Echantillonnage et estimation d’abondance des populations d’arthropodes terrestres

Isabelle Badenhausser
Centre d’Etudes Biologiques de Chizé
INRA-CNRS
Villiers-en-Bois
79360 Beauvoir sur Niort
badenh@cebc.cnrs.fr

Plan

- Introduction
- Quelques techniques de prélèvement et de collecte des arthropodes
  - Arthropodes aériens
  - Arthropodes souterrains
  - Conséquences sur l'unité d'échantillonnage
- Nature de la variable mesurée
  - Dénombrement exhaustif
  - Sous-échantillonnage
  - Méthodes visuelles: classe d'abondance, symptôme,
  - Notation présence-absence
- Plans d'échantillonnage classiques
- Echantillonnage aléatoire simple
  - Cas d'une variable qualitative
  - Cas d'une variable quantitative continue
  - Cas d'une variable quantitative discrète
- Un exemple concret
- Bibliographie

Introduction

- Quelques définitions relatives à l’échantillonnage
- Quelles caractéristiques à estimer?
- Ce qu’il faut définir au préalable pour établir un plan d’échantillonnage
- Qu’est ce qu’établir un plan d’échantillonnage?
- Qu’est ce que la précision d’un estimateur?
- Qu’est ce qu’un biais?

Quelques définitions relatives à l'échantillonnage

Procédure d'échantillonnage: consiste à observer sur un sous ensemble de la population certaines caractéristiques dans le but d'en déduire des valeurs concernant l'ensemble de la population échantillonnée

Définir la procédure de recueil des unités statistiques qui seront observées = f (objectif et des contraintes techniques)

Faisabilité, coût, analyse des résultats

Quelle question ? Quelle population?
Quelle précision?

Représentativité: tout élément de la population a une probabilité non nulle d'appartenir à l'échantillon (cadre=population fixe)

Quelles caractéristiques à estimer?

- estimation de l'abondance relative d'espèces ou de la proportion d'individus présentant telle ou telle caractéristique dans la population

- estimation de densités de population: nombre/plante, nombre/m², nombre/g de sol, nombre/piège

- estimation de la biomasse

Pour caractériser les milieux, gérer la biodiversité, évaluer l'impact des pratiques agricoles, proposer des outils de gestion du paysage agricole, gérer les populations

Ce qu'il faut définir au préalable pour établir un plan d'échantillonnage

✓L'information recherchée

✓La population (sens statistique): Notion de population finie, infinie, fixe, aléatoire (super population)

• un champ donné pour prendre une décision ponctuelle (population fixe)

• les champs d'une région pour procéder à des avertissements agricoles (population aléatoire): les champs échantillonnés ne nous intéressent pas en soi.

✓La précision:

• liée aux risques qu'on accepte de prendre de se tromper et aux conséquences

Il faut une bonne analyse des objectifs recherchés et des connaissances ou des hypothèses sur le phénomène étudié

Domaine « découpé » en unités d'échantillonnage

l'ensemble des unités d'échantillonnage constitue la population étudiée

Echantillon:

-Nombre d'unités d'échantillonnage à observer ?

-Comment les choisir dans le domaine ?

Unité d'échantillonnage

-naturelle: plante, tige, individu

-Arbitraire: volume, surface…

- Technique de collecte ?

- Variable mesurée ?

Qu’est ce qu’établir un plan d’échantillonnage ?

C’est:

-Définir la technique de collecte

-Définir l’unité d’échantillonnage

-Définir la variable mesurée

-Calculer la taille de l’échantillon à observer

-Définir la procédure de tirage des unités d’échantillonnage

-Vérifier qu’on obtient bien l’information recherchée

Qu'est ce que la précision d'un estimateur ?

La moyenne vraie et inconnue, μ, d'une population est estimée par échantillonnage, m (variable aléatoire=proportion, effectif)

On cherche à construire autour de m un intervalle de confiance I d'étendue 2i, pour un seuil choisi α (α =0.95 par ex), tel que:

$$\text{Proba}\ (m - \text{i} \leq \mu \leq m + \text{i}) = \alpha$$

α = proba que l'intervalle de confiance contienne μ

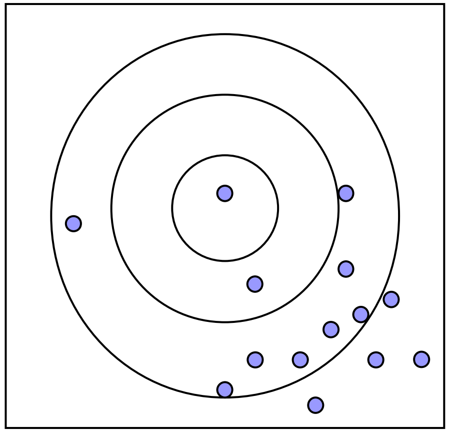
Précision absolue de m d = i

Précision relative de m D = i / m
(% de m)

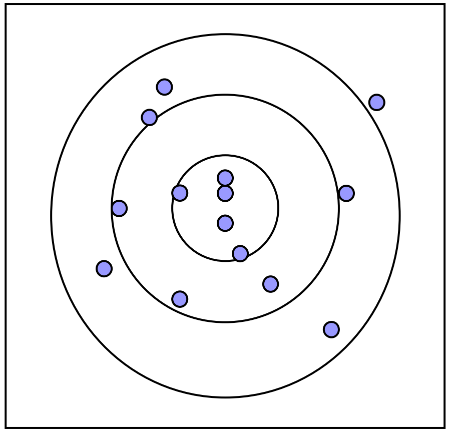
Qu'est ce qu'un biais ?

Définition: écart entre le paramètre estimé et son espérance

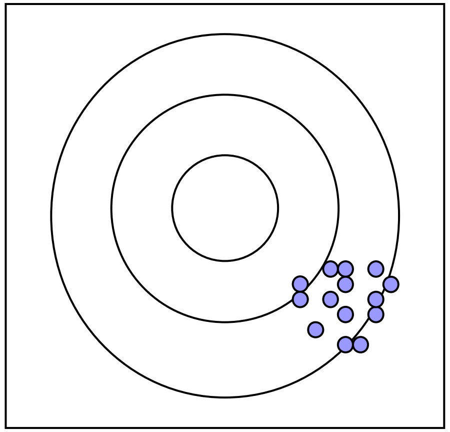
Centre de la cible=vraie valeur à estimer



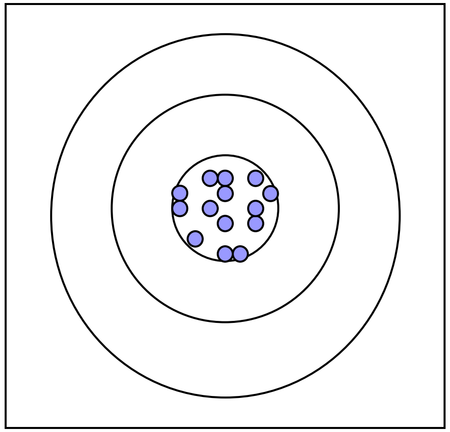
- Avec biais
- Grande variance



- Sans biais
- Grande variance



- Avec biais
- Petite variance



- Sans biais
- Petite variance

= estimateur utilisé

1. Quelques techniques de prélèvement et de collecte

- 1.1. Généralités
- 1.2. Quelques techniques de collecte d’arthropodes aériens
- 1.3. Quelques techniques de collectes d’arthropodes souterrains
- 1.4. Conséquences sur l’unité d’échantillonnage

1.1. Généralités

➢La technique de collecte détermine l'unité d'échantillonnage

L'unité statistique d'échantillonnage est :

• naturelle: une plante, un arbre

• arbitraire: 20 cm^2 de végétation, portion de plante, branche… (comment la construire?), volume de sol

Degré de visibilité de l’arthropode:

- localisation sur ou dans la plante
- effectifs
- morphologie
- mode de reproduction
- mobilité, comportement

Acyrthosiphon pisum/luzerne

Macrosiphum albifrons/lupin

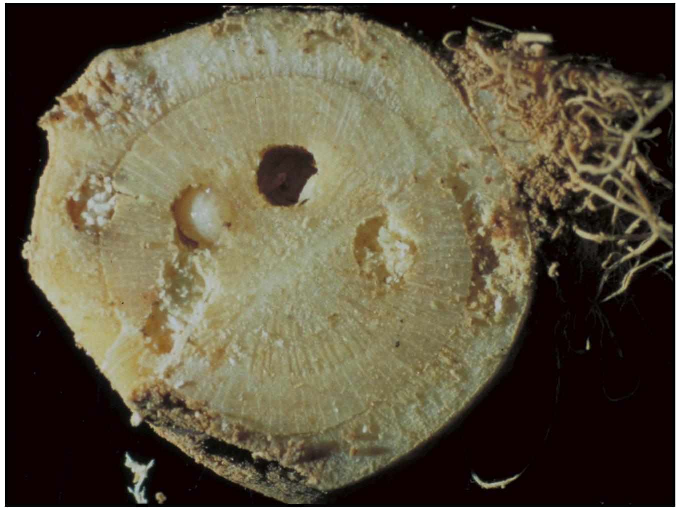
Chorthippus biguttulus/jachère

Arthropode du sol, ou se développant dans les tissus végétaux

Lobesia botrana / vigne

Isopode terrestre

Sitona lineatus / pois



Baris coerulescens / colza

Support du cycle de développement :

- Végétal:
  - port
  - densité de plantes
  - taille
  - morphologie
- Sol

Luzerne

Vigne

Jachère

1.2. Quelques techniques de collecte d'arthropodes aériens

- Observations in situ : non destructrices (répétitions possibles dans le temps)
  - Sur les mêmes unités
  - Sur d'autres unités
- Prélèvements in situ des insectes présents
  - Battage
  - Filet-fauchoir
  - Aspiration: D-vac, piège à succion
  - Enceintes de dimensions variables : biocénomètres
  - Piéges divers : méthodes relatives (pièges à eau/lumineux,mécaniques,…)
- Prélèvements de l'échantillon pour observations ultérieures: destructrices
  - Brossage/lavage/battage
  - Dissections
  - Extraction: techniques du Berlese (extraction par la chaleur)

Filet fauchoir

-Normalisation de la méthode de fauchage

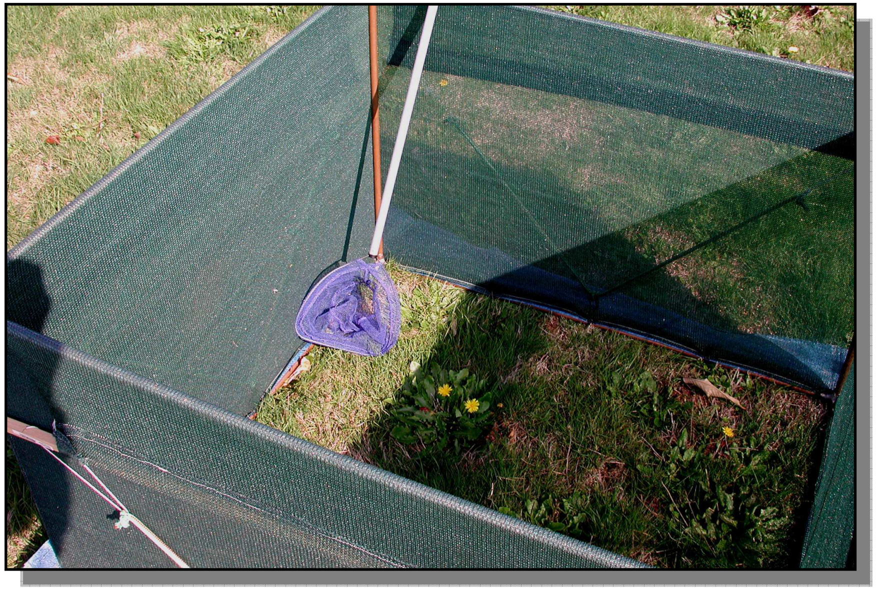
-Impose de trier les arthropodes parmi ce qui est prélevé par fauchage, et de compter

On peut chercher à établir une relation entre cette méthode et les effectifs par plante:

A.pisum sur la luzerne: Aphids/stem=-9.9873 + 0.308 SW -0.00008 SW^2

Enceinte: biocénomètre

- Enceinte dont les dimensions sont connues
- Emprisonne les arthropodes et permet les comptages
- La donnée peut être un nombre d'individus par unité de surface



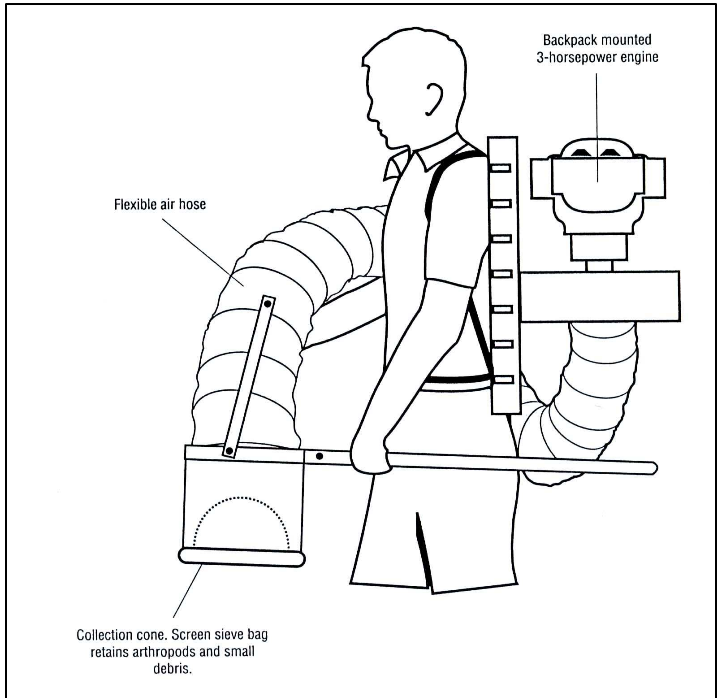
DVAC

-Aspirateur portable.

-On peut rapporter les prélèvements à une unité de surface.

-Efficacité différente selon les morphes, le stade des plantes.

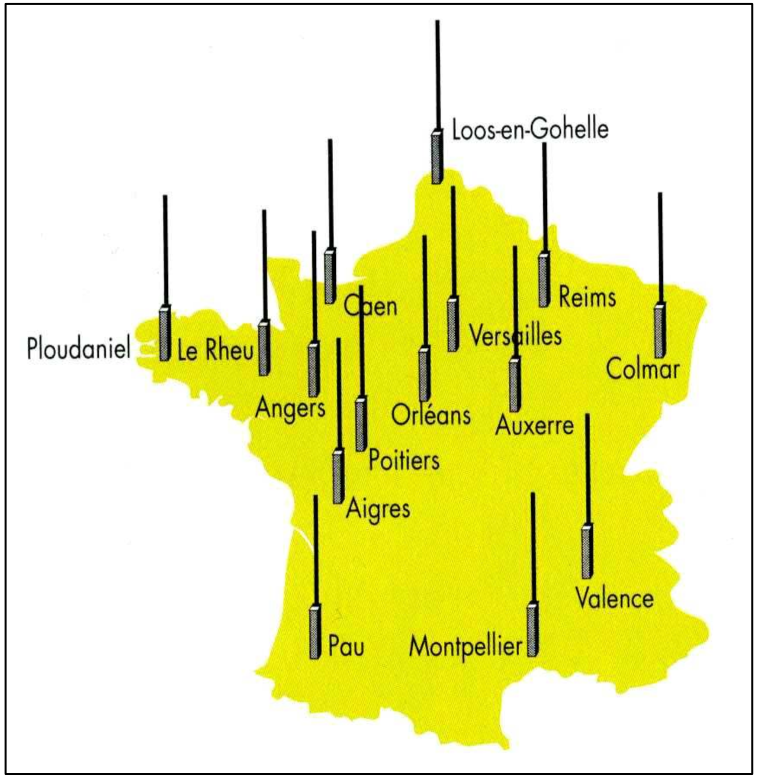
-Standardisation difficile



Drawing of a D-Vac® suction device indicating generalized construction and architecture

Piège à succion

Piège à succion: Agraphid (1978), Euraphid (60 pièges). Mesure la densité aérienne en un point



Piège jaune à eau

Echantillonnage des pucerons ailés (contaminants) dans les cultures:
mesure de la tendance à l ’atterrissage d ’une fraction de la population

Piège jaune à eau:

Normalisation nécessaire: taille, couleur, implantation

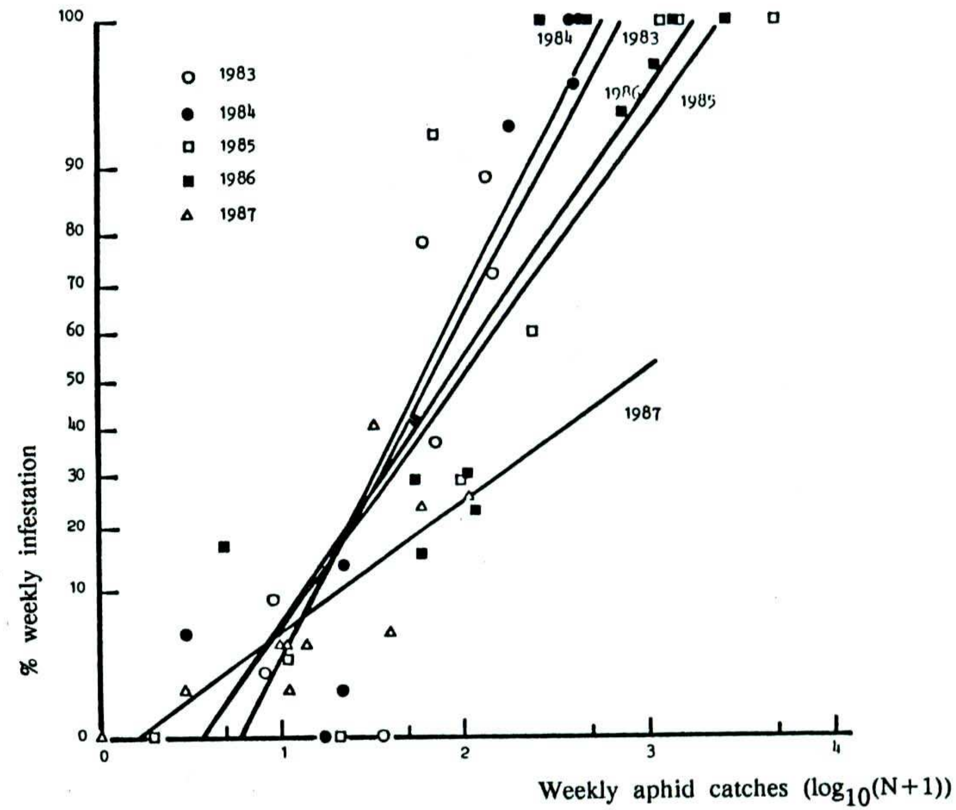
Taille : 60 cm x 60 cm, 10 cm épaisseur. Rempli aux 2/3 eau+mouillant

Nombreux facteurs de variation de l ’efficacité

Efficacité=*f*(espèce)

Ex: *R.padi*

La densité de vol reflète l ’activité des populations au niveau des plantes et permet d ’estimer les risques encourus



Relations entre les captures hebdomadaires de R. padi au piège à succion du Rheu et le pourcentage de plantes pièges infestées par cette espèce au cours des semaines correspondantes (d’après GILLET et al., 1990).

1.3. Quelques techniques de collecte d'arthropodes souterrains

➢Prélèvements in situ des arthropodes présents

- Comptages si visibles
- Pièges divers : méthodes relatives.

Pièges attractifs/mécaniques

Ex: pots pièges, et carreau

➢Prélèvements de l'échantillon pour observations ultérieures: destructrices

- Extraction mécanique :
  - dissections
  - lavages de sol/flottaison,centrifugation/sédimentation
- Extraction fondée sur certains comportements :
  - Berlese: utilisation de la chaleur et de la lumière

1.4. L'unité d'échantillonnage

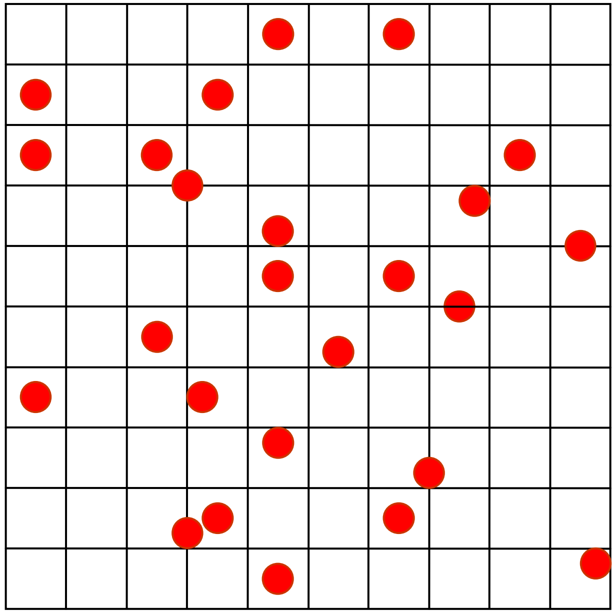
Naturelle:	Arbitraire :
-Une plante -Un arbre -Un insecte -Un poisson	- 1m^2 de prairie - 1 ha de blé - 1 litre d'eau - 1 piège
Pas de problème	Comment la construire ? Quelles conséquences ?

C'est une vraie question

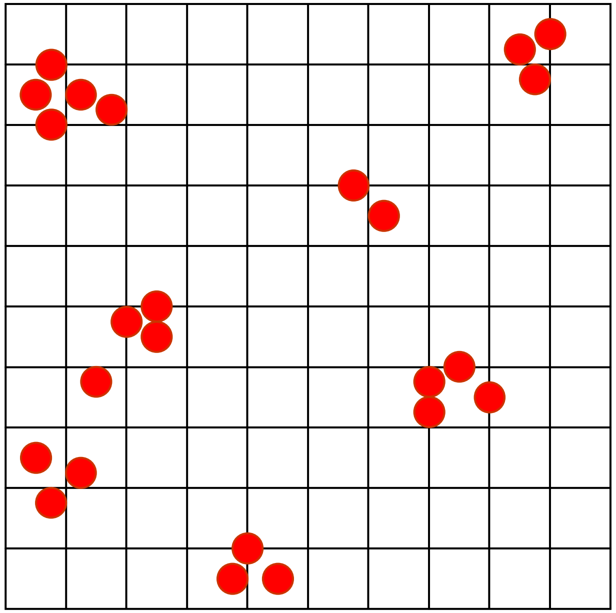
Il faut prendre en compte:

-La répartition spatiale des individus

-Le comportement des individus



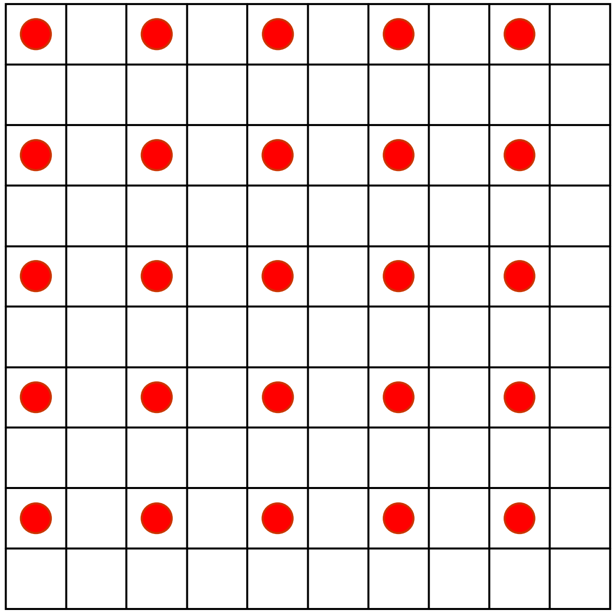
Répartition aléatoire des tâches



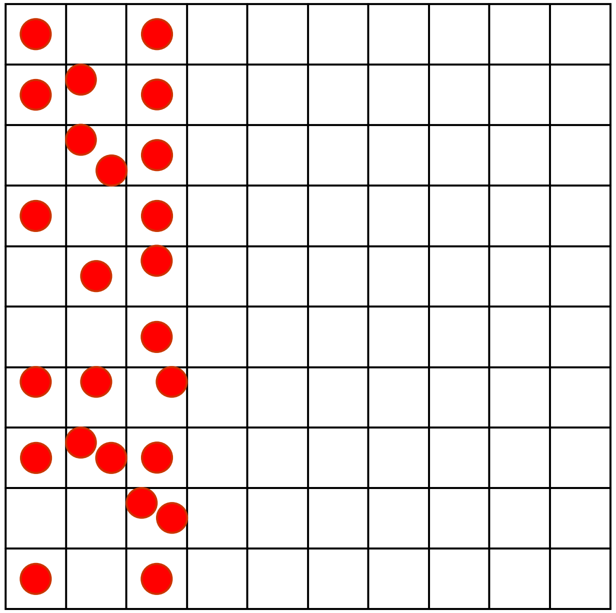
Répartition en agrégats des tâches

La taille de l'UE n'a aucune conséquence sur l'estimation du paramètre

La taille de l'UE a des conséquences sur l'estimation du paramètre



Répartition régulière des tâches



Répartition des tâches en agrégats

Si l'UE=1 m2, le taux d'unités occupées par un patch=25/100

Si l'UE=4m2, le taux d'unités occupées par un patch est 25/25

Si l'UE=1 m2, le taux d'unités occupées par un patch=21/100

Si l'UE=4m2, le taux d'unités occupées par un patch est 10/25

On aurait intérêt à stratifier l'échantillonnage

On n'a pas la même image du phénomène selon l'UE et la répartition spatiale des tâches. Quelle est la meilleure UE ?

Il faut expérimenter

On a intérêt à choisir la plus petite taille d’UE:

-À travail égal: $\searrow$ de l’erreur statistique car $\nearrow$ du nombre d’UE

-Meilleure représentativité de l’hétérogénéité du domaine car exploration plus importante du fait de l’ $\nearrow$ du nombre d’UE

Mais:

Attention : une plus petite UE a une probabilité d’inclusion plus faible

A mettre en relation avec le comportement et les effectifs de l’espèce considérée

2. Nature de la variable mesurée

- 2.1.Variable mesurée et traitement statistique
- 2.2. Dénombrement exhaustif
- 2.3. Sous-échantillonnage
- 2.4. Méthodes visuelles par classes d'abondance
- 2.5. Méthodes visuelles par notations symptomatiques
- 2.6 Notation Présence-absence

2.1. Variable mesurée et traitement statistique

- Dénombrement exhaustif
- Sous échantillonnage
- Biomasse

Variable quantitative discrète ou continue

- Estimations visuelles par classe d'abondance
- Note symptomatique

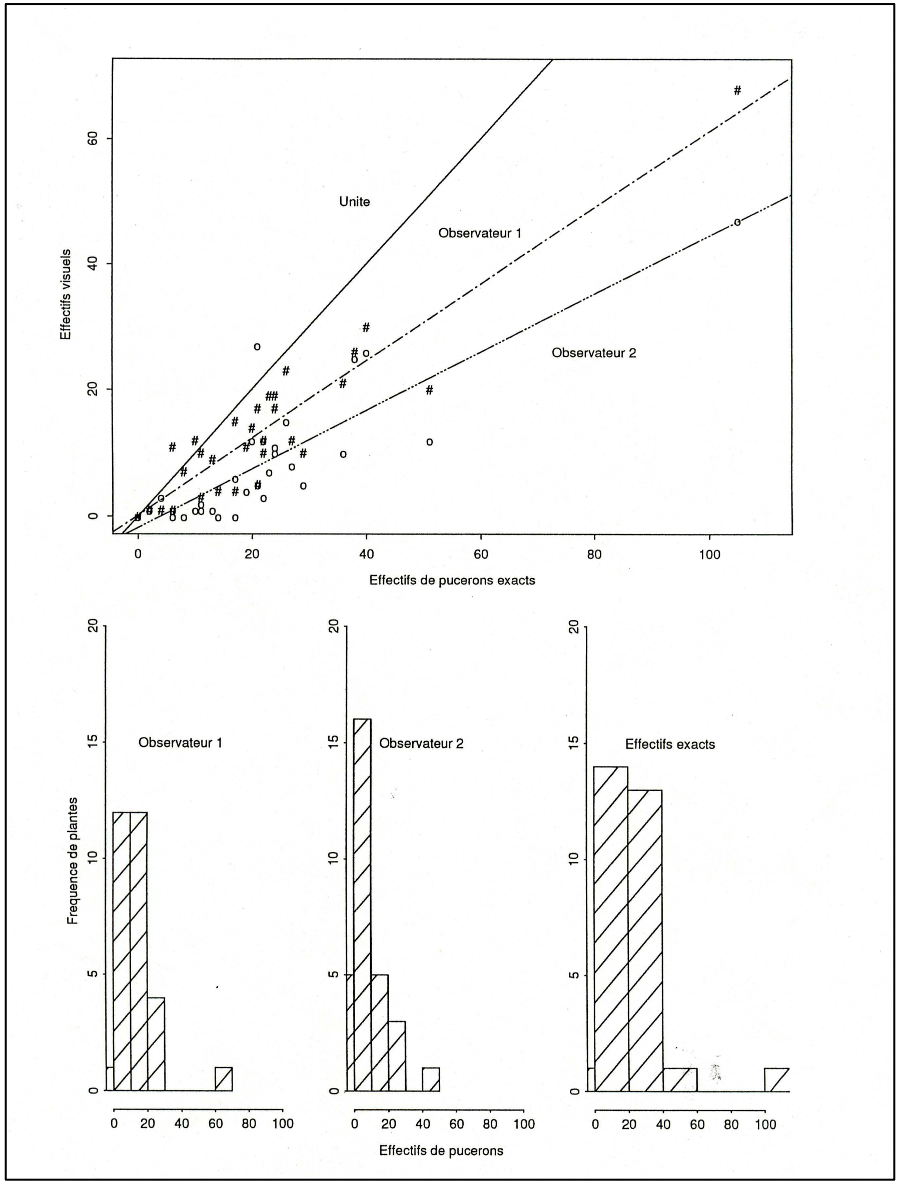
Variable Semi-quantitative

- Présence-absence

Variable binomiale

Traitement statistique!

2.2. Dénombrement exhaustif



B. helichrysi et le tournesol

Il faut évaluer:

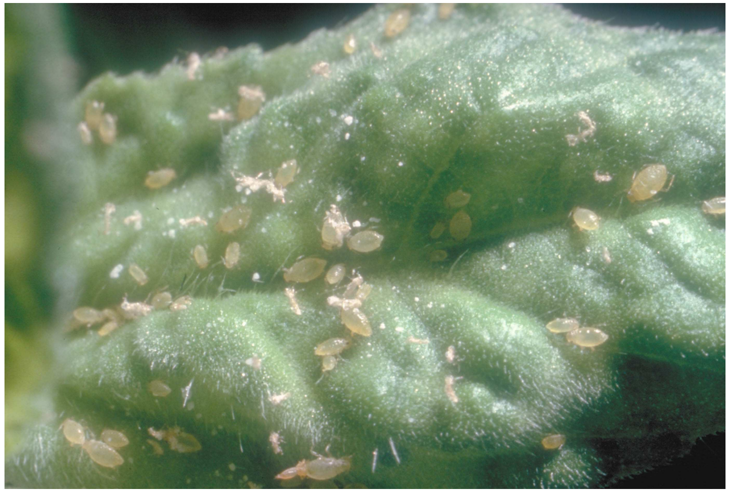
- le coût
- la précision
- la répétabilité entre observateurs
- les facteurs qui agissent sur la fiabilité du comptage: ordre d'acquisition, stade des plantes, effectif de ravageurs

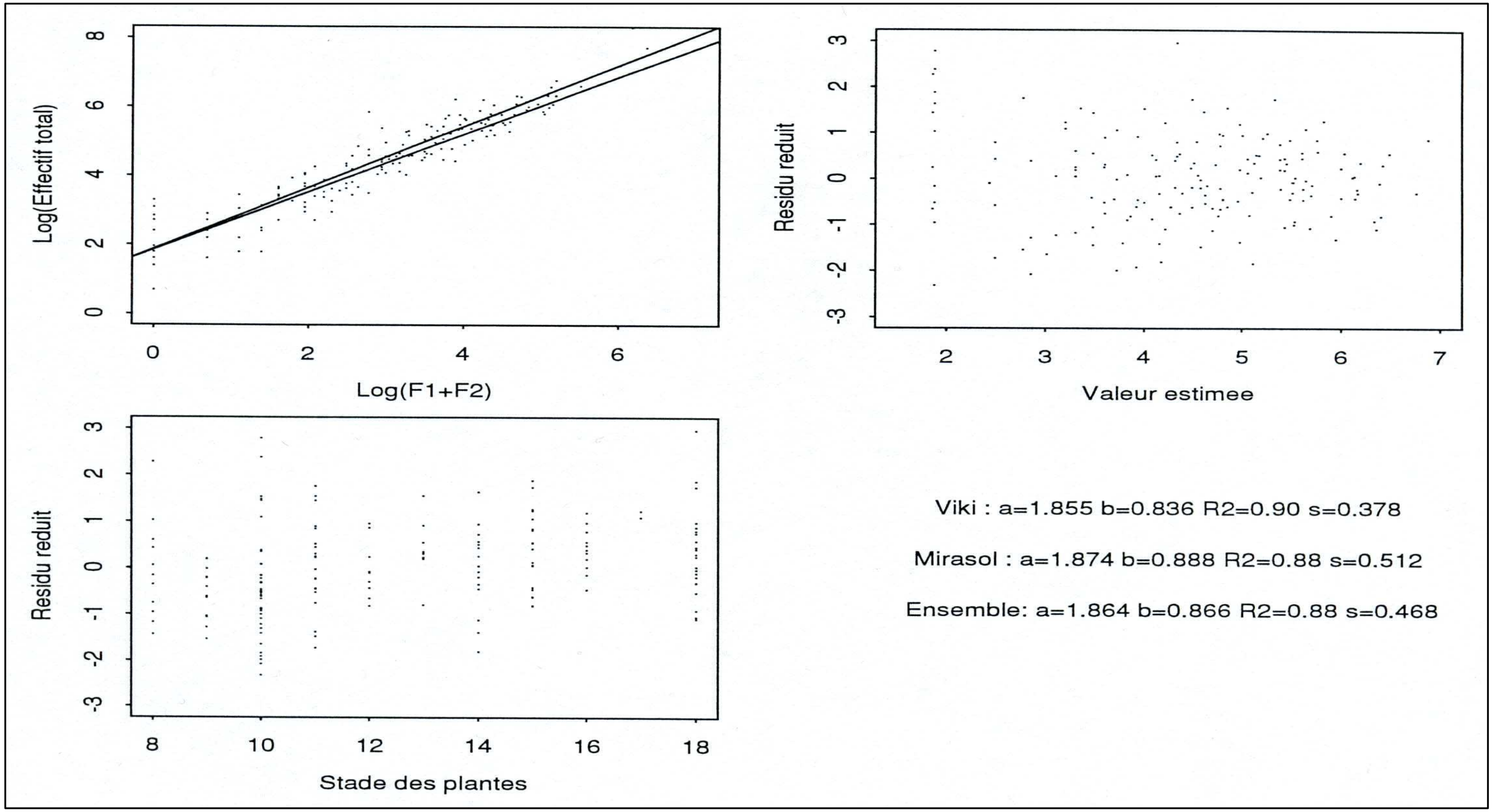
2.3. Sous-échantillonnage

- dénombrement sur un ou plusieurs organes de la plante sélectionnés selon divers procédés
- nécessite de connaître la distribution des pucerons sur la plante

Problème de l ’étalonnage, de la mesure de la fiabilité de la méthode et du calcul de la précision associée à l ’estimation de la population

Ex: mise au point d’une technique pour estimer les effectifs par plante du puceron du tournesol, *Brachycaudus helichrysi*





Dernière paire de feuilles formées >4cm= sous-échantillon

- limite le comptage à environ 20% des effectifs
- fournit une estimation indépendante du stade (de 8 à 18 feuilles)
- gamme d 'effectifs concernés: de 0 à 4000 pucerons/plante
- facile à mettre en œuvre

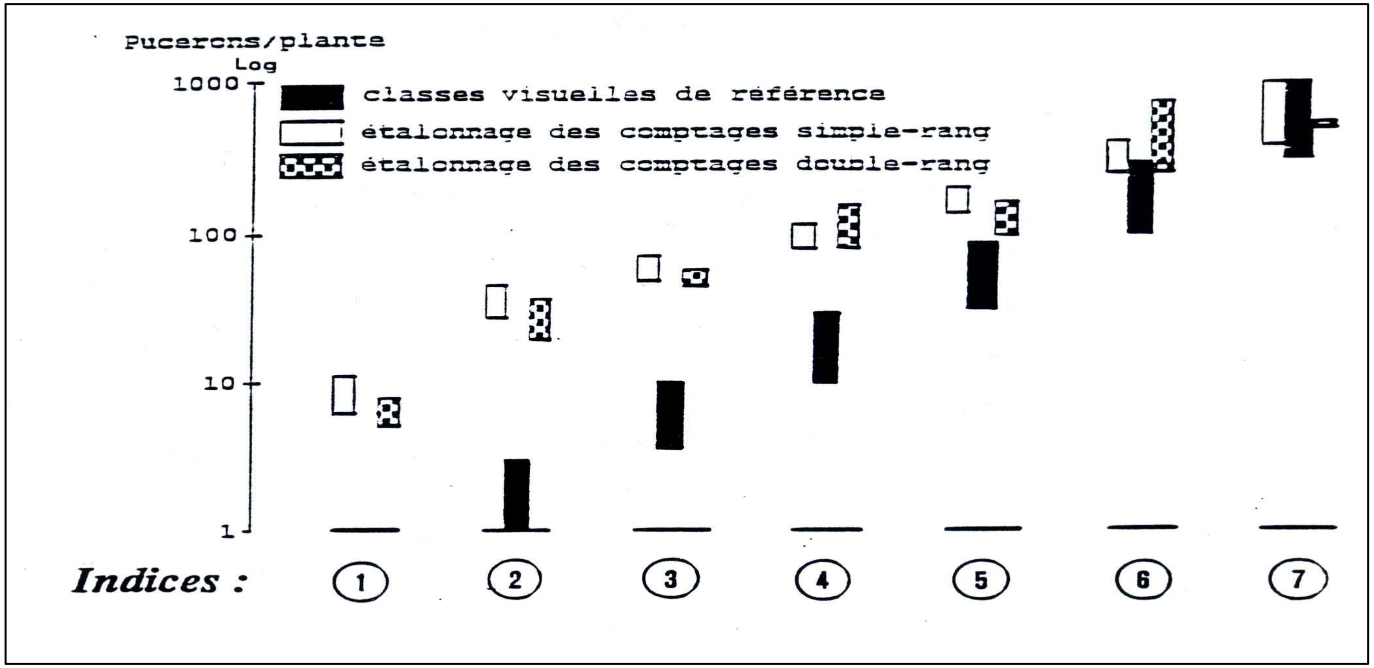
2.4. Méthodes visuelles par classe d'abondance

- classes d 'indice visuel : nul, très faible, faible, moyen, élevé
- classes pseudo-quantitatives: échelle logarithmique, puissance de 5

Problème de l 'étalonnage des classes, de la mesure de la fiabilité de la méthode et du calcul de la précision associée à l 'estimation de la population: f(degré de visibilité de l 'espèce)

Ex: Mise au point du dénombrement visuel de *Macrosiphum euphorbiae* en serre de tomates (Boll, 1991)

1: absence de pucerons	5: 31 à 100 pucerons
2: 1 à 3 pucerons	6: 101 à 300 pucerons
3: 4 à 10 pucerons	7: 301 à 1000 pucerons
4: 11 à 30 pucerons	8: 1001 à 3000 pucerons



<u>Modélisation du dénombrement</u>

Variables expliquées:

Classes d 'abondance modélisées

Variables explicatives:

classe visuelle d 'abondance du jour
classe visuelle date précédente
classe visuelle plantes voisines
dimension de la plante
niveau infestation global de la serre

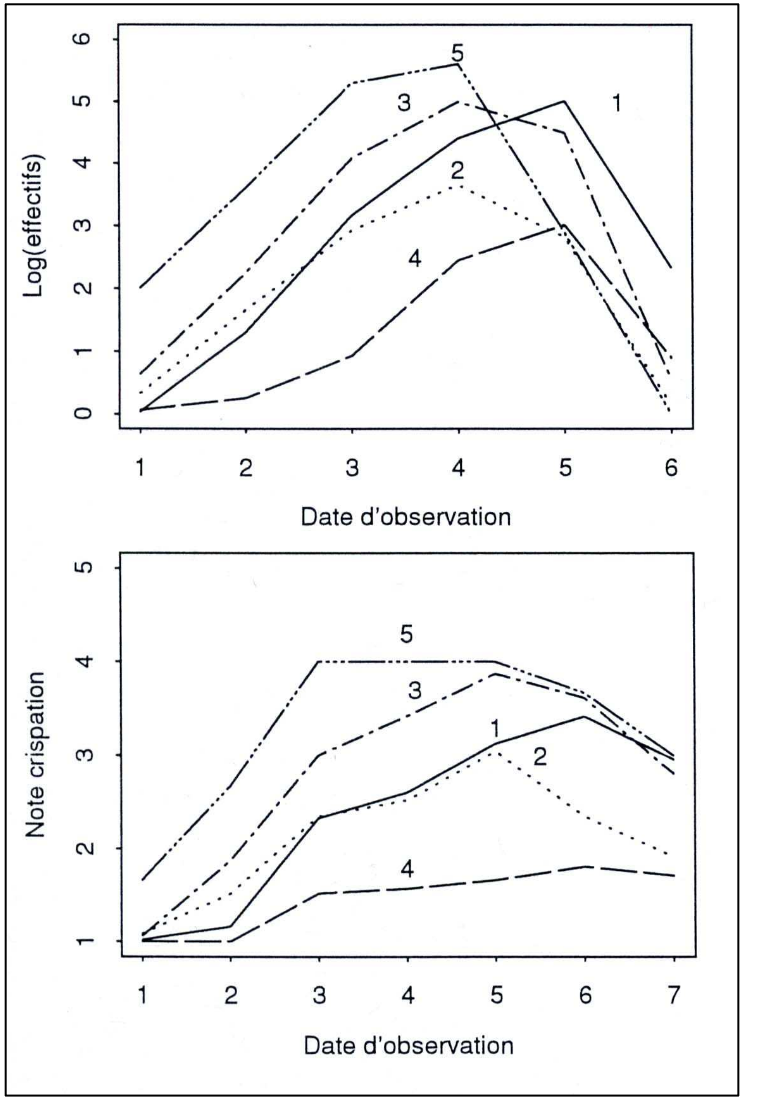
2.5. Méthodes visuelles par notation symptomatique

Exemple de symptôme = déformation du feuillage due à la toxicité de la salive de *B.helichrysi* = symptôme spécifique

Echelle selon 5 classes : proportion de feuilles parmi les 10 dernières dans un état de déformation défini: nul, qlq cloques, déformation, enroulement, redressement

Relation instantanée entre note de déformation et effectif de pucerons de qualité médiocre

Approche dynamique nécessaire:



Le suivi du symptôme en fonction du temps renseigne sur:

•la date d 'arrivée des pucerons sur la culture

•l 'évolution des populations: les symptômes s 'accroissent avec l 'accroissement des populations ou régressent avec leur extinction. Bonne corrélation entre les 2 dynamiques

•Limites: sensibilité variétale

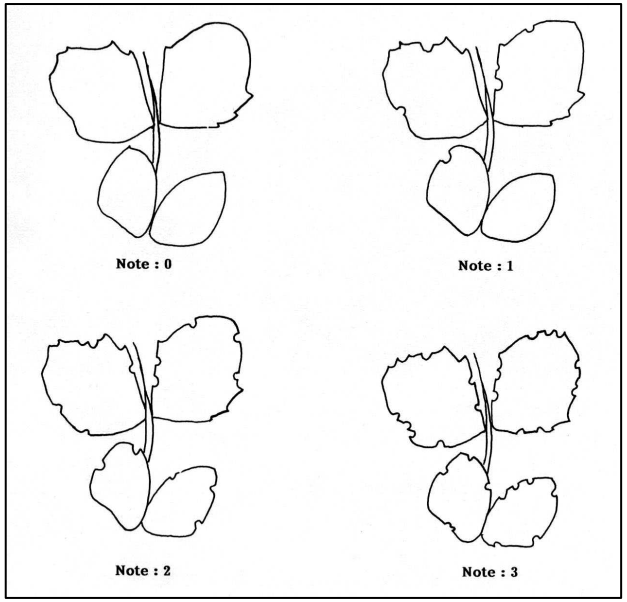
Exemple 2: Le sitone du pois

Dégâts et pertes de rendement dus à la destruction des nodosités fixatrices d'azote par les larves qui s'y développent.

Symptôme : les morsures faites par les adultes à l'origine des larves, sur les premières feuilles du pois.

Echelle selon 4 classes: porte sur l'observation des folioles de la 1ère feuille.
0: Pas d'encoche, 1:moins de 4 encoches, 2: de 4 à 10 encoches, 3: >10

Figure 1
Dégâts d'adultes au printemps.
Attack by S. lineatus *adults.*
(Photo S. Carré).



Relation entre la note et les populations larvaires:

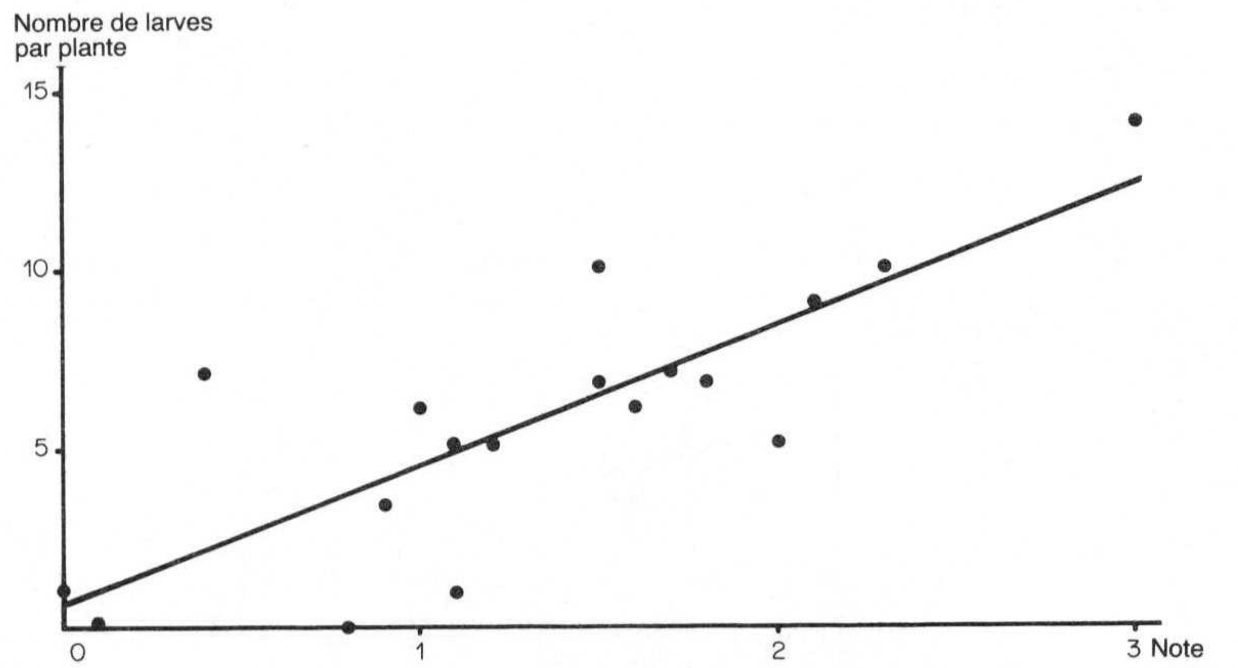


Figure 9

Relation entre la note d'attaque printanière des adultes et le nombre de larves par plante en été.

Relationship between the spring scoring and the number of larvae per plant in summer.

Problème du calcul de la précision associée à l 'estimation de la population

3. Plans d'échantillonnage classiques

- 3.1. Généralités
- 3.2. Échantillonnage non aléatoire
- 3.3. Échantillonnage aléatoire à 1 niveau
  - Simple: taille fixe ou séquentiel
  - Systématique
- 3.4. Échantillonnage aléatoire à plusieurs niveaux
  - Stratifié
  - En grappe

3.1. Généralités

Une fois la question posée, les risques choisis, la population définie, la variable choisie, on définit un plan d'échantillonnage:

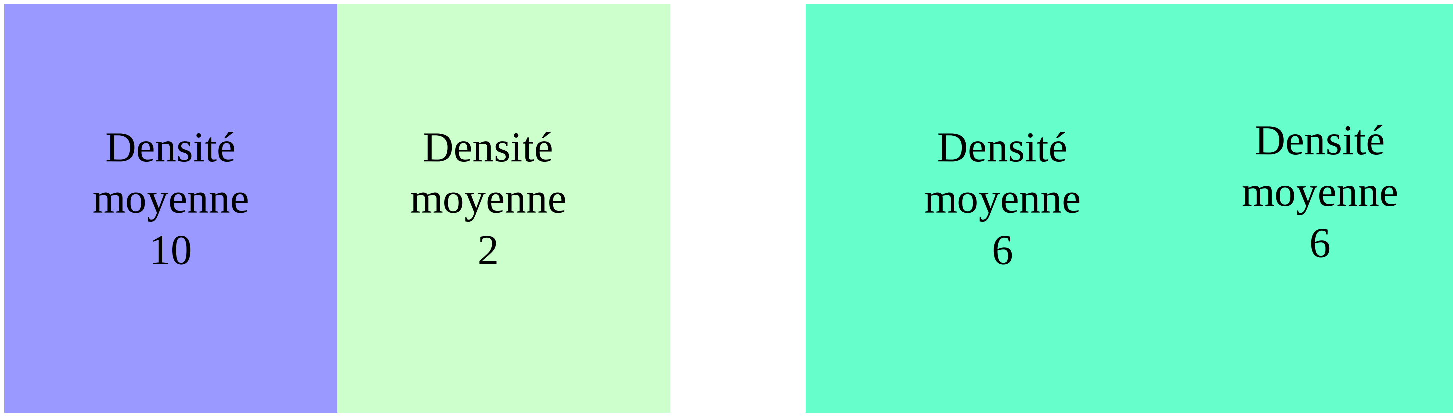
- nombre d'unités d'échantillonnage à observer ?

- comment choisir les unités d'échantillonnage (ex: au hasard, sur une ligne, au hasard dans des zones définies) ?

Etablir un plan d'échantillonnage sur une population dont on ne sait rien est impossible!

Si la population n'est pas structurée, tous les plans d'échantillonnage sont équivalents

- le nombre d’unités à observer est fonction de la variabilité entre unités, pour obtenir une précision voulue
- on peut s’éviter du travail si on montre par exemple qu’il n’y a pas de variation de densité au sein d’une parcelle, ou si la variabilité est la même entre unités voisines qu’entre les unités les plus distantes
- éviter des biais liés à des phénomènes périodiques
- s’il existe des structures spatiales nettes, par exemple des différences de densités d’infestation entre plusieurs zones d’une parcelle, on peut avoir intérêt à stratifier l’espace pour moduler par exemple des interventions.



3.2. Echantillonnage non aléatoire

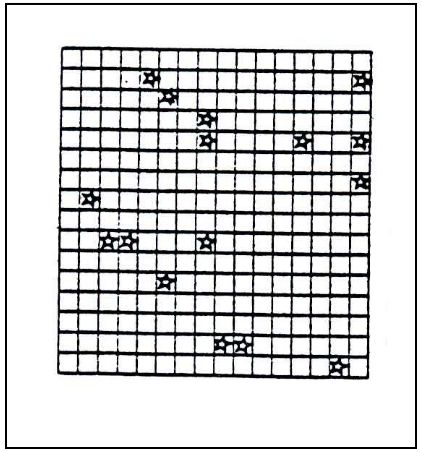
- biaisé par le choix des unités
- exhaustif = recensement complet

3.3. Echantillonnage aléatoire à un niveau

3.3.1. simple (taille fixe ou séquentiel)

passe-partout

16 unités parmi 256



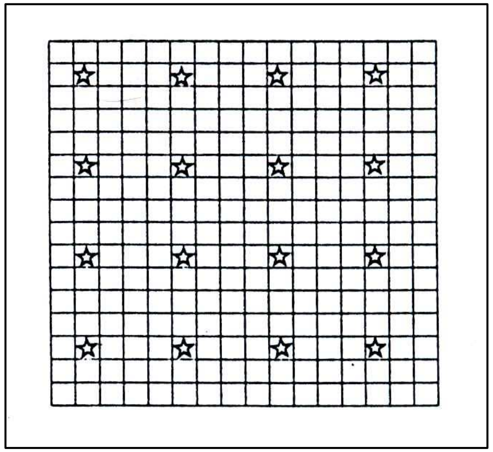
Prélèvement au hasard et indépendamment de *n* éléments dans une population qui en a *N*

•Avantages mathématiques: estimateurs non biaisés et tous tests d'hypothèses applicables, mais variance souvent plus élevée que celle d'autres plans

• mise en œuvre correcte difficile car on doit identifier tous les éléments de la population et donner à chacun la même probabilité d'être prélevé

3.3.2. Systématique

pas d'information préalable sur la population étudiée



16 unités parmi 256

1er élément au hasard puis régularité des prélèvements en une ou deux dimensions avec un pas choisi. Population statistique divisée en *n* (taille de l'échantillon) sous-ensembles de taille *t*. On a $N=nt$, et l'échantillon est constitué du *i ème* élément de chacun des sous-ensembles (*t* est par exemple une période de prélèvement, une surface)

•Bien si points autocorrélés (redondance d'informations si prélèvements voisins),

• équivalent à l'aléatoire si les éléments sont indépendants, mais moins efficace que l'aléatoire si phénomène observé périodique.

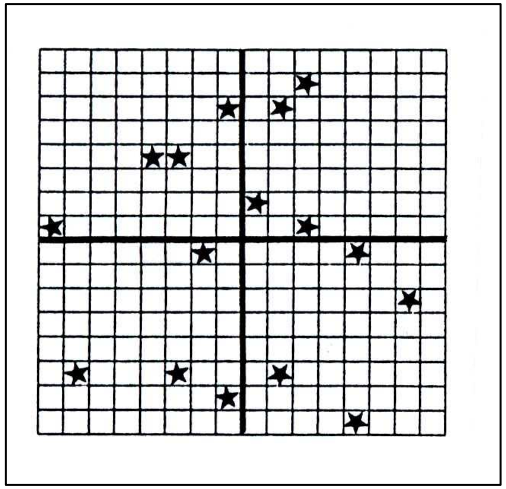
• Difficultés d'estimation de la variance des estimateurs quand la population n'est pas disposée en ordre aléatoire (sinon équivalente)

• mise en œuvre plus facile, mais nécessite d'identifier tous les éléments de la population

3.4. Echantillonnage aléatoire à plusieurs niveaux

Nécessite des informations sur la population étudiée

4 strates de 64 unités; 4 unités par strate



3.4.1. Stratifié

Subdivision de la population hétérogène en *k* strates plus homogènes, non chevauchantes et collectivement exhaustives. Somme des effectifs des strates = *N* (effectif de la population). Au sein de chaque strate on tire un échantillon aléatoire simple. 2 niveaux:

- 1er niveau = recensement et construction des strates.
  - Le critère de stratification doit minimiser la variance des estimateurs.
  - Il faut déterminer le nombre de strates (la précision s'améliore avec le nombre de strates).

- 2ème niveau = échantillonnage simple des individus dans une strate.

  - Détermination de l'effectif des échantillons : peut être variable d'une strate à l'autre. Allocation proportionnelle à *Nh*, mais ne tient pas compte du fait que la variabilité peut être différente entre les strates. Allocation dite optimale: échantillon de taille d'autant plus élevée que la variance de la strate est grande et que *Nh* est grand.

•Exemple: On veut estimer l'abondance en arthropodes d'une mare. Variable= nombre d'individus de l'espèce *i* /litre d'eau.

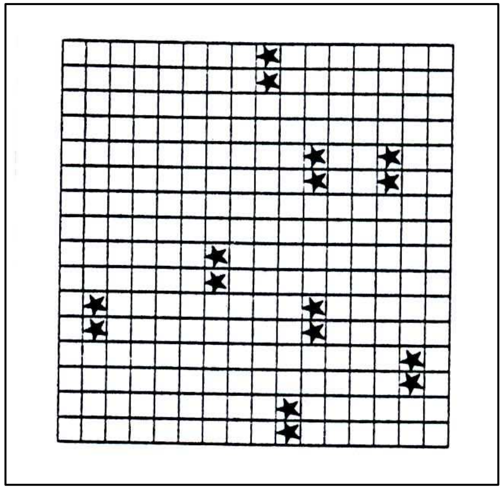
- On définit des strates correspondant à des profondeurs différentes
- On prélève *n* litres d'eau au hasard par strate
  - Estimateurs de l'abondance au sein de chaque strate
  - Estimateur de l'abondance au sein de la mare

• Plus efficace que l'aléatoire simple même quand rudimentaire

• Il faut connaître avec précision la répartition de la population (statistique) entre les différentes strates (sur l'exemple: volume de chaque strate)

3.4.2. En grappe

8 sous-populations parmi 128;
2 unités par sous-population



la population est divisée en sous populations (les grappes) et seulement quelques unes de ces sous populations sont observées exhaustivement. 2 niveaux:

- 1er niveau = <u>échantillonnage simple</u>
  - On définit les sous-populations : grappes
  - On tire certaines grappes au hasard (aléatoire simple ou systématique) parmi les *N* grappes de la population. Il faut définir le nombre de grappes à tirer.
- 2ème niveau = <u>recensement</u>
  - On examine tous les grains des grappes tirées

- Exemple:
  - on veut estimer le poids individuel des larves de processionnaires du pin dans une pinède. Variable = poids d’une larve
  - On ne peut pas inventorier toutes les larves (grains), mais on peut inventorier tous les nids (grappes)
  - On tire au hasard *n* nids, et on pèse toutes les larves des *n* nids.
- Moins efficace que aléatoire et systématique (perte de précision) quand les grains d’une grappe se ressemblent beaucoup (redondance)
- Équivalent à aléatoire et systématique si le regroupement en grappe n’a aucun rapport avec la variable étudiée
- Plus efficace si dans chaque grappe, les individus sont très dissemblables. Optimum=hétérogénéité intra-grappe aussi grande que possible
- sur le plan statistique: certains tests (anova) pas directement applicables. Difficultés pour calculer les estimateurs

Peu fréquent

4. Echantillonnage aléatoire à 1 niveau

- 4.1. Généralités
- 4.2. Echantillonnage à taille fixe
  - Cas d’une variable binomiale
  - Cas d’une variable quantitative continue
  - Cas d’une variable quantitative discrète
  - Cas particulier : échantillonnage d’une variable binomiale pour estimer une variable quantitative
- 4.3. Evaluation, simulation, validation des plans

4.1. Généralités

Hypothèse: les unités échantillonnées sont indépendantes les unes des autres

Objectif: calculer la taille d'échantillon pour atteindre une précision voulue pour l'estimateur de la moyenne

Principe: on se fonde sur la définition de l'intervalle de confiance du paramètre estimé

4.2. Echantillonnage aléatoire simple à taille fixe

4.2.1 Cas d'une variable binomiale

Distribution binomiale: Proportion p d'individus présentant l'un ou l'autre de 2 caractères opposés dans un échantillon aléatoire de n individus (proportion d'unités occupées par ex) :

- quel est l'intervalle de confiance pour α choisi (95% par ex)?

- quelle est la précision absolue, et relative sur cette proportion?

- combien d'individus récolter pour atteindre une précision choisie?

On se fonde sur la loi binomiale et ses approximations limites:

-le nombre d'individus, X, présentant l'un des caractères dans la population est une variable binomiale de moyenne $E(X)=np$ et de variance $\sigma^2(X)=npq$

-la proportion d'individus possédant le caractère dans l'échantillon est une variable aléatoire $X'=X/n$:

$$E(X/n)=E(X)/n \longrightarrow E(X/n)=p$$

$$\sigma^2(X/n)=\sigma^2(X)/n^2 \longrightarrow \sigma^2(X/n)=pq/n$$

. On doit calculer *i* de manière exacte quand *n* est petit

. Quand *n* est grand on doit utiliser des distributions approchées: la loi Normale quand np et $nq \geq 5$ (évènements non rares) ou la loi de Poisson (évènements rares)

Si n est assez grand et pour des évènements non rares, on en déduit la taille de l'échantillon n nécessaire pour atteindre une précision fixée pour l'estimateur de p :

$$i = z_{\alpha/2}\sqrt{pq/n}$$

$$\Downarrow$$

Précision	$d=i$	$D=d/p$
Calcul de n	$z^2_{\alpha/2}pq/d^2$	$z^2_{\alpha/2}q/pD^2$

Illustration :

n=50, n_A=16

1) $p=f=n_A/n$=16/50=0.32

2) Précision absolue

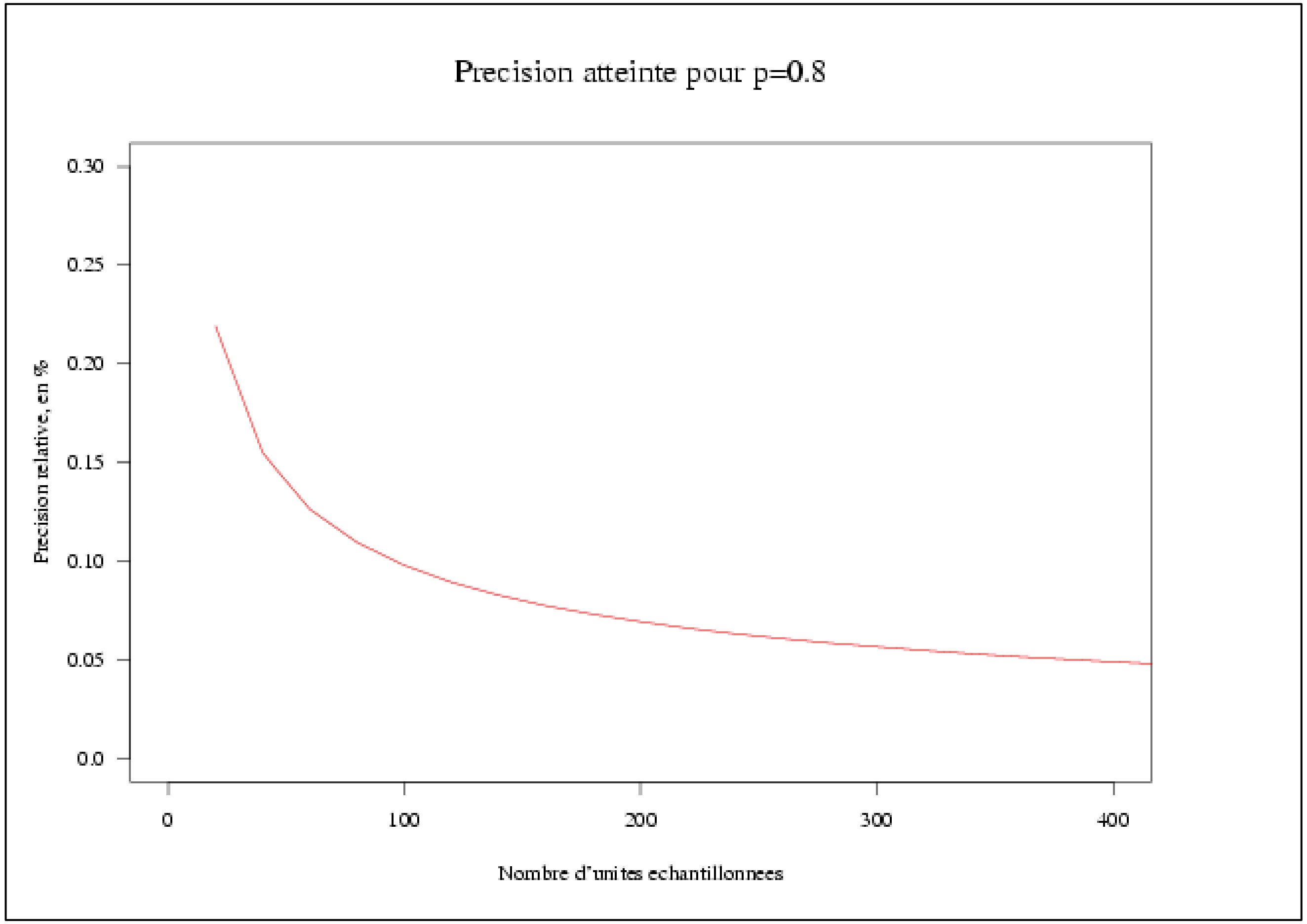
$$d = i = z_{\alpha/2}\sqrt{pq/n} = 1.96\sqrt{0.32\times 0.68/50} = 0.13$$

Précision relative D= d/p=0.13/0.32=0.40=40%

3) n' à observer pour atteindre d = 0.01 par exemple (on admet qu'on a une première estimation assez précise de la variance):

d = i=0.01 $$n' = z^2_{\alpha/2}\,pq/d^2 = 1.96^2\times 0.32\times 0.68/0.01^2 = 8359$$

On peut donc construire une courbe qui fournit pour *p* donné, la taille de l’échantillon en fonction de la précision recherchée



4.2.2. Cas d’une variable quantitative continue

Ex: Poids des larves de 4ème stade chez *Calliptamus italicus* sur la base d’un échantillonnage aléatoire de n larves: m , σ^2 estimés

-quel est l’intervalle de confiance de m pour α choisi (95% par ex) ?

-quelle est la précision absolue, et relative sur cette moyenne?

-combien d’individus récolter pour atteindre une précision choisie sur le poids moyen?

En général la variable est distribuée Normalement et donc sa variance est indépendante de sa moyenne.

Pour n >30, la moyenne observée m suit une loi Normale

Pour n<30, $m \sigma / \sqrt{n-1}$ suit un Student à n-*1* ddl

On est capable de calculer la taille de l’échantillon *n* nécessaire pour atteindre une précision fixée pour l’estimateur de la moyenne:

$$i = z_{1-\alpha/2}\sigma / \sqrt{n}$$

$$\Downarrow$$

Précision	$d=i$	$D=d/m$
Calcul de *n*	$z^2_{1-\alpha/2}\sigma^2 / d^2$	$z^2_{1-\alpha/2}\dfrac{\sigma^2}{D^2 m^2}$

Illustration:

$$m = 0.485 \quad \sigma^2 = 0.08 \quad n = 35$$

1) $$i = z_{\alpha/2}\sigma / \sqrt{n} = 1.96 \times 0.28 / 5.92 = 0.09$$

2) Précision absolue d = i=0.09,

Précision relative D= d/m=0.09/0.485=0.19=19%

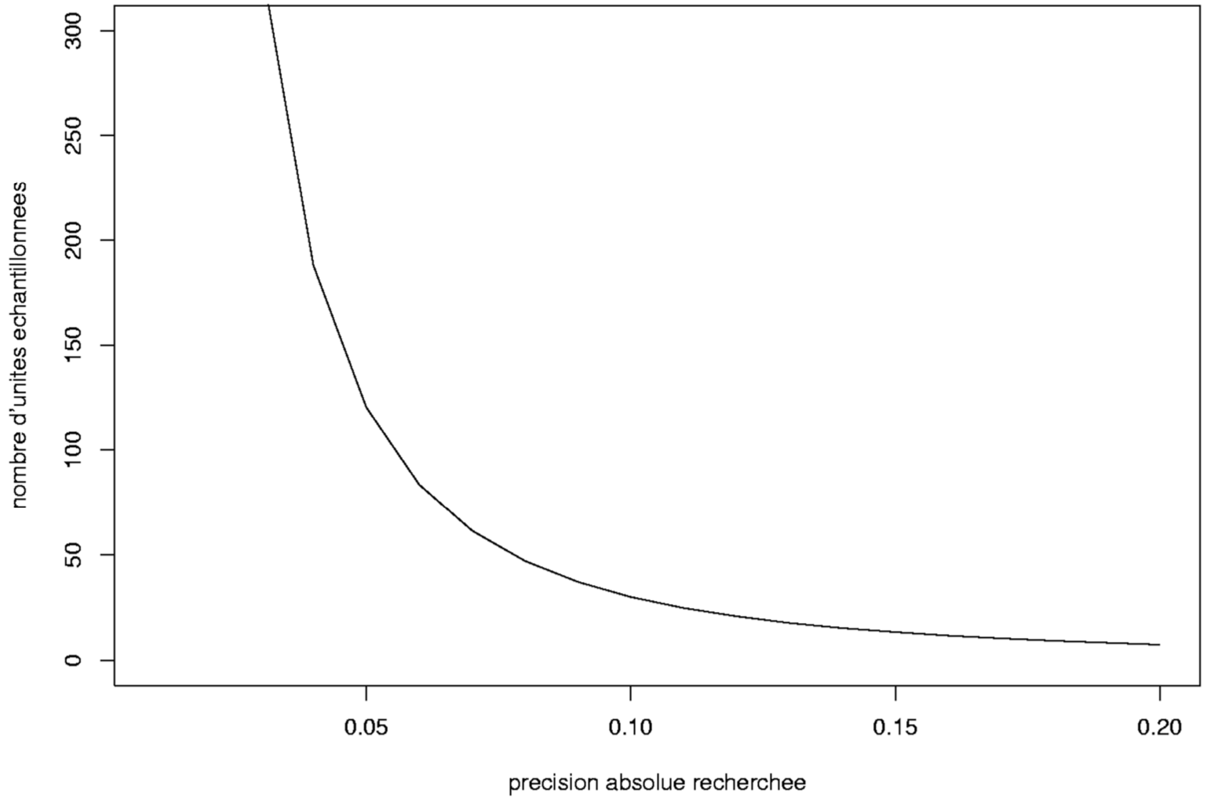
3) n' à observer pour atteindre D = 10% par exemple:

on pose D = d / m = i / m = 0.10

$$n' = \left(\frac{1.96\sigma}{\overline{m}D} \right)^2 = (1.96 \times 0.28 / 0.485 \times 0.1)^2 = 130$$

n augmente très vite quand D s'améliore

On peut construire une courbe qui fournit la taille de l'échantillon en fonction de la précision recherchée, σ^2 supposé constant = 0.08 quel que soit le poids moyen des larves



4.2.3. Cas d’une variable quantitative discrète

C’est les dénombrements.

En général phénomène d’hétéroscédasticité : la variance n’est pas indépendante de la moyenne. Il faut pouvoir calculer la variance …

Il s’ensuit que la taille de l’échantillon dépend de la moyenne

Quels sont les outils pour modéliser la variance en fonction de la moyenne ?

-loi de distribution des effectifs par unité d'échantillonnage et estimer ses paramètres pour calculer la variance.

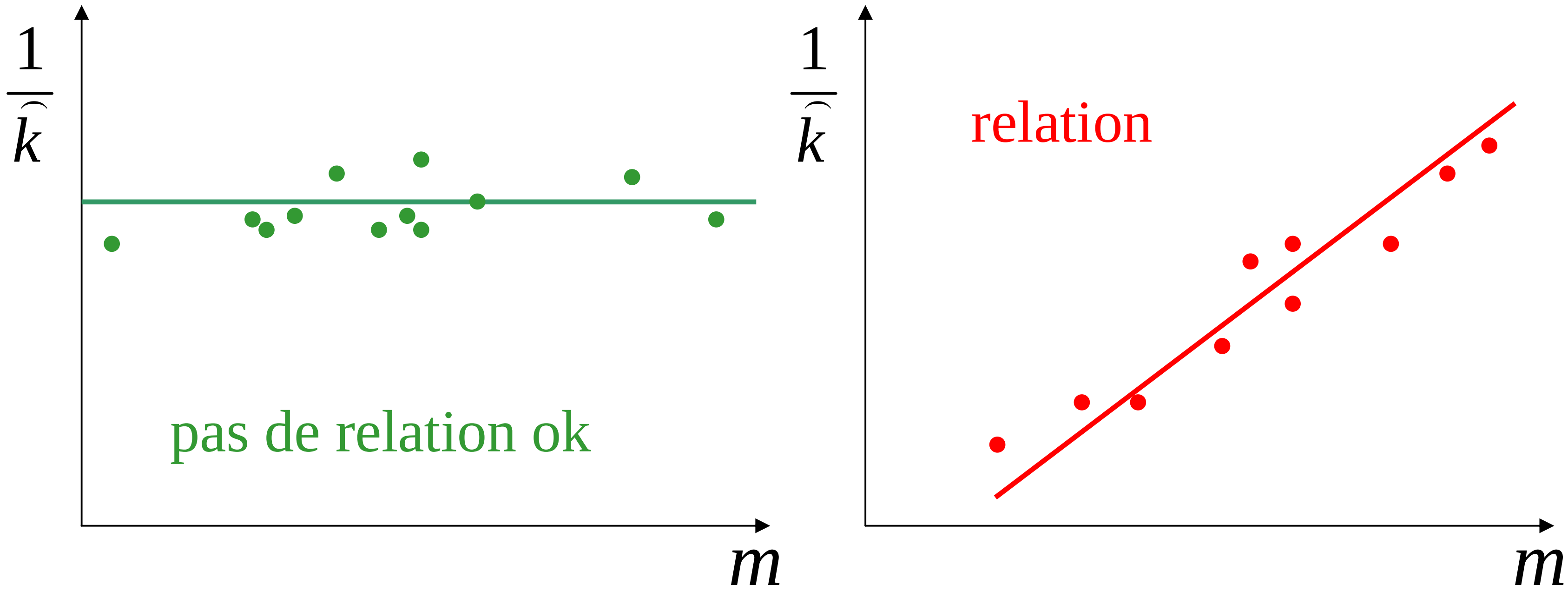
- Le nombre de lois possibles est alors important
- il s'agit de lois de probabilité discrètes

-relation empirique entre la moyenne et la variance

Lois de probabilité discrètes

- Loi de Poisson $\mathrm{var}(X) = m$
- Loi binomiale négative $\mathrm{var}(X) = m + \frac{m^2}{k}$

Il faut estimer k et vérifier qu'il est stable en fonction de la taille de l'échantillon, de l'UE et de la moyenne



Relations empiriques

- Relation de Taylor (1961)

$$\log \sigma_i^2 = \log a + b \log m_i$$

Où σ_i^2 est la variance de l 'échantillon *i* pour la v.a *X*: nombre d 'individus par Unité d 'échantillonnage, et m_i la moyenne de l 'échantillon *i* pour *X*

a est lié à l 'échantillonnage,

b est spécifique et indépendant de la moyenne, de la taille de l 'échantillon, mais pas de l '*UE*

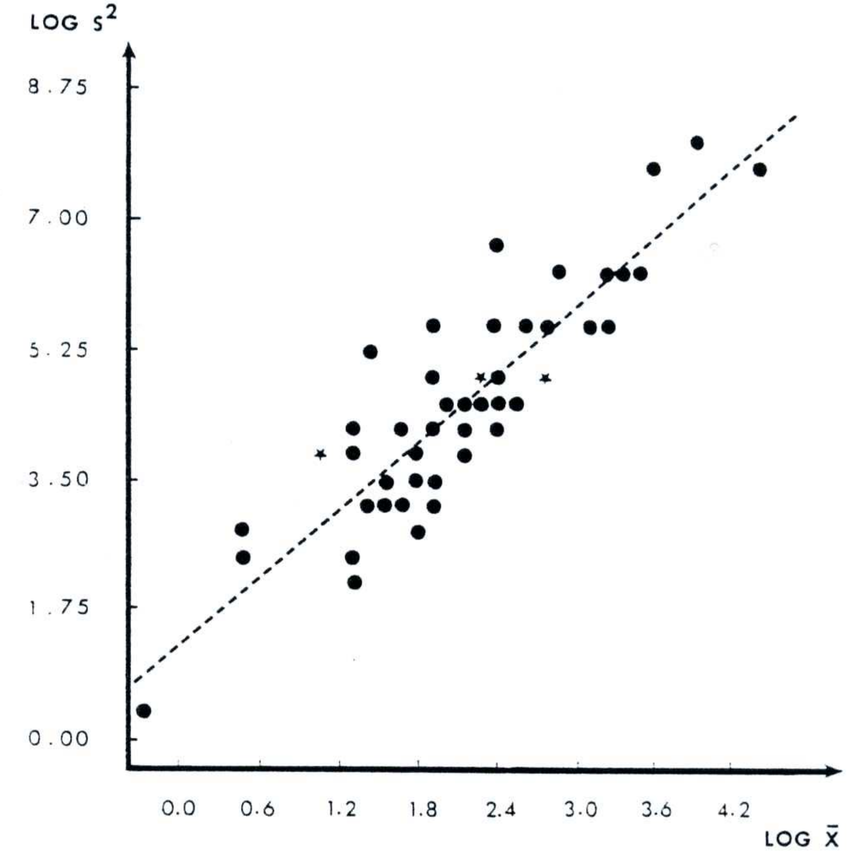
Relation robuste souvent stable

- Description de la variabilité d 'infestation entre UE : ex de la relation moyenne-variance de Taylor:

$$\sigma^2 = a\, m^{\,b}$$

Pour *A.pisum* et le pois (UE=une tige) :

a=4.05 b=1.47 R2=0.89
df=50 (52 jeux de données)



Représentation de la relation moyenne-variance (en Log) dans le cas des sondages des parcelles d'enquêtes du puceron du pois.

Calcul de la taille d'échantillon

Taille d'échantillon, *n*, pour une précision fixée à l'avance :

- Précision absolue *d* :

$$i = z_{1-\alpha/2}\sigma/\sqrt{n} = d \quad \Rightarrow \quad n = \frac{z^2_{1-\alpha/2}\sigma^2}{d^2}$$

On remplace σ^2 par son expression avec les paramètres estimés:

	Poisson	Binomiale négative	Taylor
n	$\frac{z^2_{1-\alpha/2}m}{d^2}$	$\frac{z^2_{1-\alpha/2}m + \frac{m^2}{\widehat{k}}}{d^2}$	$\frac{z^2_{1-\alpha/2}am^b}{d^2}$

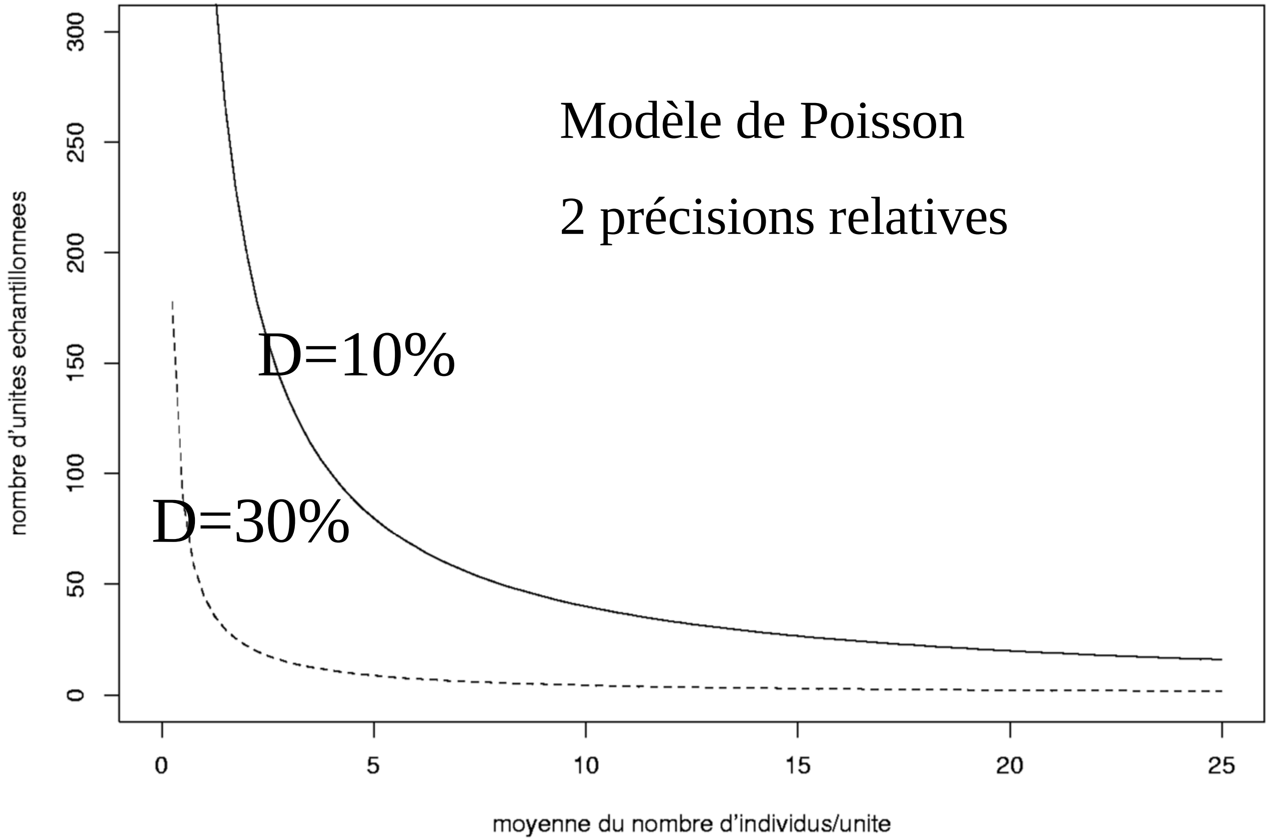
- Précision relative *D*:

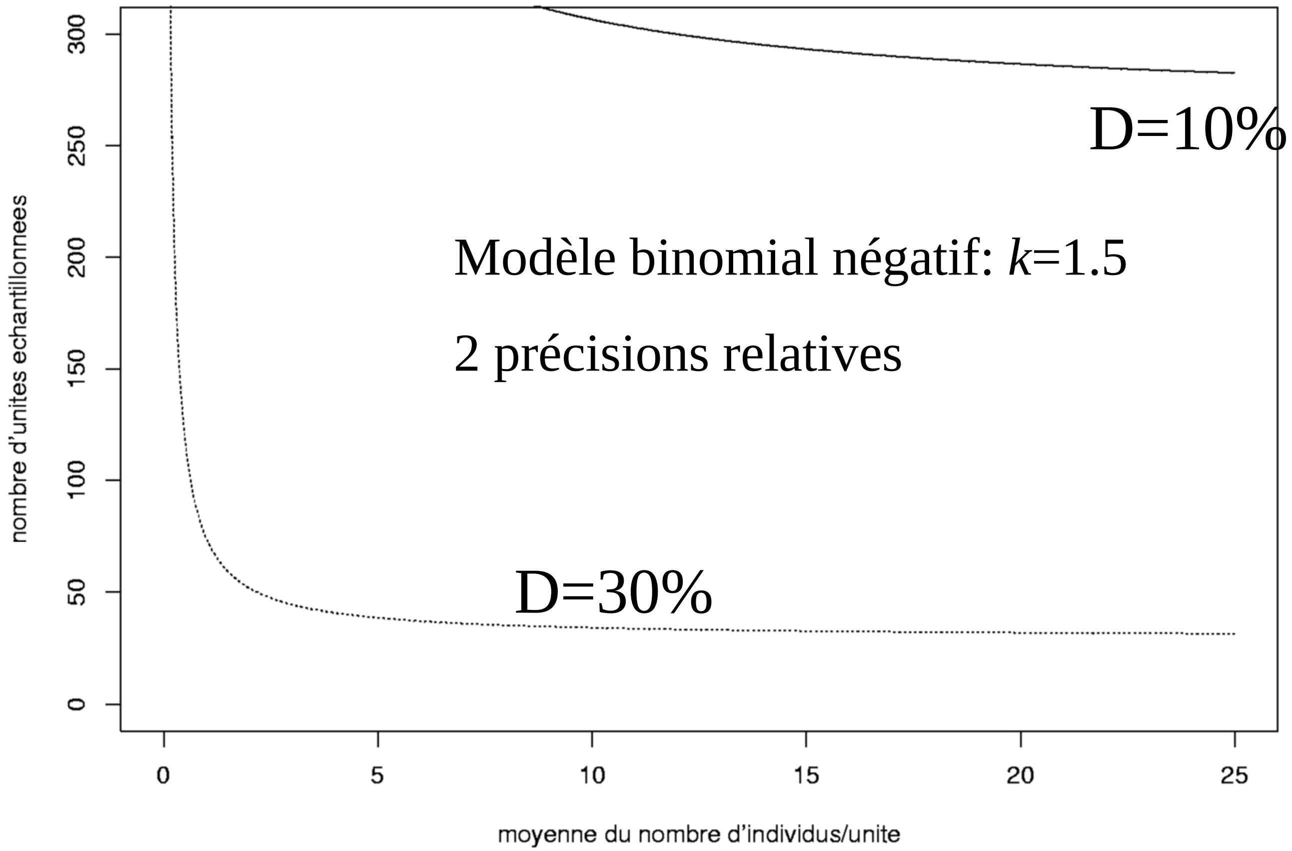
$$D = d / m \Rightarrow n = z_{1-\alpha/2}^{2} \frac{\sigma^{2}}{D^{2} m^{2}}$$

On remplace σ^2 par son expression avec les paramètres estimés:

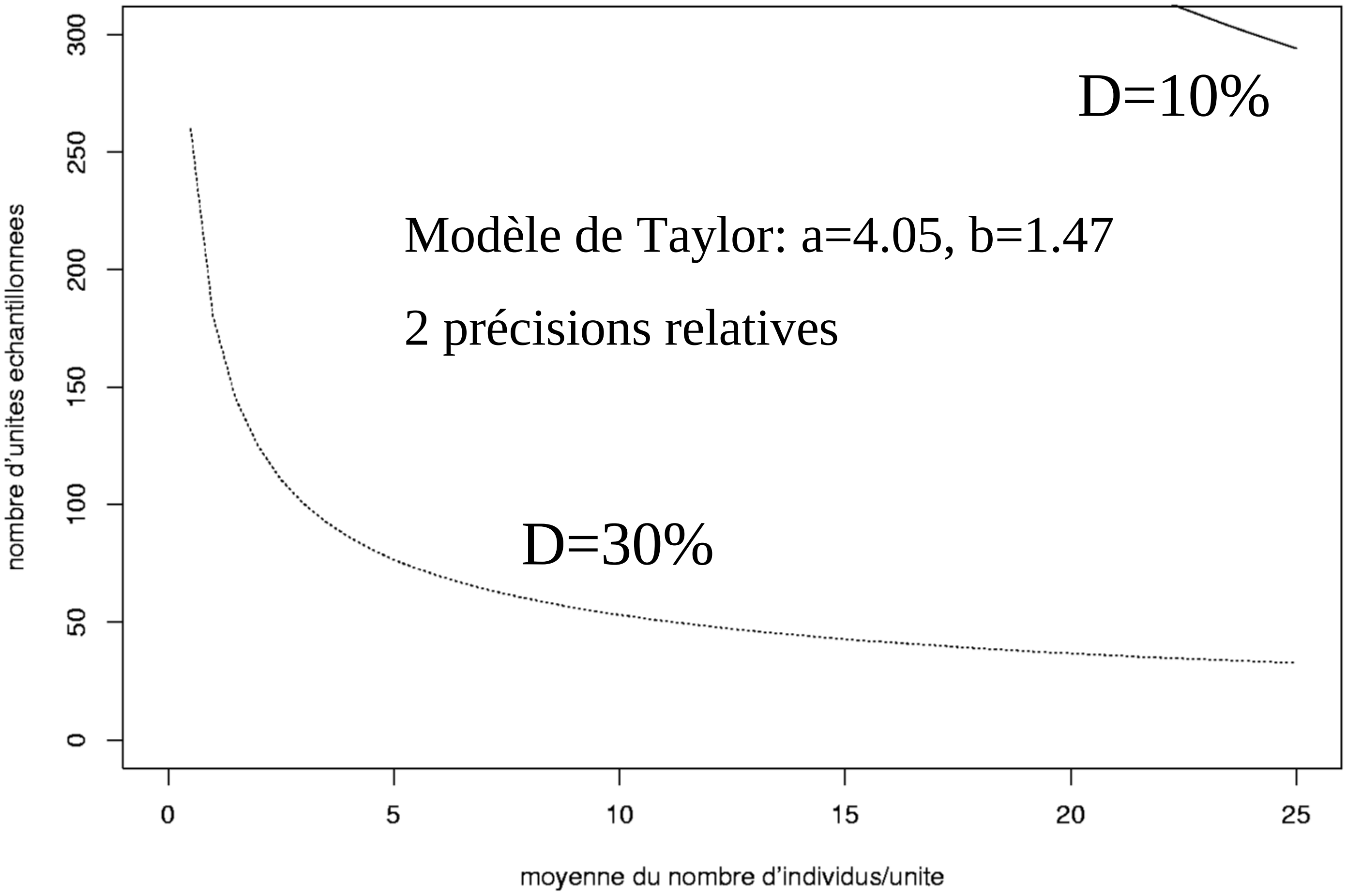
	Poisson	Binomiale négative	Taylor
n	$z_{1-\alpha/2}^{2} \frac{1}{mD^{2}}$	$z_{1-\alpha/2}^{2} \frac{1+m/k}{mD^{2}}$	$z_{1-\alpha/2}^{2} \frac{a}{D^{2}} m^{(b-2)}$

Plan d’échantillonnage pour estimer la moyenne du nombre d’individus par unité d’échantillonnage





D=10%
Modèle binomial négatif: k=1.5
2 précisions relatives
D=30%
nombre d'unites echantillonnees
moyenne du nombre d'individus/unite
0
50
100
150
200
250
300
0
5
10
15
20
25


D=10%
Modèle de Taylor: a=4.05, b=1.47
2 précisions relatives
D=30%
nombre d'unites echantillonnees
moyenne du nombre d'individus/unite
0
50
100
150
200
250
300
0
5
10
15
20
25

4.2.4. Cas particulier : échantillonnage d'une variable binomiale pour estimer une variable quantitative (modèle de Gerrard et Chiang)

Proportion P_T d'unités occupées parmi les n unités observées: moyenne m estimée par la relation de Gerrard et Chiang:

$$\ln m = \alpha + \beta \ln\left[-\ln(1 - P_T)\right]$$

- quel est l'intervalle de confiance pour α choisi (95% par ex)?
- quelle est la précision absolue, et relative sur cette moyenne?
- combien d'unités observer pour atteindre une précision choisie pour estimer m ?

1) $i = z_{1-\alpha/2}\sigma / \sqrt{n} = d$

2) $D = d / m$

$\longrightarrow$ 3) $n = z^2_{1-\alpha/2} \dfrac{\sigma^2}{D^2 m^2}$

Il suffit de disposer d'un estimateur de σ

Pour la relation de Gerrard et Chiang, on a un estimateur de la variance:

$$\sigma_m^2 = \frac{\sigma^2}{n} = m^2(c_1 + c_2 + MSE - c_3)$$

c_1, c_2, c_3, MSE expression complexe, liée aux caractéristiques de la régression. Comporte un terme avec n

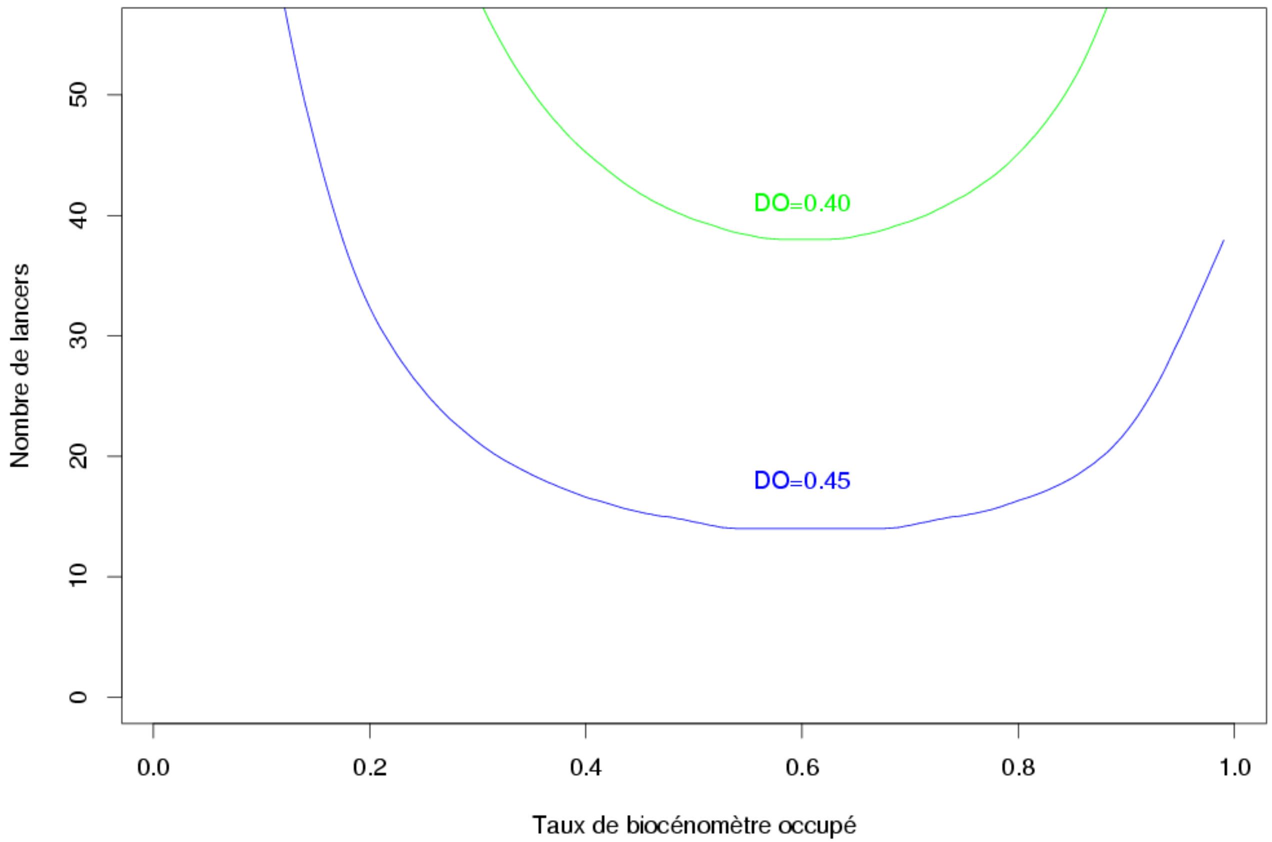
Si on fixe la précision recherchée (relative par exemple) à une valeur voulue D:

$$D = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \frac{1}{m} = z_{1-\alpha/2} (c_1 + c_2 + MSE - c_3)^{0.5}$$

On peut en déduire: 1) d 2) D 3) n

Plan d'échantillonnage:

n



P_T

Un échantillon de n unités permet d'estimer m par P_T avec la précision D_0

4.3. Evaluation, simulation et validation des procédures d'échantillonnage

Tout plan nécessite d 'être évalué et validé soit sur la base d 'observations (exhaustives ou non) sur des parcelles n 'ayant pas servi à l 'élaboration du modèle de répartition ou de variabilité, soit sur des observations simulées bootstrap

Les techniques de simulation permettent alors de calculer des caractéristiques des plans d 'échantillonnage conçus: moyenne estimée, précision atteinte, taille d 'échantillon

Caractéristiques des plans d’échantillonnage

- Pour évaluer l’effort d’échantillonnage d’un plan donné:

- Taille moyenne des échantillons observés en fonction de la moyenne d’infestation de la population

- On peut calculer d’autres caractéristiques des plans d’échantillonnage:

- La précision réellement obtenue pour l’estimation

- les biais d’estimation

- le coût de l’échantillonnage si une fonction de coût existe

- Différentes stratégies d’échantillonnage peuvent être comparées sur la base de ces indicateurs calculés ou simulés

Simulation et évaluation des qualités d'un plan d'échantillonnage basé sur les dénombrements

Principe:

-Un jeux de données qui n'a pas servi pour établir le plan d'échantillonnage

-On crée à partir de ce jeu de données un très grand jeu de données par simulation bootstrap, dont les propriétés sont supposées être celles de la population

-On calcule pour ce grand jeux de données la « vraie » valeur de l'estimateur:

$$\mu$$

-On simule l'échantillonnage sur ce grand jeux de données

-On calcule des caractéristiques à partir des résultats de la simulation:

-Estimateur obtenu: m_1

-Précision observée de l'estimateur: $d_{1obs} = \quad i = z_{1-\alpha/2}\sigma_1/\sqrt{n_1}$ ou $D_{1obs} = d_{1obs}/\ m_1$

-Coût

-On compare ces caractéristiques à: μ et D ou d

Un exemple :

A=jeux de données de 90 unités d'échantillonnage observées:

0 1 0 5 11 0 1 2 0 0 1 2 0 0 1 2 0 0 0 0 0 4 3 0 0 1 2 0 4 1 1 0 0 0 1 1 2 0 1 0 1 2 5
0 1 0 0 0 1 2 0 1 0 0 0 1 0 1 1 3 0 0 7 0 1 2 5 0 0 0 1 2 1 0 1 0 1 0 1 1 1 5 2 3 1 0
4 0 0 1

Abis=jeu de données issu de A par simulation bootstrap de 200 unités d'échantillonnage:

0 1 1 0 0 0 0 0 5 0 0 1 1 1 0 2 0 0 2 0 1 0 0 0 2 1 0 0 1 0 0 0 1 0 0 0 3 4 0 2 0 1 0
1 0 0 1 0 0 0 0 0 0 0 1 0 0 0 1 5 4 0 0 1 0 0 2 0 4 0 1 5 5 4 0 0 2 1 0 0 0 1 5 1 1 2
2 0 2 0 4 1 1 4 2 2 5 0 1 5 0 5 0 2 0 0 7 0 0 1 4 5 1 0 0 1 1 1 2 2 11 0 0 1 2 4 2 1 0
0 1 1 0 0 1 1 1 0 1 1 0 0 11 2 0 2 0 0 1 7 0 0 2 2 1 0 0 0 0 1 0 0 0 5 1 0 1 0 0 0 0 5
1 0 1 0 1 0 0 0 2 0 0 1 2 0 4 2 1 0 2 0 11 0 0 0 2 1 0 1

On simule l'échantillonnage sur ce grand jeux de données:

1) on prélève parmi Abis quelques unités (5 par ex) au hasard pour obtenir un ordre de grandeur de la moyenne à estimer

Par exemple: simulation1: 1 1 0 2 4 moyenne estimée = 1.6

2) on regarde d'après cette première approximation le nombre d'unités supplémentaires à observer pour D choisie

Pour D=30%, n à observer vaut 55 pour simulation1

3) On prélève au hasard parmi Abis le nombre d'unités à observer en plus: on obtient donc pour cette simulation un échantillon de taille *n1*

simulation1: 1 1 0 2 4 1 0 1 2 5 4 0 2 5 1 1 0 2 0 0 0 4 2 0 2 1 2 1 0 0 5 0 11 0 0 1 1 0 0 2 2 0 0 0 1 2 0 0 1 0 1 0 0 11 0

4) On calcule les caractéristiques de cet échantillon

m(simulation1)=1.49

σ(simulation1)=2.34

n1=55

Précision observée de l'estimateur m: $dobs = 1.96 \times 2.34/\sqrt{55} = 0.62$

$Dobs = 0.62/1.49 = 0.41$

Coût si on a une fonction de coût

5) On compare ces caractéristiques à: μ *de Abis* et à la précision qu'on s'était fixée *D* ou *d*

μ(Abis) = 1.20

D fixée = 30%=0.30

Conclusion pour cette simulation: la précision fixée n'est pas atteinte

6) On fait beaucoup de simulations (100 par exemple) pour ce jeux de données Abis

Simulation1, simulation2,……….simulation100

m(simulation1),…………………..m(simulation100) → distribution des moyennes estimées

n(simulation1)……………………n(simulation100) → distribution des tailles d'échantillon

D_{obs}(simulation1),………………D_{obs}(simulation100) → distribution des précisions atteintes

Il faut plusieurs jeux de données de validation pour couvrir une gamme étendue de moyennes à estimer

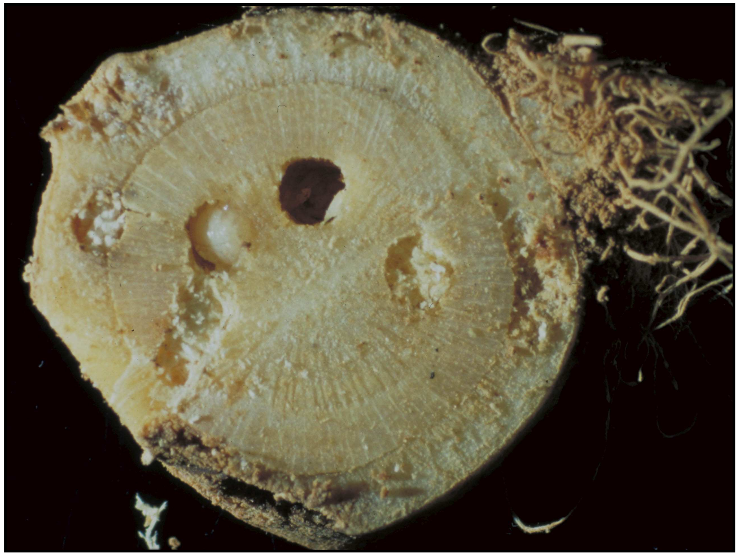
5. Un exemple concret

- 5.1. Les données dont on dispose
- 5.2. Construction d'un plan d'échantillonnage basé sur une variable de dénombrement
  - Modèle de description de la variabilité
  - Échantillonnage du nombre d'insectes/pivot
  - Simulation et évaluation des qualités de ce plan
- 5.3. Construction d'un plan d'échantillonnage basé sur une variable binomiale
  - Relation entre la moyenne et la proportion de pivots infestés avec au moins T insectes
  - Echantillonnage de la proportion de pivots infestés avec au moins 2 larves de baris
  - Simulation et évaluation des qualités de ce plan binomial

Le Baris du colza:

✓charançon endophyte à l’état larvaire

✓ il faut disséquer la tige pour trouver les larves



On veut mettre au point une procédure d’échantillonnage pour estimer le nombre moyen de larves par tige pour:

-étudier la dynamique de ses populations

-tester l’efficacité de stratégies d’intervention

On recherche une précision assez bonne : environ 20% à 30%

5.1. Les données

- ✓ 91 jeux de données d'échantillons de 20 à 93 plantes chacun (différents cv de colza, différentes années, différentes régions).
- ✓ Variable = nombre d'œufs + larves par plante (pivot).

-Quelle allure a la relation entre la moyenne et la variance d'échantillon ?

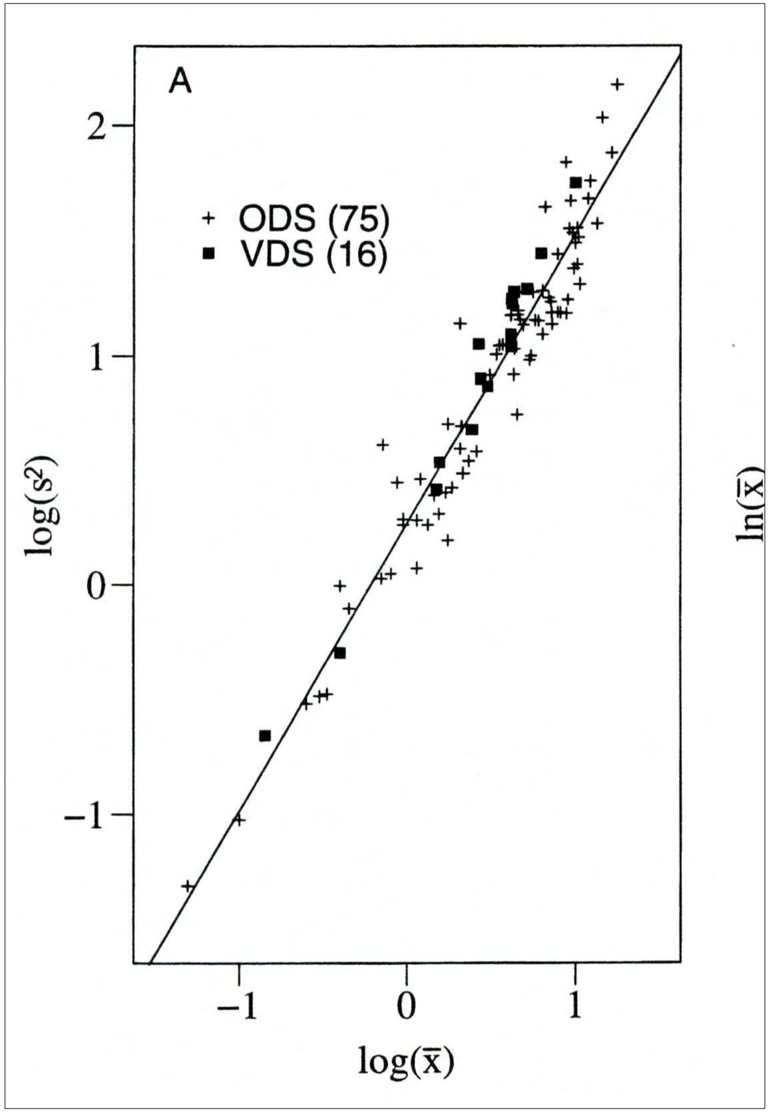
-Existe-t-il une relation entre la proportion de pivots infestés et la moyenne d'infestation ?

5.2. Construction d’un plan d’échantillonnage basé sur une variable de dénombrement

- Analyse de la relation entre la moyenne et la variance d’échantillon

- Quel plan d’échantillonnage ? ses caractéristiques ? Sa compatibilité avec les moyens dont on dispose?

5.2.1. Modèle de description de la variabilité



Variable = nombre d'œufs + larves par plante (pivot).

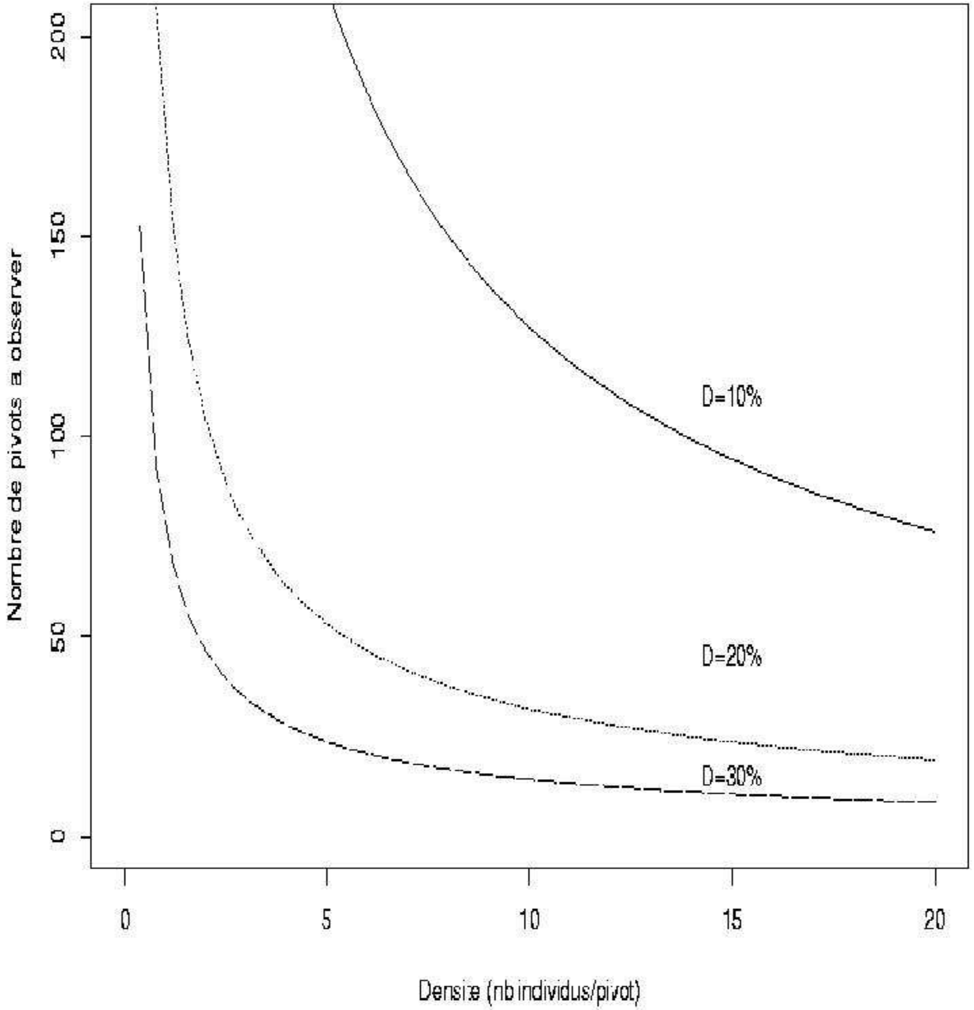
Parmi les 93 jeux de données, on en garde 16 qui serviront pour simuler les plans qu'on va proposer

Modèle testé et retenu = Relation moyenne-variance de Taylor:

$$\sigma^2 = am^b$$

a=1.82 b=1.26 R2=0.93 df=73

5.2.2. Echantillonnage du nombre d'individus par pivot



$$n = z^2_{1-\alpha/2}\frac{a}{D^2}m^{(b-2)}$$

Avec a=1.82 b=1.26

5.2.3. Simulation de ce plan d'échantillonnage

On dispose de :

- 16 jeux de données de validation de 30 à 90 plantes chacun .
- Variable disponible: nombre de Baris/pivot.
- Gamme de moyennes: 0.14 à 10.01

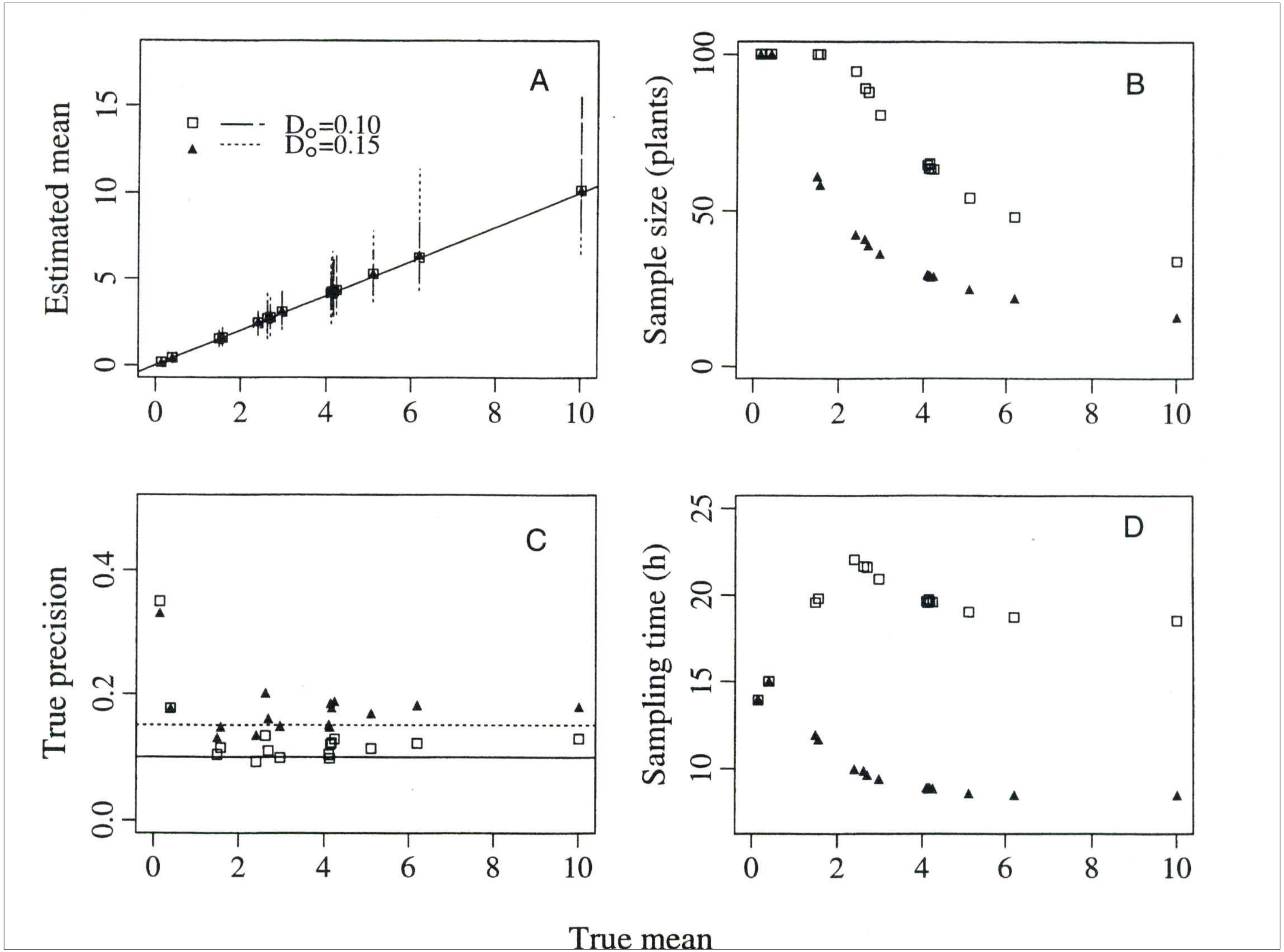
- simulation pour D=0.10 et 0.15:

  ✓ à partir de chaque jeux de données, construction par simulation d'un échantillon de 200 plantes

  ✓ pour chaque jeux de données ainsi simulé, 100 simulations de la procédure d'échantillonnage proposée:

  ✓Moyennes sur les 100 simulations de : la moyenne estimée, de la taille d'échantillon, de la précision vraiment atteinte.

  Remarque: on s'est fixé une taille d'échantillon minimale de 10 pivots et une taille d'échantillon maximale de 100 pivots.


A
B
C
D
D_o=0.10
D_o=0.15
Estimated mean
Sample size (plants)
True precision
Sampling time (h)
True mean

5.3. Construction d’un plan d’échantillonnage basé sur une variable binomiale

- Existe-t-il une relation entre la moyenne et le taux de pivots infestés ?

- Quel plan d’échantillonnage ? ses caractéristiques ? Sa compatibilité avec les moyens dont on dispose?

5.3.1. Relation entre la moyenne et la proportion de pivots infestés

75 jeux de données d'échantillons de 20 à 93 plantes chacun

Variable = nombre d'œufs + larves par plante (pivot).
On en déduit la variable: présence d'au moins 1 baris par pivot

Modèle testé = Relation de Gerrard et Chiang (tels que P ≠ 1 et P ≠ 0)

$$\ln m = \alpha + \beta \ln\left[-\ln(1-P)\right]$$

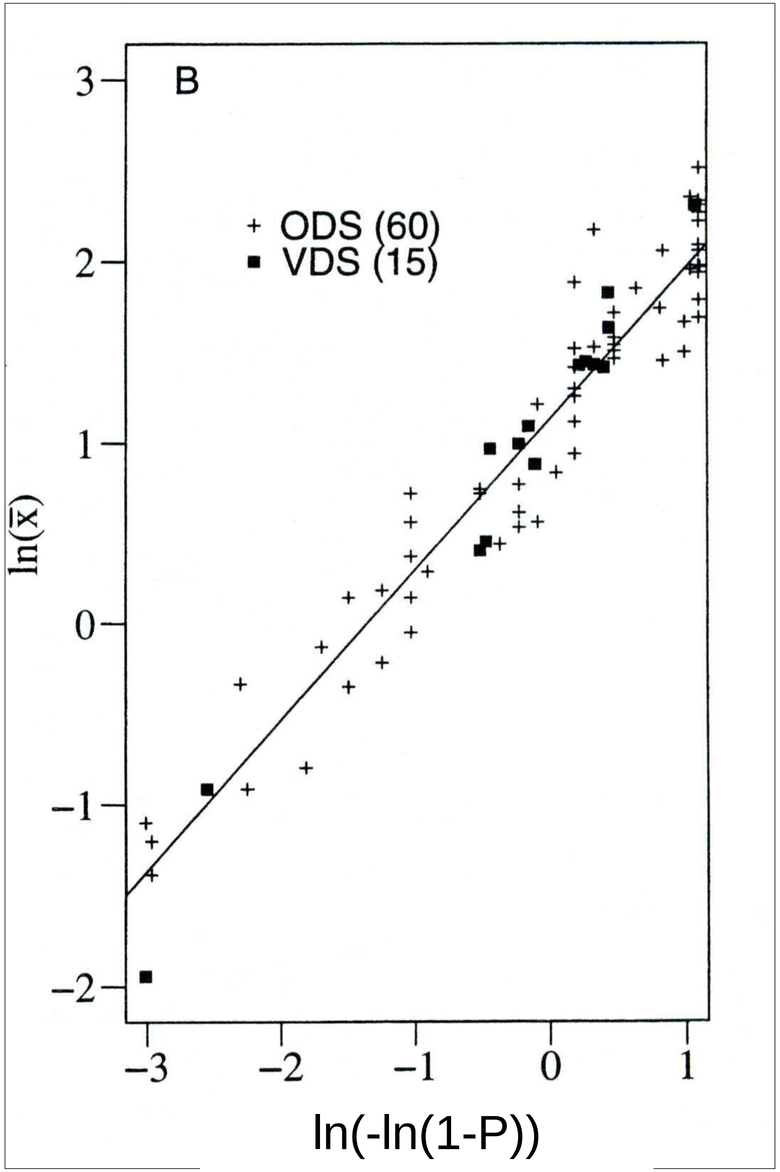
avec $$\frac{\sigma^2}{n} = m^2(c_1 + c_2 + MSE - c_3)$$

Calcul de la précision relative d'une estimation de m par P?

$$D = \frac{\sigma}{\sqrt{n}}\frac{1}{m} \longrightarrow D = (c_1 + c_2 + MSE - c_3)^{0.5}$$

c_1, c_2, c_3, MSE se calculent en fonction des paramètres estimés de la relation (α et β), de la qualité de l'ajustement au modèle, de n et P

Si on considère la proportion de pivots occupés par au moins 1 individu:
On obtient après ajustement linéaire:



α=1.11 β=0.85 R2=0.94
df=73

Précision la meilleure qu'on peut atteindre:

D= 0.25 c'est raisonnable

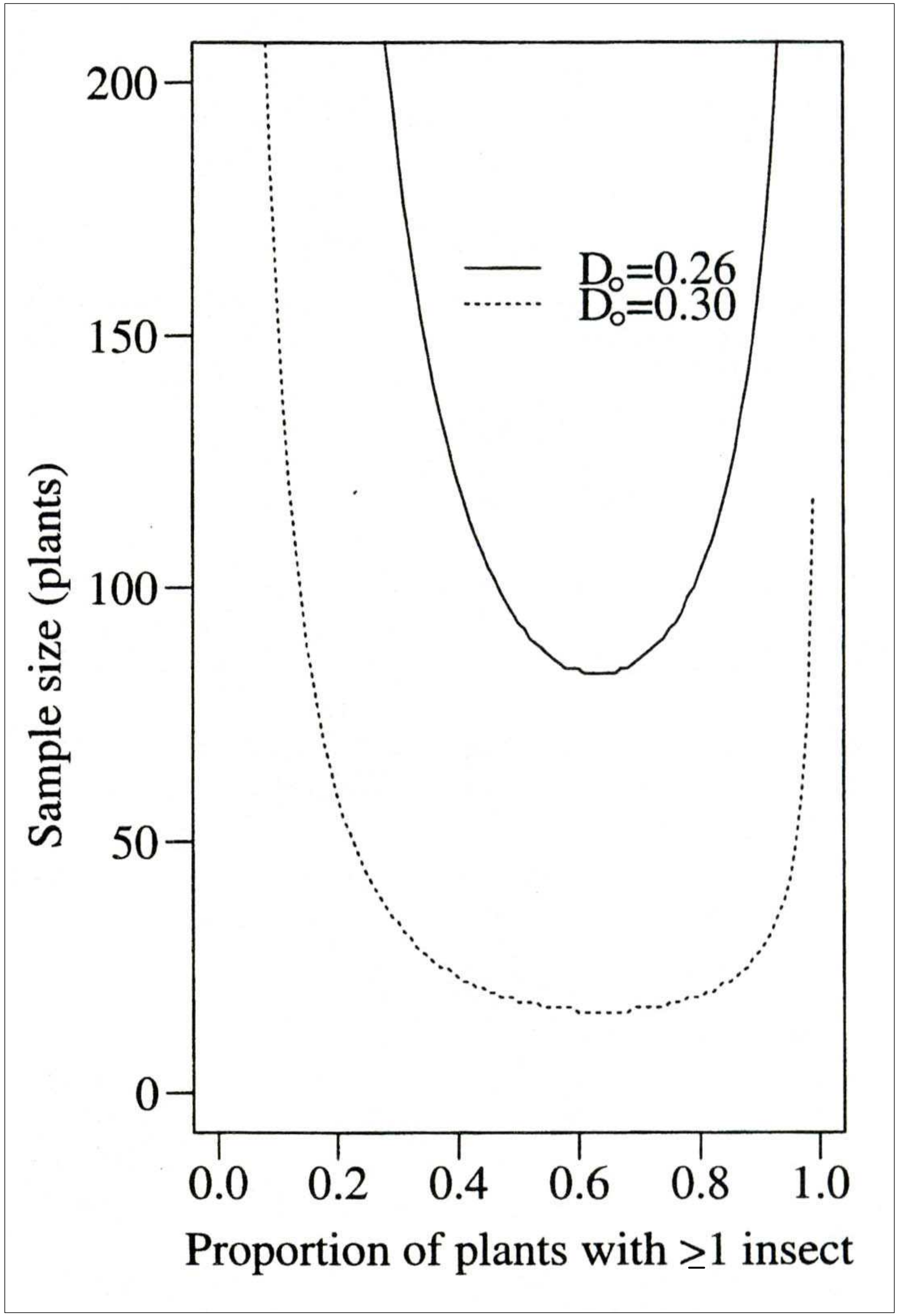
5.3.2. Echantillonnage de la proportion de pivots infestés

Compte tenu du modèle employé et de ses caractéristiques, on ne peut avoir une meilleure précision que 0.25

Les relations permettant de calculer la précision de l’estimation font intervenir *n :*

on peut donc établir une courbe qui pour toute proportion donne le nombre de pivots à observer

Procédure binomiale à taille fixe avec une première estimation de la proportion de plantes infestées



Sample size (plants)
200
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
Proportion of plants with ≥1 insect
D_o=0.26
D_o=0.30

5.3.3. Simulation et évaluation des qualités de ce plan binomial

On dispose des mêmes jeux de données de validation

- simulation pour D=0.30:
  - ✓ à partir de chaque jeux de données, construction par simulation d'un échantillon de 200 plantes
  - ✓ pour chaque jeux de données ainsi simulé, 100 simulations de la procédure d'échantillonnage proposée:
  - ✓ Moyennes sur les 100 simulations de : la proportion de plantes infestées, la moyenne estimée, la taille d'échantillon

n=10 plantes prélevées au hasard sans remise

⇓

1ère estimation de P: si 0 ou 1 STOP

⇓

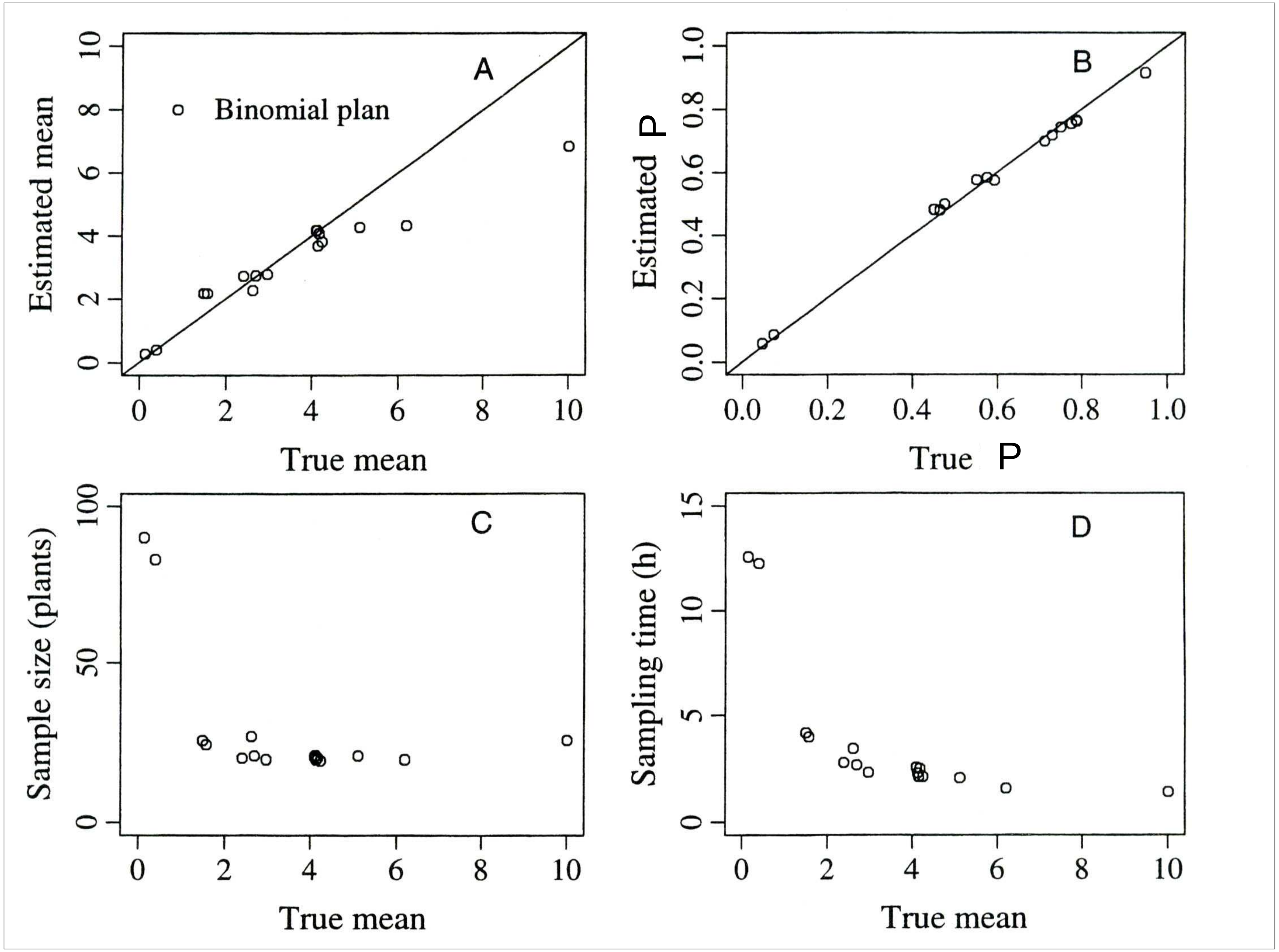
Calcul de la taille d'échantillon N

⇓

N-10 plantes prélevées au hasard sans remise

N_{max}=100

Estimation finale de P_2, et calcul de *m*


A
Binomial plan
Estimated mean
True mean
B
Estimated P
True P
C
Sample size (plants)
True mean
D
Sampling time (h)
True mean
0
2
4
6
8
10
0.0
0.2
0.4
0.6
0.8
1.0
50
100
5
15

Bibliographie sommaire

Echantillonnage

- Cochran, W.G., 1977. Sampling techniques. John Wiley & Sons, New York, third edition.
- Frontier, S., 1983. Stratégies d'échantillonnage en écologie. Masson, Paris.
- Binns, M.R., Nyrop, J.P., van der Werf, W., 2000. Sampling and monitoring in Crop protection. The theoretical basis for developing practical decision guides, CABI publishing, New York.
- Pedigo, L.P., Buntin, G.D., 2000. Handbook of sampling methods for arthropods in agriculture, CRC Press, London.