance du χ^2 contient son espérance. Nous préférons cependant diminuer l'estimation que l'augmenter, par stratégie conservatrice. Si jamais \hat{s} était négatif, on pourrait prendre une valeur non nulle aussi petite que l'on veut. Il se pourrait aussi que $J(I - 1)$ soit inférieur à l'asymptote, auquel cas l'estimation serait plus l'infini; la probabilité de ce cas de figure doit être très faible dans le cadre de notre modèle, nous le considérons comme un indice que les données ne correspondent pas au modèle envisagé : une vérification du protocole expérimental s'imposerait.

Les résultats des simulations présentées dans le tableau 4, sont tout à fait conforme à nos prévisions, en particulier en ce qui concerne la distribution de $S(\hat{p}, 0)$ sous l'hypothèse $s = 0$.

TABLEAU 4

Résultats de la simulation pour le test de $s = 0$ lorsque p est libre; on a mis entre parenthèses le résultat théorique lorsqu'il était connu

pour $s =$ pour $p =$	0 0,5	0 0,4	0,15 0,5	0,15 0,4
moyenne de $S(\hat{p}, 0)$	1,98 (2)	1,97 (2)	15,10	15,08
écart-type de $S(\hat{p}, 0)$	1,98 (2)	1,97 (2)	13,34	13,54
proportion de rejet à 5 %	0,050 (0,05)	0,051 (0,05)	0,705	

IV. QUELQUES EXEMPLES

Les données traitées sont extraites de la série de piégeages décrite dans CHAMBON *et al.* (1984); pour pouvoir considérer les pièges équivalents, n'ont été retenus que 3 pièges du même type placés en bordure du champ. Les répétitions sont des captures réalisées sur des semaines consécutives. Le tableau 5 donne pour 2 catégories d'insectes les effectifs recueillis et les résultats des procédures statistiques proposées.

TABLEAU 5

Traitement statistique de 2 cas concrets

Données			
	R ₁	R ₂	
P ₁	26	28	
P ₂	30	36	
P ₃	9	13	
supposant $p_i = 1/3$			
intervalle bilatère à 5 % [0,014, 1,815]			
intervalle unilatère à 5 % [0,019, + ∞[
estimation de $\hat{s} = 0,069$			
supposant les p_i inconnus			
$\hat{p}_1 = 0,380$; $\hat{p}_2 = 0,465$; $\hat{p}_3 = 0,155$			
S($\hat{p}, 0$) = 0,335 pour 2 ddl (non-significatif)			
estimation de $s = 0$			
analyse des données du genre <i>Metopina</i>			
Données			
	R ₁	R ₂	R ₃
P ₁	83	18	129
P ₂	106	39	85
P ₃	67	38	105
supposant $p_i = 1/3$			
intervalle bilatère à 5 % [0,0044, 0,136]			
intervalle unilatère à 5 % [0,0054, + ∞[
estimation de $s = 0,0194$			
supposant les p_i inconnus			
$\hat{p}_1 = 0,343$; $\hat{p}_2 = 0,343$; $\hat{p}_3 = 0,313$			
S($\hat{p}, 0$) = 25,47 pour 4 ddl (significatif)			
$\hat{s} = 0,0192$			
analyse des données de <i>Meligethes æneus</i>			

Pour le genre *Metopina* si on suppose les pièges équivalents, on concluerait à l'existence d'agrégativité or les estimations et le test en laissant p libre montrent qu'on ne peut rien affirmer, une hypothèse plus vraisemblable serait que le piège 3 est moins attractif que les autres. C'est une démonstration de la fragilité de la dernière procédure vis à vis des suppositions sur lesquelles elle repose.

Pour l'espèce *Meligethes æneus* on peut au contraire conclure sans arrière-pensée.

REMERCIEMENTS

Nous tenons à remercier Jean-Pierre CHAMBON qui a gracieusement mis ses données à notre disposition, Guy DECOUX qui a effectué les calculs de simulations et les analyses statistiques, Jean-Marc AZAIS et Camille DUBY qui ont relu la première version de notre manuscrit.

RÉFÉRENCES BIBLIOGRAPHIQUES

- J.P. CHAMBON, J.B. DENIS, C. COCQUEMPOT, G. GENESTIER, P. PINEAU et J. PRONOST (1984). — Etude du piégeage des arthropodes en fonction de la situation des pièges, en céréaliculture de zone ouverte. *Acta Oecologica Applicata* 5 (3), 199-210.
- D. DACUNHA-CASTELLE et M. DUFLO (1982). — *Probabilités et Statistiques 1. Problèmes à temps fixe* MASSON Paris, 214 pp.
- W. FELLER (1968). — *An introduction to probability theory and its applications*, volume I, 3^e édition John WILEY New-York, 509 pp.
- S. FRONTIER (1983). — *Stratégies d'échantillonnage en écologie*, MASSON Paris, 494 pp.
- N.L. JOHNSON et S. KOTZ (1969). — *Discrete Distributions*, John WILEY New-York, 328 pp.
- N.L. JOHNSON et S. KOTZ (1977). — *Urn models and their application*, John WILEY New-York, 402 pp.
- A. ROGERS (1974). *Statistical Analysis of Spatial Dispersion* PION limited London, 164 pp.