



université
PARIS-SACLAY

FACULTÉ
DES SCIENCES
D'ORSAY

Analyse des entrées et sorties d'un modèle: précision, élasticité et sensibilité

Marc Girondot, Université Paris Saclay
marc.girondot@universite-paris-saclay.fr

1



Plan

- 1/ Les deux composantes d'un modèle
- 2/ Incertitude sur des valeurs empiriques
 - L'écart-type, l'erreur standard et autres
- 3/ Incertitude sur les paramètres du modèle
 - L'erreur standard et la distribution par MCMC
- 4/ Prendre en compte les incertitudes sur les entrées pour générer des incertitudes sur les sorties
- 5/ Impact des paramètres sur les sorties d'un modèle
 - L'élasticité: « One at the time »
 - L'analyse de sensibilité par décomposition de la variance

2



Notions à retenir

- Robustesse
- Ecart-type, erreur standard, quantiles
- Matrice hessienne
- MCMC par méthode bayésienne et algorithme de Metropolis-Hastings
- Elasticité
- Sensibilité, Indice de Sobol

3



Partie I

LES DEUX COMPOSANTES D'UN MODÈLE

4

Qu'est-ce qu'un modèle ?

- Un modèle est constitué de deux parties:
 - Une formalisation de connaissances
 - Une paramétrisation de ces connaissances
- La formalisation des connaissances peut être verbale ou par des équations.
 - Il n'y a pas de différence fondamentale entre un modèle verbal ("Il fait plus chaud quand il y a une couverture nuageuse plus faible") ou un modèle décrit par des équations:
 $T_1 > T_2$ si $N_1 < N_2$.
 - L'utilisation d'équations permet cependant souvent d'être plus concis et précis.

5

Exemple

- Prenons comme exemple la croissance d'une population sous la forme d'un modèle logistique:

$$N_t = \frac{N_0 K}{N_0 + e^{-rt}(K - N_0)}$$

- Qui est la solution de l'équation différentielle:

$$\frac{dN_t}{dt} = r N_t \left(1 - \frac{N_t}{K}\right)$$

6

Croissance logistique

- Le modèle est donc constitué:
 - de l'équation qui correspond à une formalisation de connaissances;
 - des valeurs des paramètres N_0 , r et K qui décrivent un comportement particulier du modèle.

7

La croissance logistique

- P. F. Verhulst est élève de Adolphe Quételet, un des fondateurs des statistiques.
- Quételet connaît bien les travaux de Malthus et la croissance géométrique des populations.
 - Il faut noter que Malthus discute déjà des mécanismes écologiques qui doivent freiner la croissance. Mais il ne fournit pas d'équation.
- Quételet propose à Verhulst de trouver une fonction de freinage en s'inspirant de la résistance de l'air en aérodynamique:
 - La résistance aérodynamique s'écrit : $R_a = 1/2 \cdot \mu \cdot k \cdot v^2$ où μ représente la masse volumique de l'air, k un coefficient dépendant de la surface frontale du véhicule et de sa résistance aérodynamique et v sa vitesse.

Lambert Adolphe Jacques Quételet né à Gand le 22 février 1796 et mort à Bruxelles le 17 février 1874.



Pierre-François Verhulst, né le 28 octobre 1804 et mort le 15 février 1849 à Bruxelles



8

Lambert Adolphe Jacques Quetelet né à Gand le 22 février 1796 et mort à Bruxelles le 17 février 1874.



L'origine du modèle

- « J'ai tenté depuis longtemps de déterminer par l'analyse, la loi probable de la population ; mais j'ai abandonné ce genre de recherches parce que les données de l'observation sont trop peu nombreuses pour que les formules puissent être vérifiées, de manière à ne laisser aucun doute sur leur exactitude » (in Quételet, 1850)

9

Les fonctions de freinage

$$\frac{dp}{dt} = (m p) - \varphi(p)$$

Verhulst indique ainsi avoir testé successivement quatre fonctions retardatrices :

$$\varphi(p) = n \cdot p^2 \quad \varphi(p) = n \cdot p^3$$

$$\varphi(p) = n \cdot p^4 \quad \varphi(p) = n \cdot \log(p)$$

Il choisit la première car c'est la plus simple et il n'avait pas d'argument pour choisir les autres !

10

Croissance logistique

- La relation mécaniste entre un modèle démographique et l'équation logistique n'a été démontrée qu'en 2002 dans:




Techniques of spatially explicit individual-based models: construction, simulation, and mean-field analysis

Luděk Berec *

Department of Theoretical Biology, Institute of Entomology, Academy of Sciences of the Czech Republic, Faculty of Biological Sciences, University of South Bohemia, Branišovská 31, 370 05 České Budějovice, Czech Republic
Received 25 February 2001; received in revised form 24 August 2001; accepted 28 October 2001

11

Formalisation de connaissance

- La formalisation des connaissances sous la forme d'un modèle peut être elle-même sujette à une part d'incertitude; par exemple, on peut penser que K peut lui-même dépendre de N_t si K inclut le développement de prédateurs quand N_t est élevé... Mais alors c'est un modèle de type Lotka-Volterra !
- Cependant souvent cette incertitude n'est pas prise en compte. On considèrera alors ce modèle comme décrivant au mieux les processus que l'on cherche à modéliser.
 - C'est clairement trop optimiste et une solution est proposée sur la page suivante !

12

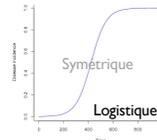
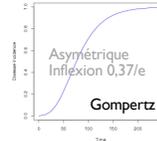
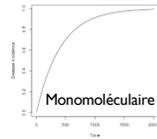
Une solution

Richards F.J. 1959. A flexible growth function for empirical use. *Journal of Experimental Botany* 29:290-300.

$$g = a(1 + b \exp(-rt))^{1/(1-k)}$$

$$g = \begin{cases} a(1 + b e^{-rt}) & \text{Monomolecular curve when } k = 0 \\ a e^{-b e^{-rt}} & \text{Gompertz's curve when } k = 1 \\ \frac{a}{1 + b e^{-rt}} & \text{Autocatalytic or logistic curve when } k = 2. \end{cases}$$

Façon astucieuse de tester différentes formulations de modèles qu'on peut changer avec un paramètre. Donc différentes catégories de modèles sont dépendantes d'un paramètre.



13

Paramètres

- L'autre composante du modèle est celle pour laquelle on considère le plus souvent l'incertitude, il s'agit des paramètres du modèle.
- Les valeurs de ces paramètres peuvent être déterminées expérimentalement et elles sont donc entachées d'incertitude.

14

Partie II

INCERTITUDE SUR DES VALEURS EMPIRIQUES (ISSUES DE L'OBSERVATION)

15

Paramètres mesurés empiriquement

- Certains paramètres peuvent être directement mesurés sur le terrain.
- Il convient cependant d'effectuer une série de mesures pour capturer la variabilité sur ce paramètre.

Comment décrire cette variabilité ?

16

Les mesures de la dispersion

- Maximum et minimum
 - L'écart-type
 - L'erreur-standard
 - Intervalle interquantile
- Toutes ces mesures sont justes mais ne représentent pas la même chose ; il faut donc bien savoir ce que l'on cherche à représenter.

17

Différence principale

- Ce qui différencie principalement ces mesures de dispersion, c'est leur comportement par rapport à des valeurs extrêmes ou aberrantes ainsi que la symétrie de l'estimateur autour de la moyenne.

18

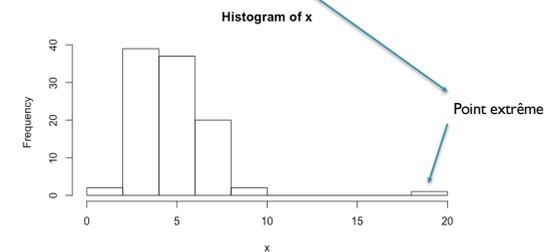
Robustesse d'une statistique

- La **robustesse** d'une statistique est sa capacité à supporter des violations d'hypothèses.
- Dans notre cas, si on veut mesurer la dispersion de nos observations, l'hypothèse est que les points soient tous tirés d'une même distribution.
- Mais si un point sort clairement du lot (erreur de mesure, autres phénomènes), il faut que la statistique décrivant la distribution ne dévie pas trop à cause de ce point.

19

Exemple

```
> x <- c(rgamma(100, shape=9, scale=0.5), 20)
> hist(x)
```



20

Minimum et maximum

- Les minimum et maximum sont des statistiques très sensibles aux valeurs extrêmes. Ce sont les statistiques les moins robustes pour décrire la dispersion.

```
> x <- rgamma(100, shape=9, scale=0.5)
> x2 <- c(x, 20)
> range(x)
[1] 1.558270 9.499228
> range(x2)
[1] 1.558270 20.00000
```

Dans x on a les points tous tirés d'une même distribution et dans x2 on rajoute un point aberrant (20).

Pour avoir un code répliquable quand on utilise des nombres aléatoires, pensez à utiliser `set.seed(x)`.

21

Ecart-type - standard deviation

- L'écart-type mesure la dispersion des valeurs.
- Dans une loi normale, 95% des observations sont situées dans l'intervalle moyenne $\pm 1,96$ SD.

```
> sd(x)
[1] 1.542434
> sd(x2)
[1] 2.179
> 100*(sd(x2)-sd(x))/sd(x)
[1] 41.27029
```

22

L'écart-type (2)

- L'écart-type est très sensible à la présence d'une valeur extrême puisqu'il augmente de 41% seulement à cause d'une valeur.

Attention: Les observations sont situées dans l'intervalle moyenne ± 1.96 SD seulement dans le cas de la loi normale.

23

L'erreur standard - standard error

- L'erreur standard mesure la dispersion de la moyenne de nos observations. Il se calcule comme:

```
> sd(x)/sqrt(length(x))
[1] 0.1542434
> sd(x2)/sqrt(length(x2))
[1] 0.2168186
```

$$SE = \frac{SD}{\sqrt{N}}$$

24

L'erreur standard (2)

- Mais l'erreur standard reste très sensible à une observation extrême puisqu'il augmente aussi de 41% sur la base de l'inclusion d'une seule valeur extrême.

```
> sd(x)/sqrt(length(x))
[1] 0.1542434
> sd(x2)/sqrt(length(x2))
[1] 0.2168186
```

25

Masquer les erreurs

- On a comme relation : $SE = \frac{SD}{\sqrt{N}}$

« Donc SE est toujours plus faible que SD; et donc autant présenter SE que SD sur les barres d'erreur, cela cachera qu'on a mal travaillé ! »

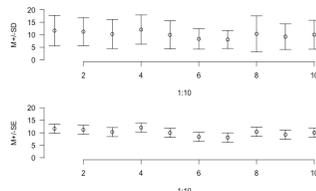
26

Représenter SE ou SD

```
M <- NULL
SD <- NULL
SE <- NULL

for (i in 1:10) {
  x <- runif(10, 5, 15)
  M <- c(M, mean(x))
  SD <- c(SD, sd(x))
  SE <- c(SE, (sd(x)/sqrt(length(x))))
}

library(HelpersMG)
par(mar=c(4, 4, 1, 1))
layout(mat = matrix(1:2, nrow=2))
plot_errbar(1:10, M, errbar.y = deux*SD, bty="n", las=1, ylab="M+/-SD", xlab="1:10", ylim=c(0, 20))
plot_errbar(1:10, M, errbar.y = deux*SE, bty="n", las=1, ylab="M+/-SE", xlab="1:10", ylim=c(0, 20))
```



27

Barres d'erreur

- En fait le terme « barre d'erreurs » serait à proscrire. Ces barres mesurent la dispersion d'un résultat et pas forcément une erreur.
- La source de dispersion des valeurs est le plus souvent due à une variabilité naturelle. Bien sûr parfois la variabilité peut être générée par des erreurs de mesure, mais c'est rare. C'est une des raisons pour laquelle on peut répliquer aussi les mesures individuelles.
- Vouloir masquer la variabilité naturelle est stupide puisque justement c'est cette variabilité naturelle qui est utile pour répondre à des questions intéressantes (ou pas) et qui est le moteur de l'évolution !

28

SD vs SE

- SD est une mesure de la dispersion des observations.
- SE est une mesure de la dispersion des moyennes, c'est à dire où se trouve la moyenne.
 - La moyenne peut être considérée déjà comme un modèle.
- **SD et SE ne sont donc pas interchangeables ; selon ce qu'on veut montrer sur un graphique, il faut choisir l'un ou l'autre.**

29

SCIENTIFIC REPORTS

OPEN TRPV4 associates environmental temperature and sex determination in the American alligator

Received: 22 June 2015
 Accepted: 20 November 2015
 Published: 18 December 2015

Ryohel Yatsu^{1,2,*}, Shinichi Miyagawa^{1,2,*}, Satomi Kohno³, Shigeru Salto^{1,3}, Russell H. Lowers⁴, Yukiko Ogino^{1,2}, Naomi Fukuta⁴, Yoshinao Katsu², Yasuhiko Ohta⁴, Makoto Tominaga^{4,5}, Louis J. Guillette Jr⁶ & Taiseni Iguchi^{1,2}

30

Expression of TRPV4 gene

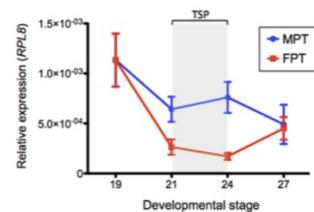
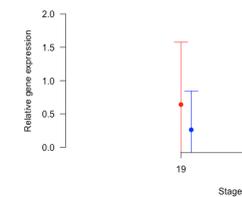


Figure 1. Developmental expression profile of American alligator TRP channels in gonad during sexual development. (A) The mRNA levels of various thermosensitive TRP channels were assessed in gonads at the onset of TSP (stage 21) incubated under MPT and FPT conditions. Gene expressions of 5 AmTRP ion channels (AmTRPV2, AmTRPV4, AmTRPA1, AmTRPM3, AmTRPM8) were observed in varying expression levels. (B) Quantitative RT-PCR analysis was performed for AmTRPV4 at various key sexual developmental stages including bipotential (stage 19; n = 13), sex determination (stage 21; n = 14, 14), sex differentiation (stage 24; n = 14, 15), and pre-hatching (stage 27; n = 14, 15) stages at both FPT and MPT temperature conditions respectively; \pm SEM. Temperature sensitive period is indicated in gray.

31

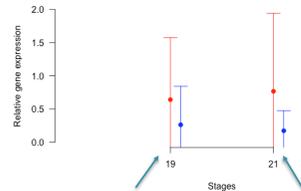
Expression of TRPV4 gene



- Les barres dans la publication indiquent \pm SEM, la dispersion au niveau individuel est donc $(14)^{-2} = 3,7$ fois plus élevée.
- Voici un graphique montrant ± 2 SD qui est la métrique correcte pour montrer que les individus MPT et FPT peuvent présenter des niveaux d'expression différents.

32

Expression of TRPV4 gene



- Notez qu'en utilisant ± 2 SD on trouve que l'expression relative pourrait être négative, cela n'a aucun sens.
- Cela est dû au fait que la distribution de l'expression relative (un rapport) n'est pas normale alors que quand on fait ± 2 SD, on suppose implicitement qu'elle est normale.

33

Le nombre magique 2

- On a tous entendu que 95% des points sont situés dans l'intervalle moyenne ± 2 SD.
- Déjà, on sait tous que ce n'est pas 2, mais 1,96:

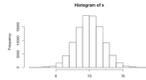
```
> (deux <- qnorm(p=0.975, 0, 1))
[1] 1.959964
```
- Si on prend 2, cela fait:

```
> pnorm(2, 0, 1)-pnorm(-2, 0, 1)
[1] 0.9544997
```
- Pas si différent si l'effectif n'est pas trop important.

34

Distribution des observations

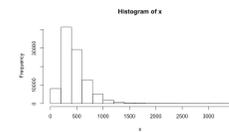
```
> x <- rnorm(100000, 10, 2)
> hist(x)
> mean(x)
[1] 10.01199
> sqrt(sum((x-mean(x))^2)/length(x))
[1] 1.997326
> sd(x)
[1] 1.997336
> sum(x < (mean(x) - deux*sd(x))) / length(x)
[1] 0.02401
> sum(x > (mean(x) + deux*sd(x))) / length(x)
[1] 0.02531
```



35

Distribution des observations

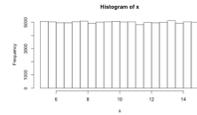
```
> x <- rlnorm(100000, 6, 0.5)
> hist(x)
> mean(x)
[1] 457.4219
> sd(x)
[1] 245.5845
> sum(x < (mean(x) - deux*sd(x))) / length(x)
[1] 0
> sum(x > (mean(x) + deux*sd(x))) / length(x)
[1] 0.04643
```



36

Distribution des observations

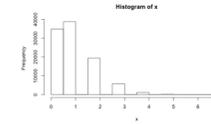
```
> x <- runif(100000, 5, 15)
> hist(x)
> mean(x)
[1] 9.994812
> sd(x)
[1] 2.890294
> sum(x < (mean(x) - deux*sd(x))) / length(x)
[1] 0
> sum(x > (mean(x) + deux*sd(x))) / length(x)
[1] 0
```



37

Distribution des observations

```
> x <- rbinom(100000, 10, 0.1)
> hist(x)
> mean(x)
[1] 1.00094
> sd(x)
[1] 0.9483818
> sum(x < (mean(x) - deux*sd(x))) / length(x)
[1] 0
> sum(x > (mean(x) + deux*sd(x))) / length(x)
[1] 0.07043
```



38

Conclusion

- 95% des observations ne sont pas dans l'intervalle ± 1.96 SD sauf dans un cas très particulier: la loi normale.

- Cela peut conduire à des situations absurdes :

```
x <- x/10
mean(x)
sd(x)
library(HelpersMG)
plot_errbar(1, mean(x),
  errbar.y = deux*sd(x), bty="n",
  ylim=c(-0.1, 1), xlab="",
  ylab="Fréquence", xaxt="n")
segments(x0=0, x1=2, y0=0, y1=0, lty=2)
```

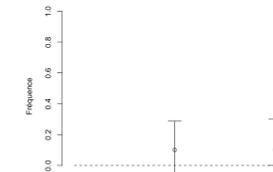


Et voici comment une fréquence peut être négative !

39

Retour aux fréquences

```
x <- rbinom(100000, 10, 0.1)
x <- x/10
mean(x)
sd(x)
l <- quantile(x = x, probs = c(0.025, 0.9)
library(HelpersMG)
plot_errbar(c(1, 2), c(mean(x), mean(x)), errbar.y.plus =
  c(deux*sd(x), l[2]-mean(x)), errbar.y.minus =
  c(deux*sd(x), mean(x)-l[1]), bty="n", ylim=c(-0.1,
  1), xlab="", ylab="Fréquence", xaxt="n", xlim=c(0,
  3))
segments(x0=0, x1=2, y0=0, y1=0, lty=2)
```



40

Les quantiles

- Un quantile q est la valeur pour laquelle une fréquence q des observations sont situées en dessous du quantile.
- On utilisera les quantiles 0.025 et 0.975 pour définir les bornes qui incluent 95% des valeurs ($0.975-0.025=0.95$).

```
> quantile(x, probs=c(0.025, 0.975))
```

```
2.5% 97.5%
```

```
2.273944 7.963572
```

```
> quantile(x2, probs=c(0.025, 0.975))
```

```
2.5% 97.5%
```

```
2.276463 8.151774
```

41

Les quantiles (2)

- Les quantiles sont des statistiques très robustes notamment aux points extrêmes mais aussi à l'asymétrie de la distribution.

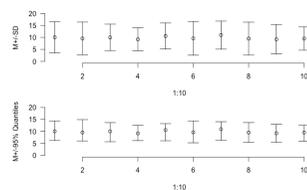
42

Les quantiles 0,025 et 0,975

```
M <- NULL; SD <- NULL; SE <- NULL; QM <- NULL; QP <- NULL
```

```
for (i in 1:10) {
  x <- runif(10, 5, 15)
  M <- c(M, mean(x))
  l <- quantile(x = x,
    probs = c(0.025, 0.975))
  QM <- c(QM, l[1])
  QP <- c(QP, l[2])
  SD <- c(SD, sd(x))
  SE <- c(SE, (sd(x)/sqrt(length(x))))
}
```

```
layout(mat = matrix(1:2, nrow=2))
plot_errbar(1:10, M, errbary = deux*SD, bty="n", las=1, ylab="M+-SD", xlab="1:10", ylim=c(0, 20))
plot_errbar(1:10, M, yplus = QP, yminus = QM, bty="n", las=1, ylab="M+/-95% Quantiles", xlab="1:10", ylim=c(0, 20))
```



43

Conclusion

- Pour mesurer la dispersion des points, nous utiliserons les quantiles 0.025 et 0.975.
- Pour mesurer la dispersion de la moyenne nous utiliserons l'erreur standard sachant qu'il y a 95% de chance que la vraie moyenne soit entre moyenne ± 1.96 SE
 - A noter que la distribution de la moyenne est, elle, normale.

44

D'où vient cette affirmation ?

- Le théorème central limite (Laplace, 1809) établit la convergence en loi de la somme d'une suite de variables aléatoires vers la loi normale.
- La moyenne est une somme divisée par une constante (N), donc le théorème central limite s'applique mais que dit-il en clair:
Définition - On dit que la suite $(X_n)_{n \geq 1}$ converge en loi vers X si, pour toute fonction φ continue bornée sur \mathbb{R} , à valeurs dans \mathbb{R} :



23 mars 1749, Beaumont-en-Auge
5 mars 1827, Paris.

Pierre-Simon Laplace, « Mémoire sur les approximations des formules qui sont fonctions de très-grands nombres, et sur leur application aux probabilités », Mémoires de la Classe des sciences mathématiques et physiques de l'Institut de France, 1809, p. 353-415

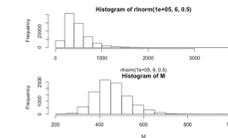
$$\lim_n \mathbb{E}[\varphi(X_n)] = \mathbb{E}[\varphi(X)].$$

45

Distribution de la moyenne

```
M <- NULL; SD <- NULL; SE <- NULL
```

```
for (i in 1:10000) {
  x <- rlnorm(10, 6, 0.5)
  M <- c(M, mean(x))
  SD <- c(SD, sd(x))
  SE <- c(sd(x)/sqrt(length(x)))
}
```

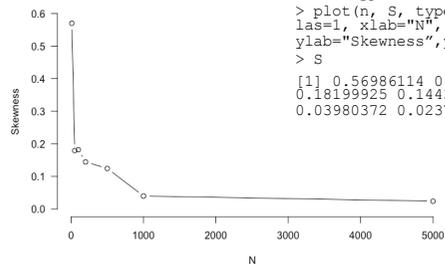


```
layout(mat = matrix(1:2, nrow=2))
hist(rlnorm(100000, 6, 0.5))
hist(M, xlim=c(200, 1000))
```

```
> mean(((M-mean(M))/sd(M))^3)
[1] 0.5624483
> library(e1071)
> skewness(M, type=3)
[1] 0.5624483
```

46

Quel n est nécessaire pour que la convergence en loi normale soit une approximation correcte ?



```
> N <- c(10, 50, 100, 200, 500,
1000, 5000)
> S <- skewness(rlnorm(n, 6,
0.5), type=3)
> plot(N, S, type="b", bty="n",
las=1, xlab="N",
ylab="Skewness", ylim=c(0,0.6))
> S
[1] 0.56986114 0.17927412
0.18199925 0.14438459 0.12433417
0.03980372 0.02376757
```

47

La quête du graal

- Le théorème central limite nous dit que les estimateurs sont distribués normalement mais on ne l'atteint réellement que quand un nombre très grand d'observations utilisés pour estimer les distributions.
- Qu'en est-il pour les observations elles-mêmes ?
- Revenons aux caractéristiques de la loi normale:
 - Non bornée
 - Symétrique
- Est-ce réaliste: non !
- La loi normale admet des valeurs négatives alors que la plupart des métriques utilisée en biologie n'en admettent pas.

48

La quête du graal

Donc pourquoi continuer à utiliser une distribution que l'on sait non-appropriée ?

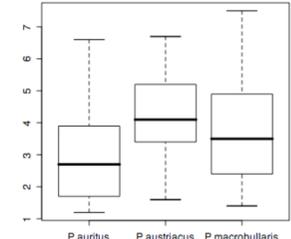
- On a d'autres distributions qui sont flexibles et plus appropriées:
 - La binomiale pour les proportions
 - La binomiale négative pour les comptages
 - La gamma pour les valeurs >0
 - La lognormale pour les valeurs >0
 - La différence entre une gamma et une lognormale est subtile et tient à l'épaisseur de la queue de la distribution.
- Bien sûr, on peut s'en sortir en parlant de tests robustes et continuer à utiliser la loi normale... mais ce n'est pas satisfaisant.



49

Autre façon de représenter une distribution univariée

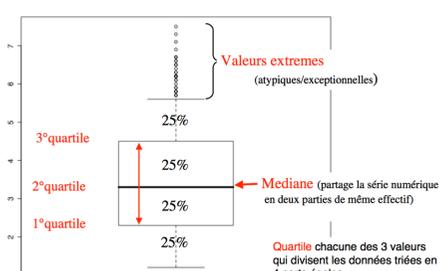
`boxplot(d[,2]-d[,1])`



50

Les moustaches: les 4 quartes et un peu plus

"Variations autour des boîtes à moustaches"



3^equartile
2^equartile
1^equartile

Valeurs extrêmes (atypiques/exceptionnelles)

Mediane (partage la série numérique en deux parties de même effectif)

Quartile chacune des 3 valeurs qui divisent les données triées en 4 parts égales

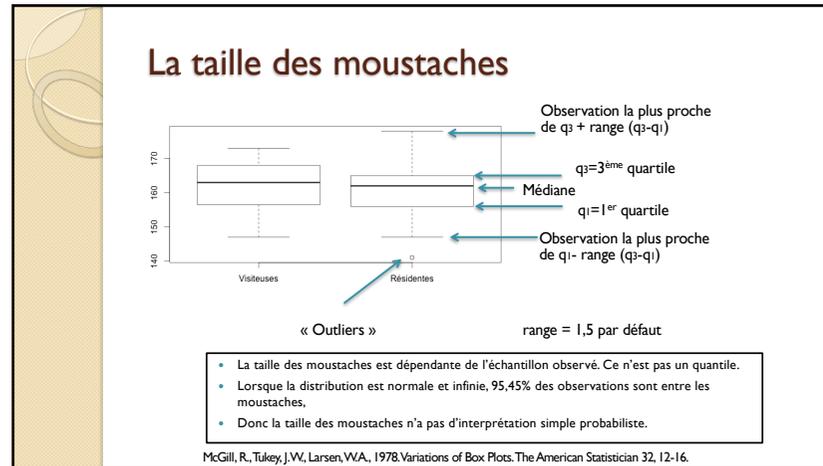
51

Marcel Pagnol, réplique extraite du film "Marius" (1931)

Les tiers et les quartes



52



53

Partie 3

INCERTITUDE SUR LES PARAMÈTRES AJUSTÉS D'UN MODÈLE

54

Ajustement d'un modèle à des observations

- Il arrive assez fréquemment qu'on cherche à ajuster des paramètres d'un modèle à partir d'observation.
- Cela revient à chercher la valeur que peut prendre un paramètre.
- Mais cette valeur a elle-même une incertitude.
- Appelons x_1, \dots, x_n les données de l'observation et $\theta_1, \dots, \theta_k$ les paramètres du modèle à estimer.

55

Maximum de vraisemblance

- Etant donné un échantillon observé x et une fonction de vraisemblance f , la vraisemblance quantifie la probabilité que les observations proviennent effectivement d'un échantillon (théorique) de la loi f .
- La fonction de vraisemblance, notée $L(x_1, \dots, x_n | \theta_1, \dots, \theta_k)$ est une fonction de probabilités conditionnelles qui décrit les valeurs x_i (observations) d'une loi statistique en fonction des paramètres θ_j
- Estimer un paramètre θ_j par la méthode du maximum de vraisemblance, c'est proposer comme valeur de ce paramètre celle qui rend maximale la vraisemblance, à savoir la probabilité d'observer les données comme réalisation d'un échantillon de la loi f .

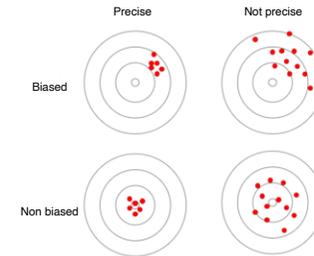
56

L'estimation au ML

- Les paramètres $\theta_1, \dots, \theta_k$ qui maximisent la vraisemblance $L(x_1, \dots, x_n | \theta_1, \dots, \theta_k)$ correspondent à une estimation ponctuelle (*point estimation*).
- Ils représentent la combinaison qui maximise la probabilité d'observer les données comme réalisation d'un échantillon de la loi f mais ils ne nous donnent pas d'information sur la précision à laquelle on les connaît.

57

Toujours bien de se rappeler...



- Il est difficile de déterminer qu'un estimateur n'est pas précis ou biaisé car cela implique de connaître la vraie valeur;
- Il y a souvent un compromis entre précision et biais. On ne sait pas forcément ce qui devrait être privilégié: estimateur précis ou non biaisé?

58

Estimer la précision d'un paramètre

- Estimer la précision à laquelle est connue un paramètre est très important car cela permet de calculer son intervalle de confiance.
- On pourra alors conclure par exemple s'il est différent ou non de 0, ou bien si deux conditions expérimentales donnent des résultats significativement ou non ou si deux populations sont significativement différentes pour le caractère mesuré.

59

L'ERREUR-STANDARD

60

Exemple avec une distribution binomiale

- La loi binomiale modélise un évènement répété où il y a deux sorties possibles (A et B). Cette loi a un paramètre, p qui désigne la probabilité d'avoir un A (et donc $q=1-p$ est la probabilité d'avoir un B).
- La probabilité d'une réalisation particulière d'obtention de k objets A sur un total de n s'écrit:

$$\binom{n}{k} = C_n^k = \frac{n!}{k!(n-k)!} \quad P(x = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

61

Autres lois possibles

- On va considérer ici que le modèle binomial n'est pas une variable en soi.
- Notez quand même qu'il existe d'autres lois possibles pour décrire ces données, par exemple la loi bêta-binomiale lorsque le paramètre p de la binomiale est aléatoire et donné par une loi bêta:
 - [https://fr.wikipedia.org/wiki/Loi_bêta-binomiale](https://fr.wikipedia.org/wiki/Loi_b%C3%a9ta-binomiale)

62

Exemple avec une distribution binomiale

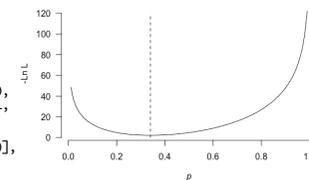
```
> k <- 17 # Nombre de A
> n <- 50 # Nombre total de tirage
> p <- 0.1 # probabilité de tirer un A
> choose(n, k)*p^k*(1-p)^(n-k)
[1] 3.043151e-06
> dbinom(x=k, size=50, prob = p, log=FALSE)
[1] 3.043151e-06
> dbinom(x=k, size=50, prob = p, log=TRUE) # même chose en log
[1] -12.70262
> -dbinom(x=k, size=50, prob = p, log=TRUE) # même chose en -log
[1] 12.70262
```

63

Profil de vraisemblance

Quelle est la probabilité d'observer $k=17/n=50$ dans un modèle binomial ($p, 1-p$).
Le modèle est défini à la fois par la loi binomiale et la valeur de p .

```
p <- seq(from=0.01, to=0.99, by=0.01)
LnL <- -dbinom(x=k, size=50, prob = p, log=TRUE)
plot(p, LnL, xlab=expression(italic("p")), ylab="-Ln L", bty="n", las=1, type="l")
segments(x0=p[which.min(LnL)], y0=0, y1=120, lty=2)
```



64

Maximum de vraisemblance

```

p <- seq(from=0.01, to=0.99, by=0.01)
LnL <- -dbinom(x=k, size=50, prob = p, log=TRUE)
plot(p, LnL, xlab=expression(italic("p")), ylab="-Ln L",
     bty="n", las=1, type="l")
segments(x0=p[which.min(LnL)], y0=0, y1=120, lty=2)
    
```

- On va utiliser un algorithme qui va rechercher les valeurs des paramètres $(p, q=1-p)$ qui maximisent la vraisemblance des observations dans une loi binomiale. Ceci est fait avec la fonction `optim()` de r.


```

> negdbinom <- function(x, size, prob, log = FALSE) return(-dbinom(x, size,
prob, log))
> o <- optim(par = c(prob=0.1), fn=negdbinom, size=50, x=k, log=TRUE,
method = "Brent", lower=0, upper=1)
> o$par # C'est la valeur de p qui maximise la vraisemblance
[1] 0.34
            
```

65

Le profil de vraisemblance dépend de la taille de l'échantillon

```

k <- 17 # Nombre de A
n <- 50 # Nombre total de tirage
p <- seq(from=0.01, to=0.99, by=0.01)
LnL <- -dbinom(x=k, size=n, prob = p, log=TRUE)
plot(p, LnL, xlab=expression(italic("p")), ylab="-Ln L",
     ylim=c(0, 130),
     bty="n", las=1, type="l")
segments(x0=p[which.min(LnL)], y0=0, y1=1300, lty=2)

k <- 34 # Nombre de A
n <- 100 # Nombre total de tirage
p <- seq(from=0.01, to=0.99, by=0.01)
LnL <- -dbinom(x=k, size=n, prob = p, log=TRUE)
lines(p, LnL, col="red")

legend(x=0.4, y=120, legend = c("n=50", "n=100"),
      lty=1, col=c("black", "red"))
    
```

66

Le profil de vraisemblance dépend de la taille de l'échantillon

Plus l'échantillon est grand, plus le profil de vraisemblance est aigu au niveau du point ML.
 Cela signifie que quand on change la valeur du paramètre p , la vraisemblance se dégrade plus rapidement lorsque la taille de l'échantillon est importante que lorsque la taille de l'échantillon est plus faible.
 On va utiliser une mesure de cette courbure au point ML pour décrire la précision avec laquelle on connaît le paramètre p .

67

Mesurer le changement de la vraisemblance au point de ML ?

- Petit rappel de maths:
 - La dérivée première en un point mesure la pente de la tangente en ce point; elle est nulle pour un extremum
 - La dérivée seconde en un extremum mesure la courbure en cet extremum.

Test de la dérivée seconde

Si la dérivée première s'annule, la tangente est horizontale.

Si, de plus, $f''(a) < 0$, la courbe est concave vers le bas et $f(x)$ a un maximum relatif à $x = a$.

Si, de plus, $f''(a) > 0$, la courbe est concave vers le haut et $f(x)$ a un minimum relatif à $x = a$.

68

Erreur standard de p

- L'erreur standard de p est alors la racine carrée de l'inverse de la dérivée seconde de la vraisemblance L au point ML:

$$SE(p) = \sqrt{\frac{1}{\frac{d^2 L}{d p^2}}}$$

69

Effet de la taille de l'effectif sur SE

```
> k <- 17 # Nombre de A
> n <- 50 # Nombre total de tirage
> o <- optim(par = c(prob=0.1), fn=negdbinom,
size=n, x=k, log=TRUE, method = "Brent",
lower=0, upper=1, hessian = TRUE)
> sqrt(1/o$hessian)
0.0669921

> k <- 34 # Nombre de A
> n <- 100 # Nombre total de tirage
> o <- optim(par = c(prob=0.1), fn=negdbinom,
size=n, x=k, log=TRUE, method = "Brent",
lower=0, upper=1, hessian = TRUE)
> sqrt(1/o$hessian)
0.04737057
```

70

Ludwig Otto Hesse (22 avril 1811 à Königsberg, Prusse –
4 août 1874 à Munich, Allemagne)



Extension dans un système à n variables

- La matrice hessienne (ou simplement la hessienne), Hessian matrix en anglais, d'une fonction numérique f est la matrice carrée, notée $H(f)$, de ses dérivées partielles secondes.

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

En pratique, la fonction de vraisemblance n'est souvent pas dérivable et on utilise des approximations numériques de cette matrice.

71

Extension dans un système à n variables

- L'utilisation de la matrice hessienne permet de prendre en compte les covariances entre paramètres qui souvent ne sont pas nulles.

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

72

Extension dans un système à n variables

- Par exemple, ici, les covariances sont supposées nulles :
 - Girondot, M. (1999) Statistical description of temperature-dependent sex determination using maximum likelihood. *Evolutionary Ecology Research*, 1, 479-486.
 - Godfrey, M.H., Delmas, V. & Girondot, M. (2003) Assessment of patterns of temperature-dependent sex determination using maximum likelihood model selection. *Ecoscience*, 10, 265-272.
- Alors qu'en réalité elles ne le sont pas - corrigé ici :
 - Hulin, V., Delmas, V., Girondot, M., Godfrey, M.H. & Guillon, J.-M. (2009) Temperature-dependent sex determination and global change: Are some species at greater risk? *Oecologia*, 160, 493-506.

$$s_j^{-1} = \sqrt{\frac{\partial^2 \ln L(\theta_1, \dots, \theta_k)}{\partial \theta_j^2}} \quad (7)$$

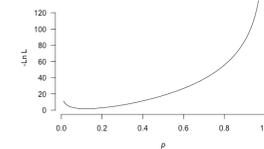
$$s_j^{-1} = \sqrt{\frac{\partial^2 \ln L(\theta_1, \dots, \theta_k)}{\partial \theta_j^2}} \quad (8)$$

73

Problèmes de cette méthode

- Si le profil de vraisemblance est plat autour du point ML, la dérivée seconde est nulle et SE ne peut pas être estimé.
 - Un pis-aller pour s'en sortir est de postuler que $1/0=0$. C'est ce que l'on appelle une matrice inversée généralisée (par exemple Moore-Penrose generalized inverse). Voir https://en.wikipedia.org/wiki/Moore-Penrose_inverse
- La méthode suppose que le profil de vraisemblance est symétrique de part et d'autre du point ML.

```
k <- 6
n <- 50
p <- seq(from=0.01, to=0.99,
        by=0.01)
lnL <- rdbinom(x=k, size=n, prob =
p, log=TRUE)
plot(p, lnL,
     xlab=expression(italic("p")),
     ylab="-ln L",
     ylim=c(0, 130),
     bty="n", las=1, type="l")
```



74

De plus...

- La fonction de vraisemblance, notée $L(x_1, \dots, x_n | \theta_1, \dots, \theta_k)$ est une fonction de probabilités conditionnelles qui décrit les valeurs x_i (observations) d'une loi statistique en fonction des paramètres θ_j
- Notez l'opérateur $|$ qui indique « sachant que ». Donc ici sachant le modèle... or c'est justement le modèle qu'on cherche. Donc ce qu'on veut c'est plutôt: $P(\theta_1, \dots, \theta_k | x_1, \dots, x_n)$ qu'on appelle la plausibilité
- Donc quand on utilise le maximum de vraisemblance, on prend le problème à l'envers !

Lê Nguyễn, H. *La Formule du Savoir: Une philosophie unifiée du savoir fondée sur le théorème de Bayes*; Paris, France, 2018.

75

Deux types de statistiques

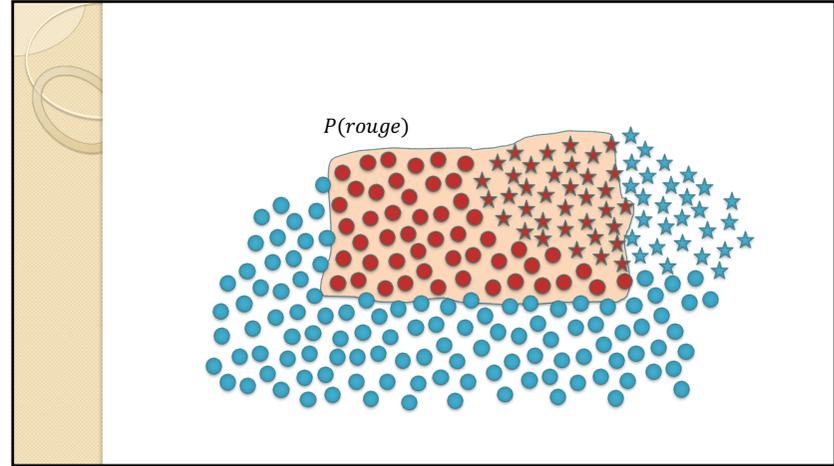
- | | |
|---|---|
| <ul style="list-style-type: none"> Statistiques fréquentistes On mesure la probabilité des observations selon un modèle $p(x_1, \dots, x_n \theta_1, \dots, \theta_k)$ | <ul style="list-style-type: none"> Statistiques bayésiennes On mesure la probabilité d'un modèle selon des observations $p(\theta_1, \dots, \theta_k x_1, \dots, x_n)$ |
|---|---|

Voir par exemple: <https://www.youtube.com/watch?v=x-2uVNze56s>

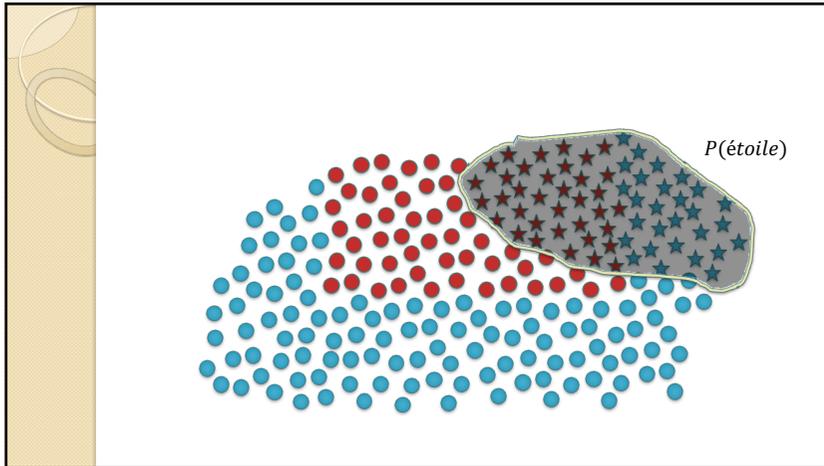
76

LE BAYÉSIEN

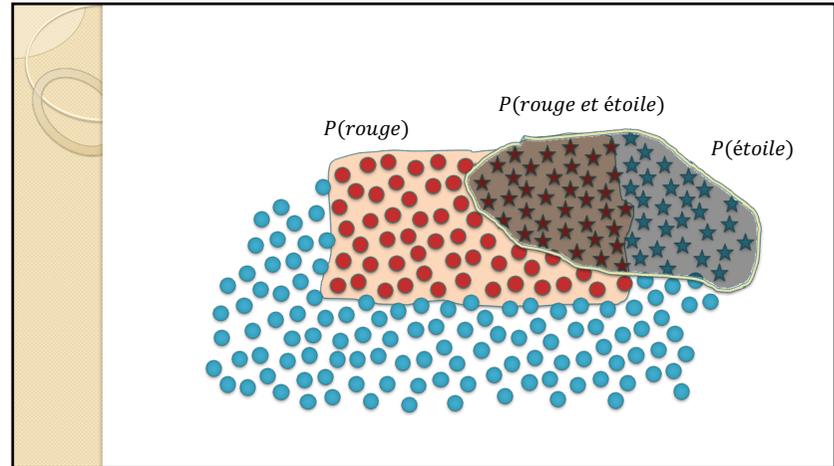
77



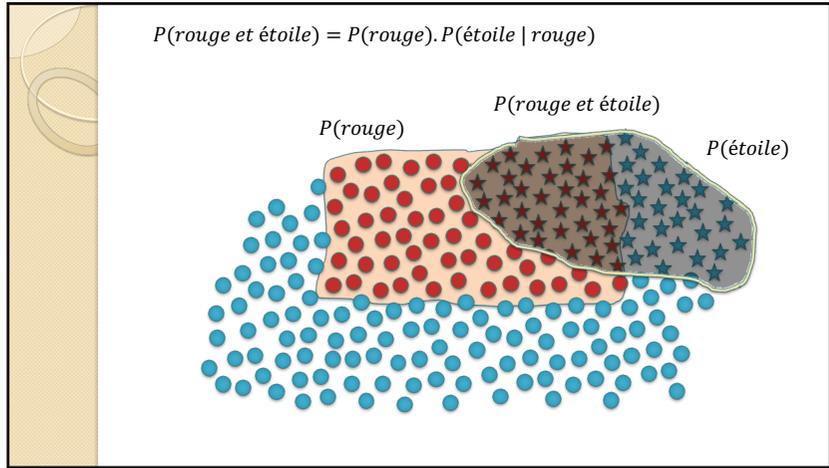
78



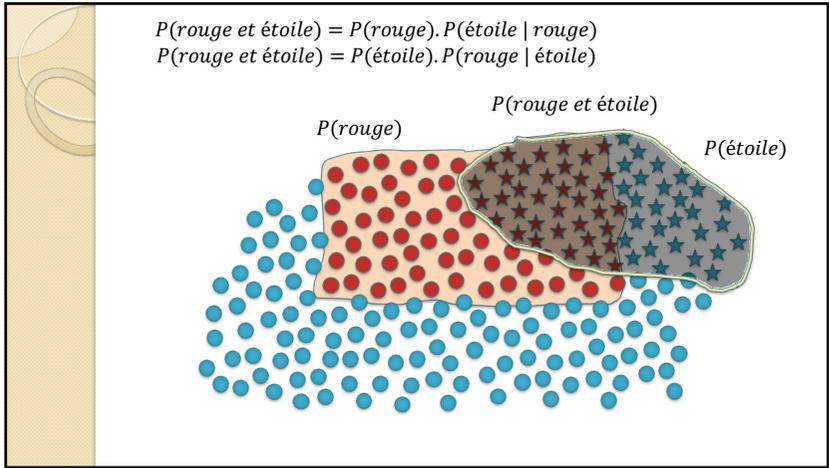
79



80



81



82

Règle de Bayes puis de Laplace

$P(\text{rouge et étoile}) = P(\text{rouge}) \cdot P(\text{étoile} | \text{rouge})$
 $P(\text{rouge et étoile}) = P(\text{étoile}) \cdot P(\text{rouge} | \text{étoile})$

$P(A \text{ et } B) = P(A) \cdot P(B | A)$
 $P(A \text{ et } B) = P(B) \cdot P(A | B)$

$P(B) \cdot P(A | B) = P(A) \cdot P(B | A)$

$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$



Thomas Bayes (1702-1761)



Pierre Simon Laplace, 1774
 23 mars 1749, Beaumont-en-Auge
 5 mars 1827, Paris

Thomas Bayes (1764). Essay Towards Solving a Problem in the Doctrine of Chances (1764). Philosophical Transactions of the Royal Society of London
 Pierre Simon Laplace. *Théorie Analytique des Probabilités*, Second Edition; Mme Veuve Courcier: Paris, France, 1814.

83

Règle de Bayes

- Permet de calculer la probabilité d'un événement (le *posterior*) en prenant en compte des données $p(Y/q)$ et une information que l'on a sur cet événement (le *prior*).

84

La bayésien en pratique

- Pendant longtemps, on n'a pas **su** utiliser le bayésien car on ne savait pas calculer $p(Y)$ sauf dans des cas très simples.
- Pendant longtemps, on n'a pas **voulu** utiliser le bayésien à cause d'une incompréhension de ce que sont les *priors*, notamment en relation avec l'objectivité scientifique qui date de la méthode scientifique vue par Roger Bacon (1220-1292).
- Et puis le MCMC bayésien a été inventé:
 - Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **1953**, *21*, 1087-1092.
 - Hastings, W.K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **1970**, *57*, 97-109.

85

Chaîne de Markov

Andrei A. Markov (1866-1922)



- Un processus de Markov est un processus stochastique possédant la propriété de Markov :
 - l'information utile pour la prédiction du futur est contenue dans l'état présent du processus.



86

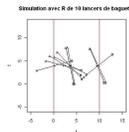
Monte Carlo est un quartier de Monaco, une ville-État.

Monte Carlo



- Le terme méthode de Monte-Carlo désigne une famille de méthodes algorithmiques visant à calculer une valeur numérique approchée en utilisant des procédés aléatoires, c'est-à-dire des techniques probabilistes.

L'aiguille de Buffon est une expérience de probabilité proposée en 1733 par Georges-Louis Leclerc de Buffon, au XVIIIe siècle. Cette expérience fournit une approximation du nombre π .

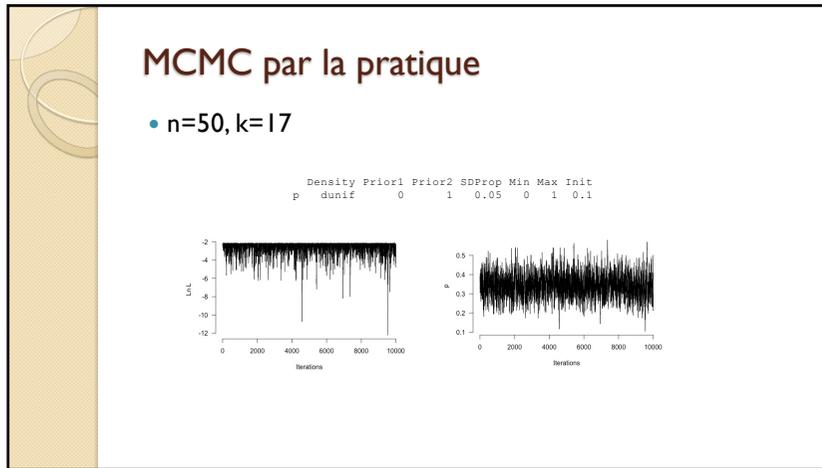


87

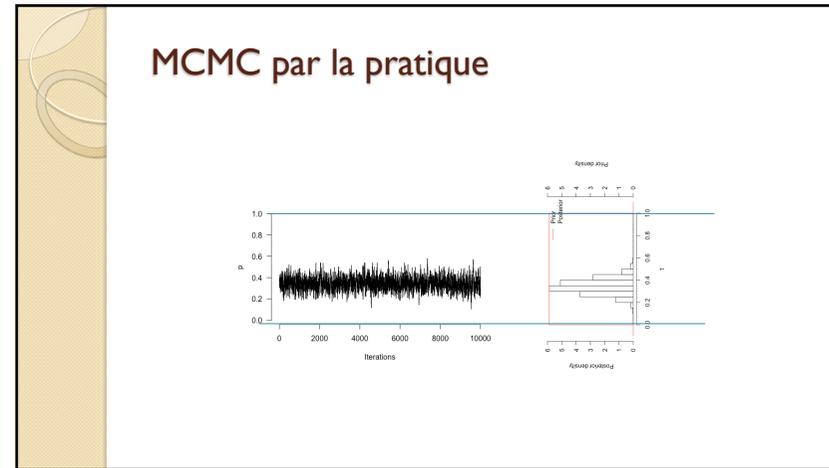
Revenons à notre exemple

- On cherche la distribution p . On a les données $n=50$ et $k=17$. On doit choisir un *prior* pour p . qu'a-t-on comme information sur p ? Rien. Donc on choisira un *prior* de p dans une loi uniforme comprise dans $]0, 1[$.
- On commence une chaîne de Markov avec une valeur $p_0=0.1$ par exemple.
- On calcule la vraisemblance L_0 . On tire au hasard une valeur p_1 (le processus de Monte-Carlo). On calcule la vraisemblance L_1 avec cette valeur. Si L_1 est meilleure que L_0 , on la garde sinon on peut quand même la garder avec une probabilité dépendant de sa probabilité *posterior* et de celle du point précédent: si elle est vraiment très peu probable, on a peu de chance de la garder mais ce n'est jamais nul quand même (algorithme de Metropolis-Hastings). Cela génère une chaîne de Markov.

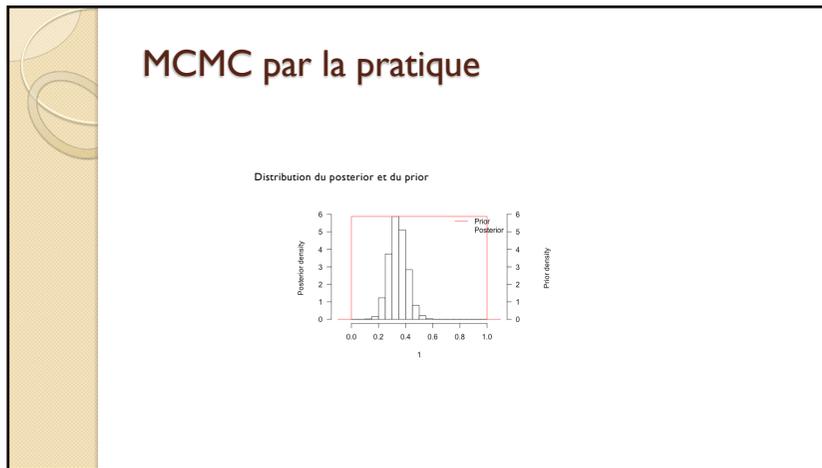
88



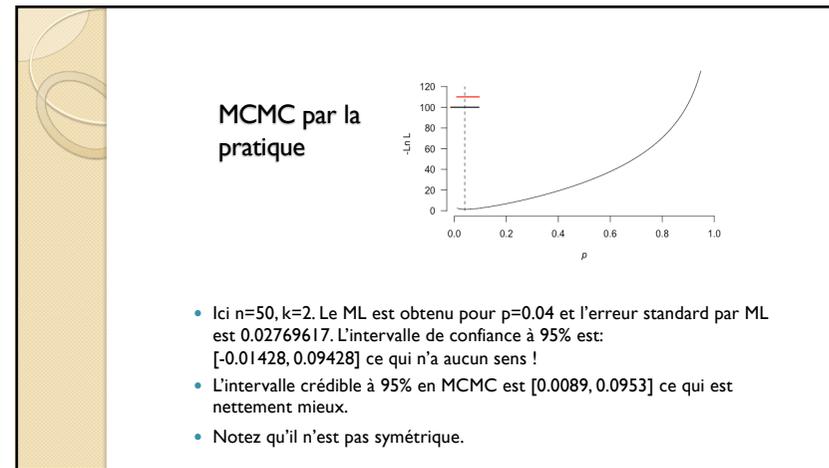
89



90



91



92

Partie 4

PRENDRE EN COMPTE L'INCERTITUDE DES ENTRÉES

93

Soit un modèle M

- Dans ce modèle, il y a k paramètres qui sont connus avec une certaine incertitude, soit décrite par une erreur standard, soit par une distribution interquartiles, soit une distribution de valeurs possibles.
- On veut savoir comment cette incertitude se reflète sur l'incertitude des sorties.
- On utilisera une erreur standard si l'incertitude est distribuée normalement ou au moins est symétrique, on utilisera l'intervalle interquartiles pour une distribution asymétrique et une liste de valeurs pour un résultat de MCMC.

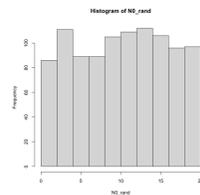
94

Modèle logistique

- Le modèle logistique est décrit par 3 paramètres, N_0 , r et K .
- A chacun de ces paramètres est associé une incertitude. On va générer un set de m valeurs tirées dans la gamme des possibles.
- Par exemple:

```
N0_rand <- runif(1000, min=0.1, max=20)
hist(N0_rand)
```

Attention au biais sur l'histogramme car la première colonne n'a pas la même largeur que les suivantes car elle n'inclut que des valeurs $> 0,1$ et non >0 .



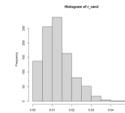
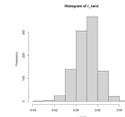
95

Une incertitude tirée d'une distribution normale...

- Si l'écart-type est du même ordre de grandeur que la moyenne, on est obligé de tronquer la distribution pour ne pas avoir de valeurs négatives:

```
r_rand <- rnorm(1000, mean=0.01, sd=0.01)
hist(r_rand)
while(any(r_rand <= 0)) {
  pos <- which(r_rand <= 0)
  r_rand[pos] <- rnorm(length(pos), mean=0.01, sd=0.001)
}
hist(r_rand)
mean(r_rand); sd(r_rand)
```

Attention, la moyenne et le SD de la distribution tronquée sont biaisés (la moyenne est plus forte que mean et l'écart-type plus faible que sd).
Ce n'est pas la bonne façon de procéder. On va plutôt utiliser des distributions toujours positives.



96

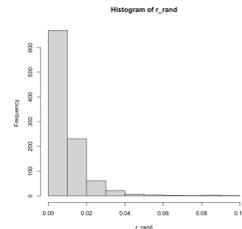
Avec une distribution lognormale

```
# Les valeurs que je veux dans une lognormale; moyenne et variance
mean <- 0.01
v <- 0.01^2

# Les paramètres de la lognormale nécessaires - démonstration
# mu + sigma^2/2=log(mean)
# 2 mu + sigma^2=2 log(mean)
# log(exp(sigma^2)-1)+ 2 mu + sigma^2 = log(v)
# log(exp(sigma^2)-1) + 2 log(mean) = log(v)
# log(exp(sigma^2)-1) = log(v) - 2 log(mean)
# exp(sigma^2)-1 = exp(log(v) - 2 log(mean))
# exp(sigma^2) = exp(log(v) - 2 log(mean)) + 1
# exp(sigma^2) = v/mean^2 + 1
# sigma^2 = log(v/mean^2 + 1)

sigma <- sqrt(log(v / mean ^ 2 + 1))
mu <- log(mean) - sigma ^ 2 / 2

r_rand <- rlnorm(1000, meanlog = mu, sdlog=sigma)
hist(r_rand)
mean(r_rand); sd(r_rand)
```



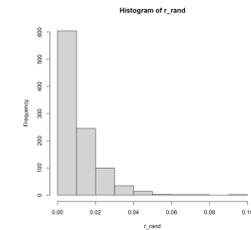
97

Avec une distribution gamma

```
# Les valeurs que je veux dans une distribution gamma
mean <- 0.01
v <- 0.01^2

# Les paramètres de la gamma nécessaires - démonstration
# shape * scale = mean
# shape * scale^2 = v
# shape = mean / scale
# v = scale^2 * mean / scale
# v = scale * mean
# scale = v / mean
# shape = mean / 1 / (v / mean)
# shape = mean^2 / v

r_rand <- rgamma(1000, shape=mean^2 / v, scale = v / mean)
hist(r_rand)
mean(r_rand); sd(r_rand)
```



98

Finalement

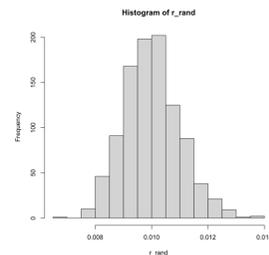
- On va prendre une distribution gamma avec une variance plus faible pour limiter un peu la variation de r:

```
# Les valeurs que je veux dans une gamma
mean <- 0.01
v <- 0.001^2

# Les paramètres de la gamma nécessaires
r_rand <- rgamma(1000, shape=mean^2 / v,
                scale = v / mean)

hist(r_rand)
mean(r_rand); sd(r_rand)
```

Notez la forme de la distribution très différente en fonction de SD.



99

K tiré dans une distribution de Cauchy

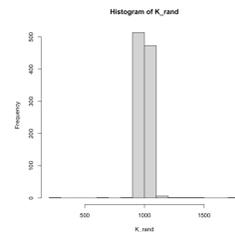
- La distribution de Cauchy ressemble à une distribution normale mais a une queue plus grosse. Elle peut être utile quand on veut définir un prior symétrique qui n'exclut pas trop des valeurs extrêmes. Mais attention, le fait qu'elle ait de grosses queues fait que si on l'utilise pour générer une distribution de nombre aléatoire, on peut avoir facilement des nombres négatifs.
- La distribution de Cauchy apparaît naturellement quand on fait le ratio de deux distributions normales.

[https://fr.wikipedia.org/wiki/Loi_de_Cauchy_\(probabilités\)](https://fr.wikipedia.org/wiki/Loi_de_Cauchy_(probabilités))

100

K tiré dans une distribution de Cauchy

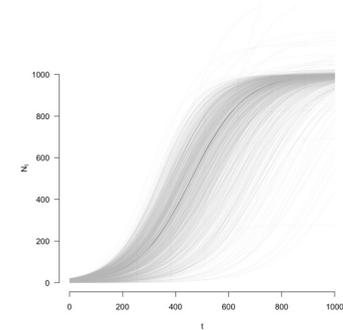
```
K_rand <- rcauchy(1000, location=1000, scale=2)
hist(K_rand)
# plot(density(K_rand))
while (any(K_rand <= 0)) {
  pos <- which(K_rand <= 0)
  K_rand[pos] <- rcauchy(length(pos),
                        location=1000,
                        scale=2)
}
hist(K_rand)
```



101

Et finalement, on affiche les 1000 réplicats

```
plot(x = c(1:1000), y=logistique(1:1000,
NO=10, r=0.01, K=1000), type="l",
bty="n",
      las=1, xlab="t",
ylab=bquote("N"[t]), ylim=c(0, 1100))
mat <- matrix(data = NA, nrow = 1000,
ncol=1000)
for (i in 1:1000) {
  K_ec <- K_rand[i]
  r_ec <- r_rand[i]
  NO_ec <- NO_rand[i]
  x <- 1:1000
  y <- logistique(x, NO=NO_ec, r=r_ec,
K=K_ec)
  lines(x=x, y=y, col=rgb(red = 0.7, green
= 0.7, blue=0.7, alpha = 0.1))
  mat[i, ] <- y
}
```



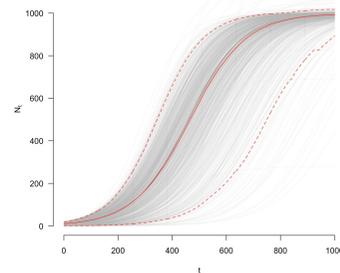
102

Et l'intervalle interquantiles à 95%

```
# Là on calcule l'intervalle
interquantiles des sorties
md <- rep(NA, 1000)
lw <- rep(NA, 1000)
up <- rep(NA, 1000)

for (t in 1:1000) {
  md[t] <- median(mat[, t])
  lw[t] <- quantile(mat[, t], prob=0.025)
  up[t] <- quantile(mat[, t], prob=0.975)
}

lines(x=x, y=md, col="red")
lines(x=x, y=lw, col="red", lty=2)
lines(x=x, y=up, col="red", lty=2)
```



103

Partie 5

ANALYSE DE SENSIBILITÉ ET D'ÉLASTICITÉ

104

Introduction

- En étudiant comment la réponse du modèle réagit aux variations de ses variables d'entrée, l'analyse de sensibilité permet de répondre à un certain nombre de questions.
 - 1. Le modèle est-il bien fidèle au phénomène/processus modélisé ?
 - 2. Quelles sont les variables qui contribuent le plus à la variabilité de la réponse du modèle ?
 - 3. Quelles sont au contraire les variables les moins influentes ?
 - 4. Quelles variables, ou quels groupes de variables, interagissent avec quelles (quels) autres ?

Jacques J. (2001) Pratique de l'analyse de sensibilité : comment évaluer l'impact des entrées aléatoires sur la sortie d'un modèle mathématique. *PUB. IRMA*, 71 (II), 1-19.

105

Pourquoi une analyse de sensibilité ?

- 1. Le modèle est-il bien fidèle au phénomène/processus modélisé ?
 - En effet, si l'analyse exhibe une forte influence d'une variable d'entrée habituellement connue comme non influente, il sera nécessaire de remettre en cause la qualité du modèle ou (et) la véracité de nos connaissances sur l'impact réel des variables d'entrée.

106

Pourquoi une analyse de sensibilité ?

- 2. Quelles sont les variables qui contribuent le plus à la variabilité de la réponse du modèle ?
 - Si cette variabilité est synonyme d'imprécision sur la valeur prédite de la sortie, il sera alors possible d'améliorer la qualité de la réponse du modèle à moindre coût en concentrant les efforts sur la réduction des variabilités des entrées les plus influentes.
 - Mais cela n'est pas toujours possible, notamment lorsque la variabilité d'une variable d'entrée est intrinsèque à la nature de la variable et non due à un manque d'information ou à des imprécisions de mesures.

107

Pourquoi une analyse de sensibilité ?

- 3. Quelles sont au contraire les variables les moins influentes ?
 - Il sera possible de les considérer comme des paramètres déterministes, en les fixant par exemple à leur espérance, et obtenir ainsi un modèle plus léger avec moins de variables d'entrée.

108

Pourquoi une analyse de sensibilité ?

- 4. Quelles variables, ou quels groupes de variables, interagissent avec quelles (quels) autres ?
 - L'analyse de sensibilité peut permettre de mieux appréhender et comprendre le phénomène modélisé, en éclairant les relations entre les variables d'entrée.

109

Contexte général

- Elasticité et sensibilité sont deux termes utilisés pour décrire la relation qui existe entre une ou des variables d'entrée d'un modèle et la ou les sorties de ce modèle.

110

Origine des concepts

- À la fin du XVII^e siècle, en particulier, Gregory King, puis Charles D'Avenant notent qu'une baisse de l'offre de blé conduit à un renchérissement du prix de cette denrée qui n'est pas en proportion de la baisse.

Gregory King (né le 15 décembre 1648 - mort le 29 août 1712) est un haut fonctionnaire britannique et l'un des premiers grands statisticiens économiques du monde moderne.
Charles d'Avenant (1656-1714), économiste anglais



111

Attention, en marketing ou physique

- L'**élasticité** des prix est un principe économique créé en 1890 par l'économiste Alfred Marshall. Lorsque le prix d'un produit change, la volonté des clients de l'acheter change également.
- La **sensibilité** au prix, quant à elle, indique dans quelle mesure un client est sensible aux variations de prix et s'avère précieuse pour comparer différents segments afin de déterminer lequel, le cas échéant, est le plus sensible au prix.
- On trouve aussi des concepts d'élasticité ou de sensibilité en physique des matériaux sans qu'il n'y ait de relation avec ceux décrits ci-après.

112

Analyse de type "One at a time" (OAT)

ELASTICITÉ

113

Sensibilité (*sensitivity* en anglais)

- La sensibilité est la variation d'une variable y par rapport à la variation d'une variable x .

$$S(y, x) = \frac{\Delta(y)}{\Delta(x)}$$

- Notez que la sensibilité a l'unité de celle de $y \cdot x^{-1}$

Attention, faux ami : En anglais, « *sensitivity* » est « an understanding of or ability to decide about what is good or valuable, especially in connection with artistic or social activities ».

114

Elasticité (*elasticity* en anglais)

- L'élasticité est le coefficient de variation d'une variable y par rapport à la variation d'une variable x .
- Le pourcentage de variation de y se notant $\frac{\Delta(y)}{y} \times 100$ et celui de x se notant $\frac{\Delta(x)}{x} \times 100$, on obtient, pour x et y non nuls, la formule suivante (les 100 se simplifient) :

$$E(y, x) = \frac{\frac{\Delta(y)}{y}}{\frac{\Delta(x)}{x}} = \frac{x}{y} \cdot \frac{\Delta(y)}{\Delta(x)}$$

- Notez que l'élasticité n'a pas d'unité.

115

Analyse de sensibilité appliquée aux modèles matriciels

- Un outil important dans l'analyse des modèles matriciels de population consiste à comprendre comment les probabilités de transition et de permanence de chaque classe affectent la croissance de la population. Les valeurs qui expriment cela sont appelés sensibilité et élasticité.
- Ce sont des outils tant pour la compréhension des différentes stratégies de cycle biologique que pour la gestion des populations menacées, voire pour leur utilisation durable.

http://ecovirtual.lib.usp.br/doku.php?id=en:ecovirt:roteiro:pop_str:psr_mtr

116

Elasticité et sensibilité

- La sensibilité est la contribution directe de chaque transition à λ . En termes mathématiques, la sensibilité de l'élément a_{ij} de la matrice de projection A correspond à la variation de λ due à une petite variation de cet élément ($\delta\lambda/\delta a_{ij}$).
- L'élasticité, en revanche, est une mesure de sensibilité proportionnelle à l'effet. Elle permet de déterminer si un élément de transition peut doubler et avoir peu d'impact sur le lambda, ou si cet élément peut doubler et avoir un effet fort sur le lambda, quelles que soient les valeurs initiales.

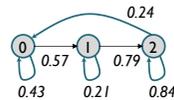
117

Exemple pour un modèle matriciel

- $S_{i,j} = \frac{\lambda_{perturbé} - \lambda_{original}}{a_{perturbé\ i,j} - a_{original\ i,j}}$
- Pour le calcul de l'élasticité, il suffit de diviser chaque différence de la formule ci-dessus par la valeur d'origine afin que les différences soient exprimées en proportion:
- $E_{i,j} = \frac{(\lambda_{perturbé} - \lambda_{original}) / \lambda_{original}}{(a_{perturbé\ i,j} - a_{original\ i,j}) / a_{original\ i,j}}$
- Une autre solution consiste à multiplier la valeur de la sensibilité par le rapport entre le taux original et le lambda original :
- $E_{i,j} = \frac{a_{original\ i,j}}{\lambda_{original}} S_{i,j}$
- Notez que, pour le calcul matriciel, on peut calculer les dérivées de façon analytique sans faire d'approximation.

118

Exemple « Girafe »



- Soit une matrice

```
mat <- matrix(c(0, 0, 0.24, 0.57, 0, 0, 0, 0.79, 0.84),
             nrow=3, byrow=TRUE,
             dimnames=list(c("calf", "subadult", "adult"),
                          c("calf", "subadult", "adult")))
```

	calf	subadult	adult
calf	0.00	0.00	0.24
Subadult	0.57	0.00	0.00
adult	0.00	0.79	0.84

Reproduction, Croissance + survie, Survie

<https://compadre-db.org/Education/article/sensitivity-and-elasticity-matrices>

119

Sensibilité

Quand on ajoute l'option zero=FALSE, on considère que les cases où il n'y a pas de flèche pourraient en avoir.

```
Library(popbio)
sensitivity(mat, zero = FALSE)
>
> calf      calf      subadult      adult
> calf      0.09871191 0.05874458 0.3939444
> subadult  0.16587133 0.09871191 0.6619676
> adult     0.20110408 0.11967932 0.8025762
```

```
sensitivity(mat, zero = TRUE)
>
> calf      calf      subadult      adult
> calf      0.00000000 0.00000000 0.3939444
> subadult  0.1658713 0.00000000 0.0000000
> adult     0.00000000 0.1196793 0.8025762
```

Reproduction, Croissance + survie, Survie

120

Élasticité

```
Library(popbio)
(emat <- elasticity(mat))
>
> calf      0.00000000 0.00000000 0.09871191
> subadult 0.09871191 0.00000000 0.00000000
> adult     0.00000000 0.09871191 0.70386427
```

- La somme des élasticités fait 1; ce sont des valeurs relatives.

```
> sum(emat)
[1] 1
```

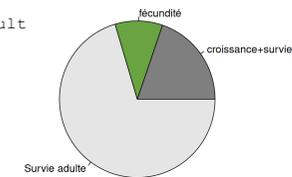
Reproduction, Croissance + survie, Survie

121

Elasticité

```
> calf      calf  subadult  adult
> calf      0.00000000 0.00000000 0.09871191
> subadult 0.09871191 0.00000000 0.00000000
> adult     0.00000000 0.09871191 0.70386427
```

```
emat <- elasticity(mat)
sumfec <- sum(emat[1,3])
sumgrow <- sum(emat[2,1],emat[3,2])
sumsurvieadult <- emat[3,3]
par(mar=c(2, 2, 2, 2))
pie(c(sumgrow,sumfec,sumsurvieadult), col = c("gray50", "#6AA341",
"gray90"),
labels=c("croissance+survie",
"fécundité", "Survie adulte"))
```



122

Cas général de calcul de l'élasticité

- Soit un modèle avec k paramètres. On fait varier chacun des k paramètres de $\pm p\%$ (0,1%, 1% ou 10%) et on regarde de combien varie la sortie du modèle.
- On fait alors le rapport des taux de variation qui est l'élasticité, une valeur sans unité.
- En général on choisit p% petit pour se rapprocher de l'élasticité en un point particulier.

123

Exemple avec le modèle logistique

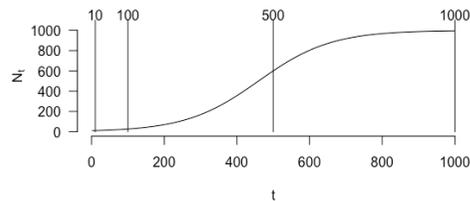
$$N_t = \frac{N_0 K}{N_0 + e^{-rt} (K - N_0)}$$

- On va chercher à évaluer l'impact de K, r et N_0 sur N_{10} , N_{100} , N_{500} et N_{1000} avec $K=1000$, $r=0,01$ et $N_0=10$.

124

Etat initial

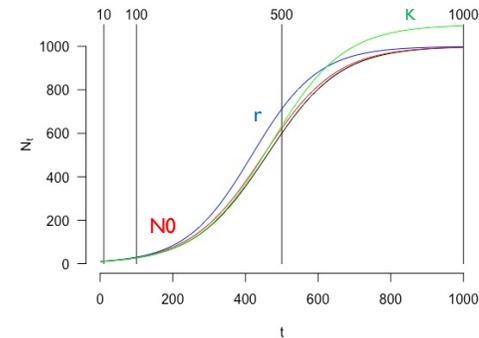
```
> logistique <- function(t, N0, r, K) {(N0*K)/(N0+exp(-r*t)*(K-N0))}
> plot(x = c(1:1000), y=logistique(1:1000, N0=10, r=0.01, K=1000), type="l",
      bty="n", las=1, xlab="t", ylab=bquote("N[t]"), ylim=c(0, 1100))
> segments(x0=c(10, 100, 500, 1000), y0=0, y1=1100)
> par(xpd=TRUE); text(x=c(10, 100, 500, 1000), y=1150,
                    labels = c("10", "100", "500", "1000"))
```



125

Effet des paramètres +10%

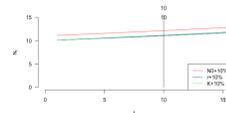
Notez que le +10% a été choisi pour que l'effet soit visible sur le graphique.



126

Elasticité pour t=10 avec +1%

```
> N10Ini <- logistique(10, N0=10, r=0.01, K=1000)
> N10_N0_1pc <- logistique(10, N0=10+10*0.01, r=0.01, K=1000)
> Dyy <- (N10_N0_1pc-N10Ini)/N10Ini
> Dyy/0.01
Nb [1] 0.9989389
> N10_r_1pc <- logistique(10, N0=10, r=0.01+0.01*0.01, K=1000)
> Dyy <- (N10_r_1pc-N10Ini)/N10Ini
> Dyy/0.01
r [1] 0.09894436
> N10_K_1pc <- logistique(10, N0=10, r=0.01, K=1000+0.01*1000)
> Dyy <- (N10_K_1pc-N10Ini)/N10Ini
> Dyy/0.01
K [1] 0.001040213
```



127

Synthèse

- Si l'élasticité >0 , cela signifie qu'une augmentation sur le paramètre induit une augmentation sur la sortie
- Si l'élasticité <0 , cela signifie qu'une augmentation sur le paramètre induit une baisse sur la sortie

128

Synthèse, si l'élasticité est >0

- Si l'élasticité = 0, cela signifie qu'1% d'augmentation sur le paramètre n'induit aucun changement sur la sortie
- Si l'élasticité est comprise entre 0 et 1, cela signifie qu'1% d'augmentation conduit à une augmentation de moins de 1% en sortie.
- Si l'élasticité = 1, cela signifie qu'1% d'augmentation sur le paramètre induit 1% d'augmentation sur la sortie
- Si l'élasticité est supérieure à 1, cela signifie qu'1% d'augmentation conduit à une augmentation de plus de 1% en sortie.

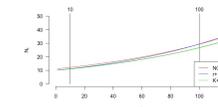
129

Elasticité pour $t=100$ avec +1%

```

> N100Ini <- logistique(100, N0=10, r=0.01, K=1000)
> N100_N0_1pc <- logistique(100, N0=10+10*0.01, r=0.01, K=1000)
> Dyy <- (N100_N0_1pc-N100Ini)/N100Ini
> (eN0_t100 <- Dyy/0.01)
Nb [1] 0.9829414
> N100_r_1pc <- logistique(100, N0=10, r=0.01+0.01*0.01, K=1000)
> Dyy <- (N100_r_1pc-N100Ini)/N100Ini
> (er_t100 <- Dyy/0.01)
r [1] 0.9778964
> N100_K_1pc <- logistique(100, N0=10, r=0.01, K=1000+0.01*1000)
> Dyy <- (N100_K_1pc-N100Ini)/N100Ini
> (eK_t100 <- Dyy/0.01)
K [1] 0.0167281

```



130

Elasticité pour $t=500$ avec +1%

```

> N500Ini <- logistique(500, N0=10, r=0.01, K=1000)
> N500_N0_1pc <- logistique(500, N0=10+10*0.01, r=0.01, K=1000)
> Dyy <- (N500_N0_1pc-N500Ini)/N500Ini
> (eN0_t500 <- Dyy/0.01)
Nb [1] 0.4017883
> N500_r_1pc <- logistique(500, N0=10, r=0.01+0.01*0.01, K=1000)
> Dyy <- (N500_r_1pc-N500Ini)/N500Ini
> (er_t500 <- Dyy/0.01)
r [1] 1.99035
> N500_K_1pc <- logistique(500, N0=10, r=0.01, K=1000+0.01*1000)
> Dyy <- (N500_K_1pc-N500Ini)/N500Ini
> (eK_t500 <- Dyy/0.01)
K [1] 0.5934193

```

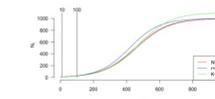
131

Elasticité pour $t=1000$ avec +1%

```

> N1000Ini <- logistique(1000, N0=10, r=0.01, K=1000)
> N1000_N0_1pc <- logistique(1000, N0=10+10*0.01, r=0.01, K=1000)
> Dyy <- (N1000_N0_1pc-N1000Ini)/N1000Ini
> (eN0_t1000 <- Dyy/0.01)
Nb [1] 0.00447513
> N1000_r_1pc <- logistique(1000, N0=10, r=0.01+0.01*0.01, K=1000)
> Dyy <- (N1000_r_1pc-N1000Ini)/N1000Ini
> (er_t1000 <- Dyy/0.01)
r [1] 0.04259847
> N1000_K_1pc <- logistique(1000, N0=10, r=0.01, K=1000+0.01*1000)
> Dyy <- (N1000_K_1pc-N1000Ini)/N1000Ini
> (eK_t1000 <- Dyy/0.01)
K [1] 0.9954353

```



132

Choix de la déviation

- Le choix de l% n'est pas neutre:

```
> N500_r_1pc <- logistique(500, N0=10, r=0.01+0.01*0.01, K=1000)
> Dyy <- (N500_r_1pc-N500Ini)/N500Ini
> (er_t500 <- Dyy/0.01)
[1] 1.99035
> N500_r_10pc <- logistique(500, N0=10, r=0.01+0.1*0.01, K=1000)
> Dyy <- (N500_r_10pc-N500Ini)/N500Ini
> Dyy/0.1
[1] 1.868633
```

133

Symétrie de l'effet

- L'effet n'est pas strictement symétrique:

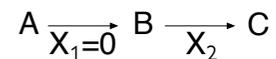
```
> N500_r_10pc <- logistique(500, N0=10, r=0.01+0.1*0.01, K=1000)
> Dyy <- (N500_r_10pc-N500Ini)/N500Ini
> Dyy/0.1
[1] 1.868633
> N500_r_10pc <- logistique(500, N0=10, r=0.01-0.1*0.01, K=1000)
> Dyy <- (N500_r_10pc-N500Ini)/N500Ini
> Dyy/-0.1
[1] 2.060843
```

On pourra prendre l'élasticité moyenne pour $\pm x\%$

134

Autres problèmes (1)

- Comment prendre en compte les interactions entre paramètres ?
 - Exemple: Le changement d'un paramètre ne change pas la sortie du modèle mais quand il change alors qu'un autre change aussi, l'effet est très important.
 - On trouvera cet effet lorsqu'il y a des chaînes de causalité:



- Changer X_2 n'aura aucun effet car B est absent, mais changer à la fois X_1 et X_2 pourra avoir un effet très fort.

135

Autres problèmes (2)

- Comment prendre en compte l'incertitude sur les paramètres ?
 - Cette méthode calcule l'incertitude en un point mais ce point est lui-même souvent mal connu et on risque d'explorer l'élasticité dans une zone de l'espace des paramètres qui ne correspond pas à la zone où se trouve la vraie valeur.

136

Gérer l'incertitude en OAT

- Morris (1991) propose d'explorer l'espace des paramètres toujours dans une stratégie OAT pour y détecter trois comportements de variables:
 - Variables sans effet sur la sortie
 - Variables avec effet fort sur la sortie
 - Variables avec effet non-linéaire

M.D. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33:161-174, 1991.

Mais en pratique ça ne résout pas les problèmes décrits précédemment.

137

Conclusion

- C'est souvent la première analyse que l'on fait car elle est très peu coûteuse en temps de calcul.
- Mais il ne faut pas s'arrêter là car on risque de passer à côté d'effets importants ou intéressants.

138

ANALYSE DE SENSIBILITÉ BASÉE SUR LA DÉCOMPOSITION DE LA VARIANCE

Sobol I.M. (1990) Sensitivity estimates for nonlinear mathematical models. *Matematicheskoe Modelirovanie*, 2, 112-118.
Sobol I.M. (1993) Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiments*, 1, 407-414.
Sobol I.M. (2001) Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55, 271-280.

139

Sensibilité de Sobol

- Une nouvelle méthode basée sur la décomposition d'une fonction a été proposée en 1990 par I. M. Sobol, un mathématicien russe.
- Cette méthode est devenue la méthode de référence même si elle est encore assez peu utilisée:
 - **Nossent J., Elsen P. and Bauwens W.** (2011) Sobol' sensitivity analysis of a complex environmental model. *Environmental Modelling & Software*, 26(12), 1515-1525.
 - **Chevallier, D., Mourrain, B. & Girondot, M.** (2020) Modelling leatherback biphasic indeterminate growth using a modified Gompertz equation. *Ecological Modelling*, 426, 109037



Илья Меерович Соболев

140

Rappels

- Soit une variable aléatoire x de variance s_x^2

- Alors $s_{x+k}^2 = s_x^2$
et $s_{k \cdot x}^2 = k^2 \cdot s_x^2$

```
> x <- rnorm(20, mean=10, sd=2)
> sd(x)^2
[1] 2.12796
> sd(10+x)^2
[1] 2.12796
> sd(10*x)^2
[1] 212.79
```

141

Rappels

- Soit une variable aléatoire x_1 de variance $s_{x_1}^2$ et x_2 de variance $s_{x_2}^2$, la covariance de x_1 et x_2 étant nulle:
- Alors $s_{x_1+x_2}^2 = s_{x_1}^2 + s_{x_2}^2$
et $s_{k_1 \cdot x_1 + k_2 \cdot x_2}^2 = k_1^2 \cdot s_{x_1}^2 + k_2^2 \cdot s_{x_2}^2$
- La variance totale est $s_{k_1 \cdot x_1 + k_2 \cdot x_2}^2$ qui peut être décomposée en deux termes $k_1^2 \cdot s_{x_1}^2$ et $k_2^2 \cdot s_{x_2}^2$
- $k_1^2 \cdot s_{x_1}^2 / (s_{k_1 \cdot x_1 + k_2 \cdot x_2}^2)$ est la contribution de x_1 à la variance totale et $k_2^2 \cdot s_{x_2}^2 / (s_{k_1 \cdot x_1 + k_2 \cdot x_2}^2)$ est la contribution de x_2 à la variance totale avec leur somme égale à 1.

142

Généralisons

- Supposons que le modèle étudié soit linéaire, et qu'il s'écrive sous la forme suivante :

$$Y = \beta_0 + \sum_{i=1}^p \beta_i X_i.$$

- Si les variables X_i sont supposées indépendantes, la variance de Y s'écrit alors :

$$V(Y) = \sum_{i=1}^p \beta_i^2 V(X_i).$$

Et $\beta_i^2 V(X_i)$ est la part de la variance totale due à la variable X_i .

143

L'indice SRC

- La sensibilité de Y à X_i peut donc simplement être quantifiée par le rapport de la part de variance due à X_i sur la variance totale.
- On définit ainsi l'indice de sensibilité SRC (Standardized Regression Coefficient) :

$$SRC_i = \frac{\beta_i^2 V(X_i)}{V(Y)}.$$

144

L'apport de Ilya Sobol

- Un modèle est décrit comme $Y=f(\mathbf{X})$ avec \mathbf{X} un vecteur $\{X_1, X_2, \dots, X_d\}$.
- Sobol (1990, 1993) a proposé une méthode pour écrire cette fonction $f(\mathbf{X})$ comme étant une décomposition linéaire :

$$f(\mathbf{X}) = f_0 + \sum_{i=1}^d f_i(X_i) + \sum_{i<j}^d f_{ij}(X_i, X_j) + \dots + f_{1,2,\dots,d}(X_1, X_2, \dots, X_d)$$

avec $X_i \in [0, 1]$ et $i=1, 2, \dots, d$

\mathbf{X} représente donc un hypercube de côté 1.

145

f comme une somme de fonctions

- La fonction f du modèle peut donc être décomposée en une somme de fonctions de dimensions croissantes :

$$Y = f(X_1, \dots, X_p) \\ = f_0 + \sum_{i=1}^p f_i(X_i) + \sum_{1 \leq i < j \leq p} f_{ij}(X_i, X_j) + \dots + f_{1, \dots, p}(X_1, \dots, X_p).$$

où

$$\begin{aligned} f_0 &= \mathbf{E}[Y], \\ f_i(X_i) &= \mathbf{E}[Y|X_i] - \mathbf{E}[Y], \\ f_{i,j}(X_i, X_j) &= \mathbf{E}[Y|X_i, X_j] - \mathbf{E}[Y|X_i] - \mathbf{E}[Y|X_j] + \mathbf{E}[Y], \\ f_{i,j,k}(X_i, X_j, X_k) &= \mathbf{E}[Y|X_i, X_j, X_k] - \mathbf{E}[Y|X_i, X_j] - \mathbf{E}[Y|X_i, X_k] \\ &\quad - \mathbf{E}[Y|X_j, X_k] \dots \end{aligned}$$

146

Variance du modèle

- La variance du modèle à entrées indépendantes se décompose en :

$$V = \sum_{i=1}^p V_i + \sum_{1 \leq i < j \leq p} V_{ij} + \dots + V_{1, \dots, p},$$

où

$$\begin{aligned} V_i &= V(\mathbf{E}[Y|X_i]), \\ V_{ij} &= V(\mathbf{E}[Y|X_i, X_j]) - V_i - V_j, \\ V_{ijk} &= V(\mathbf{E}[Y|X_i, X_j, X_k]) - V_{ij} - V_{ik} - V_{jk} - V_i - V_j - V_k, \\ &\dots \\ V_{1, \dots, p} &= V - \sum_{i=1}^p V_i - \sum_{1 \leq i < j \leq p} V_{ij} - \dots - \sum_{1 \leq i_1 < \dots < i_{p-1} \leq p} V_{i_1, \dots, i_{p-1}} \end{aligned}$$

147

Décomposition linéaire de $f(x)$

- Lorsque la fonction $f(x)$ est intégrable, il est possible de calculer la décomposition de façon analytique;
- Sinon (ce qui est le plus souvent le cas dans des cas réels), on effectue une décomposition par une méthode itérative basée sur une méthode de (quasi)-Monte-Carlo.

◦ https://en.wikipedia.org/w/index.php?title=Variance-based_sensitivity_analysis

148

Indices de Sobol

- I. M. Sobol se base sur cette décomposition pour définir des indices de sensibilité d'ordre un: $S_i = V_i/V$.
- Les indices de sensibilité d'ordre deux, $S_{ij} = V_{ij}/V$, expriment la sensibilité de la variance de Y à l'interaction des variables X_i et X_j , c'est-à-dire la sensibilité de Y aux variables X_i et X_j qui n'est pas prise en compte dans l'effet des variables seules.
- On peut aussi définir des indices d'ordre 3 jusqu'à p.

149

Interprétation des indices

- L'interprétation de ces indices est facile, puisque leur somme est égale à 1, et étant tous positifs, plus l'indice sera grand (proche de 1), plus la variable aura d'importance pour expliquer la variance de la sortie.

150

Indices composites

- On peut aussi définir la contribution d'une variable à la variance totale comme étant la somme de tous les indices où elle apparaît, par exemple, pour S_i :

$$S_i + S_{i(1:p)} + S_{i(1:p)(1:p)} + \dots + S_{1:p}$$

Attention, la somme des S_i définis de cette façon est supérieure à 1.

151

Estimation des indices

- Il existe plusieurs méthodes permettant leur estimation.
- La méthode originale, nommée ici Sobol'93 permet l'estimation de toutes les interactions mais les estimateurs sont assez instables.
- On utilisera aussi celle de Jansen'99 plus stable qui donne la contribution à la variance totale de chaque paramètre pris isolément, S_i , et de chaque paramètre S_i plus l'ensemble de ses interactions nommé $S_{i\cdot}$.

M.J.W. Jansen, 1999, Analysis of variance designs for model output, Computer Physics Communication, 117, 35-43.

152

Génération de l'hypercube

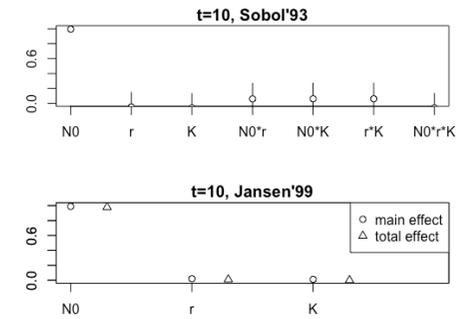
- Les paramètres doivent tous se situer dans un hypercube de côté 1; il convient donc d'effectuer un changement de variable pour les utiliser dans un modèle logistique avec $N_0=10$, $r=0.01$ et $K=1000$:

```

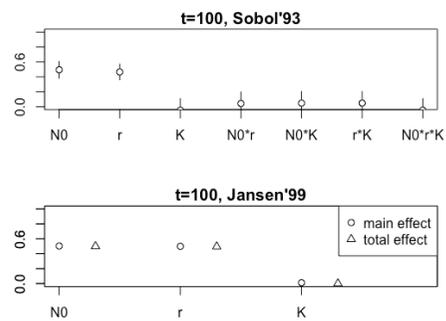
N0 <- ((runif(1)*2)-1)*0.1*10+10
r <- ((runif(1)*2)-1)*0.1*0.01+0.01
K <- ((runif(1)*2)-1)*0.1*1000+1000

```

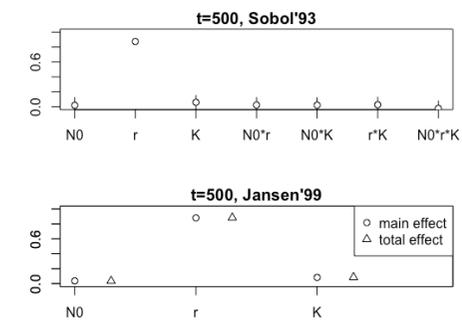
153



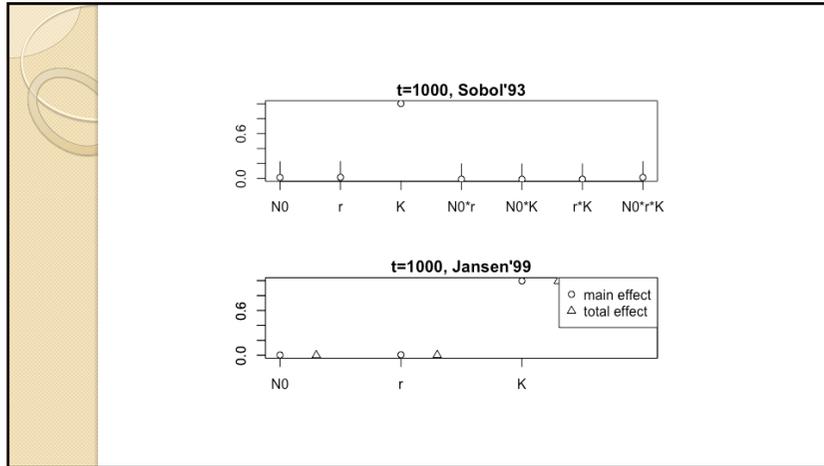
154



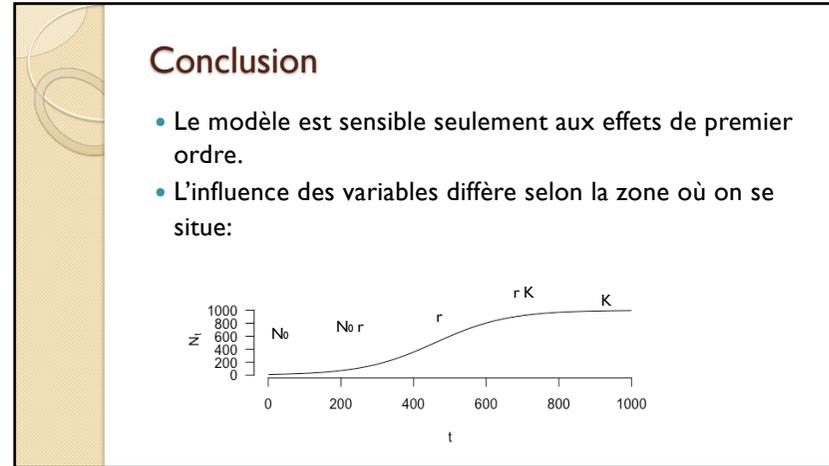
155



156



157



158