

## ADN fécal et comptage de coyotes

D. Chessel, A.B. Dufour & J.R. Lobry


---

Étude d'un problème ouvert

### Table des matières

<b>1</b>	<b>Résumé</b>	<b>1</b>
<b>2</b>	<b>La situation</b>	<b>2</b>
<b>3</b>	<b>La courbe de raréfaction</b>	<b>3</b>
<b>4</b>	<b>L'estimation de l'effectif</b>	<b>4</b>
4.1	Courbe de raréfaction observée . . . . .	6
4.2	Ajustement à la courbe observée . . . . .	7
4.3	Procédure . . . . .	8
<b>5</b>	<b>Perspectives</b>	<b>13</b>
<b>6</b>	<b>Objections</b>	<b>14</b>
<b>7</b>	<b>Paramètre de la moyenne ou moyenne du paramètre</b>	<b>19</b>
	<b>Références</b>	<b>22</b>

### 1 Résumé

La fiche donne des indications sur le dénombrement des individus d'une population à partir de l'identification des déjections par séquençage de l'ADN fécal. Le problème est posé dans [2]. La solution proposée est discutée. Le problème est assez difficile. La fiche montre comment on peut se servir de  pour aborder une question ouverte. Elle peut être utilisée pour redéfinir un exercice plus contrôlé.

## 2 La situation

N. Valière demande où trouver une procédure de minimisation de fonction (pour estimer les paramètres d'une courbe du type  $y = \frac{ax}{b+x}$ ). D'où vient le problème? d'une histoire de comptage de coyotes! Ah? Résumons :  $n$  individus d'une espèce vivent sur un territoire. Pendant un intervalle de temps donné, ces individus déposent des fèces accessibles sur ce territoire. Les effectifs de fèces sont :

$$f_1 + \dots + f_n = f$$

Ces fèces sont ramassées et rangées dans un ordre aléatoire. Les  $p$  premières forment un échantillon aléatoire simple de l'ensemble. Quand on a sous-ensemble de  $p$ , on peut indiquer à combien d'individus différents ils appartiennent (*faecal typing by DNA extractions*). La première déjection identifie un individu, chacune des suivantes permet ou non d'identifier un nouvel individu. L'échantillon a ainsi fourni progressivement un effectif observé d'individus différents :

$$1 = x_1 \leq x_2 \leq \dots, x_p = x$$

La question est d'inférer  $f$  à partir de la suite  $1 = x_1, x_2, \dots, x_p$ . Dans l'article cité, les auteurs ont prélevé 651 fèces, les ont tirées une par une au hasard et se sont arrêté quand 30 échantillons consécutifs n'ont fourni qu'un seul nouvel individu. Ainsi 238 échantillons ont été analysés. 115 d'entre eux ont été attribué à l'espèce étudiée, ce qui leur a permis de trouver 30 génotypes distincts et d'estimer la population totale autour de 38 individus.

Pour simuler des données de ce type on prend une population de 40 coyotes ayant déposé un nombre de fèces suivant une loi de Poisson de paramètre 5 :

```
urne <- rpois(40, 5)
urne

[1] 6 5 4 1 5 6 9 5 4 3 6 9 2 4 2 4 8 3 5 4 7 5 2 5 4 6 5 1 6 6 6 6 2 5 7 6 3 3 6 3

urne <- rep(1:40, urne)
urne

[1] 1 1 1 1 1 1 2 2 2 2 2 3 3 3 3 4 5 5 5 5 5 6 6 6 6 6
[27] 6 7 7 7 7 7 7 7 7 7 8 8 8 8 8 9 9 9 9 10 10 10 11 11 11 11
[53] 11 11 12 12 12 12 12 12 12 12 12 13 13 14 14 14 14 15 15 16 16 16 16 17 17 17
[79] 17 17 17 17 17 18 18 18 18 19 19 19 19 19 20 20 20 21 21 21 21 21 21 21 22 22
[105] 22 22 22 23 23 24 24 24 24 24 25 25 25 25 26 26 26 26 26 26 27 27 27 27 27 28
[131] 29 29 29 29 29 30 30 30 30 30 30 31 31 31 31 31 32 32 32 32 32 32 33 33
[157] 34 34 34 34 34 35 35 35 35 35 35 36 36 36 36 36 36 37 37 37 38 38 38 39 39
[183] 39 39 39 39 40 40 40

length(unique(urne))

[1] 40

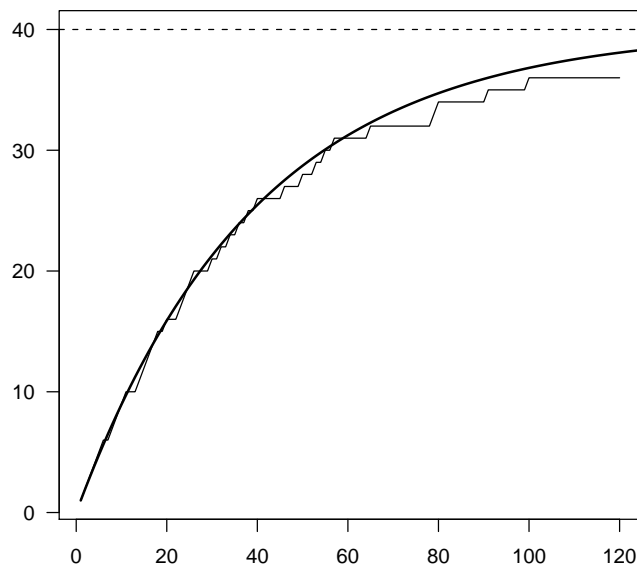
length(urne)

[1] 189
```

Dans cet exemple nous avons  $f = 189$  et  $n = 40$

### 3 La courbe de raréfaction

```
flocal1 <- function(pop, j) {
  return(length(unique(pop[1:j])))
}
flocal2 <- function(nind, n) {
  return(nind - nind * ((1 - 1/nind)^(1:n)))
}
f1 <- function(ur = urne, nessai = 10, nfaesel = length(ur)/3) {
  nfae <- length(ur)
  nind <- flocal1(ur, nfae)
  plot(0, 0, xlim = c(1, nfaesel), ylim = c(1, nind), type = "n",
       las = 1, xlab = "", ylab = "")
  y <- matrix(1:nfaesel, ncol = 1)
  abline(h = nind, lty = 2)
  for (k in 1:nessai) {
    x <- sample(ur, nfaesel, replace = F)
    z <- apply(y, 1, flocal1, pop = x)
    points(1:nfaesel, z, type = "l")
  }
  points(1:nfae, flocal2(nind, nfae), type = "l", lwd = 2)
}
f1(nfaesel = 120, nessai = 1)
```



Quand on prend au hasard un échantillon de  $p$  fèces on trouve un nombre croissant d'individus qui tend vers  $n$ . les auteurs utilisent une courbe du type :

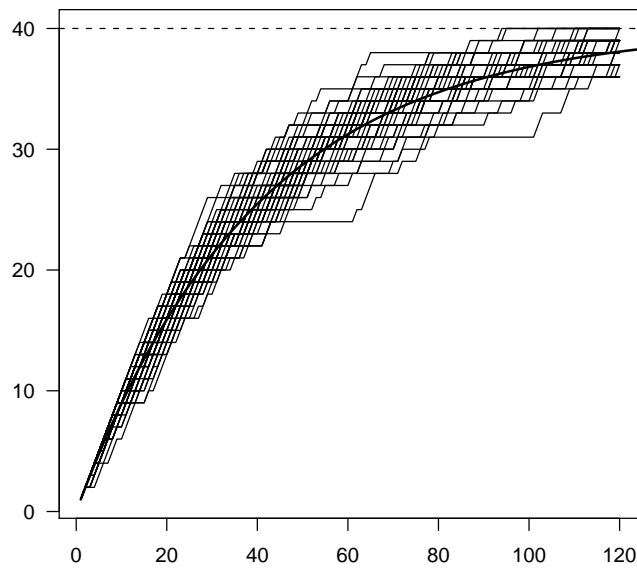
$$y = \frac{ax}{b+x}$$

pour décrire ce type de phénomène (ce qui est purement formel). Quand  $a$  (l'asymptote, donc  $n$  dans les notations présentes) est connue, on peut penser à l'approximation :

$$y = a - a \left(1 - \frac{1}{a}\right)^x$$

qui est l'espérance du nombre de cases occupées quand on distribue au hasard  $x$  boules dans  $a$  cases. Cette approximation est raisonnable si le nombre d'objets reçus par une boîte n'approche pas le nombre total possible qui est borné (alors que dans le *classical occupancy problem* ce n'est pas le cas. Voir Feller [1]).

```
f1(nfaesel = 120, nessai = 50)
```



## 4 L'estimation de l'effectif

Prenons un échantillon et inversons la question :

```
echa <- sample(urne, 70, replace = F)
echa
```

```
[1] 20 17 7 34 21 27 31 22 10 31 17 38 16 39 6 24 27 35 12 5 38 26 20 34 13 12 38
[28] 15 36 29 5 23 28 19 3 17 32 11 34 32 12 11 37 40 36 24 16 21 19 6 6 2 14 1
[55] 40 12 14 29 35 17 17 8 16 2 30 25 1 32 11 24
```

```
table(echa)
```

```
echa
 1  2  3  5  6  7  8 10 11 12 13 14 15 16 17 19 20 21 22 23 24 25 26 27 28 29 30 31
 2  2  1  2  3  1  1  1  3  4  1  2  1  3  5  2  2  2  1  1  3  1  1  2  1  2  1  2
32 34 35 36 37 38 39 40
 3  3  2  2  1  3  1  2
```

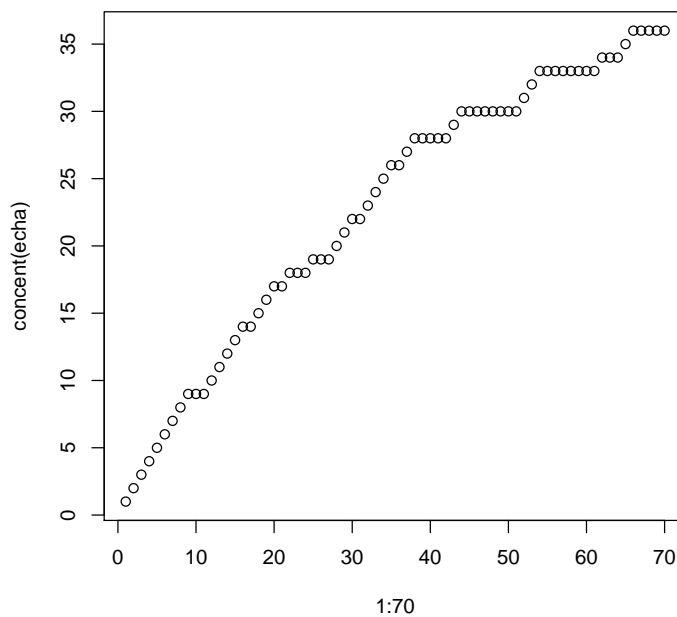
```
length(unique(echa))
```

[1] 36

L'échantillon de crottes représente 36 individus. Combien y en a t'il au total ?  
On a une courbe de raréfaction observée.

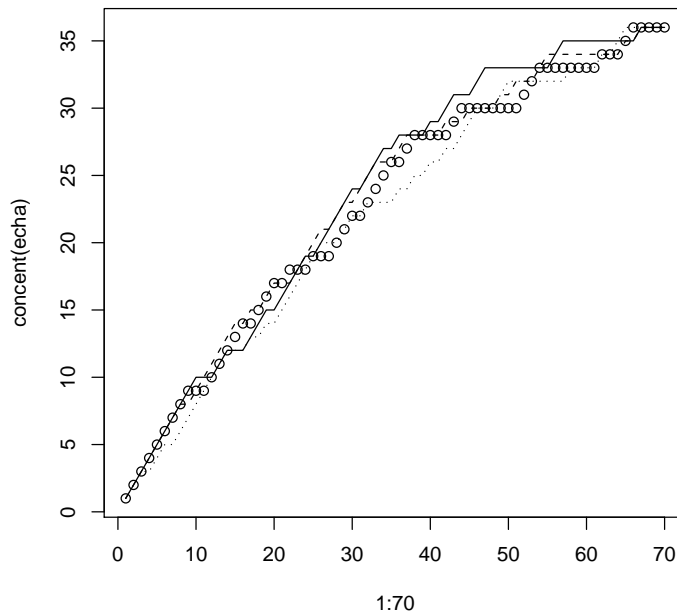
```
concent <- fonction(echa) apply(matrix(1:length(echa), ncol = 1),
  1, flocal1, pop = echa)

plot(1:70, concent(echa), type = "b")
```



Mais le même échantillon présenté dans un ordre différent donne une autre courbe :

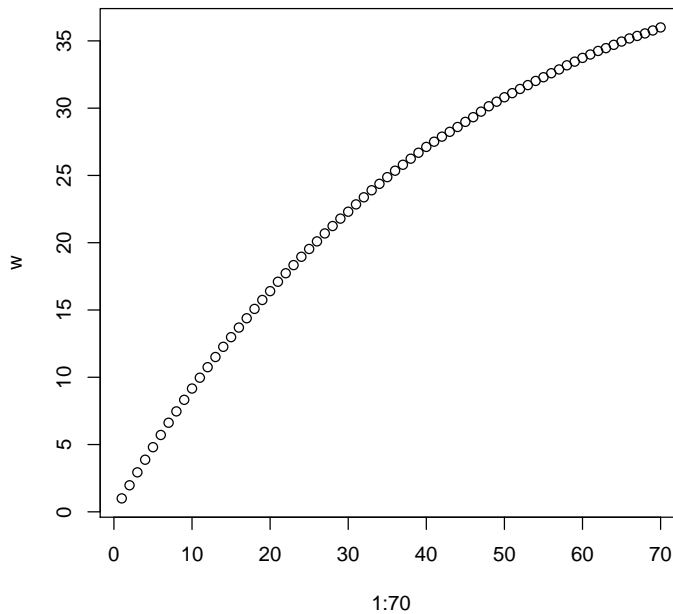
```
plot(1:70, concent(echa), type = "b")
lines(1:70, concent(sample(echa, 70, replace = F)), lty = 1)
lines(1:70, concent(sample(echa, 70, replace = F)), lty = 2)
lines(1:70, concent(sample(echa, 70, replace = F)), lty = 3)
```



La variabilité d'échantillonnage est donc double. Elle provient d'une part de l'échantillonnage dans l'ensemble des fèces récoltées, de l'autre de l'ordre de présentation des échantillons. La seconde peut être éliminée par randomisation, la première est irréductible.

#### 4.1 Courbe de raréfaction observée

```
w <- rep(0, 70)
for (i in 1:500) w <- w + concent(sample(echa, 70, replace = F))
w <- w/500
plot(1:70, w, type = "b")
```



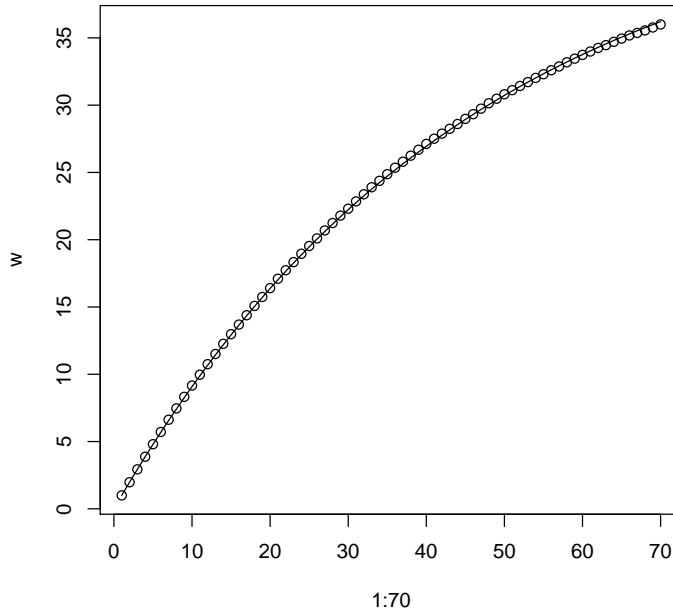
## 4.2 Ajustement à la courbe observée

```
dxy <- cbind.data.frame(1:70, w)
names(dxy) <- c("x", "y")
nls0 <- nls(y ~ a - a * ((1 - 1/a)^x), data = dxy, start = list(a = max(dxy$y)))
nls0
```

```
Nonlinear regression model
  model: y ~ a - a * ((1 - 1/a)^x)
  data: dxy
      a
46.18
residual sum-of-squares: 0.625

Number of iterations to convergence: 4
Achieved convergence tolerance: 5.117e-08
```

```
plot(1:70, w, type = "b")
lines(1:70, predict(nls0))
```



On a trouvé 46.18 coyotes au lieu des 40 de départ.

### 4.3 Procédure

Si on résume les informations précédentes, on peut dire que la statistique de base est une suite  $m_1, m_2, \dots, m_t$  de  $t$  individus découverts respectivement  $m_1, m_2, \dots, m_t$  fois par sélection aléatoire. Cette statistique donne une courbe de raréfaction du type  $y = a - a \left(1 - \frac{1}{a}\right)^x$  qui est ajustée par régression non linéaire et donne une valeur estimée de la taille totale.

```
n <- table(echa)
n

echa
 1 2 3 5 6 7 8 10 11 12 13 14 15 16 17 19 20 21 22 23 24 25 26 27 28 29 30 31
 2 2 1 2 3 1 1 1 3 4 1 2 1 3 5 2 2 2 1 1 3 1 1 2 1 2 1 2
32 34 35 36 37 38 39 40
 3 3 2 2 1 3 1 2

echa

[1] 20 17 7 34 21 27 31 22 10 31 17 38 16 39 6 24 27 35 12 5 38 26 20 34 13 12 38
[28] 15 36 29 5 23 28 19 3 17 32 11 34 32 12 11 37 40 36 24 16 21 19 6 6 2 14 1
[55] 40 12 14 29 35 17 17 8 16 2 30 25 1 32 11 24

estima <- function(n, n500 = 500) {
  concent <- function(echa) apply(matrix(1:length(echa), ncol = 1),
    1, flocal1, pop = echa)
  flocal1 <- function(pop, j) length(unique(pop[1:j]))
  indsel <- length(n)
  if (indsel <= 1)
    stop("n>1 expected")
}
```



```

faesel <- sum(n)
echa <- rep(1:indsel, n)
w <- rep(0, faesel)
for (i in 1:n500) w <- w + concent(sample(echa, faesel, replace = F))
w <- w/n500
dxy <- cbind.data.frame(1:faesel, w)
names(dxy) <- c("x", "y")
nls0 <- nls(y ~ a - a * ((1 - 1/a)^x), data = dxy, start = list(a = max(dxy$y)))
return(coefficients(nls0)[1])
}
estima(n)

```

46.40023<sup>a</sup>

estima(n)

46.35852<sup>a</sup>

estima(n, 100)

46.55977<sup>a</sup>

estima(n, 100)

45.88676<sup>a</sup>

estima(n, 100)

46.33470<sup>a</sup>

Exemple : 100 coyotes à 20 crottes en moyenne, 200 crottes analysées :

```

tot <- rpois(100, 20)
tot

```

```

[1] 18 19 27 25 14 15 19 20 28 21 18 18 22 28 23 23 25 19 18 23 22 21 18 28 23 32
[27] 18 13 14 20 18 12 19 20 26 24 20 16 17 28 13 19 24 21 23 17 18 18 14 18 13 29
[53] 17 31 21 19 21 25 28 23 14 24 14 21 25 25 25 18 29 27 12 18 20 20 22 29 18 15
[79] 14 20 17 23 18 22 24 17 16 25 18 18 14 20 23 23 25 25 18 18 29 18

```

```

crottes <- rep(1:100, tot)
crotsel <- sample(crottes, 200, replace = FALSE)
table(crotsel)

```

```

crotsel
 1  2  3  4  5  6  8  9 10 11 13 14 15 16 17 18 19 21 22 23 24
2  1  2  3  2  2  1  3  1  1  4  2  3  3  2  2  4  2  2  1  1
25 26 27 28 29 30 31 32 34 35 36 37 38 39 40 43 44 45 46 48 49
 4  3  1  1  1  2  1  2  5  3  2  1  2  2  4  4  1  1  3  1  2
50 51 52 53 54 55 56 57 58 59 60 61 64 65 66 67 68 69 70 71 72
 3  3  3  1  1  4  4  5  2  2  3  2  3  4  2  1  1  3  1  1  2
73 74 75 76 77 78 79 80 81 82 84 85 86 87 88 89 91 92 93 94 95
 2  1  1  2  1  2  2  1  2  2  2  4  3  1  3  2  2  2  4  1
96 97 98 99 100
 3  2  4  6  2

```

```
estima(table(crottsel))
```

```
105.1561a
```

Exemple : 200 coyottes à 10 crottes, 200 crottes analysées :

```
crottes <- rep(1:200, 10)
crottsel <- sample(crottes, 200, replace = F)
estima(table(crottsel))
```

```
207.1999a
```

Exemple : 500 coyottes à 5 crottes, 500 crottes analysées :

```
crottes <- rep(1:500, 5)
crottsel <- sample(crottes, 500, replace = F)
estima(table(crottsel))
```

```
607.9873a
```

Exemple : 50 coyottes 25 à 5 crottes et 25 à 20 crottes, 100 crottes analysées :

```
a <- rep(c(2, 20), 25)
b <- rep(1:50, a)
head(b)
```

```
[1] 1 1 2 2 2 2
```

```
c <- sample(b, 100, replace = F)
c
```

```
[1] 36 2 6 8 46 30 34 47 4 42 42 48 24 32 8 48 4 28 16 36 24 6 7 40 18 50
[27] 48 2 34 38 34 6 24 44 22 44 12 40 1 40 32 4 22 12 40 10 42 10 30 10 14 16
[53] 40 14 22 46 28 46 22 10 40 42 18 19 47 6 46 4 30 28 24 42 17 28 50 38 11 38
[79] 20 48 18 3 20 6 18 13 38 24 14 8 44 4 19 14 12 8 26 18 10 37
```

```
d <- table(c)
d
```

```
c
 1  2  3  4  6  7  8 10 11 12 13 14 16 17 18 19 20 22 24 26 28 30 32 34 36 37 38 40
 1  2  1  5  5  1  4  5  1  3  1  4  2  1  5  2  2  4  5  1  4  3  2  3  2  1  4  6
42 44 46 47 48 50
 5  3  4  2  4  2
```

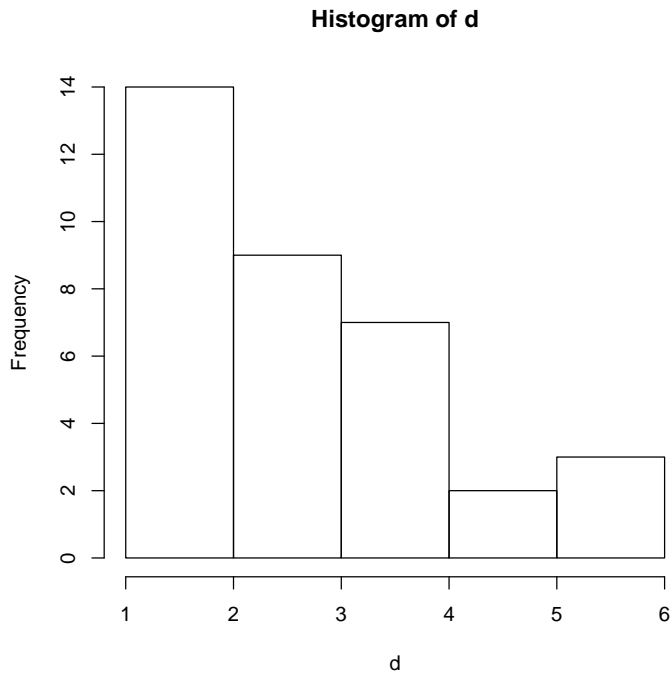
```
estima(d)
```

```
35.69874a
```

```
estima(d)
```

```
35.7436a
```

```
hist(d)
```



L'estimation est bien moins bonne quand la distribution est hétérogène.  
Exemple : 50 coyottes à 5 crottes, 100 crottes analysées :

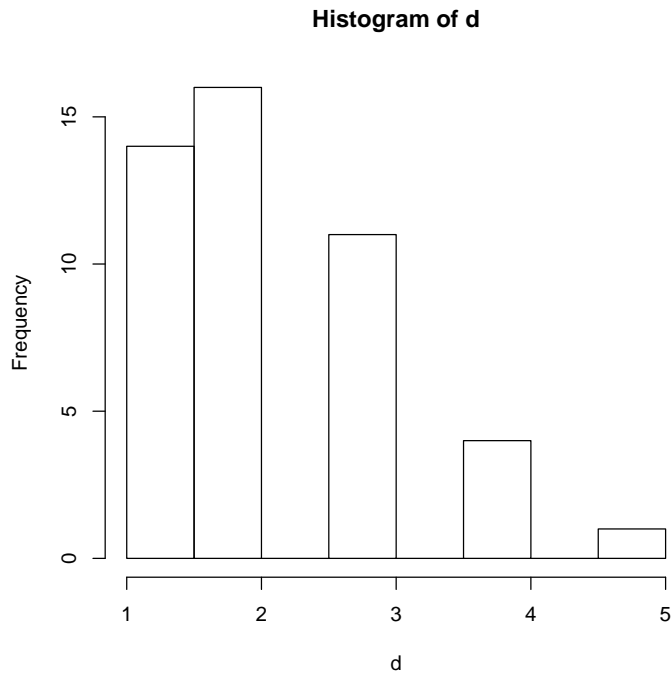
```
b <- rep(1:50, 5)
c <- sample(b, 100, replace = F)
d <- table(c)
d
```

```
c
 2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 22 23 24 25 26 27 28 29 30
 2  1  2  3  3  2  1  1  1  3  3  4  1  5  2  2  2  3  4  2  2  3  1  3  1  2  2  2
31 32 34 35 37 38 39 40 41 42 44 45 46 47 48 49 50
 2  3  3  3  2  2  3  1  1  2  2  2  3  2  1  3  2
```

```
estima(d)
```

```
54.77591a
```

```
hist(d)
```



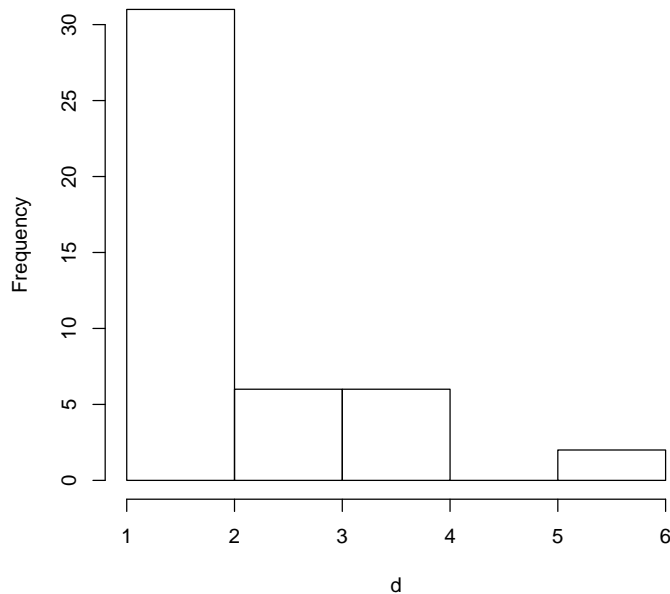
Exemple : 50 coyottes à 50 crottes, 100 crottes analysées :

```
b <- rep(1:50, 50)
c <- sample(b, 100, replace = F)
d <- table(c)
estima(d)
```

50.31934<sup>a</sup>

```
hist(d)
```

Histogram of d



## 5 Perspectives

Globalement, si l'hypothèse d'homogénéité du nombre de déjections par animal est vérifiée, ces estimations sont remarquablement précises. Il semble que la substitution du modèle  $y = a - a \left(1 - \frac{1}{a}\right)^x$  au modèle  $y = \frac{ax}{b+x}$  soit très utile. On pourra étudier en détail, par simulation, la qualité de l'estimateur dans une circonstance donnée.

```
estima(c(6, 5))
```

```
2.026431a
```

```
estima(c(2, 2, 1, 1, 1, 1, 1, 1))
```

```
20.64777a
```

```
estima(c(1,1,1,1,1,1,1,1)) # Normalement on ne peut plus rien dire
```

Et comme on s'y attend :

```
Error in nls(y ~ a - a * ((1 - 1/a)^x), data = dxy, start = list(a = max(dxy$y))) :
step factor 0.000488281 reduced below 'minFactor' of 0.000976562
```

## 6 Objections

Nous voulons travailler sur les vraies données. D. Pontier écrit à l'auteur. La réponse montre que nous n'avons pas lu correctement l'article :

"You could get these numbers though from the figure showing the map and genotypes. Just count the number of times you see individual A, B, C...etc on the map, that is the number of times these individuals were found."

Les coyotes ont les noms :

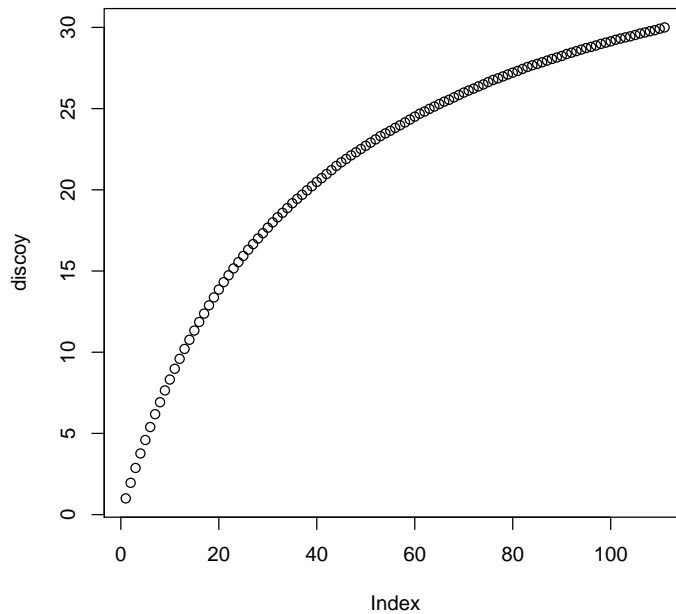
```
labcoy <- c("A", "B", "C", "D", "F", "H", "I", "J", "K", "L", "N",  
           "O", "P", "Q", "R", "T", "W", "X", "Y", "AB", "AC", "AD", "E",  
           "G", "M", "S", "U", "V", "Z", "AA")
```

Ils ont fourni chacun des crottes avec les effectifs :

```
ncoy <- c(6, 5, 3, 2, 5, 3, 6, 3, 2, 7, 5, 5, 11, 3, 2, 2, 8, 9,  
         9, 2, 3, 2, 1, 1, 1, 1, 1, 1, 1, 1)
```

On écrit la fonction qui extrait la courbe de concentration :

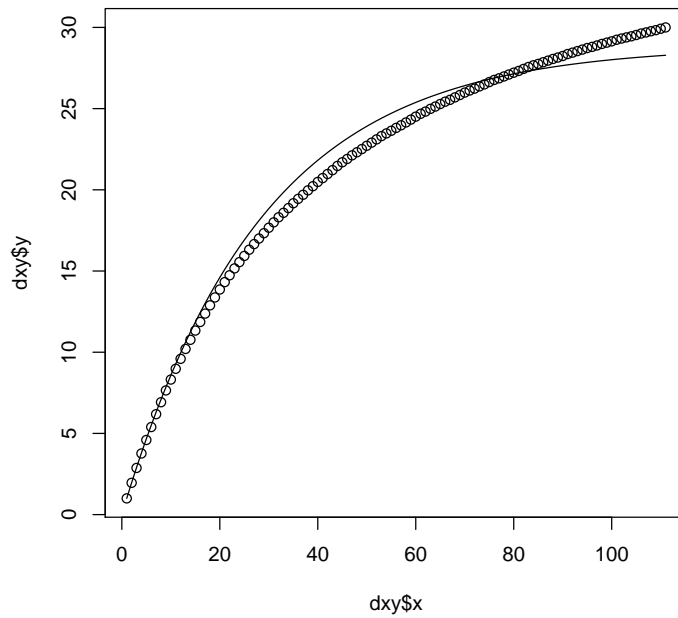
```
discover <- function(n, n500 = 500) {  
  concent <- function(echa) apply(matrix(1:length(echa), ncol = 1),  
    1, flocal1, pop = echa)  
  flocal1 <- function(pop, j) length(unique(pop[1:j]))  
  indsel <- length(n)  
  if (indsel <= 1)  
    stop("n>1 expected")  
  faesel <- sum(n)  
  echa <- rep(1:indsel, n)  
  w <- rep(0, faesel)  
  for (i in 1:n500) w <- w + concent(sample(echa, faesel, replace = F))  
  w <- w/n500  
  return(w)  
}  
discoy <- discover(ncoy, 1000)  
plot(discoy)
```



On extrait la fonction qui ajuste la courbe :

```
estim1 <- function(w) {  
  faesel <- length(w)  
  dxy <- cbind.data.frame(1:faesel, w)  
  names(dxy) <- c("x", "y")  
  nls0 <- nls(y ~ a - a * ((1 - 1/a)^x), data = dxy, start = list(a = max(dxy$y)))  
  plot(dxy$x, dxy$y)  
  lines(dxy$x, predict(nls0))  
  return(coefficients(nls0)[1])  
}  
estim1(discoy)
```

28.85968<sup>a</sup>

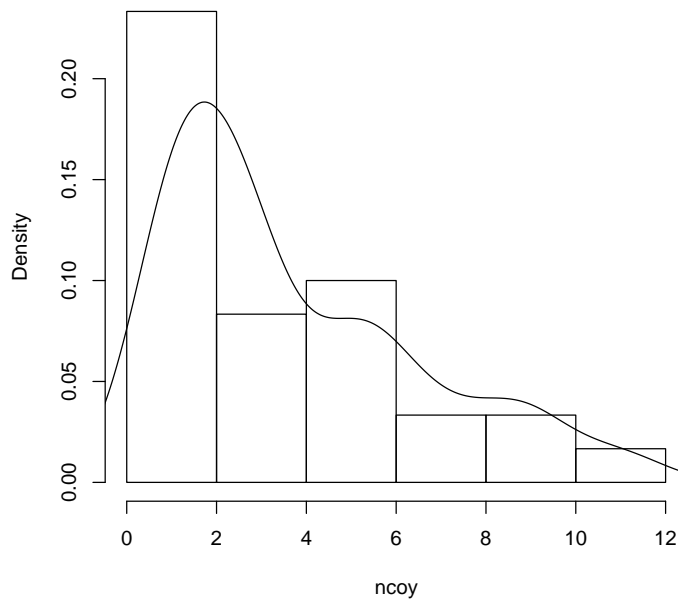


Il y a une grosse difficulté. On estime l'effectif à 29 coyotes alors qu'on en a déjà 30 dans l'échantillon! Ce n'est pas la bonne fonction. Pourquoi?

```
hist(ncoy, proba = T)  
lines(density(ncoy, 1))
```



Histogram of ncoy



```
mean(ncoy)
```

```
[1] 3.7
```

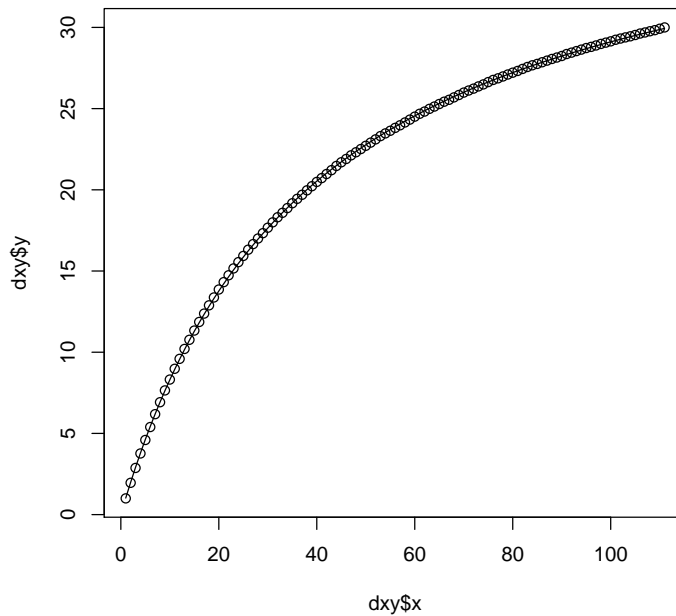
```
var(ncoy)
```

```
[1] 8.07931
```

Nous ne sommes pas dans les hypothèses utilisées. Certains individus sont largement sur représentés. Si on prend les propositions des auteurs :

```
estim2 <- function(w) {
  faesel <- length(w)
  dxy <- cbind.data.frame(1:faesel, w)
  names(dxy) <- c("x", "y")
  nls0 <- nls(y ~ a * x/(b + x), data = dxy, start = list(a = max(dxy$y),
    b = 1))
  plot(dxy$x, dxy$y)
  lines(dxy$x, predict(nls0))
  return(coefficients(nls0))
}
estim2(discoy)
```

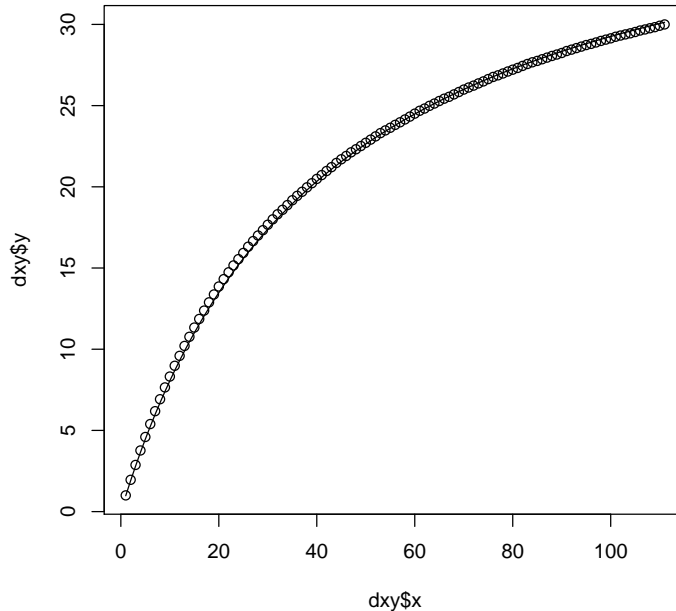
```
      a      b
40.35263 38.61231
```



C'est bien meilleur. Ce n'était pas la peine de faire le malin. On essaye  $a = b$  :

```
estim3 <- function(w) {  
  faesel <- length(w)  
  dxy <- cbind.data.frame(1:faesel, w)  
  names(dxy) <- c("x", "y")  
  nls0 <- nls(y ~ a * x / (a + x), data = dxy, start = list(a = max(dxy$y)))  
  plot(dxy$x, dxy$y)  
  lines(dxy$x, predict(nls0))  
  return(coefficients(nls0))  
}  
estim3(discoy)
```

41.35754<sup>a</sup>



C'est purement formel mais guère moins bon.

Finally, population size was estimated as about 41 (95% CI, 38-45) using the mark-recapture model by Burnham & Overton (1979). The model allows for animals to differ in probability of capture, which was evident in our sample. (p. 661).

L'argument est percutant. On n'a pas de raison théorique d'utiliser la courbe  $y = \frac{ax}{b+x}$  mais elle s'adapte trop bien aux données pour qu'on puisse s'en passer.

## 7 Paramètre de la moyenne ou moyenne du paramètre

On revient sur la stratégie utilisée par les auteurs. Nous avons permuté l'ordre de présentation des fèces pour obtenir une courbe moyenne puis estimer le paramètre de la courbe moyenne. Les auteurs estiment le paramètre sur une présentation permutée puis la moyenne des estimations. Obtient-on la même chose ?

```
tot <- rep(labcoy, ncoy)
tot
```

```
[1] "A" "A" "A" "A" "A" "A" "B" "B" "B" "B" "B" "C" "C" "C" "D" "D"
[17] "F" "F" "F" "F" "F" "H" "H" "H" "H" "I" "I" "I" "I" "I" "J" "J"
[33] "J" "K" "K" "L" "L" "L" "L" "L" "L" "N" "N" "N" "N" "N" "O" "O"
[49] "O" "O" "O" "O" "P" "P" "P" "P" "P" "P" "P" "P" "P" "P" "Q" "Q"
[65] "Q" "Q" "R" "R" "T" "T" "W" "W" "W" "W" "W" "W" "W" "W" "X" "X"
[81] "X" "X" "X" "X" "X" "X" "Y" "Y" "Y" "Y" "Y" "Y" "Y" "Y" "Y" "Y"
[97] "AB" "AB" "AC" "AC" "AC" "AD" "AD" "E" "G" "M" "S" "U" "V" "Z" "AA" "AA"
```

```
echa <- sample(tot, sum(ncoy), replace = F)
echa
```

```
[1] "E" "AA" "Y" "Y" "U" "X" "Q" "X" "G" "B" "L" "L" "L" "Y" "X" "B"
[17] "I" "W" "A" "Q" "W" "W" "R" "T" "H" "P" "Q" "Z" "L" "N" "Y" "L"
[33] "P" "H" "P" "X" "A" "I" "C" "D" "P" "S" "P" "K" "P" "D" "Y" "X"
[49] "F" "Y" "B" "F" "B" "P" "V" "I" "O" "T" "W" "P" "C" "X" "O" "AD"
[65] "N" "N" "AC" "A" "AB" "F" "A" "W" "X" "O" "AC" "B" "X" "Y" "W" "W"
[81] "H" "Y" "F" "F" "Y" "X" "K" "N" "O" "L" "J" "A" "W" "I" "R" "AD"
[97] "A" "I" "I" "P" "M" "P" "L" "P" "AC" "O" "J" "AB" "N" "J" "C"
```

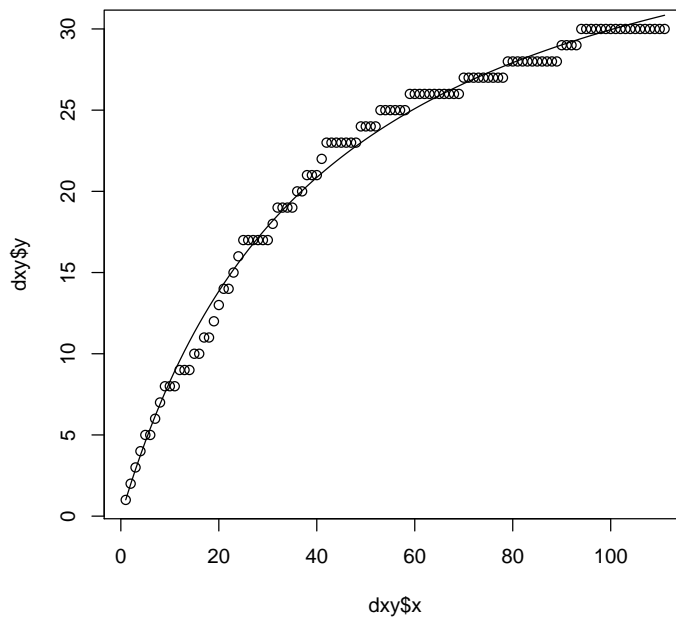
```
concent(echa)
```

```
[1] 1 2 3 3 4 5 6 6 7 8 9 9 9 9 9 10 11 12 12 12 12 13 14 15 16
[27] 16 17 17 18 18 18 18 18 18 18 18 18 18 19 20 20 21 21 22 22 22 22 23 23 23
[53] 23 23 24 24 25 25 25 25 25 25 26 26 26 27 27 28 28 28 28 28 28 28 28 28
[79] 28 28 28 28 28 28 28 28 28 28 28 28 28 29 29 29 29 29 29 29 29 29 29 29
[105] 30 30 30 30 30 30 30
```

```
estim2(concent(echa))
```

```

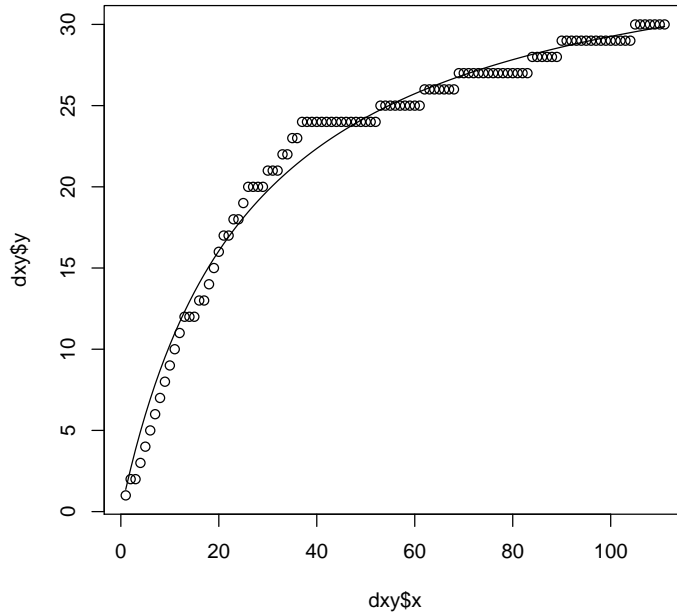
      a      b
44.76790 49.38246
```



```
estim2(concent(sample(tot, sum(ncoy), replace = F)))
```

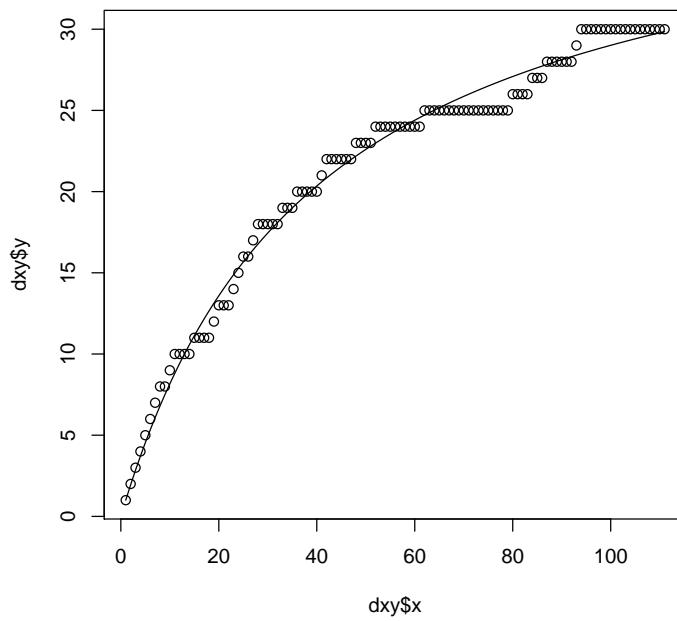
```

      a      b
42.46401 40.89057
```



```
estim2(concent(sample(tot, sum(ncoy), replace = F)))
```

43.52542<sup>a</sup> 44.44989<sup>b</sup>



```
simul1 <- function(w = ncoy, nsim = 10) {
  n0 <- length(w)
  tot <- rep(1:n0, w)
  ns <- sum(w)
  a.vec <- rep(0, nsim)
  for (i in 1:nsim) a.vec[i] <- estim2red(concent(sample(tot,
    ns, replace = F)))
  return(a.vec)
}
estim2red <- function(w) {
  faesel <- length(w)
  dxy <- cbind.data.frame(1:faesel, w)
  names(dxy) <- c("x", "y")
  nls0 <- nls(y ~ a * x/(b + x), data = dxy, start = list(a = max(dxy$y),
    b = 1))
  return(coefficients(nls0)[1])
}
simul1()

[1] 41.74071 46.06351 35.54260 38.66058 44.62992 44.20420 41.91102 38.77194 38.92885
[10] 40.22632

asim <- simul1(nsim = 1000)
mean(asim)
```

```
[1] 41.11391
```

```
hist(asim)
```

Nous ne retrouvons pas les résultats de la figure 2 de l'article. Mais plus de cohérence. L'estimation sur la moyenne est 40.32 et la moyenne des estimations est 41. C'est une erreur que de croire que la variabilité ainsi définie par l'ordre d'introduction est la variabilité d'échantillonnage.

On pourrait approfondir la question en analysant la variabilité d'échantillonnage par bootstrap sur l'ensemble des fèces. On gardera pour le moment la qualité des ajustements avec la courbe  $y = \frac{ax}{b+x}$ . Une raison pour cela serait bienvenue.

## Références

- [1] W. Feller. *An introduction to probability theory and its applications. Vol 1.* Wiley & Sons, New York, 1957.
- [2] M.H. Kohn, E.C. York, D.A. Kamradt, G. Haught, R.M. Sauvajot, and R.K. Wayne. Estimating population size by genotyping faeces. *Proceedings of the Royal Society of London B*, 266 :657–663, 1999.