
Problèmes d'estimation dans un processus de ramification

Mémoire présenté comme exigence partielle
de la maîtrise en mathématiques

Mémoire n° 383

2^e édition

Alain L^AT_EX_R

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
DÉPARTEMENT DE MATHÉMATIQUES

UQÀM

Préface

Avec une satisfaction particulière, je me suis replongé dans les pages jaunies de mon mémoire de maîtrise, rédigé il y a près de 50 ans en 1978. Ce document représente mes premiers pas significatifs en mathématiques et en l'informatique.

En parcourant mes archives, je suis tombé sur une copie soigneusement reliée de ce document, un témoignage d'une époque où les mémoires étaient méticuleusement alignés sur les étagères des bibliothèques de l'Université du Québec à Montréal ([UQAM](#)). Sa lecture a ravivé les souvenirs d'une période révolue.

Un nom me revient en mémoire : Peter Jones. D'une grande gentillesse et très disponible, Peter était un informaticien talentueux qui avait développé un formateur de texte révolutionnaire à cette époque.

Probablement conçu pendant les temps libres de Peter, cet outil a permis de formater mon mémoire de manière impressionnante, trouvais-je, intégrant même les signes diacritiques pour les lettres minuscules. C'était une avancée incroyable, ouvrant de nouvelles possibilités pour rédiger des textes en français avec aisance, ou presque!

Je crois que, à l'[UQAM](#), le mémoire numéro 383 a été le précurseur de la mise en page assistée par ordinateur. L'utilisation de cet outil a, bien sûr, nécessité des efforts supplémentaires pour ajouter manuellement les formules mathématiques. Chaque formule était préalablement « tapée » à la machine à écrire, certaines permettant l'introduction de caractères grecs et de symboles mathématiques spécifiques. Un extrait du document de 1978 est présenté dans la figure 1. On peut voir ce que devient cet extrait à la page 18 de cette nouvelle édition.

La comparaison entre cette mise en page et celle des documents actuels réalisés par des systèmes tels que \LaTeX et reproduits par les imprimantes laser contemporaines est sans appel! L'avancement remarquable des outils et des techniques depuis cette époque témoigne d'une évolution significative la mise en page des mathématiques.

Notons que cette récurrence fut utilisée dans les différents programmes conçus pour l'analyse des différents estimateurs. Le prochain résultat, dû encore une fois à Gilbert Labelle, fut nécessaire dans la recherche de l'estimateur à vraisemblance maximale.

Proposition 2.4:

$$\frac{\partial l_{m,n}}{\partial p_k} = m l_{m-1,n-k}$$

Démonstration:

Pour prouver cette assertion, il suffit de dériver $\Phi(s,t)$ par rapport à p_k . En effet,

$$\begin{aligned} \frac{\partial \Phi(s,t)}{\partial p_k} &= s \left(\frac{\partial \zeta(t)}{\partial p_k} \right) e^{s \zeta(t)} \\ &= s t^k \left(\sum_{m,n \geq 0} l_{m,n} \frac{s^m}{m!} t^n \right) \end{aligned}$$

Figure 1 : Échantillon du texte produit en 1978

Aujourd'hui, une variété d'outils sont à notre disposition. Des environnements tels que R-Studio facilitent grandement le travail, permettant la création de graphiques d'une grande beauté en quelques minutes. L'importation du contenu d'un fichier Excel se fait aisément, dans un format impeccable. Les fichiers sources de programmes peuvent être formatés de manière très attrayante. Même les programmes Fortran paraissent jolis! Ces innovations, dont nous aurions aimé disposer dans les années 70, ont considérablement simplifié le travail.

Dans les années 1980, ce mémoire a disparu des rayons de la bibliothèque, tout comme ceux de certains de mes amis de l'époque. Un curieux hasard, sans doute. Pour pallier partiellement cette absence, je prévois offrir cette nouvelle édition à la bibliothèque de l'[UQAM](#).

Dans les quelques lignes de cette préface, j'ai voulu présenter au lecteur le contexte dans lequel nous évoluions dans les années 70. Ce bref aperçu dévoile les débuts captivants de ma carrière, caractérisés par l'innovation technique de l'époque.

L'étude des mathématiques, source d'un plaisir immense, a été au cœur de cette période. J'ai eu la chance de naître à une époque où l'éducation demeurait un puissant ascenseur social. J'ai exercé une profession qui m'a permis de continuer à apprendre tout au long de ma vie, ce qui a été un véritable bonheur.

J'espère que ce mémoire, revisité après toutes ces années, continue d'inspirer les esprits curieux et de témoigner des progrès remarquables des moyens de production.

A. Latour – 17 février 2024

Résumé

Ce travail porte sur l'estimation des probabilités initiales et de la moyenne d'un processus de Galton-Watson, noté $\{Z_t, t = 0, \dots, n\}$.

Dans le premier chapitre, nous introduisons la notion combinatoire de partage afin d'expliciter la fonction de vraisemblance du processus. Ensuite, nous démontrons simplement que $\hat{m} = (Z_1 + \dots + Z_n)/(Z_0 + \dots + Z_{n-1})$ est l'estimateur à vraisemblance maximale de $m = E[Z_1]$, lorsque seule la connaissance de Z_t , pour $t = 0, \dots, n$, est disponible.

Le deuxième chapitre aborde le concept de fonction génératrice, en présentant deux fonctions génératrices des probabilités de transition $(\ell_{m,n})$ de la chaîne de Markov associée au processus de ramification. Ces fonctions sont ensuite utilisées pour démontrer certaines propriétés de ces probabilités. Nous identifions une récurrence satisfaite par les $\ell_{m,n}$, ainsi qu'une autre récurrence pour leurs dérivées. Ces récurrences font intervenir les probabilités de la loi initiale, spécifiquement $p_k = \Pr[Z_1 = k]$, où $k = 0, 1, \dots$.

Le troisième chapitre aborde les problèmes d'estimation, introduisant une suite qui a permis de délimiter notre étude. On propose onze méthodes d'estimation des probabilités initiales, en tenant compte du fait qu'un individu ne peut générer plus de deux descendants. Les principaux algorithmes utilisés dans nos programmes sont détaillés dans ce chapitre.

Dans le quatrième chapitre, nous apportons des précisions sur le type d'ordinateur utilisé et fournissons des informations relatives à l'utilisation des procédures programmées. Ensuite, nous commentons les résultats obtenus à partir des programmes. Enfin, nous élaborons un modèle simple et explicatif du biais de \hat{m} en fonction de la taille de l'échantillon, c'est-à-dire du nombre de générations observées.

Le cinquième chapitre établit le lien entre deux méthodes d'estimation suggérées, offrant ainsi des pistes pour réduire significativement les calculs impliqués

dans une telle recherche. Nous concluons en présentant une propriété intéressante de l'estimation à entropie maximale.

Remerciements

Je tiens à exprimer ma profonde gratitude envers mes directeurs de recherche, Messieurs Jean-Pierre Dion et Gilbert Labelle, pour leur patience, leur disponibilité et la grande liberté qu'ils m'ont accordée dans mon travail.

Mes remerciements s'adressent également au Conseil National de la Recherche du Canada, qui m'a généreusement octroyé une bourse d'études pour les années universitaires 1976-1977 et 1977-1978.

Un remerciement particulier revient à ma compagne, Louise Laforest, pour son soutien précieux et quotidien au cours de ces années, ainsi que pour son investissement dans la révision et la mise en page de ce travail.

Je souhaite également exprimer ma reconnaissance envers mon amie Louise Régnier pour son assistance dans la réalisation des figures de la version originale¹. J'ai fait la connaissance de Louise R. en 1972, et elle demeure présente en 2024.

1. Elle a eu la gentillesse de réaliser les graphiques présentant les courbes de niveau à la page 43.

Table des matières

Préface	i
Résumé	v
Remerciements	vii
Introduction	1
1 Fonction de vraisemblance du processus de Galton-Watson	3
1.1 Introduction	3
1.2 Fonction de vraisemblance	4
Notion de partage	4
Forme explicite	5
1.3 Estimation à vraisemblance maximale pour la moyenne	6
2 Introduction aux fonctions génératrices	11
2.1 Introduction	11
2.2 Fonctions génératrices	12
Concept de base	12
Opérations sur les fonctions génératrices	13
Fonctions génératrices à deux variables	14
2.3 Résultats relatifs aux probabilités de transition	15

3	Estimation	21
3.1	Introduction	21
3.2	Les limites de l'étude	22
	La suite $\{s(n,k)\}$	22
	Les estimateurs	28
3.3	Les algorithmes	33
	Maximisation de la fonction de vraisemblance	33
	Génération des partages d'un entier	37
4	Analyse des résultats	41
4.1	Introduction	41
4.2	Les calculs	42
	L'outil de calcul	42
	Le choix du point initial pour la maximisation	42
	La procédure MAXLIK	44
	La procédure NEWTON	45
	Les autres procédures	46
4.3	Analyse des résultats	47
	Les distributions du type $p_0 = 0, p_1 \neq 0$ et $p_2 \neq 0$	50
	Les distributions du type $p_0 \neq 0, p_1 = 0, p_2 \neq 0$	51
	Les distributions du type $p_0 \neq 0, p_1 \neq 0$ et $p_2 \neq 0$	51
4.4	Étude du biais de l'estimateur de la moyenne	53
5	Deux résultats importants	57
5.1	Introduction	57
5.2	Le lien entre le huitième et le neuvième estimateur	57
	L'équation de la droite	62
5.3	L'arbre le plus vraisemblable	66

5.4	L'entropie maximale	68
6	Conclusion	73
	Bibliographie	75
A	Programmes utilisés	77
A.1	Procédure MAXLIK	78
A.2	Procédure PART	79
A.3	Procédure INIT	80
A.4	Procédure SUCC.	80
A.5	Procédure MULT.	81
A.6	Procédure TRI.	81
A.7	Procédure RECH	82
A.8	Procédure NEWTON	82
A.9	Procédure EVAL	83
A.10	Procédure LIKE	85
A.11	Procédure GRAD	85
A.12	Procédure DLIKE	86
A.13	Procédure CONTRA	86
A.14	Procédure ANGLE	90
B	Exemples d'estimation	91
C	Tableaux et graphiques des biais	97
C.1	Objectifs.	97
C.2	Tableaux des biais	98
C.3	Graphiques des biais	108

Liste des figures

3.1	Deux arbres possibles pour la réalisation $\langle 1, 1, 2, 1, 2, 4, 5 \rangle$	31
3.2	Exemple de trajectoire empruntée par l'algorithme du gradient . . .	34
3.3	Méthode <i>regula falsi</i> : détermination de t_i à partir de $(t_{i-2}, F(t_{i-2}))$ et $(t_{i-1}, F(t_{i-1}))$	36
3.4	Déplacement dans la direction du gradient	36
4.1	Surface et courbes de niveau de deux fonctions de vraisemblance .	43
4.2	Distributions choisies pour l'étude des estimateurs	48
5.1	Différentes estimations pour la réalisation $\langle 1, 2, 3, 4, 4 \rangle$	59
5.2	Différentes estimations pour la réalisation $\langle 1, 1, 2, 4, 4 \rangle$	60
C.1	Biais en fonction de la taille pour $\mathbf{p} = (0.0, 0.1, 0.9)$	108
C.2	Biais en fonction de la taille pour $\mathbf{p} = (0.0, 0.3, 0.7)$	108
C.3	Biais en fonction de la taille pour $\mathbf{p} = (0.0, 0.5, 0.5)$	109
C.4	Biais en fonction de la taille pour $\mathbf{p} = (0.0, 0.7, 0.3)$	109
C.5	Biais en fonction de la taille pour $\mathbf{p} = (0.0, 0.8, 0.2)$	110
C.6	Biais en fonction de la taille pour $\mathbf{p} = (0.0, 0.9, 0.1)$	110
C.7	Biais en fonction de la taille pour $\mathbf{p} = (0.1, 0.0, 0.9)$	111
C.8	Biais en fonction de la taille pour $\mathbf{p} = (0.3, 0.0, 0.7)$	111
C.9	Biais en fonction de la taille pour $\mathbf{p} = (0.4, 0.0, 0.6)$	112

C.10 Biais en fonction de la taille pour $\mathbf{p} = (0.2, 0.1, 0.7)$	112
C.11 Biais en fonction de la taille pour $\mathbf{p} = (0.1, 0.2, 0.7)$	113
C.12 Biais en fonction de la taille pour $\mathbf{p} = (0.1, 0.7, 0.2)$	113
C.13 Biais en fonction de la taille pour $\mathbf{p} = (0.1, 0.6, 0.3)$	114
C.14 Biais en fonction de la taille pour $\mathbf{p} = (0.3, 0.4, 0.3)$	114
C.15 Biais en fonction de la taille pour $\mathbf{p} = (0.2, 0.5, 0.3)$	115
C.16 Biais en fonction de la taille pour $\mathbf{p} = (0.4, 0.1, 0.5)$	115
C.17 Biais en fonction de la taille pour $\mathbf{p} = (0.3, 0.3, 0.4)$	116
C.18 Biais en fonction de la taille pour $\mathbf{p} = (0.3, 0.2, 0.5)$	116
C.19 Biais en fonction de la taille pour $\mathbf{p} = (0.2, 0.3, 0.5)$	117
C.20 Biais en fonction de la taille pour $\mathbf{p} = (0.1, 0.4, 0.5)$	117
C.21 Biais en fonction de la taille pour $\mathbf{p} = (0.0, 0.1, 0.9)$	118
C.22 Biais en fonction de la taille pour $\mathbf{p} = (0.0, 0.3, 0.7)$	118
C.23 Biais en fonction de la taille pour $\mathbf{p} = (0.0, 0.5, 0.5)$	119
C.24 Biais en fonction de la taille pour $\mathbf{p} = (0.0, 0.7, 0.3)$	119
C.25 Biais en fonction de la taille pour $\mathbf{p} = (0.0, 0.8, 0.2)$	120
C.26 Biais en fonction de la taille pour $\mathbf{p} = (0.0, 0.9, 0.1)$	120
C.27 Biais en fonction de la taille pour $\mathbf{p} = (0.1, 0.0, 0.9)$	121
C.28 Biais en fonction de la taille pour $\mathbf{p} = (0.3, 0.0, 0.7)$	121
C.29 Biais en fonction de la taille pour $\mathbf{p} = (0.4, 0.0, 0.6)$	122
C.30 Biais en fonction de la taille pour $\mathbf{p} = (0.2, 0.1, 0.7)$	122
C.31 Biais en fonction de la taille pour $\mathbf{p} = (0.1, 0.2, 0.7)$	123
C.32 Biais en fonction de la taille pour $\mathbf{p} = (0.1, 0.7, 0.2)$	123
C.33 Biais en fonction de la taille pour $\mathbf{p} = (0.1, 0.6, 0.3)$	124
C.34 Biais en fonction de la taille pour $\mathbf{p} = (0.3, 0.4, 0.3)$	124
C.35 Biais en fonction de la taille pour $\mathbf{p} = (0.2, 0.5, 0.3)$	125
C.36 Biais en fonction de la taille pour $\mathbf{p} = (0.4, 0.1, 0.5)$	125

C.37	Biais en fonction de la taille pour $\mathbf{p} = (0.3, 0.3, 0.4)$	126
C.38	Biais en fonction de la taille pour $\mathbf{p} = (0.3, 0.2, 0.5)$	126
C.39	Biais en fonction de la taille pour $\mathbf{p} = (0.2, 0.3, 0.5)$	127
C.40	Biais en fonction de la taille pour $\mathbf{p} = (0.1, 0.4, 0.5)$	127

Liste des tableaux

3.1	Tableau de quelques valeurs des $n(u, j, k)$, pour $k = 2$	23
3.2	Tableau de quelques valeurs des $n(u, j, k)$, pour $k = 3$	23
3.3	Cardinalité de $R(u, k)$ pour de petites valeurs de u et k	27
3.4	Énumération des partages pour la réalisation $\langle 1, 1, 2, 1, 2, 4, 5 \rangle$. . .	28
3.5	Partages de 15 en au plus 5 parts toutes inférieures ou égales à 7 . .	39
4.1	Estimation des coefficients α et β , valeur de r_{XY} et de m	54
4.2	Corrélation entre les différents paramètres	55
4.3	Corrélation partielle entre les différents paramètres après avoir éliminé l'effet de σ	56
4.4	Corrélation partielle entre les différents paramètres après avoir éliminé l'effet de m	56
5.1	Énumération des partages pour la réalisation $\langle 1, 2, 3, 4, 4 \rangle$	58
5.2	Énumération des partages pour la réalisation $\langle 1, 1, 2, 4, 4 \rangle$	59
B.1	Énumération des partages pour la réalisation $\langle 1, 2, 2, 2, 2 \rangle$	92
B.2	Valeurs des estimateurs pour la réalisation $\langle 1, 2, 2, 2, 2 \rangle$	92
B.3	Énumération des partages pour la réalisation $\langle 1, 2, 4, 5, 7 \rangle$	93
B.4	Valeurs des estimateurs pour la réalisation $\langle 1, 2, 4, 5, 7 \rangle$	93
B.5	Énumération des partages pour la réalisation $\langle 1, 2, 3, 4, 4 \rangle$	94

B.6	Valeurs des estimateurs pour la réalisation $\langle 1, 2, 3, 4, 4 \rangle$	94
B.7	Énumération des partages pour la réalisation $\langle 1, 1, 2, 4, 4 \rangle$	95
B.8	Valeurs des estimateurs pour la réalisation $\langle 1, 1, 2, 4, 4 \rangle$	95
C.1	Biais des estimateurs pour $p_0 = 0.0, p_1 = 0.1, p_2 = 0.9$	98
C.2	Biais des estimateurs pour $p_0 = 0.0, p_1 = 0.3, p_2 = 0.7$	98
C.3	Biais des estimateurs pour $p_0 = 0.0, p_1 = 0.5, p_2 = 0.5$	99
C.4	Biais des estimateurs pour $p_0 = 0.0, p_1 = 0.7, p_2 = 0.3$	99
C.5	Biais des estimateurs pour $p_0 = 0.0, p_1 = 0.8, p_2 = 0.2$	100
C.6	Biais des estimateurs pour $p_0 = 0.0, p_1 = 0.9, p_2 = 0.1$	100
C.7	Biais des estimateurs pour $p_0 = 0.1, p_1 = 0.0, p_2 = 0.9$	101
C.8	Biais des estimateurs pour $p_0 = 0.3, p_1 = 0.0, p_2 = 0.7$	101
C.9	Biais des estimateurs pour $p_0 = 0.4, p_1 = 0.0, p_2 = 0.6$	102
C.10	Biais des estimateurs pour $p_0 = 0.2, p_1 = 0.1, p_2 = 0.7$	102
C.11	Biais des estimateurs pour $p_0 = 0.1, p_1 = 0.2, p_2 = 0.7$	103
C.12	Biais des estimateurs pour $p_0 = 0.1, p_1 = 0.7, p_2 = 0.2$	103
C.13	Biais des estimateurs pour $p_0 = 0.1, p_1 = 0.6, p_2 = 0.3$	104
C.14	Biais des estimateurs pour $p_0 = 0.3, p_1 = 0.4, p_2 = 0.3$	104
C.15	Biais des estimateurs pour $p_0 = 0.2, p_1 = 0.5, p_2 = 0.3$	105
C.16	Biais des estimateurs pour $p_0 = 0.4, p_1 = 0.1, p_2 = 0.5$	105
C.17	Biais des estimateurs pour $p_0 = 0.3, p_1 = 0.3, p_2 = 0.4$	106
C.18	Biais des estimateurs pour $p_0 = 0.3, p_1 = 0.2, p_2 = 0.5$	106
C.19	Biais des estimateurs pour $p_0 = 0.2, p_1 = 0.3, p_2 = 0.5$	107
C.20	Biais des estimateurs pour $p_0 = 0.1, p_1 = 0.4, p_2 = 0.5$	107

Introduction

Cette recherche, suggérée par Jean-Pierre Dion, a pour but d'étudier différentes méthodes d'estimation, à partir de petits échantillons, dans le contexte d'un processus de ramification. Comme le soulignait DION (1972), bien peu d'études traitant des problèmes d'estimation dans un tel processus avaient été faites jusqu'à ce jour. Toutefois, ces dernières années, beaucoup de résultats asymptotiques ont été obtenus. On peut trouver la majorité de ceux-ci dans « *Statistical Inference in Branching Processes* » de DION et KEIDING (1978). Ces derniers ont réuni dans leur texte les développements des cinq dernières années et nous font sentir la nécessité d'étudier le problème des petits échantillons dans les termes suivants :

« Very little thought has been given to small-sample theory, including a suitable conditionality arguments for defining reasonable reference sampling distributions ... there exist practically no small-sample distribution theory for \hat{m} and the interesting question of possible conditional inference has hardly been touched upon. »

Nous ne croyons pas pouvoir satisfaire les espérances de ces auteurs, mais notre recherche se veut un premier pas en ce sens.

Le problème qui donna naissance à la théorie probabiliste des processus de Galton-Watson fut posé de façon mathématique en 1873 par Francis Galton, un cousin du célèbre Charles Darwin, père de la théorie de l'évolution. En réalité, les recherches récentes de HEYDE et SENETA (1972) nous apprennent que dès 1845, Bienaymé connaissait le théorème fondamental reliant la moyenne et la probabilité d'extinction de ces processus.

Il est recommandé au lecteur de se référer à KARLIN et TAYLOR (1975) en ce qui concerne les résultats fondamentaux. Définissons en termes simples le processus stochastique qui nous intéresse. Un individu a, à la fin de sa vie, un nombre aléatoire ξ de descendants suivant une loi de probabilité donnée par : $\Pr[\xi = k] = p_k, k \geq 0$. Nous supposons que les individus de la population naissent indépendamment les uns des autres et que leur progéniture respective est identiquement distribuée

comme ξ . On suppose de plus que tous les individus ont la même durée de vie. Dénotons par Z_n la taille de la population à la n^e génération. Le processus ainsi défini est Markovien.

Définissons maintenant ce processus en termes probabilistes. Soit (Ω, \mathcal{F}, P) un espace de probabilité et soit $\{Z_t\}_{t=0}^\infty$ le processus stochastique suivant :

1. $Z_0 = 1$;
2. $\Pr\{\omega : Z_1(\omega) = k\} = p_k \geq 0, \quad k = 0, 1, \dots \text{ et } \sum_{k \geq 0} p_k = 1$;
3. $\Pr\{\omega : Z_t(\omega) = k \mid Z_{t-1}(\omega) = j\} = p_{jk}$

$$= \begin{cases} \delta^{0,k}, & \text{si } j = 0; \\ \Pr\{\omega : \xi_1(\omega) + \dots + \xi_j(\omega) = k\}, & \text{si } j \neq 0. \end{cases}$$

où $\delta^{0,k}$ est le delta de Kronecker et où les ξ_i sont des variables aléatoires indépendantes sur (Ω, \mathcal{F}, P) de même loi que ξ .

Le travail se divise en cinq parties. La première donne une démonstration simple du fait que $\hat{m} = (Z_1 + \dots + Z_n) / (Z_0 + \dots + Z_{n-1})$ est l'estimateur à vraisemblance maximale pour la moyenne de ξ . Dans la seconde, nous démontrons, en utilisant des fonctions génératrices, les résultats fondamentaux relativement aux probabilités de transition. En troisième lieu, viennent les définitions des méthodes d'estimation. Dans la quatrième partie, nous analysons les résultats obtenus et nous élaborons un modèle explicatif du biais de l'estimateur à vraisemblance maximale de la moyenne de ξ . Dans la dernière partie, nous exposons le lien existant entre deux de nos méthodes d'estimation ainsi qu'une propriété dont jouit l'estimation à entropie maximale.

A. Latour – décembre 1978

Fonction de vraisemblance du processus de Galton-Watson

§1.1 INTRODUCTION

Un des buts de ce chapitre est de donner une démonstration simple du fait que

$$\hat{m} = \frac{\sum_{i=1}^n z_i}{\sum_{i=0}^{n-1} z_i}$$

est l'estimateur à vraisemblance maximale de m , le nombre moyen de descendants d'un individu dans un processus de Galton-Watson. On sait que HARRIS (1948) avait démontré un résultat analogue dans le cas où nous connaissons Z_{mr} , c'est-à-dire le nombre de fois qu'un individu de la m^{e} génération donne naissance à r descendants. Cet échantillonnage se révèle plus riche que celui obtenu par une simple observation du processus de ramification jusqu'à la n^{e} génération. FEIGIN (1977) a récemment proposé une démonstration plus simple que celle de DION (1972). L'intérêt de la démonstration présentée dans ce chapitre réside dans l'utilisation d'outils simples et d'un argument géométrique.

La définition 1.2.1 introduit le concept de partage ultérieurement exploité pour

donner une forme explicite à la fonction de vraisemblance du processus. Enfin, nous démontrerons que les coordonnées du point où cette fonction atteint son maximum satisfont une équation particulière, nous permettant ainsi de parvenir au résultat recherché.

§1.2 FONCTION DE VRAISEMBLANCE

Notion de partage

Voici la définition de partage que nous retrouvons dans la majorité des textes de combinatoire, par exemple, (RIORDAN, 1958, Cchapitre 6)

Définition 1.2.1 (Partage) Soit n et p deux entiers positifs. Soit $\alpha_1, \alpha_2, \dots, \alpha_p$, une suite d'entiers positifs tels que :

1. $n = \alpha_1 + \alpha_2 + \dots + \alpha_p$;
2. $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_p$.

Cette suite d'entiers positifs est dite partage de l'entier n en p parts. Les α_i , $i = 1, \dots, p$, sont appelés « parts. » \square

Considérons les partages de l'entier n , en au plus p parts, dont aucune n'est plus grande que m (il ne faut pas confondre ce m avec la moyenne de ξ). Remarquons que si nous posons $m = n = p$, nous retrouvons tous les partages de l'entier n .

Aussi, une suite satisfaisant les conditions énoncées peut être représentée comme suit :

$$\underbrace{m, \dots, m}_{\lambda_m}, \underbrace{m-1, \dots, m-1}_{\lambda_{m-1}}, \dots, \underbrace{2, \dots, 2}_{\lambda_2}, \underbrace{1, \dots, 1}_{\lambda_1}$$

Cependant, certains λ_i peuvent être nuls. Posons $\lambda_0 = p - \lambda_1 - \dots - \lambda_m$. Nous aurons :

1. $\sum_{i=0}^m \lambda_i = p$;
2. $\sum_{i=0}^m i \lambda_i = n$.

Ainsi, à chaque suite définissant un partage de n en plus p parts, aucune n'étant plus grande que m , nous pouvons associer de façon bijective la suite suivante :

$$(\lambda_0, \dots, \lambda_m) = \boldsymbol{\lambda},$$

que nous dénoterons de façon plus compacte par :

$$0^{\lambda_0} 1^{\lambda_1} \dots m^{\lambda_m} = \boldsymbol{\lambda}.$$

De plus, puisque dans la majorité des cas, il n'y a pas d'ambiguïté possible, il est fréquent d'utiliser la notation suivante : $0^{\lambda_0} 1^{\lambda_1} \dots = \boldsymbol{\lambda}$.

Forme explicite

Considérons $Z_0 = 1, Z_1 = z_1, \dots, Z_n = z_n$, une réalisation du processus de Galton-Watson. Évidemment, puisque la suite des $\{Z_i\}$ est une chaîne de Markov, la vraisemblance de cette observation est :

$$L(z_0, \dots, z_n; p_0, p_1, \dots) = \prod_{i=1}^n \ell_{z_{i-1}z_i}$$

où $\ell_{z_{i-1}z_i} = \Pr(Z_i = z_i | Z_{i-1} = z_{i-1})$. Il suffit donc d'expliciter les termes $\ell_{z_{i-1}z_i}$. Nous sommes ainsi amenés à examiner toutes les façons dont les z_{i-1} individus peuvent donner naissance à z_i descendants.

Numérotons temporairement les individus de la $(i-1)^{\text{e}}$ génération de 1 à z_{i-1} . Deux cas doivent être considérés. Si nous savons *a priori* qu'un individu ne peut donner naissance à plus de M descendants, le nombre maximal de descendants qu'un individu de la génération $i-1$ aurait pu avoir est évidemment le minimum entre z_i et M . Cependant, si le contexte ne donne pas cette information, on voit qu'il suffit de prendre M égal au maximum des z_i .

Dénotons par M_i , le nombre maximal de descendants d'un individu de la $(i-1)^{\text{e}}$ génération. Ainsi, tout partage de z_i en au plus z_{i-1} parts, dont aucune n'est supérieure à M_i , est une transition admissible de z_{i-1} à z_i individus au sens suivant : les λ_0^i premiers individus n'ont généré aucun descendant, les λ_1^i suivants n'en ont généré qu'un seul, etc. On voit alors qu'un partage fixe correspond à

$$C(\boldsymbol{\lambda}^i) = \left(\sum_{j=0}^{M_i} \lambda_j^i \right)! \left(\prod_{j=0}^{M_i} \lambda_j^i! \right)^{-1}$$

passages possibles d'une génération de z_{i-1} individus à une génération de z_i individus. Soit \mathcal{P}_i , l'ensemble de tous les partages admissibles. Les éléments de \mathcal{P}_i seront dénotés par $\boldsymbol{\lambda}^i$. Ainsi

$$\Pr(Z_i = z_i | Z_{i-1} = z_{i-1}) = \sum_{\boldsymbol{\lambda}^i \in \mathcal{P}_i} C(\boldsymbol{\lambda}^i) p^{\boldsymbol{\lambda}^i}$$

où $\mathbf{p}^{\lambda^i} = p_0^{\lambda_0^i} p_1^{\lambda_1^i} \cdots$. Nous sommes maintenant en mesure d'énoncer la proposition 1.2.1.

Proposition 1.2.1 (Vraisemblance) *La vraisemblance du processus de Galton-Watson est donnée par la fonction suivante :*

$$L(z_0, \dots, z_n; p_0, p_1, \dots) = \prod_{i=1}^n \left(\sum_{\lambda^i \in \mathcal{P}_i} C(\lambda^i) \mathbf{p}^{\lambda^i} \right). \quad (1.1)$$

C'est à l'aide de (1.1) que nous retrouverons \hat{m} , l'estimateur à vraisemblance maximale du nombre moyen de descendants par individu dans un processus de Galton-Watson.

§1.3 ESTIMATION À VRAISEMBLANCE MAXIMALE POUR LA MOYENNE

Dans la recherche du maximum de la fonction de vraisemblance nous devons considérer une section d'hyperplan définie comme suit :

$$S = \left\{ \mathbf{x} = (x_0, \dots, x_M) : M = \max_{1 \leq i \leq n} \{M_i\}, x_i \geq 0, i = 0, \dots, M, \sum_{i=0}^M x_i = 1 \right\}$$

où les M_i ont été définis précédemment.

Pour démontrer le résultat annoncé dans l'introduction relativement à \hat{m} , nous allons faire une preuve par cas. En fait, nous allons classer les différentes situations selon nombre de composantes nulles du vecteur \mathbf{p} , point où la fonction L atteint son maximum.

Supposons qu'il n'y a aucune composante nulle dans le vecteur \mathbf{p} , c'est-à-dire que le maximum est atteint en un point intérieur (au sens topologique du terme) du simplexe S .

Pour maximiser la fonction de vraisemblance, on prend son logarithme et on introduit un multiplicateur de Lagrange pour satisfaire la contrainte $\sum_{i=0}^M x_i = 1$.

Considérons la fonction suivante :

$$\varphi(\mathbf{x}) = \sum_{i=1}^n \ln \left(\sum_{\lambda^i \in \mathcal{P}_i} C(\lambda^i) \mathbf{x}^{\lambda^i} \right) + \gamma \left(1 - \sum_{i=0}^M x_i \right).$$

Dérivons cette fonction par rapport à x_j :

$$\begin{aligned} \frac{\partial \varphi(\mathbf{x})}{\partial x_j} &= \sum_{i=1}^n \left(\sum_{\lambda^i \in \mathcal{P}_i} C(\lambda^i) \mathbf{x}^{\lambda^i} \right)^{-1} \left(\sum_{\lambda^i \in \mathcal{P}_i} \frac{1}{x_j} \lambda_j^i C(\lambda^i) \mathbf{x}^{\lambda^i} \right) - \gamma \\ &= \frac{1}{x_j} \sum_{i=1}^n \frac{1}{\ell_{z_{i-1}z_i}} \left(\sum_{\lambda^i \in \mathcal{P}_i} \lambda_j^i C(\lambda^i) \mathbf{x}^{\lambda^i} \right) - \gamma \end{aligned}$$

et posons ces dérivées égales à zéro. Nous obtenons :

$$\gamma x_j = \sum_{i=1}^n \frac{1}{\ell_{z_{i-1}z_i}} \left(\sum_{\lambda^i \in \mathcal{P}_i} \lambda_j^i C(\lambda^i) \mathbf{x}^{\lambda^i} \right). \quad (1.2)$$

Sommons à gauche et à droite, sur toutes les valeurs de j , cette dernière équation :

$$\gamma \sum_{j=0}^M x_j = \sum_{j=0}^M \left[\sum_{i=1}^n \frac{1}{\ell_{z_{i-1}z_i}} \left(\sum_{\lambda^i \in \mathcal{P}_i} \lambda_j^i C(\lambda^i) \mathbf{x}^{\lambda^i} \right) \right].$$

Mais puisque $x \in S$, nous avons $\sum_{j=0}^M x_j = 1$ et en interchangeant l'ordre des sommations, nous trouvons :

$$\gamma = \sum_{i=1}^n \frac{1}{\ell_{z_{i-1}z_i}} \left(\sum_{\lambda^i \in \mathcal{P}_i} C(\lambda^i) \mathbf{x}^{\lambda^i} \right) \sum_{j=0}^M \lambda_j^i.$$

Sachant que $0^{\lambda_0^i} 1^{\lambda_1^i} \dots$ est un partage de z_i en au plus z_{i-1} parts, nous avons :

$$\begin{aligned} \gamma &= \sum_{i=1}^n \left(\sum_{\lambda^i \in \mathcal{P}_i} C(\lambda^i) \mathbf{x}^{\lambda^i} \right)^{-1} \left(\sum_{\lambda^i \in \mathcal{P}_i} C(\lambda^i) \mathbf{x}^{\lambda^i} \right) \sum_{j=0}^M \lambda_j^i \\ &= \sum_{i=1}^n z_{i-1} \end{aligned}$$

Ainsi, en reportant la valeur de γ dans l'équation (1.2), on obtient :

$$x_j = \left(\sum_{i=0}^{n-1} z_i \right)^{-1} \left[\sum_{i=1}^n \frac{1}{\ell_{z_{i-1}z_i}} \left(\sum_{\lambda^i \in \mathcal{P}_i} \lambda_j^i C(\lambda^i) \mathbf{x}^{\lambda^i} \right) \right] \quad (1.3)$$

Proposition 1.3.1 (Estimation de \hat{m}) L'estimateur de vraisemblance maximale de m , le nombre moyen de descendants par individu dans un processus de Galton-Watson, est donné par :

$$\hat{m} = \left(\sum_{i=1}^n z_i \right) \left(\sum_{i=0}^{n-1} z_i \right)^{-1}.$$

Preuve Supposons que $\hat{\mathbf{p}}$ soit l'estimateur à vraisemblance maximale de la fonction de masse de ξ . Tout d'abord, supposons qu'aucune composante de $\hat{\mathbf{p}}$ soit nulle. Alors \hat{p}_j satisfait l'équation (1.3). C'est-à-dire :

$$\hat{p}_j = \left(\sum_{i=0}^{n-1} z_i \right)^{-1} \sum_{i=1}^n \frac{1}{\ell_{z_{i-1}z_i}} \left(\sum_{\lambda^i \in \mathcal{P}_i} \lambda_j^i C(\lambda^i) \hat{\mathbf{p}}^{\lambda^i} \right).$$

Aussi, il va de soi que si $\hat{\mathbf{p}}$ est l'estimateur de vraisemblance maximale de \mathbf{p} , $\hat{m} = \sum_{j=0}^M j \hat{p}_j$ sera l'estimateur de vraisemblance maximale de m (voir GIRI (1977)). En interchangeant l'ordre des sommations, nous obtenons

$$\begin{aligned} \hat{m} &= \sum_{j=0}^M j \hat{p}_j \\ &= \left(\sum_{i=0}^{n-1} z_i \right)^{-1} \left(\sum_{i=1}^n \sum_{\lambda^i \in \mathcal{P}_i} \sum_{j=0}^M \frac{j \lambda_j^i C(\lambda^i) \hat{\mathbf{p}}^{\lambda^i}}{\ell_{z_{i-1}z_i}} \right) \\ &= \left(\sum_{i=0}^{n-1} z_i \right)^{-1} \left[\sum_{i=1}^n \frac{1}{\ell_{z_{i-1}z_i}} \left(\sum_{\lambda^i \in \mathcal{P}_i} C(\lambda^i) \hat{\mathbf{p}}^{\lambda^i} \right) \sum_{j=0}^M j \lambda_j^i \right] \end{aligned}$$

Mais puisque $\lambda^i \in \mathcal{P}_i$, on a $\sum_{j=0}^M j \lambda_j^i = z_j$ et ainsi

$$\begin{aligned} \hat{m} &= \left(\sum_{i=0}^{n-1} z_i \right)^{-1} \left(\sum_{i=1}^n \frac{1}{\ell_{z_{i-1}z_i}} \ell_{z_{i-1}z_i} z_i \right) \\ &= \left(\sum_{i=0}^{n-1} z_i \right)^{-1} \left(\sum_{i=1}^n z_i \right) \end{aligned}$$

Maintenant, s'il y a exactement une composante nulle dans le vecteur $\hat{\mathbf{p}}$, les autres composantes satisfont l'équation (1.3). Ainsi, dans la recherche du maximum, nous ne sommes plus intéressés par la variation de la fonction dans le sens de la composante nulle. Dans la sommation $\sum_{j=0}^M j \hat{p}_j$, la composante nulle n'intervient pas.

Le problème se ramène donc, en fixant la composante en question à zéro, au cas précédent dans un espace à M dimensions (nous en avons $M + 1$) et la preuve reste valide. Un argument semblable la rend aussi valable si k composantes de $\hat{\mathbf{p}}$ sont nulles, dans la mesure où nous nous ramenons dans un espace de dimension $M - k + 1$.

Finalement, si toutes les composantes sont nulles, à l'exception d'une seule, alors le maximum est atteint en un sommet du simplexe S . Donc, il existe un k tel que $\hat{p}_k = 1$ et $p_j = 0 \forall j \neq k$. Ainsi, dans l'expression de la fonction de vraisemblance, tout terme faisant intervenir des $\hat{p}_j, j \neq k$, sera nul. Donc, chaque sommation ne contient qu'un terme qui ne peut être que $\hat{p}_k^{z_{i-1}}$, c'est-à-dire les z_{i-1} individus ont tous donné naissance à k descendants. Ainsi, la réalisation est $Z_0 = 1, Z_1 = k, Z_2 = k^2, \dots, Z_n = k^n$. Donc

$$\hat{m} = \sum_{j=0}^M j \hat{p}_j = k$$

$$\sum_{i=1}^n z_i = \frac{k^n - k}{k - 1}$$

et

$$\sum_{i=0}^{n-1} z_i = \frac{k^{n-1} - 1}{k - 1}$$

Donc,

$$\left(\sum_{i=0}^{n-1} z_i \right)^{-1} \left(\sum_{i=1}^n z_i \right) = k = \hat{m}.$$

□

Introduction aux fonctions génératrices

§2.1 INTRODUCTION

D'après KNUTH (1969), dès le début du XVII^e siècle, le concept de fonction génératrice était déjà présent. Cet outil de calcul fut utilisé par plusieurs mathématiciens parmi lesquels Abel (1881), Bernoulli (1728), Euler (1740) et Stirling (1730). Certaines techniques maintenant classiques sont dues à Laplace et développées dans sa « *Théorie analytique des probabilités* » parue en 1812.

BERGE (1968) et RIORDAN (1958) font une utilisation significative de cet outil dans leurs textes de combinatoire. D'autre part, DOUBILET et collab. (1974) se sont fixé pour objectif de développer une théorie des fonctions génératrices englobant non seulement toutes les algèbres classiques de fonctions génératrices telles que celles des fonctions génératrices ordinaires, exponentielles, de Dirichlet, etc., mais également une technique systématique permettant d'établir d'autres algèbres de fonctions génératrices liées à des problèmes spécifiques d'énumération. Le lecteur intéressé par le sujet peut se référer aux textes mentionnés précédemment ainsi qu'à l'article de BENDER et GOLDMAN (1971).

Nous définirons de manière concise le concept de fonction génératrice et

examinerons brièvement la pertinence de discuter de la convergence des séries impliquées. Par la suite, nous présenterons deux fonctions génératrices pour les $\ell_{m,n} = \Pr(Z_i = n \mid Z_{i-1} = m)$. La première sera principalement fournie à titre informatif. En revanche, la seconde sera largement exploitée dans les démonstrations des résultats concernant les probabilités de transition de la chaîne de Markov impliquée dans un processus de Galton-Watson.

§2.2 FONCTIONS GÉNÉRATRICES

Concept de base

Pour introduire le sujet, nous débuterons en examinant les fonctions génératrices à une variable.

Définition 2.2.1 (Fonction génératrice) Soit $\{a_k\}$ une suite de nombres réels.

La série

$$\varphi(t) = \sum_{k \geq 0} a_k t^k$$

sera dite fonction génératrice ordinaire de la suite $\{a_k\}$ et la série

$$\psi(t) = \sum_{k \geq 0} a_k \frac{t^k}{k!}$$

sera dite fonction génératrice exponentielle de la suite $\{a_k\}$. □

Il est clair que l'on peut munir les fonctions génératrices ordinaires d'opérations pour obtenir une algèbre, connue sous le nom d'algèbre de Cauchy. Ce principe s'applique également aux fonctions génératrices exponentielles, et dans ce contexte, on parle du « calcul de Blissard ». Citons (RIORDAN, 1958, p. 20) :

Formal operations on generating functions, such as addition, multiplication, differentiation and integration with respect to t , serve to express definitions and relations in this algebra, in particular by equating coefficients of t^k after completing these operations. From this point of view, convergence of the sum is irrelevant.

Dans un tel contexte, la variable t de la définition 2.2.1 est purement formelle. Cependant, certains auteurs préfèrent définir t comme une variable réelle ou complexe. Par conséquent, des questions de convergence émergent, visant à garantir l'existence de la fonction génératrice. Le lemme suivant, que nous ne démontrerons pas ici, est fréquemment employé pour établir la convergence de certaines séries.

Lemme 2.2.1 *S'il existe une constante positive et réelle M telle que pour tout n on a $|a_n| \leq M$, alors :*

1. $\varphi(t)$ converge pour tout t de $(-1; 1)$;
2. $\psi(t)$ converge pour tout t réel.

Considérons la fonction génératrice des probabilités initiales du processus de Galton-Watson, notée $\zeta(t)$, définie par l'expression $\zeta(t) = \sum_{k \geq 0} p_k t^k$. Puisque $p_k \leq 1$ pour tout $k \geq 0$, il suffit de poser $M = 1$ et d'appliquer le lemme précédent pour constater que $\zeta(t)$ existe.

Il est courant d'utiliser des fonctions génératrices comme outil exploratoire pour résoudre un problème donné. Une fois que la réponse est connue et qu'il est possible de fournir une démonstration indépendante et potentiellement plus simple du résultat, rien ne nous empêche de le faire.

Considérons, par exemple, les différentes démonstrations du fait qu'un ensemble de n éléments a 2^n sous-ensembles. Nous ne reprendrons pas les diverses démonstrations classiques telles que la récurrence ou l'argument combinatoire, notons cependant que la démonstration basée sur le développement binomial de $(1 + t)^n$ repose clairement sur le concept de fonction génératrice.

Opérations sur les fonctions génératrices

Nous allons maintenant donner les différentes opérations et l'opérateur principal intervenant dans l'algèbre de Cauchy. Pour ce, soit $f(t) = \sum_{k \geq 0} a_k t^k$ et $g(t) = \sum_{k \geq 0} b_k t^k$, deux fonctions génératrices. On définit :

1. L'addition de deux fonctions génératrices :
Si $h(t) = f(t) + g(t)$, alors $h(t) = \sum_{k \geq 0} (a_k + b_k) t^k$.
2. La multiplication de deux fonctions génératrices :
Si $h(t) = f(t)g(t)$, alors $h(t) = \sum_{k \geq 0} \left(\sum_{j=0}^k a_j b_{k-j} \right) t^k$.
3. Le produit d'une fonction génératrice par un scalaire :
Si $h(t) = cf(t)$, alors $h(t) = \sum_{k \geq 0} ca_k t^k$.
4. La dérivée de $f(t)$, dénotée par $f'(t)$ est
 $f'(t) = \sum_{k \geq 0} (k + 1) a_{k+1} t^k$.

Une conséquence directe de la définition du produit de deux fonctions génératrices est que si $f_j(t) = \sum_{k \geq 0} a_{jk} t^k$ sont des fonctions génératrices, on peut évaluer le produit $h(t) = \prod_{j \geq 0} f_j(t)$ de la manière suivante :

$$h(t) = \sum_{n \geq 0} \left(\sum_{\substack{k_0, k_1, \dots \geq 0 \\ k_0 + k_1 + \dots = n}} a_{0k_0} a_{1k_1} \dots \right) t^n$$

Fonctions génératrices à deux variables

Jusqu'ici, nous n'avons considéré que des fonctions à une variable. Cependant, dans les sections à venir, nous aurons à manipuler des suites doublement indicées. Il sera alors pratique de considérer des fonctions génératrices à deux variables.

Exemple 2.2.1

1. $f(t_1, t_2) = \sum_{m, n \geq 0} a_{m, n} t_1^m t_2^n;$
2. $g(t_1, t_2) = \sum_{m, n \geq 0} a_{m, n} \frac{t_1^m}{m!} t_2^n;$

□

Le lemme suivant nous assurera de la convergence des séries que nous considérerons par la suite.

Lemme 2.2.2 *S'il existe une constante positive et réelle M telle que $\forall m$ et $\forall n$ $|a_{m, n}| \leq M$, alors*

1. $f(t_1, t_2)$ converge $\forall t_1 \forall t_2$ tels que $|t_1| < 1$ et $|t_2| < 1$;
2. $g(t_1, t_2)$ converge $\forall t_1 \in \mathbb{R}$, et $\forall t_2$ tel que $|t_2| < 1$.

Preuve Nous ne démontrerons que la première assertion, puisque la seconde se démontre d'une façon analogue. Soit

$$S_{k, j} = \sum_{m=0}^k \sum_{n=0}^j |a_{m, n} t_1^m t_2^n|$$

alors :

$$\begin{aligned} S_{k, j} &= \sum_{m=0}^k \sum_{n=0}^j |a_{m, n}| |t_1|^m |t_2|^n \\ &\leq \sum_{m=0}^k \sum_{n=0}^j M |t_1|^m |t_2|^n \\ &= M \sum_{m=0}^k |t_1|^m \sum_{n=0}^j |t_2|^n \\ &= M \frac{1 - |t_1|^{k+1}}{1 - |t_1|} \times \frac{1 - |t_2|^{j+1}}{1 - |t_2|} \end{aligned}$$

Ainsi, si $|t_1| < 1$ et si $|t_2| < 1$, nous aurons, pour $\forall k$ et $\forall j$,

$$S_{k, j} \leq M \frac{1}{(1 - |t_1|)} \frac{1}{(1 - |t_2|)} < \infty$$

Donc la série converge absolument. Par conséquent, elle converge. \square

§2.3 RÉSULTATS RELATIFS AUX PROBABILITÉS DE TRANSITION

Nous avons vu au chapitre 1 que :

$$\Pr(Z_i = n \mid Z_{i-1} = m) = \sum_{\lambda \in \mathcal{P}_i} C(\lambda) \mathbf{p}^\lambda = \ell_{m,n}$$

Rappelons que $\lambda \in \mathcal{P}_i$ si, et seulement si, $\sum_{i=0}^m \lambda_i = m$ et $\sum_{i=0}^m i\lambda_i = n$.

Les deux fonctions génératrices suivantes peuvent être considérées comme celles des $\ell_{m,n}$:

1. $\Lambda(s,t) = \sum_{m,n \geq 0} \ell_{m,n} s^m t^n$
2. $\Phi(s,t) = \sum_{m,n \geq 0} \ell_{m,n} \frac{s^m}{m!} t^n$

De même, il convient de rappeler que la fonction génératrice de la variable aléatoire ξ est $\zeta(t) = \sum_{k \geq 0} p_k t^k$. Nous allons à présent expliciter les deux fonctions Λ et Φ .

Proposition 2.3.1

$$\Lambda(s,t) = \frac{1}{1 - s\zeta(t)}$$

Preuve Puisque

$$\frac{1}{1-x} = \sum_{m \geq 0} x^m$$

on a

$$\begin{aligned} \frac{1}{1 - s\zeta(t)} &= \sum_{m \geq 0} (s\zeta(t))^m \\ &= \sum_{m \geq 0} s^m \left[\sum_{k \geq 0} p_k t^k \right]^m \\ &= \sum_{m \geq 0} s^m \left[\sum_{n \geq 0} \left(\sum_{\substack{k_0, k_1, \dots \geq 0 \\ k_0 + k_1 + \dots = n}} \underbrace{p_{k_0} p_{k_1} \dots}_{m \text{ termes}} \right) t^n \right] \end{aligned}$$

La dernière égalité est obtenue par une application directe du produit de plusieurs fonctions génératrices.

Dans le produit $p_{k_0}p_{k_1}\dots$, il y a m termes, la série étant élevée à la puissance m . Notons λ_j le nombre de fois que p_j apparaît dans ce produit, pour tout $j \geq 0$. Remarquons que certains λ_j peuvent être nuls. Ainsi, ce produit peut être exprimé sous la forme $p_0^{\lambda_0}p_1^{\lambda_1}\dots$.

Évidemment, $\lambda_0 + \lambda_1 + \dots + \lambda_k = m$ puisque le produit contient m termes. De plus, p_j étant le coefficient de t^j , il s'ensuit que $0\lambda_0 + 1\lambda_1 + \dots = n$.

Finalement, la sommation délimitée par les parenthèses comprend exactement $(\sum_{j \geq 0} \lambda_j)! / (\prod_{j \geq 0} \lambda_j!)$ termes où p_j apparaît λ_j fois, pour $j = 0, 1, 2, \dots$. Ce qui nous montre que

$$\begin{aligned} \frac{1}{1 - s\zeta(t)} &= \sum_{m \geq 0} s^m \left[\sum_{n \geq 0} \left(\sum_{\substack{\lambda_0 + \lambda_1 + \dots = m \\ 0\lambda_0 + \lambda_1 + \dots = n}} \frac{(\sum_{j \geq 0} \lambda_j)!}{\prod_{j \geq 0} \lambda_j!} p_0^{\lambda_0} p_1^{\lambda_1} \dots \right) t^n \right] \\ &= \sum_{m \geq 0} s^m \sum_{n \geq 0} \ell_{m,n} t^n \\ &= \sum_{m,n \geq 0} \ell_{m,n} s^m t^n \\ &= A(s,t) \end{aligned}$$

□

Proposition 2.3.2

$$\Phi(s,t) = e^{s\zeta(t)}$$

Preuve Un simple calcul suffit :

$$\begin{aligned} e^{s\zeta(t)} &= \sum_{m \geq 0} \frac{s^m}{m!} \zeta^m(t) \\ &= \sum_{m \geq 0} \frac{s^m}{m!} \left(\sum_{k \geq 0} p_k t^k \right)^m \\ &= \sum_{m,n \geq 0} \ell_{m,n} \frac{s^m}{m!} t^n \\ &= \Phi(s,t) \end{aligned}$$

□

Le détail des calculs de la dernière démonstration a été omis puisque ces calculs sont essentiellement les mêmes que ceux de la proposition précédente. Voici des formules de récurrence pour les $\ell_{m,n}$.

Proposition 2.3.3 Dans un processus de Galton-Watson, les probabilités de transition satisfont les équations suivantes :

1. $\ell_{0,n} = \delta^{0,n}$ (delta de Kronecker);
2. $\ell_{m+1,n} = p_0 \ell_{m,n} + p_1 \ell_{m,n-1} + \cdots + p_n \ell_{m,0}$.

Preuve Par définition du processus de Galton-Watson, 1. est évident, Pour prouver 2., il suffit de dériver $\Phi(s,t)$ par rapport à s . En effet,

$$\frac{\partial \Phi(s,t)}{\partial s} = \frac{\partial e^{s\zeta(t)}}{\partial s} = \zeta(t) e^{s\zeta(t)} = \zeta(t) \Phi(s,t)$$

D'autre part, par définition de la dérivée d'une fonction génératrice (écrite sous forme de série), nous avons :

$$\begin{aligned} \frac{\partial \Phi(s,t)}{\partial s} &= \sum_{m,n \geq 0} m \ell_{m,n} \frac{s^{m-1}}{m!} t^n \\ &= \sum_{m,n \geq 0} m \ell_{m+1,n} \frac{s^m}{m!} t^n \end{aligned} \quad (2.1)$$

De plus,

$$\begin{aligned} \zeta(t) \Phi(s,t) &= \sum_{k \geq 0} p_k t^k \sum_{m,k \geq 0} \ell_{m,k} \frac{s^m}{m!} t^k \\ &= \sum_{m,n \geq 0} \left(\sum_{k_1+k_2=n} p_{k_1} \ell_{m,k_2} \right) \frac{s^m}{m!} t^n \end{aligned} \quad (2.2)$$

Ainsi, en identifiant les coefficients de $\frac{s^m}{m!} t^n$ dans (2.1) et (2.2), nous obtenons :

$$\begin{aligned} \ell_{m+1,n} &= \sum_{k_1+k_2=n} p_{k_1} \ell_{m,k_2} \\ &= p_0 \ell_{m,n} + p_1 \ell_{m,n-1} + \cdots + p_n \ell_{m,0} \end{aligned}$$

□

Les démonstrations des trois dernières propositions sont essentiellement dues à Gilbert Labelle. L'auteur du présent mémoire s'est limité à fournir les détails facilitant la compréhension. Pour la dernière proposition, on peut donner une preuve fort simple.

Considérons une union d'évènements disjoints. Pour ce, prenons un individu de la génération des parents, et intéressons-nous à sa progéniture. Un seul des évènements suivants a pu être réalisé :

1. cet individu n'a eu aucun descendant les m autres en ont eu globalement n ;
2. cet individu a eu un descendant et les m autres en ont eu globalement $n - 1$;
- ⋮
3. cet individu a eu n descendants et les m autres en ont eu aucun.

Évidemment, $\ell_{m+1,n}$ sera la somme des différentes probabilités associées aux événements décrits. Or, la probabilité du $(i + 1)^{\text{e}}$ est $p_i \ell_{m,n-i}$. D'où le résultat cherché.

Notons que cette récurrence fut utilisée dans les programmes conçus pour l'étude du biais des différents estimateurs. Le prochain résultat, dû encore une fois à Gilbert Labelle, fut nécessaire dans la recherche de l'estimateur à vraisemblance maximale.

Proposition 2.3.4

$$\frac{\partial \ell_{m,n}}{\partial p_k} = m \ell_{m-1,n-k}$$

Preuve Pour prouver cette assertion, il suffit de dériver la fonction génératrice $\Phi(s,t)$ par rapport à la variable p_k . En effet,

$$\begin{aligned} \frac{\partial \Phi(s,t)}{\partial p_k} &= s \left(\frac{\partial \zeta(t)}{\partial p_k} \right) e^{s\zeta(t)} \\ &= st^k \left(\sum_{m,n \geq 0} \ell_{m,n} \frac{s^m}{m!} t^n \right) \\ &= \sum_{m,n \geq 0} \ell_{m,n} \frac{s^{m+1}}{m!} t^{n+k} \\ &= \sum_{m,n \geq 0} (m+1) \ell_{m,n} \frac{s^{m+1}}{(m+1)!} t^{n+k} \end{aligned}$$

Or, si $m = 0$, on a $\ell_{0,n}$ qui est $\delta^{0,n}$, d'où :

$$\begin{aligned} \frac{\partial \Phi(s,t)}{\partial p_k} &= \sum_{n \geq 0} \ell_{0,n} st^{n+k} + \sum_{m \geq 1} \sum_{n \geq 0} (m+1) \ell_{m,n} \frac{s^{m+1}}{(m+1)!} t^{n+k} \\ &= st^k + \sum_{m \geq 2} \sum_{n \geq 0} m \ell_{m-1,n} \frac{s^m}{m!} t^{n+k} \\ &= \sum_{m \geq 1} \sum_{n \geq 0} m \ell_{m-1,n} \frac{s^m}{m!} t^{n+k} \\ &= \sum_{m,n \geq 0} m \ell_{m-1,n} \frac{s^m}{m!} t^{n+k} \end{aligned}$$

$$= \sum_{m,n \geq 0} m \ell_{m-1,n-k} \frac{s^m}{m!} t^n$$

en supposant que $\ell_{m,k} = 0$, si $k < 0$. D'autre part, par définition de la dérivée d'une fonction génératrice écrite sous forme de série, nous avons

$$\frac{\partial \Phi(s,t)}{\partial p_k} = \sum_{m,n \geq 0} \frac{\partial \ell_{m,n}}{\partial p_k} \frac{s^m}{m!} t^n$$

Si on identifie les coefficients de $\frac{s^m}{m!} t^n$ dans les deux dernières équations, on trouve :

$$\frac{\partial \ell_{m,n}}{\partial p_k} = m \ell_{m-1,n-k}$$

On remarque que $\frac{\partial \ell_{0,n}}{\partial p_k}$ n'a de sens que si $n = 0$, et dans ce cas la dérivée vaut zéro.

□

On obtient une preuve facile de la dernière proposition en utilisant les relations de la proposition 2.3.3. On procède alors par récurrence. Si $m = 0$, il n'y a rien à montrer et si $m = 1$, cela est évident. Supposons alors $m \geq 2$. On a :

$$\begin{aligned} \ell_{m,n} &= \sum_{j=0}^n p_j \ell_{m-1,n-j} \\ \frac{\partial \ell_{m,n}}{\partial p_k} &= \sum_{j=0}^n \frac{\partial}{\partial p_k} (p_j \ell_{m-1,n-j}) \\ &= \sum_{j=0}^{n-k} p_j (m-1) \ell_{m-2,n-k-j} + \ell_{m-1,n-k} \\ &= (m-1) \ell_{m-1,n-k} + \ell_{m-1,n-k} \\ &= m \ell_{m-1,n-k} \end{aligned}$$

Ce dernier résultat revêt un intérêt substantiel. Étant donné que nous disposons d'une récurrence pour les $\ell_{m,n}$, il devient aisé de calculer les dérivées partielles, voire le gradient, en exploitant la même récurrence.

Estimation

§3.1 INTRODUCTION

Dans cette section, nous explorerons diverses méthodes visant à estimer les p_j . Plus particulièrement, nous présenterons onze estimateurs, parmi lesquels l'estimateur de vraisemblance maximale figurera.

Avant d'entrer dans les détails des méthodes d'estimation, nous introduirons une suite doublement indexée notée $\{s(n,k)\}$. Cette suite permettra d'évaluer l'ampleur des calculs nécessaires à la mise en œuvre des différents estimateurs.

Une fois cette suite connue, nous définirons les paramètres de notre étude en fixant le nombre maximal de descendants par individu dans le processus. De plus, nous spécifierons la profondeur temporelle de notre observation en déterminant le nombre de générations que nous prendrons en considération.

Enfin, nous aborderons la procédure de maximisation qui sera utilisée pour estimer les p_j . Nous détaillerons également la procédure permettant de générer les partages d'un entier, un élément clé dans l'évaluation de la vraisemblance.

§3.2 LES LIMITES DE L'ÉTUDE

La suite $\{s(n, k)\}$

Afin d'appréhender l'ampleur des calculs à réaliser, nous envisageons de déterminer la cardinalité des ensembles suivants :

$$R(u, k) = \{(z_0, \dots, z_u) : z_0 = 1, z_{i+1} \leq kz_i, i = 1, \dots, u-1\}$$

En effet, $R(u, k)$ représente l'ensemble de toutes les réalisations qui pourraient éventuellement se produire si on admettait que $p_i \neq 0, 0 \leq i \leq k$ et $p_i = 0$ pour tout i plus grand que k . Les cas où, pour certains $i, 0 \leq i \leq k, p_i = 0$ ne sont que des sous-ensembles de $R(u, k)$.

Comme mentionné dans l'introduction de ce chapitre, nous avons 11 estimateurs à analyser, dont un est l'estimateur à vraisemblance maximale. Cela signifie que pour chaque élément de $R(u, k)$, nous devons maximiser une fonction. Les fonctions en question sont d'une complexité arithmétique évidente et presque toutes différentes les unes des autres. Étant donné que la recherche du maximum d'une fonction à plusieurs variables nécessite un temps d'ordinateur significatif, il est nécessaire de connaître dès le départ l'ampleur des calculs impliqués.

Notons donc par $n(u, j, k)$ le nombre d'éléments de $R(u, k)$ ayant $z_u = j, 0 \leq j \leq k^u$. Évidemment, nous aurons :

$$\text{card } R(u, k) = \sum_{j=0}^{k^u} n(u, j, k)$$

Proposition 3.2.1 Les $n(u, j, k)$ satisfont les relations suivantes :

1. $n(0, j, k) = \delta^{1, j}$
2. $n(u, 0, k) = \sum_{i=0}^{k^{u-1}} n(u-1, i, k) = \text{card } R(u-1, k)$
3. $n(u, j, k) = \begin{cases} \sum_{i=\lfloor (j+k-1)/k \rfloor}^{k^{u-1}} n(u-1, i, k) & 0 < j \leq k^u; \\ 0 & \text{sinon.} \end{cases}$

où $\lfloor x \rfloor$ est la partie entière de x .

Preuve Les relations a. et b. sont évidentes. Pour démontrer la troisième, il suffit de voir que pour générer $mk+1, mk+2, \dots, mk+k$ descendants, il faut avoir au moins $m+1$ pères. Mais puisque $\lfloor (mk+j+k-1)/k \rfloor = m+1$ pour $j = 1, 2, \dots, k$, la proposition est démontrée. \square

TABLEAU 3.1 – Tableau de quelques valeurs des $n(u, j, k)$, pour $k = 2$

$k = 2$	j									
u	0	1	2	3	4	5	6	7	8	9
0	0	1								
1	1	1	1							
2	3	2	2	1	1					
3	9	6	6	4	4	2	2	1	1	

TABLEAU 3.2 – Tableau de quelques valeurs des $n(u, j, k)$, pour $k = 3$

$k = 3$	j														
u	0	1	2	3	4	5	6	...	22	23	24	25	26	27	
0	0	1													
1	1	1	1	1											
2	4	3	3	3	2	2	2								
3	22	18	18	18	15	15	15	...	2	2	2	1	1	1	

Dans les tableaux 3.1 et 3.2 on donne quelques-uns des $n(u, j, k)$ pour $k = 2, 3$ et $u = 0, 1, 2, 3$.

Nous supposons toujours que $k \geq 2$. Les cas où $k = 0$ ou $k = 1$ sont dénués d'intérêt. On peut considérer les lignes des tableaux des $n(u, j, k)$ comme des suites finies. Ainsi, en excluant les $n(u, 0, k)$, chaque élément de ces suites se répète successivement k fois. Ceci est facile à vérifier en utilisant la troisième relation de la proposition 3.2.1.

Pour un k fixé, si l'on écrit les termes des suites en faisant varier l'indice j de k^u à 1, la suite correspondant à un certain u est simplement une troncation de la suite correspondant à $u + 1$. Le fait de laisser tomber les $n(u, 0, k)$ ne nous gênera pas, car nous pouvons les retrouver à l'aide de la deuxième relation de la proposition 3.2.1.

Exemple 3.2.1

$$k = 2 \quad u = 3 \mapsto (1, 1, 2, 2, 4, 4, 6, 6)$$

$$u = 2 \mapsto (1, 1, 2, 2)$$

$$k = 3 \quad u = 3 \mapsto (1, 1, 1, 2, 2, 2, 3, 3, 3, 5, 5, 5, \dots, 18, 18, 18)$$

$$u = 2 \mapsto (1, 1, 1, 2, 2, 2, 3, 3, 3)$$

□

L'analyse des nouvelles suites ainsi formées peut donc se concentrer sur les termes dont l'indice est un multiple de k . Étant donné que les suites ne sont que des troncations, pour un k fixé, d'une même suite (à toutes fins pratiques), nous allons une fois de plus simplifier notre étude en négligeant l'indice u . Nous considérerons ainsi des suites doublement indexées que nous noterons $s(n, k)$

Exemple 3.2.2

$$s(n, 2) = \{1, 2, 4, 6, 10, 14, 20, 26, \dots\}$$

$$s(n, 3) = \{1, 2, 3, 5, 7, 9, 12, 15, \dots\}$$

□

Nous donnerons ultérieurement une récurrence permettant de calculer les termes de ces suites. Pour le moment, procédons à la démonstration du lemme 3.2.1.

Lemme 3.2.1 Soit m un entier tel que $0 \leq m < k^u$. Alors

$$n(u, k^u - m, k) = n(k^u - \lfloor m/k \rfloor k, k)$$

Preuve Soit m un entier satisfaisant l'hypothèse. Alors, il existe un entier i , $0 \leq i < k^{u-1}$ tel que $m = ik + r$, où r est tel que $0 \leq r < k$. D'une part,

$$\lfloor (k^u - ik - r + k - 1)/k \rfloor = k^{u-1} - i$$

et

$$\lfloor (k^u - \lfloor (ik - r)/k \rfloor k + k - 1)/k \rfloor = k^{u-1} - i$$

D'autre part, en utilisant la troisième relation de la proposition 3.2.1, on voit que

$$n(u, k^u - m, k) = n(u, k^u - \lfloor m/k \rfloor k, k).$$

Ceci termine la démonstration du lemme. □

Proposition 3.2.2 Les suites $s(n, k)$ satisfont les relations suivantes :

1. $s(0, k) = 1$
2. $s(n, k) = s(n - 1, k) + s(\lfloor n/k \rfloor, k)$

Preuve Tout d'abord, considérons $R(1, k)$. Le lien entre les $s(j, k)$ et les $n(u, i, k)$ s'exprime par :

$$s(j, k) = n(u, k^u - jk, k), \quad j = 0, 1, \dots, k^{u-1} - 1$$

Ainsi, $s(0, k) = n(1, k, k) = 1$. Ceci démontre la relation α .

Si $n = 1$, $s(1, k) = n(2, k(k-1), k)$. Puisque $\lfloor (k(k-1) + k - 1)/k \rfloor = k - 1$, en utilisant la troisième relation de la proposition 3.2.1, on obtient $n(2, k(k-1), k) = n(1, k-1, k) + n(1, k, k)$. Or, $n(1, k-1, k) = n(1, k, k) = 1$. Donc $s(1, k) = 2$. De plus, sachant que $s(0, k) = 1$, $s(0, k) + s(0, k) = 2$. Ceci démontre la validité de la deuxième relation lorsque $n = 1$.

Pour un entier non-négatif donné m , sélectionnons un entier u tel que $m < k^{u-1}$. Il est établi que, en examinant les u premières générations, la valeur maximale z_u demeure inférieure ou égale à k^u pour chaque élément de $R(u, k)$. En analysant la séquence $n(u, j, k)$ associée à u , on constate la présence de k^{u-1} termes appartenant à la suite $\{s(n, k)\}$, numérotés de 0 à $k^{u-1} - 1$.

Écrivons m sous la forme $k^{u-1} - q = m$ et posons $j = kq$. Nous aurons alors $j + k = (q + 1)k$. Comme précédemment,

$$n(u, j, k) = \sum_{i=q}^{k^{u-1}} n(u-1, i, k)$$

$$n(u, j+k, k) = \sum_{i=q+1}^{k^{u-1}} n(u-1, i, k)$$

À l'aide des ces deux équations on voit :

$$n(u-1, q, k) = n(u, j, k) - n(u, j+k, k)$$

$$n(u, j, k) = n(j, j+k, k) + n(u-1, q, k)$$

En soustrayant et additionnant k^u à l'indice du centre obtient :

$$n(u, k^u - k(k^{u-1} - q), k) = n(u, k^u - k(k^{u-1} - q - 1), k) + n(u-1, q, k)$$

En écrivant cette égalité en termes des $s(j, k)$, on trouve :

$$s(k^{u-1}, k) = s(k^{u-1} - q - 1, k) + n(u-1, q, k)$$

Mais en remplaçant $k^{u-1} - q$ par m on obtient :

$$s(m, k) = s(m-1, k) + n(u-1, q, k)$$

Cependant, le lemme 3.2.1 nous indique que

$$n(u-1, q, k) = n(u-1, k^{u-1} - m, k)$$

$$= n(u-1, k^{u-1} - \lfloor m/k \rfloor k, k)$$

En utilisant la relation entre les $n(u, j, k)$ et les $s(j, k)$, le dernier terme n'est rien d'autre que $s(\lfloor m/k \rfloor, k)$, Donc,

$$s(m, k) = s(m - 1, k) + s(\lfloor m/k \rfloor, k)$$

□

Le lemme 3.2.2 sera nécessaire à la démonstration de la proposition 3.2.3.

Lemme 3.2.2

$$s(k^{u-1}, k) = k \sum_{i=0}^{k^{u-1}-1} s(i, k) \quad u \geq 2$$

Preuve

$$\begin{aligned} s(k^u - 1, k) &= n(u + 1, k^{u+1} - (k^u - 1)k, k) \\ &= n(u + 1, k, k) \\ &= \sum_{i=1}^{k^u} n(u, i, k) \\ &= k \sum_{i=0}^{k^u-1} s(i, k) \end{aligned}$$

□

Le résultat suivant facilite le calcul de la cardinalité des ensembles $R(u, k)$.

Proposition 3.2.3

$$\text{card } R(u, k) = s(k^u, k) - 1$$

Preuve Si $u = 1$, on doit montrer que $\text{card } R(1, k) = s(k, k) - 1$. On voit aisément que $\text{card } R(1, k) = k + 1$. Aussi, en appliquant la récurrence de la proposition 3.2.2, on trouve :

$$\begin{aligned} s(k, k) &= s(k - 1, k) + s(1, k) \\ &= s(k - 2, k) + s(0, k) + s(1, k) \\ &= s(k - 3, k) + 2s(0, k) + s(1, k) \\ &\vdots \\ &= s(0, k) + (k - 1)s(0, k) + s(1, k) \\ &= ks(0, k) + s(1, k) = k + 2 \end{aligned}$$

TABLEAU 3.3 – Cardinalité de $R(u, k)$ pour de petites valeurs de u et k

k	u				
	1	2	3	4	5
2	3	9	35	201	1 827
3	4	22	238	5 827	342 382
4	5	45	1 085	79 325	18 583 581
5	6	81	3 706	642 456	$4,46021 \times 10^8$
6	7	133	10 339	3 649 345	$6,18812 \times 10^9$

Soit $u \geq 2$, quelconque, et procédons par récurrence :

$$\begin{aligned}
 \text{card } R(u, k) &= \sum_{i=0}^{k^u} n(u, i, k) \\
 &= \sum_{i=1}^{k^u} n(u, i, k) + n(u, 0, k) \\
 &= k \sum_{i=0}^{k^{u-1}-1} s(i, k) + s(k^{u-1}, k) - 1
 \end{aligned}$$

car, par l'hypothèse d'induction, $n(u, 0, k) = \text{card } R(u-1, k) = s(k^{u-1}, k) - 1$. La transformation faite à la sommation est justifiée par la construction des $s(j, k)$.
Donc

$$\begin{aligned}
 \text{card } R(u, k) &= s(k^u - 1, k) + s(k^{u-1}, k) - 1 \\
 &= s(k^u, k) - 1
 \end{aligned}$$

par le lemme 3.2.2 et la récurrence des $s(j, k)$. □

Pour ceux qui souhaitent approfondir l'étude des $s(j, k)$, l'espoir de découvrir de nouvelles relations peut être alimenté par le fait que la fonction génératrice f_k des $s(j, k)$ pour un k fixé est donnée par :

$$f_k(t) = \frac{1}{(1-t)^2(1-t^k)(1-t^{k^2})(1-t^{k^3}) \dots}$$

Cette observation a été faite par Gilbert Labelle, bien que la démonstration ne soit pas fournie ici.

Le tableau 3.3 donne la cardinalité de $R(u, k)$ pour de petites valeurs de u et k . Pour faire une étude des estimateurs à l'aide des distributions exactes, nous avons décidé de limiter à deux le nombre maximal de descendants par individu. Cependant, sachant que $\text{card}, R(6, 2) = 27,337$, un nombre relativement élevé, nous limitons notre étude à seulement 5 générations.

Les estimateurs

Pour chaque réalisation $\langle 1, z_1, \dots, z_n \rangle$ d'un processus de Galton-Watson, où la distribution de la variable aléatoire ξ est telle que $P[\xi \geq 3] = 0$, nous pouvons bâtir un tableau semblable au tableau 3.4.

TABLEAU 3.4 – Énumération des partages pour la réalisation $\langle 1, 1, 2, 1, 2, 4, 5 \rangle$

i	Réalisation		Partages admissibles			Coefficient multinomial
	z_i	z_{i+1}	$\lambda_{2,k}^{i+1}$	$\lambda_{1,k}^{i+1}$	$\lambda_{0,k}^{i+1}$	
0	1	1	0	1	0	1
1	1	2	1	0	0	1
2	2	1	0	1	1	2
3	1	1	1	0	0	1
4	2	4	2	0	0	1
5	4	5	$\frac{2}{1}$	$\frac{1}{3}$	$\frac{1}{0}$	$\frac{12}{4}$

Les informations nécessaires au calcul des estimateurs suivants sont disponibles dans le tableau 3.4. Posons

$$\mathcal{P}_i = \left\{ (0^{\lambda_0^{i+1}} 1^{\lambda_1^{i+1}} 2^{\lambda_2^{i+1}}) : \sum_{j=0}^2 \lambda_j^{i+1} = z_i, \sum_{j=0}^2 j \lambda_j^{i+1} = z_{i+1}, \right\}$$

Alors, \mathcal{P}_i est l'ensemble de tous les partages admissibles pouvant donner lieu à la transition de z_i à z_{i+1} . Soit \mathcal{P} l'ensemble suivant :

$$\mathcal{P} = \{ (\lambda_{0,k}^{i+1}, \lambda_{1,k}^{i+1}, \lambda_{2,k}^{i+1}), k = 1, \dots, c_i, i = 0, \dots, n-1 \}$$

où c_i est la cardinalité de \mathcal{P}_i .

Premier estimateur

Notre première méthode d'estimation consistera à évaluer

$$n_i = \sum_{j=0}^{n-1} \sum_{k=1}^{c_j} \lambda_{i,k}^{j+1}, \quad i = 0, 1, 2$$

et poser

$$\hat{p}_i = \frac{n_i}{n_0 + n_1 + n_2}, \quad i = 0, 1, 2$$

Indubitablement, il s'agit d'un estimateur extrêmement basique et, par conséquent, rudimentaire.

Deuxième estimateur

L'estimateur que nous venons de définir présente, *a priori*, l'inconvénient évident d'attribuer le même poids à chacun des éléments de \mathcal{P} . Ne serait-il pas pertinent de pondérer la somme en fonction de la taille de la population-mère ? C'est précisément ce que nous allons faire. À cet effet, considérons :

$$n_i = z_i \sum_{j=0}^{n-1} \sum_{k=1}^{c_j} \lambda_{i,k}^{j+1}, \quad i = 0, 1, 2$$

et

$$\hat{p}_i = \frac{n_i}{n_0 + n_1 + n_2}, \quad i = 0, 1, 2$$

Troisième estimateur

Il serait peut-être bon de donner plus de poids à une transition lorsque celle-ci peut se faire de multiples façons. Il s'agirait alors de pondérer par le coefficient multinomial associé. Dans ce cas, nous aurions :

$$n_i = \sum_{j=0}^{n-1} \sum_{k=1}^{c_j} \frac{(\lambda_{0,k}^{j+1} + \lambda_{1,k}^{j+1} + \lambda_{2,k}^{j+1})!}{\lambda_{0,k}^{j+1}! \lambda_{1,k}^{j+1}! \lambda_{2,k}^{j+1}!} \lambda_{i,k}^{j+1}, \quad i = 0, 1, 2$$

et

$$\hat{p}_i = \frac{n_i}{n_0 + n_1 + n_2}, \quad i = 0, 1, 2$$

Quatrième estimateur

Un quatrième estimateur proposé est celui obtenu en combinant les deux dernières méthodes,

$$n_i = z_i \sum_{j=0}^{n-1} \sum_{k=1}^{c_j} \frac{(\lambda_{0,k}^{j+1} + \lambda_{1,k}^{j+1} + \lambda_{2,k}^{j+1})!}{\lambda_{0,k}^{j+1}! \lambda_{1,k}^{j+1}! \lambda_{2,k}^{j+1}!} \lambda_{i,k}^{j+1}, \quad i = 0, 1, 2$$

et

$$\hat{p}_i = \frac{n_i}{n_0 + n_1 + n_2}, \quad i = 0, 1, 2$$

Cinquième estimateur

Maintenant, nous chercherons à extraire davantage d'informations. Considérons par exemple le cas où la population-mère ne compte qu'un seul individu. Dans une telle situation, nous sommes assurés que p_k est différent de zéro si k est le nombre de descendants ($k = 0, 1, 2$). En fait, plus le coefficient d'une transition est élevé, moins nous sommes certains de la façon exacte dont elle s'est produite. Ainsi, nous allons pondérer par l'inverse du coefficient multinomial. On pose alors,

$$n_i = \sum_{j=0}^{n-1} \sum_{k=1}^{c_j} \frac{\lambda_{0,k}^{j+1}! \lambda_{1,k}^{j+1}! \lambda_{2,k}^{j+1}!}{(\lambda_{0,k}^{j+1} + \lambda_{1,k}^{j+1} + \lambda_{2,k}^{j+1})!} \lambda_{i,k}^{j+1}, \quad i = 0, 1, 2$$

et

$$\hat{p}_i = \frac{n_i}{n_0 + n_1 + n_2}, \quad i = 0, 1, 2$$

Sixième estimateur

Cet estimateur sera du même type que le précédent, mais nous pondérerons aussi par la taille de la population-mère. Nous poserons alors;

$$n_i = z_i \sum_{j=0}^{n-1} \sum_{k=1}^{c_j} \frac{\lambda_{0,k}^{j+1}! \lambda_{1,k}^{j+1}! \lambda_{2,k}^{j+1}!}{(\lambda_{0,k}^{j+1} + \lambda_{1,k}^{j+1} + \lambda_{2,k}^{j+1})!} \lambda_{i,k}^{j+1}, \quad i = 0, 1, 2$$

et

$$\hat{p}_i = \frac{n_i}{n_0 + n_1 + n_2}, \quad i = 0, 1, 2$$

Septième estimateur

Considérons maintenant \mathcal{A} , l'ensemble de tous les arbres (sachant qu'un individu ne peut avoir plus de deux descendants) qui auraient pu donner lieu à la réalisation observée du processus. Dans l'exemple du début de la section, il est facile de voir qu'il y a 32 arbres possibles. Il est aussi clair que la fonction de vraisemblance d'un arbre a est donnée par :

$$L(a) = p_0^{n_0} p_1^{n_1} p_2^{n_2}$$

où n_i est le nombre de fois où dans l'arbre un individu a eu i descendants, D'autre part, chacun des 32 arbres dénombrés trouve sa fonction de vraisemblance parmi les deux fonctions

$$L_1(p_0, p_1, p_2) = p_0^2 p_1^3 p_2^6$$

$$L_2(p_0, p_1, p_2) = p_0^1 p_1^5 p_2^5$$

Nous dirons donc que deux arbres sont V -équivalents si et seulement si, ils ont la même vraisemblance. Il est connu que $p_0^{n_0} p_1^{n_1} p_2^{n_2}$ atteint son maximum en $\hat{p}_i = n_i / (n_0 + n_1 + n_2)$. Dénotons par V la relation d'équivalence que nous venons de définir. Nous allons travailler sur \mathcal{A}/V , l'ensemble \mathcal{A} quotienté par V . On remarque que l'estimation à vraisemblance maximale donne la même estimation pour chaque élément d'une même classe. Elle est donc indépendante du représentant. Dans l'exemple que nous étudions, il y a essentiellement deux arbres :

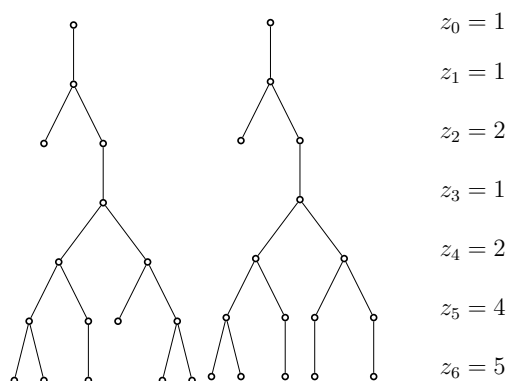


FIGURE 3.1 – Deux arbres possibles pour la réalisation $\langle 1, 1, 2, 1, 2, 4, 5 \rangle$

Pour chacun d'eux faire une estimation à vraisemblance maximale de p_i , $i = 0, 1, 2$.

Nous proposons une méthode complètement opposée aux méthodes usuelles d'estimation. Dans un certain sens, elle consisterait à laisser le bénéfice du doute à la véritable distribution de Z_1 . De toutes les estimations à vraisemblance maximale faites sur le niveau de \mathcal{A}/V , nous retenons celle qui a la plus grande entropie. En d'autres mots, nous cherchons à maximiser la fonction suivante :

$$H(p_0, p_1, p_2) = - \sum_{i=0}^2 p_i \ln(p_i)$$

Cette méthode non-usuelle d'estimer un paramètre est très discutable. Mais nous voulons tout de même regarder comment elle se comporte pour des raisons qui seront exposées au chapitre 5.

Huitième estimateur

De toutes les estimations faites au niveau de \mathcal{A}/V , nous conserverons l'estimation qui aura donné la plus grande valeur à la fonction de vraisemblance associée au processus de Galton-Watson, telle que définie au chapitre 1.

Neuvième estimateur

Il s'agit simplement d'effectuer une estimation à vraisemblance maximale en fonction de la réalisation observée du processus. En d'autres termes, nous cherchons à maximiser la fonction de vraisemblance définie au chapitre 1, sur l'ensemble du simplexe S . Cette estimation prend en considération le fait que nous sommes conscients qu'un individu ne peut générer plus de deux descendants.

Dixième estimateur

Une autre approche pour l'estimation consisterait à générer tous les arbres ayant pu donner lieu à la réalisation du processus, effectuer une estimation à vraisemblance maximale au sein de chacun de ces arbres, et finalement, à titre d'estimation finale, prendre la moyenne arithmétique simple de ces estimations.

Onzième estimateur

Il s'agit de faire un travail analogue à celui de la dixième méthode, mais en travaillant uniquement sur \mathcal{A}/V ,

Tandis que nous étudierons le comportement de ces estimateurs à l'aide de distributions exactes, nous en profiterons pour étudier aussi celui de \hat{m} . Cette dernière statistique deviendra notre point de référence, car nous savons que sous certaines conditions elle vers m .

§3.3 LES ALGORITHMES

Nous donnerons au prochain chapitre certains détails relatifs à l'utilisation des différents programmes créés pour mener à terme notre étude. Toutefois, nous allons préalablement expliciter les fondements théoriques de deux des principales procédures utilisées.

Maximisation de la fonction de vraisemblance

Nous voulons maximiser la fonction suivante

$$L(z_0, \dots, z_n; p_0, p_1, p_2) = \prod_{i=1}^n \left(\sum_{\lambda^i \in \mathcal{P}_i} C(\lambda^i) p^{\lambda^i} \right).$$

compte tenu du fait que $p_k = 0$, pour tout $k > 2$ et donc que $p_0 + p_1 + p_2 = 1$.

En analyse numérique, une technique couramment utilisée pour optimiser une fonction à plusieurs variables est la « *méthode du gradient* ». Elle est souvent assimilée à la « *méthode de la montée la plus abrupte* » en raison du fait que le gradient pointe dans la direction de la plus forte augmentation locale de la fonction. Cette méthode nécessite cependant la connaissance d'un point où une certaine dérivée directionnelle s'annule. Pour trouver ce point nous utiliserons la méthode « *regula falsi* ». ¹

Dans le cas qui nous intéresse la recherche du maximum se fera sur le simplexe suivant :

$$S = \{(x, y) : x \geq 0, y \geq 0, x + y \leq 1\}$$

Il s'agit de partir d'un point initial (x_0, y_0) et de se déplacer dans la direction du gradient. Soit G , une fonction à maximiser sur S (nous parlerons plus tard de la fonction de vraisemblance). Pour le point (x_0, y_0) , dénotons par $\nabla G(x_0, y_0)$, le gradient de G évalué au point (x_0, y_0) . Nous allons trouver un point (x_1, y_1) sur la droite

$$d : (x_0, y_0) + \nabla G(x_0, y_0) t$$

1. On dit parfois : « *méthode de la fausse position*. »

satisfaisant les contraintes qui définissent S , pour lequel la fonction G prendra sa plus grande valeur. Si (x_1, y_1) n'est pas le point où G est maximale sur tout le simplexe, nous prendrons (x_1, y_1) comme nouveau point initial et nous ferons une nouvelle itération. Bien sûr, du point de vue numérique, le point où G sera maximale ne sera connu qu'à un certain nombre de décimales près.

Pour trouver le point (x_1, y_1) appartenant à la droite d où la fonction G atteint son maximum, il peut être nécessaire d'appliquer la méthode de la *regula falsi*. Cette méthode vise à annuler la dérivée dans la direction de la droite d afin d'optimiser le point recherché.

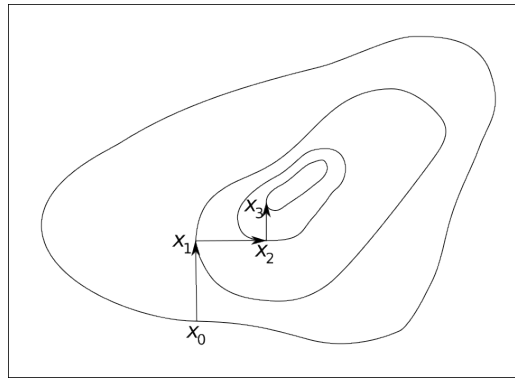


FIGURE 3.2 – Exemple de trajectoire empruntée par l'algorithme du gradient

Sur le graphique de la figure 3.2, nous illustrons le type de trajectoire parcourue en utilisant l'algorithme du gradient. Élaborons maintenant l'algorithme dans le cas particulier qui nous intéresse.

Lemme 3.3.1 Soit f_i , $i = 1, \dots, n$, n fonctions dérivables définies par rapport p_k , $k = 0, \dots, m$. Alors,

$$\frac{\partial}{\partial p_k} \left(\prod_{i=1}^n f_i \right) = \sum_{j=1}^n \left[\prod_{i=1}^n \left(\frac{\partial}{\partial p_k} \right)^{\delta^{ij}} f_i \right]$$

où $\frac{\partial}{\partial p_k}$ est considéré comme un opérateur et δ^{ij} est le delta de Kronecker.

L'opérateur est appliqué si et seulement si $i = j$. Nous ne croyons pas qu'il soit nécessaire de démontrer ce lemme. Posons

$$L^*(z_0, \dots, z_n; p_0, p_1) = L(z_0, \dots, z_n; p_0, p_1, 1 - p_0 - p_1)$$

Il nous reste à trouver le gradient de L^* . Par une simple dérivée en chaîne, nous obtenons :

$$\frac{\partial L^*}{\partial p_k} = \frac{\partial L}{\partial p_k} - \frac{\partial L}{\partial p_2}, \quad k = 0, 1$$

Nous étudierons L^* restreinte à $p_0 \geq 0, p_1 \geq 0$ et $p_0 + p_1 \leq 1$.

Nous avons montré au chapitre 1 que la dérivée par rapport à p_k de $\ell_{m,n}$ est $m\ell_{m-1,n-k}$. Donc par le lemme 3.3.1, nous avons :

$$\frac{\partial L}{\partial p_k} = \sum_{j=0}^{n-1} \left[\prod_{i=0}^{n-1} z_i^{\delta^{ij}} \ell_{z_i - \delta^{ij}, z_{i+1} - k\delta^{ij}} \right], \quad k = 0, 1, 2$$

Ainsi, malgré la complexité apparente des expressions, les $\partial L^* / \partial p_k, k = 0, 1$, seront facilement calculés... par ordinateur !

Au début du prochain chapitre, nous aborderons la question du choix du point initial. Pour l'instant, soit $X_0 = (x_0, y_0)$, le point initial. Posons

$$a_i = \left. \frac{\partial L^*}{\partial p_i} \right|_{(x_0, y_0)} = \left. \frac{\partial L}{\partial p_i} \right|_{(x_0, y_0, 1-x_0-y_0)} - \left. \frac{\partial L}{\partial p_2} \right|_{(x_0, y_0, 1-x_0-y_0)} \quad i = 0, 1$$

et soit $\vec{A}_0 = (a_0, a_1)$. Le vecteur \vec{A}_0 donne la direction de l'accroissement maximal. Nous devons maximiser $H(t) = L^*(X_0 + t\vec{A}_0)$ par rapport à t . Posons $F(t) = H'(t)$. Nous aimerions trouver un t tel que $F(t) = 0$. Pour ce, nous appliquons la méthode *regula falsi*. Il s'agit en fait d'une méthode d'interpolation inverse. Supposons que t_{i-1} et t_{i-2} soient des points tels que $F(t_{i-2}) > 0$ et $F(t_{i-1}) < 0$. Nous obtenons une nouvelle approximation t en posant :

$$t_i = t_{i-1} - \frac{F(t_{i-1})}{F(t_{i-1}) - F(t_{i-2})} \times (t_{i-1} - t_{i-2})$$

La figure 3.3 illustre ce à quoi cela correspond. Dans ses grandes lignes, l'algorithme peut être décrit comme suit :

- 1) Choisir un point initial.
- 2) Évaluer le gradient en ce point.
- 3) Voulant avancer dans la direction du gradient, ∇L^* , identifier sur quelle contrainte on pourrait éventuellement se buter. Voir la figure 3.4. (Par opposition au point initial, nous appellerons ce point, le point terminal).
- 4) Diviser l'intervalle déterminé par le point initial et le point terminal en dix parties égales. Onze points, X_0, \dots, X_{10} , sont ainsi déterminés.

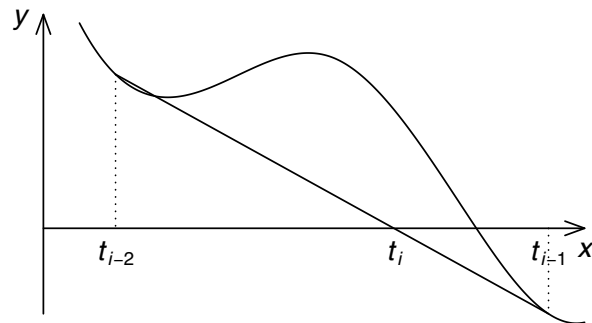


FIGURE 3.3 – Méthode *regula falsi* : détermination de t_i à partir de $(t_{i-2}, F(t_{i-2}))$ et $(t_{i-1}, F(t_{i-1}))$

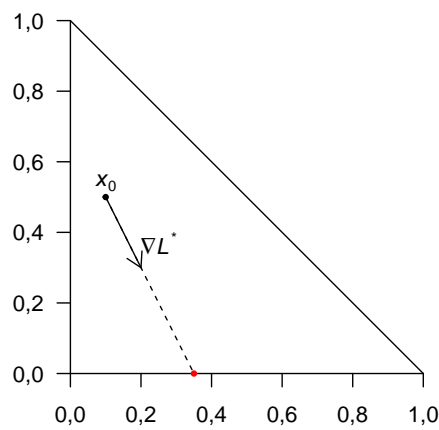


FIGURE 3.4 – Déplacement dans la direction du gradient
Le point rouge indique le point final.

5) Faire varier i de 0 à 10. De deux choses l'une

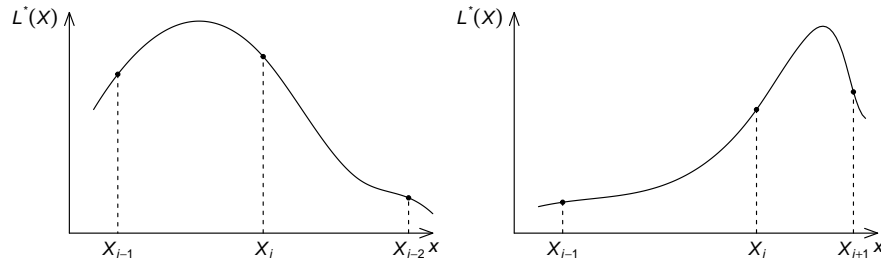
- (i) $L^*(X_i)$ croît avec i ;
- (ii) Il existe un i tel $L^*(X_i) \geq L^*(X_{i+1})$

Dans le premier cas, nous admettrons que sur le segment de droite considéré, le point où L^* est maximale est X_{10} . On passe à l'étape 6).

Dans le second cas,

- a) Si $L^*(X_i) = L^*(X_{i+1})$, on admettra que le maximum sur la droite est atteint entre X_i et X_{i+1} ;

- b) Si $L^*(X_i) > L^*(X_{i+1})$, nous essayons de déterminer si le maximum se situe avant ou après X_i .



Pour le vérifier, nous utilisons la dérivée directionnelle évaluée en x_i . Si elle est négative, le maximum se trouve entre X_{i-1} et X_i . Si elle est positive, le maximum se situe entre X_i et X_{i+1} .

- 6) Si nous sommes satisfaits du résultat obtenu, nous arrêtons les itérations. Sinon, nous retournons en 2 en utilisant le point trouvé comme nouveau point initial.

Il est donc possible de trouver numériquement l'estimateur à vraisemblance maximale, si nous sommes relativement près de ce point. Lorsque nous aboutissons sur une contrainte et que le gradient pointe vers l'extérieur de la région, il faut alors avancer dans la direction de la projection du gradient sur la contrainte. Cette projection est nulle lorsque nous sommes au point maximum.

Génération des partages d'un entier

La connaissance de tous les partages admissibles pour chacune des transitions de Z_i à Z_{i+1} est nécessaire pour le calcul des autres estimateurs. Nous expliquons maintenant l'algorithme utilisé pour les générer.

Supposons que l'on veuille connaître tous les partages de n en au plus p parts, aucune n'étant plus grande que m . On remarque :

$$n = \lfloor n/m \rfloor m + (n - \lfloor n/m \rfloor m)$$

Posons $\lambda_m = \lfloor n/m \rfloor$ et $r = n - m\lambda_m$. Si $r = 0$, posons $\lambda_r = 0$, sinon posons $\lambda_r = 1$. Finalement, soit $\lambda_0 = p - \lambda_m - \lambda_r$. Ceci nous amène à considérer le partage :

$$0^{\lambda_0} 1^{\lambda_1} \dots (r-1)^0 r^{\lambda_r} (r+1)^0 \dots m^{\lambda_m}$$

Il s'agit clairement d'un partage admissible lors de la transition de $z_i = p$ à $z_{i+1} = n$, où chaque individu ne peut engendrer plus de m descendants. On

pourrait dénommer ce partage : « *partage le plus grossier*. » Il s'agit de l'élément minimal relativement à l'ordre lexicographique décroissant.

Nous allons maintenant montrer comment générer le successeur (relativement à la relation d'ordre mentionnée plus haut) du partage $0^{\lambda_0} 1^{\lambda_1} \dots m^{\lambda_m}$. Le partage ayant au plus p parts, il peut être représenté à l'aide d'un vecteur de dimension p :

$$(\alpha_0, \dots, \alpha_p)$$

Il s'agit de trouver le premier indice j tel que :

$$n' = \sum_{k=j+1}^p \alpha_k + 1$$

sera « partageable » en au plus $p - j$ parts, aucune ne devant être supérieure à $\alpha_j - 1$. Le partage suivant $(\alpha_1, \dots, \alpha_p)$ est obtenu en concaténant $(\alpha_1, \dots, \alpha_{j-1})$ avec le « sous-partage » le plus grossier de n' en au plus $p - j$ parts, toutes plus petites ou égales à α_{j-1} . Une telle procédure s'arrête lorsque nous avons le partage $(\alpha_1, \dots, \alpha_p)$ satisfaisant la relation suivante :

$$\sum_{k=2}^p \alpha_k + 1 > (\alpha_1 - 1)(p - 1)$$

La procédure génère une et une seule fois tous les partages désirés.

Exemple 3.3.1 Soit $p = 5$, $m = 7$ et $n = 15$. On trouve comme premier partage $\langle 7, 7, 1, 0, 0 \rangle$. Pour illustrer la méthode permettant de trouver le successeur, cherchons celui de $\langle 6, 5, 2, 1, 1 \rangle$. On voit que que 2 est le premier indice trouvé conforme à nos exigences. Dans ce cas, $n' = 5$, et on doit le partager en au plus trois parts, aucune ne devant être supérieure à 4. Ce qui donne $\langle 4, 1, 0 \rangle$. En concaténant avec $\langle 6, 4 \rangle$, on obtient $\langle 6, 4, 4, 1, 0 \rangle$. La liste de tous les partages de 15 en au plus 5 parts toutes inférieures ou égales à 7 se retrouve dans le tableau 3.5. \square

TABLEAU 3.5 – Partages de 15 en au plus 5 parts toutes inférieures ou égales à 7

α_1	α_2	α_3	α_4	α_5	α_1	α_2	α_3	α_4	α_5	α_1	α_2	α_3	α_4	α_5
7	7	1	0	0	6	6	1	1	1	5	5	2	2	1
7	6	2	0	0	6	5	4	0	0	5	4	4	2	0
7	6	1	1	0	6	5	3	1	0	5	4	4	1	1
7	5	3	0	0	6	5	2	2	0	5	4	3	3	0
7	5	2	1	0	6	5	2	1	1	5	4	3	2	1
7	5	1	1	1	6	4	4	1	0	5	4	2	2	2
7	4	4	0	0	6	4	3	2	0	5	3	3	3	1
7	4	3	1	0	6	4	3	1	1	5	3	3	2	2
7	4	2	2	0	6	4	2	2	1	4	4	4	3	0
7	4	2	1	1	6	3	3	3	0	4	4	4	2	1
7	3	3	2	0	6	3	3	2	1	4	4	3	3	1
7	3	3	1	1	6	3	2	2	2	4	4	3	2	2
7	3	2	2	1	5	5	5	0	0	4	3	3	3	2
7	2	2	2	2	5	5	4	1	0	3	3	3	3	3
6	6	3	0	0	5	5	3	2	0					
6	6	2	1	0	5	5	3	1	1					



Analyse des résultats

§4.1 INTRODUCTION

Nous allons maintenant fournir des détails supplémentaires sur l'implémentation des algorithmes utilisés. De plus, en nous appuyant sur les résultats obtenus, nous essaierons d'évaluer la performance des estimateurs.

Dans un premier temps, nous ferons une brève description de la configuration informatique utilisée, notamment en ce qui concerne le type d'ordinateur et le type d'arithmétique pour les calculs. Ensuite, nous aborderons l'approche adoptée pour sélectionner le point initial de la procédure de maximisation. Vous trouverez à l'annexe A les procédures les plus significatives mises en œuvre.

En conclusion de ce chapitre, nous analyserons les résultats issus des programmes pour tirer des conclusions sur la performance globale des estimateurs.

§4.2 LES CALCULS

L'outil de calcul

Toutes les procédures utilisées ont été mises au point sur le PDP-10, produit de la Digital Equipment Company, du centre de calcul de l'Université du Québec à Montréal. Il s'agit d'un ordinateur dont les mots ont une longueur de 36 bits. Nous avons vite constaté certains problèmes de convergence. Mais en utilisant l'arithmétique en « *double précision* », ces problèmes ont disparu. Cependant, cet ordinateur est relativement lent lorsqu'on le compare au Cyber du centre commun du réseau de l'Université du Québec. Toutes les procédures ont donc été implantées sur le Cyber. Nous avons conservé l'utilisation de l'arithmétique en double précision. Un mot ayant 60 bits de longueur, nos calculs en double précision ont été exécutés en utilisant 120 bits par nombre. Notons ici que tous les programmes ont été écrits en Fortran et que nous avons utilisé le compilateur FTN.

Le choix du point initial pour la maximisation

Nous avons annoncé dans le chapitre 3 notre intention d'étudier la vraisemblance sur le triangle délimité par les points $(0,0)$, $(0,1)$, et $(1,0)$ dans \mathbb{R}^2 . En conséquence, les fonctions que nous avons examinées sont définies de \mathbb{R}^2 vers \mathbb{R} et peuvent être aisément représentées graphiquement à l'aide de surfaces.

Le choix du point initial ne poserait pas de problème si le maximum absolu était le seul point où le gradient s'annule. En d'autres termes, si nous disposions d'une fonction convexe, c'est-à-dire que pour deux points arbitraires (x_0, y_0) et (x_1, y_1) dans le simplexe,

$$f(t(x_0, y_0) + (1 - t)(x_1, y_1)) \geq tf(x_0, y_0) + (1 - t)f(x_1, y_1), \forall t \in (0, 1).$$

En résumé, le choix du point initial ne poserait pas de problème si nos fonctions étaient sympatiques. Mais, ce n'est pas le cas.

Avant d'entamer la recherche du maximum des fonctions de vraisemblance, nous avons demandé à l'ordinateur de tracer près de deux cents fonctions, ce qui nous a permis de bien localiser les maxima. En effet, nous avons fait varier p_0 et p_1 de 0 à 1 par incréments de 0,01. Les fonctions présentaient une régularité notable.

Dans les cas les plus complexes, nous avons observé la présence de deux maxima locaux distincts. La figure 4.1 présente la surface et les courbes de niveau de deux fonctions de vraisemblance, illustrant de tels scénarios.

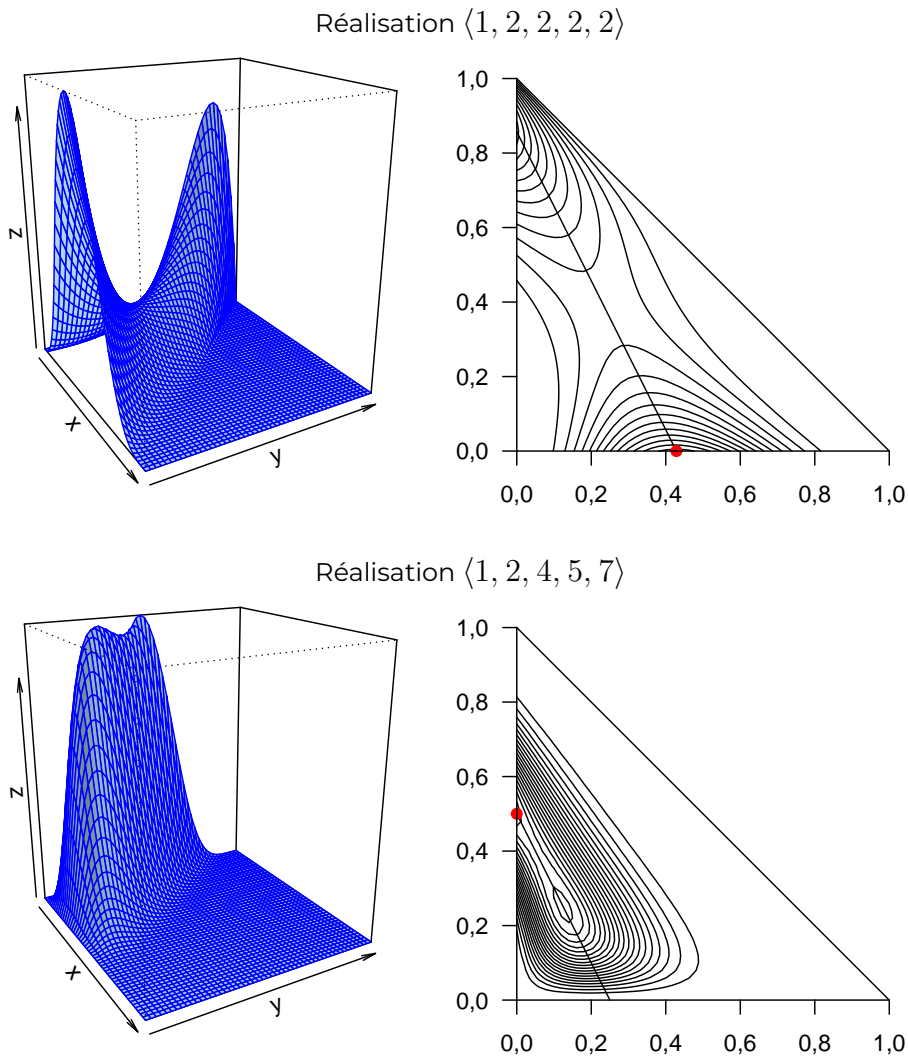


FIGURE 4.1 – Surface et courbes de niveau de deux fonctions de vraisemblance

En plus de tracer le graphique de la fonction de vraisemblance, le programme fournissait également les coordonnées du point où il avait identifié la valeur maximale de la fonction. Pour la réalisation $\langle 1, 2, 2, 2, 2 \rangle$, parmi les 5 151 évaluations effectuées, le point retenu était $(p_0, p_1) = (0.43, 0.00)$. Pour la réalisation $\langle 1, 2, 4, 5, 7 \rangle$, le point retenu était $(p_0, p_1) = (0.0, 0.5)$.¹

1. En effectuant une recherche plus précise du maximum pour la réalisation $\langle 1, 2, 2, 2, 2 \rangle$, le maximum est situé en $(0.4286, 0.0)$. Pour la réalisation $\langle 1, 2, 4, 5, 7 \rangle$, $(0.0, 0.5)$ reste le point où le maximum est atteint.

À cette époque, nous disposions d'environ deux cents scénarios où nous connaissions, avec une précision au centième près, l'emplacement du maximum. Les graphiques nous permettaient également de déterminer aisément le voisinage du point maximal.

Nous avons décidé d'utiliser le huitième estimateur défini au chapitre 3 comme point initial. En utilisant la procédure de maximisation décrite en annexe A, nous avons cherché le maximum pour toutes les fonctions pour lesquelles nous avons les graphiques. Dans tous les cas, le maximum trouvé correspondait à nos attentes. Bien entendu, la pertinence de ce choix peut être remise en question dans certains cas. Le chapitre 5 abordera cette question et montrera que ce choix est probablement judicieux. Cependant, nous ne pouvons garantir que l'algorithme a convergé vers le bon point dans toutes les situations.

La procédure MAXLIK

La procédure MAXLIK, utilisée conjointement avec d'autres telles que NEWTON, CONTRA, LIKE, DLIKE et GRAD, nous permet de localiser le maximum de la fonction de vraisemblance. Un appel à cette procédure s'effectue selon le format suivant :

```
CALL MAXLIK(LIKE, DLIKE, NSIG, MAXFN, X, F, IER, IT)
```

Les deux premiers paramètres, LIKE et DLIKE, sont des procédures dont les détails sont fournis ultérieurement. Elles doivent apparaître dans une déclaration EXTERNAL. NSIG est une variable entière, donnée en entrée, indiquant le critère d'arrêt normal des itérations. L'exécution de la procédure se termine lorsque les deux dernières itérations satisfont aux conditions suivantes :

$$\begin{aligned} |x_i - x_{i-1}| &< 0.5 \times 10^{-\text{NSIG}} \\ |y_i - y_{i-1}| &< 0.5 \times 10^{-\text{NSIG}} \end{aligned}$$

Pour prévenir un nombre excessif d'itérations, nous fournissons également en entrée la variable MAXFN, qui détermine le nombre maximal d'itérations autorisées.

Afin de gérer d'éventuels problèmes survenant pendant l'exécution, nous définissons un paramètre d'erreur, IER. À la sortie de la procédure, IER peut prendre l'une des trois valeurs suivantes :

- 0 : si tout s'est déroulé correctement ;
- 1 ; si le nombre d'itérations requis est supérieur à MAXFN ;
- 2 : si une entrée incorrecte a été fournie à la procédure NEWTON.

Dès que IER prend une valeur autre que 0, un message d'erreur est affiché sur la console, et l'exécution est interrompue. Tous les appels à MAXLIK ont été effectués

avec NSIG = 10 et MAXFN = 500. Au cours des calculs, IER n'a été égal à 1 que dans deux cas sur l'ensemble des évaluations.

Après avoir examiné de près les fonctions de vraisemblance correspondantes, nous estimons que les résultats obtenus sont satisfaisants. Le problème est survenu lors de l'analyse de Z_0 à Z_5 . Malgré cela, les performances de l'algorithme restent bonnes, puisque l'erreur ne s'est produite que dans 0.09% des cas analysés. Ce problème est véritablement négligeable et ne peut en aucun cas altérer de manière significative les résultats finaux.

Le paramètre X est un vecteur de dimension 2. À l'entrée, il contient les coordonnées du point initial, et à la sortie, celles du point où le maximum est atteint. F contient la valeur de la fonction en ce point. XPRIME représente la valeur du gradient au point maximal, et IT nous indique le nombre d'itérations effectuées.

La procédure NEWTON

Cette procédure utilise la méthode *regula falsi* pour localiser le point où la dérivée directionnelle (dans le sens du gradient) s'annule. À l'origine, notre intention était d'adopter la méthode de Newton pour cette tâche, en raison de sa potentialité à converger plus rapidement. Cependant, l'utilisation de la méthode de Newton requiert la connaissance des dérivées secondes de la fonction de vraisemblance. L'expression algébrique des dérivées partielles d'ordre deux est extrêmement complexe. Par conséquent, nous avons abandonné cette approche au profit de la *regula falsi*. Le nom attribué à la procédure demeure NEWTON, préservant ainsi une trace de l'approche initialement envisagée.

Un appel à cette procédure se fait de la manière suivante :

```
CALL NEWTON(XL, XR, EPS, SIP, IER)
```

La procédure NEWTON reçoit en entrée cinq paramètres. Les deux premiers, XL et XR, sont déterminés par la routine MAXLIK. Ce sont deux points tels que sur le segment qu'ils définissent, la dérivée directionnelle change de signe. Le troisième, EPS, représente simplement l'erreur tolérée sur l'évaluation de la dérivée. Autrement dit, lorsque nous avons une nouvelle approximation XAPP du point où la dérivée directionnelle s'annule, nous devons vérifier si la valeur FXAPP de la dérivée en ce point est, en valeur absolue, inférieure à EPS. Le paramètre SIP est tout simplement le vecteur de direction. Évidemment, XL, XR et SIP sont des vecteurs à deux composantes et EPS est un scalaire. Le dernier paramètre est IER, dont le rôle a été défini précédemment.

Il est à noter que tous les appels à la procédure NEWTON ont été effectués avec

$EPS = 0.5 \times 10^{-10}$. Vous trouverez la procédure à l'annexe A.

Les autres procédures

Les procédures MAXLIK et NEWTON font appel directement ou indirectement, aux procédures EVAL, LIKE, GRAD, DLIKE et CONTRA que nous allons brièvement décrire.

La procédure EVAL On appelle cette procédure à l'aide d'un énoncé de la forme :

CALL EVAL(M, N, F, P, MAX, NM)

où

M : Nombre de pères (en entrée);

N : Nombre de fils (en entrée);

F : Probabilité que M pères donnent naissance à N fils (en sortie);

p : Vecteur des probabilités (p_0, p_1, \dots) (en entrée);

MAX : Nombre maximal de descendants (en entrée);

MN : Dimension du vecteur des probabilités (en entrée).

Le rôle de cette procédure est de calculer la probabilité $\ell_{m,n}$. On y utilise la récurrence montrée au chapitre 2. Pour le problème qui nous intéresse, on fixe MAX = 2 et MN = 3. Aucune autre procédure n'est nécessaire à l'exécution de celle-ci.

La procédure LIKE On appelle cette procédure à l'aide d'un énoncé de la forme :

CALL LIKE(X, F)

où :

X : un vecteur de dimension 2 contenant les coordonnées du point où nous aimerions connaître la vraisemblance (en entrée);

F : la valeur de la fonction de vraisemblance (en sortie) évaluée au point X.

Le rôle de cette procédure est évidemment de calculer la vraisemblance des observations. Elle nécessite inévitablement l'utilisation la procédure EVAL.

La procédure GRAD Cette procédure nous permet de calculer le gradient de la fonction de vraisemblance L^* (voir chapitre 2). Elle est invoquée de la manière suivante :

CALL GRAD(X, XPRIME)

où

X : un vecteur de dimension 2 contenant les coordonnées du point où nous voulons connaître le gradient de la fonction L^* (en entrée);

XPRIME : à la sortie, ce vecteur de dimension 2 contient la valeur du gradient de L^* au point X.

Cette procédure est élaborée en se basant sur la relation entre les dérivées des $\ell_{m,n}$ et ces mêmes $\ell_{m,n}$ (voir chapitre 2).

La procédure DLIKE Un appel à cette procédure se fait avec un énoncé du type :
CALL DLIKE(X, FX, SUP)

où

X : les coordonnées du point qui nous intéresse, donc un vecteur de dimension 2 (en entrée);

FX : la valeur de la dérivée directionnelle de L^* dans la direction de SUP (en sortie). Il s'agit donc du produit scalaire du gradient de L^* avec SUP.

SUP : un vecteur de direction, de dimension 2 et de norme 1 (en entrée),

Pour connaître le gradient de L^* , on doit évidemment faire appel à la procédure GRAD.

La procédure CONTRA L'appel à cette procédure s'effectue selon la syntaxe suivante :

CALL CONTRA(X, XPRIME, AIG)

On a recours à cette procédure lorsque, pour un point X donné, nous souhaitons déterminer la contrainte à laquelle on pourrait éventuellement se heurter en se déplaçant dans le sens du vecteur XPRIME. Une fois cette information obtenue, nous sélectionnons le pas d'avancement ainsi que l'orientation à adopter lorsque nous nous trouvons sur la frontière du simplexe. Pour des détails approfondis, le lecteur est invité à consulter l'annexe A.

§4.3 ANALYSE DES RÉSULTATS

Lorsque nous évaluons les performances d'un estimateur, nous nous concentrons principalement sur les caractéristiques relatives au biais, à la variance et à la linéarité. Idéalement, nous souhaitons que l'estimateur proposé soit sans biais, linéaire si possible, et avec une variance minimale. Cependant, comme nous le verrons par la suite, aucun des estimateurs suggérés ne présente de telles qualités.

Pour analyser la performance des estimateurs, nous avons choisi 20 distributions *a priori* (p_0, p_1, p_2) . Parmi celles-ci, une a une moyenne égale à 1, tandis que

les 19 autres ont une moyenne supérieure à 1 et sont des cas surcritiques. Le choix de ces distributions est complètement arbitraire. Nous avons simplement cherché à sélectionner des représentants de diverses zones du domaine des distributions admissibles dans les cas surcritiques.

Nous indiquons sur le graphique de la figure 4.2 les distributions retenues. Pour

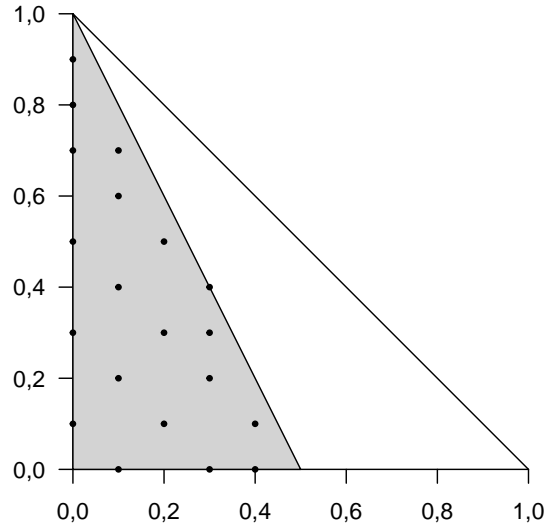


FIGURE 4.2 – Distributions choisies pour l'étude des estimateurs

chacune de ces distributions, et pour chacun des 11 estimateurs définis au chapitre 3, nous avons calculé la moyenne et la matrice des variances et des covariances. Nous avons repris ce travail une seconde fois, mais conditionnellement au fait que $z_{n-1} > 0$, et une troisième fois, conditionnellement au fait que $z_n > 0$. Nous avons aussi profité de l'occasion pour étudier la distribution de \hat{m} .

Dans tous les cas, nous constatons un biais chez tous les estimateurs. Nous savons que \hat{m} est un estimateur consistant lorsque $z_n > 0$ puisqu'évidemment il s'agit d'un comportement asymptotique, c'est-à-dire lorsque :

$$\lim_{n \rightarrow \infty} P[|\hat{m} - m| > \varepsilon \mid z_n > 0] = 0, \forall \varepsilon > 0$$

Nous souhaiterions trouver un estimateur qui se comporte comme le fait \hat{m} dans la distribution conditionnelle à $z_n > 0$. Bien naïvement, nous nous posons la question : « La distorsion semble-t-elle s'atténuer lorsque le nombre de générations observées augmente ? » Dans cette distribution conditionnelle, \hat{m} devient en quelque sorte notre point de repère.

Il est évident que le terme « s'atténuer » est vague et soulève en lui-même la question du choix d'une métrique. Puisque $\hat{p}_0 = 1 - \hat{p}_1 - \hat{p}_2$, nous avons d'abord écarté la métrique euclidienne classique de \mathbb{R}^3 . Nous préférons trouver une métrique où seules les variables \hat{p}_1 et \hat{p}_2 interviennent.

L'idée d'utiliser la célèbre métrique de Mahalanobis nous a effleuré l'esprit. Cependant, puisqu'elle est de la forme $\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}$, où \mathbf{x} est un vecteur et $\boldsymbol{\Sigma}$ est la matrice des covariances, elle implique l'inversion d'une matrice. Or, il s'avère que même en ignorant l'une des composantes (qui est manifestement une fonction linéaire des deux autres), les matrices $\boldsymbol{\Sigma}$ sont parfois singulières. Nous avons tenté les calculs dans les cas où elle s'appliquait, en posant x comme étant le vecteur des biais sur les deux autres composantes, mais elle ne reflète pas du tout ce que nous avons pu observer en termes de tendance du biais lorsque le nombre de générations augmente. Même avec cette métrique, \hat{m} semble parfois diverger. Par conséquent, nous avons mis de côté cette métrique. Finalement, nous avons convenu de prendre comme métrique la fonction suivante :

$$\sqrt{\frac{(E[\hat{p}_1] - p_1)^2 + (E[\hat{p}_2] - p_2)^2}{2}}$$

Il s'agit d'un choix discutable, et nous ne pouvons le justifier rigoureusement. Nous divisons par 2 la somme des carrés des biais sur les deux dernières composantes afin que cette métrique rende les distances entre éléments de \mathbb{R}^2 comparables à des distances entre nombres réels. Cette inspiration vient du test classique sur deux variances où nous faisons le rapport de deux χ^2 préalablement divisées par leurs degrés de liberté respectifs. Ici, par analogie, (\hat{p}_1, \hat{p}_2) aurait deux degrés de liberté tandis que \hat{m} en aurait un.

Nous avons effectivement trois classes de distributions *a priori*.

1. La première classe est constituée de distributions où $p_0 = 0$, $p_1 \neq 0$, et $p_2 \neq 0$.
2. La deuxième classe comprend des distributions où $p_0 \neq 0$, $p_1 = 0$, et $p_2 \neq 0$.
3. La troisième classe inclut des distributions où $p_0 \neq 0$, $p_1 \neq 0$, et $p_2 \neq 0$.

Puisque nous préférons nous limiter aux cas surcritiques, aucune étude n'a été effectuée en fixant $p_2 = 0$, car dans ce cas, nous aurions eu $m < 1$. Le lecteur trouvera à l'annexe C les graphiques nécessaires à la compréhension du texte qui suit. Nous fournirons peu de détails sur le comportement de \hat{m} , réservant une section complète à l'étude de son biais, car jusqu'à présent, aucune étude ne nous avait donné une évaluation de celui-ci. Nous proposerons même une forme fonctionnelle exprimant le biais de \hat{m} en fonction du nombre de générations observées.

Les distributions du type $p_0 = 0, p_1 \neq 0$ et $p_2 \neq 0$

Au fil de la discussion, lorsque nous évoquons le biais d'un estimateur différent de \hat{m}_i , nous nous référons à la métrique précédemment définie.

Sous l'hypothèse $\mathbf{p} = (0.0, 0.1, 0.9)$, il est notable que le biais de \hat{m} augmente lors du passage d'une à deux générations observées, pour ensuite décroître de manière continue. Cependant, les estimateurs 5, 6 et 11 manifestent une propension à accroître leur distorsion avec l'augmentation du nombre d'observations. L'estimateur 10 maintient une performance acceptable jusqu'à la quatrième génération, mais le biais connaît une augmentation remarquable lors du passage de la quatrième à la cinquième génération.

En utilisant la courbe du biais de \hat{m} comme référence, nous préférons les estimateurs 8 et 9, dont les courbes respectives se situent en deçà de celle de \hat{m} . Les estimateurs 1 et 2 présentent un comportement acceptable, avec des courbes de biais dans une plage relativement restreinte par rapport à celle de \hat{m} . En revanche, les estimateurs 3 et 4, bien qu'ils ne montrent pas de tendance à accroître leur biais, sont moins attractifs car leurs courbes respectives se situent au-dessus de celle de \hat{m} .

Sous la distribution $\mathbf{p} = (0.0, 0.3, 0.7)$, tous les estimateurs, à l'exception des estimateurs 8 et 9, présentent un biais supérieur à celui de \hat{m} jusqu'à la cinquième génération. Il est toutefois à noter que seuls les estimateurs 1, 5, 6, 7 et 10 manifestent une tendance claire à accroître leur biais.

Les estimateurs 3, 4 et 11 affichent une conduite convenable, étant comparable à celle de \hat{m} . Le comportement du deuxième estimateur demeure ambigu. Il ne s'aligne ni sur le premier, ni sur le onzième. Finalement, les estimateurs 8 et 9 se distinguent par un biais nettement inférieur à celui de \hat{m} .

Par rapport à la distribution $\mathbf{p} = (0.0, 0.1, 0.9)$, les observations précédentes restent valides en ce qui concerne \hat{m} .

Sous la distribution $\mathbf{p} = (0.0, 0.5, 0.5)$, les constatations demeurent essentiellement les mêmes qu'avec la distribution $\mathbf{p} = (0.0, 0.3, 0.7)$. Néanmoins, il convient d'ajouter que tous les estimateurs, à l'exception de 8, 9 et \hat{m} , montrent une tendance à s'éloigner des valeurs à estimer.

Pour les trois distributions $\mathbf{p} = (0.0, 0.7, 0.3)$, $\mathbf{p} = (0.0, 0.8, 0.2)$ et $\mathbf{p} = (0.0, 0.9, 0.1)$, nous procéderons à une discussion commune. Trois classes distinctes peuvent être identifiées. L'estimateur \hat{m} , notre point de référence, conserve un comportement souhaitable dans tous les cas. Sous ces trois distributions *a priori*, les estimateurs 5, 6, 8 et 9 se distinguent positivement, leurs courbes de biais se

situant en dessous de celle de \hat{m} . En revanche, il est à noter que les estimateurs de la deuxième classe (1, 2 et 10) démontrent des performances médiocres, tandis que ceux de la troisième classe présentent les pires résultats, avec une tendance encore plus marquée à accroître leur biais que ceux de la deuxième classe.

En conclusion, sous des distributions du type $p_0 = 0, p_1 \neq 0$ et $p_2 \neq 0$, les estimateurs 8 et 9 affichent un comportement très satisfaisant. Leur distorsion est comparable à celle de \hat{m} , et les variations du biais n'incitent pas à penser qu'il augmenterait avec le nombre de générations observées. Au contraire, il semble plausible que ces estimateurs acquièrent une certaine consistance.

Les distributions du type $p_0 \neq 0, p_1 = 0, p_2 \neq 0$

Traisons simultanément les trois cas étudiés. L'estimateur \hat{m} démontre un comportement très satisfaisant, avec un biais décroissant de manière quasi-exponentielle. Cette tendance est également observée pour les estimateurs 8 et 9. En fait, à l'exception de la distribution $\mathbf{p} = (0.1, 0.0, 0.9)$, où les estimateurs 5 et 6 ne suivent pas le comportement des estimateurs autres que 8, 9 et \hat{m} , on peut affirmer que les estimateurs suggérés (à l'exclusion évidente de 8, 9 et \hat{m}) ont tendance à accroître leur distorsion avec l'augmentation du nombre de générations observées.

Sous ce type de distribution *a priori*, nous pouvons conclure que les estimations impliquant un maximum de vraisemblance sont préférables.

Les distributions du type $p_0 \neq 0, p_1 \neq 0$ et $p_2 \neq 0$

La discussion amorcée ici s'avère relativement peu stimulante, son unique objectif étant de souligner qu'aucune méthode d'estimation n'a démontré une supériorité incontestable dans tous les cas. Le lecteur comprendra à la fin de cette section qu'avant de préconiser une méthode plutôt qu'une autre, une étude approfondie sera nécessaire. La discussion s'appuiera sur l'observation des graphiques associés à ces distributions.

Sous la distribution $\mathbf{p} = (0.2, 0.1, 0.7)$, similairement à toutes les autres du troisième type, le biais sur \hat{m} semble décroître rapidement. Les estimateurs 8 et 9 présentent le biais le plus faible parmi toutes les méthodes suggérées. Il est à espérer que la distorsion diminue davantage avec l'observation de plus de cinq générations.

Pour tous les autres estimateurs, à l'exception du cinquième et du sixième dont les comportements sont encourageants, on observe une tendance à la hausse du

biais. Si ce n'était pas pour le fait que les estimateurs 5 et 6 présentent un biais plus prononcé que les estimateurs 8 et 9, on pourrait envisager de les préférer.

Sur le graphique correspondant à la distribution $\mathbf{p} = (0.1, 0.2, 0.7)$, on constate que seuls les estimateurs 5, 6 et 7 ont un biais qui tend à s'amenuiser à partir de la troisième génération, tout au moins. Les estimateurs 8 et 9, bien que relativement peu biaisés, montrent une tendance à augmenter leur distorsion. Les estimateurs 1, 2, 3, 4 et 11 affichent une croissance du biais avec le nombre de générations observées. Soulignons toutefois que les estimateurs 2, 3 et 4 réduisent quelque peu leur biais lors du passage de la quatrième à la cinquième génération.

Nous aborderons simultanément les distributions $\mathbf{p} = (0.1, 0.7, 0.2)$ et $\mathbf{p} = (0.1, 0.6, 0.3)$, car les graphiques obtenus sont structurés de manière similaire pour toutes fins pratiques. En considérant uniquement les trois premières générations, tous les estimateurs ont une tendance à devenir moins biaisés. Cependant, à partir de la troisième génération, les estimateurs 3, 4, 7 et 11 manifestent une nette tendance à accroître leur distorsion. Ce phénomène se reproduit également lors du passage de quatre à cinq générations pour les estimateurs 1, 2, 8, 9 et 10, mais de manière moins prononcée. Les estimateurs 5 et 6 ont tendance, une fois de plus, à devenir moins biaisés. À noter que parmi ces deux méthodes, la cinquième nous offre une meilleure estimation.

Sous la distribution critique $\mathbf{p} = (0.3, 0.4, 0.3)$, tous les estimateurs présentent un comportement satisfaisant. Ils affichent un biais nettement moindre que celui observé pour \hat{m} , et aucun d'entre eux ne montre de tendance à accroître son biais. De plus, les estimateurs 1, 2, 3, 4, 10 et 11 semblent atténuer leur propre biais.

Pour la distribution $\mathbf{p} = (0.2, 0.5, 0.3)$, la situation est comparable à celle décrite précédemment. Cependant, certains estimateurs montrent une légère augmentation de leur biais, bien que cette augmentation ne soit pas préoccupante.

Sous la distribution $\mathbf{p} = (0.4, 0.1, 0.5)$, la situation est plus simple. Tous les estimateurs, à l'exception des estimateurs 8 et 9, ont tendance à augmenter leur biais. Les estimateurs 7, 8, 9 et 10 conservent toutefois une distorsion inférieure à celle de \hat{m} .

La situation rencontrée sous $\mathbf{p} = (0.3, 0.3, 0.4)$ est assez similaire à celle de $\mathbf{p} = (0.2, 0.5, 0.3)$. Tous les estimateurs, à l'exception du sixième, maintiennent un biais inférieur à celui de \hat{m} . Il est à noter que l'estimateur 7 présente d'excellents résultats ici. Ceci n'est pas surprenant car cette méthode d'estimation avait pour critère de retenir l'élément de \mathcal{A}/V qui maximisait son entropie. Or, cette distribution n'est pas très éloignée de $\mathbf{p} = (1/3, 1/3, 1/3)$, qui possède la plus grande entropie.

Sous la distribution $\mathbf{p} = (0.3, 0.2, 0.5)$, seuls les estimateurs 7, 8 et 9 conservent

un biais inférieur à celui de \hat{m} . Toutefois, tous les estimateurs autres que le septième présentent au moins une légère tendance à accroître leur biais.

Sous la distribution $\mathbf{p} = (0.2, 0.3, 0.5)$, les observations relatives au graphique sont essentiellement les mêmes que pour la distribution précédente. L'estimateur 10 a cependant tendance à diminuer son biais. Soulignons que les estimations 8, 9 et 10 conservent, jusqu'à la quatrième génération, une distorsion inférieure à celle de \hat{m} . Pour sa part, l'estimateur 7 continue à présenter le meilleur comportement, son biais demeurant constamment inférieur à celui de \hat{m} et diminuant avec le nombre de générations observées.

La situation est assez confuse sous $\mathbf{p} = (0.1, 0.4, 0.5)$. Les estimateurs 1, 2, 5, 6 et 10 laissent penser qu'ils deviennent de moins en moins biaisés. Pour les autres estimateurs, leur biais oscille dans ses variations. L'estimateur 10 se démarque ici comme le meilleur.

Suite à ces discussions sur les différentes situations étudiées, le lecteur conviendra probablement que nous n'avons pas suggéré de méthode supérieure aux autres dans tous les cas. Malgré cela, plusieurs d'entre elles adoptent un comportement quelque peu inquiétant dans plusieurs situations. Néanmoins, nous avons tendance à croire que l'estimation à vraisemblance maximale pourrait donner de bons résultats. Le huitième estimateur présente un certain intérêt, étant presque toujours aussi proche des paramètres que le neuvième, mais avec l'avantage marqué de nécessiter très peu d'évaluations de la fonction. Il serait sans doute judicieux de poursuivre l'étude et peut-être de recourir à la simulation pour examiner plus de cinq générations.

Au chapitre 4 nous donnerons un résultat qui nous permettra de réduire considérablement les calculs. Dès lors, nous pourrons espérer poursuivre l'étude facilement jusqu'à dix générations observées et peut-être même vingt. À l'annexe C, le lecteur peut trouver les graphiques des courbes du biais pour les estimateurs 1, 8 et 9, et pour \hat{m} . On voit que dans aucune situation le comportement des estimateurs 8 et 9 est véritablement alarmant.

§4.4 ÉTUDE DU BIAIS DE L'ESTIMATEUR DE LA MOYENNE

Comme évoqué précédemment, nous allons à présent définir une forme fonctionnelle qui nous permettra d'obtenir une estimation assez précise du biais de \hat{m} . Il est important de noter que le modèle proposé n'est valable que si $p_0 \neq 0$. En examinant les graphiques où $p_0 \neq 0$, nous avons observé une relation potentielle entre le biais (B) de \hat{m} et le nombre de générations observées (N), s'exprimant comme suit : $B = aN^{-b}$, où a et b sont des constantes positives à déterminer.

Afin de linéariser le modèle, il suffit de prendre le logarithme de B et N , et de poser :

$$Y = \alpha + \beta X$$

où $\alpha = \ln(a)$, $\beta = -b$, $Y = \ln(B)$ et $X = \ln(N)$.

Pour déterminer les constantes α et β , nous faisons usage de la méthode des moindres carrés. Le coefficient de corrélation linéaire de Pearson nous permet d'évaluer la qualité de l'ajustement du modèle. Ce coefficient est défini par :

$$\frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sqrt{\left[n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i \right)^2 \right] \left[n \sum_{i=1}^n Y_i^2 - \left(\sum_{i=1}^n Y_i \right)^2 \right]}}$$

Après avoir fait les calculs nous avons obtenu le tableau 4.1.

TABLEAU 4.1 – Estimation des coefficients α et β , valeur de r_{XY} et de m

p_0	p_1	α	β	r_{XY}	m
0.1	0.0	-1.382249	-2.514572	-0.875886	1.8
0.1	0.2	-1.617048	-2.116669	-0.876615	1.6
0.2	0.1	-0.603076	-2.117943	-0.940076	1.5
0.1	0.4	-1.545948	-2.156795	-0.965292	1.4
0.3	0.0	-0.333064	-1.573636	-0.979139	1.4
0.2	0.3	-0.988588	-1.226541	-0.979796	1.3
0.1	0.6	-1.927900	-1.057978	-0.987312	1.2
0.3	0.2	-0.590634	-1.003660	-0.991580	1.2
0.4	0.0	-0.158513	-1.139066	-0.995172	1.2
0.1	0.7	-2.086647	-0.641672	-0.998492	1.1
0.2	0.5	-1.257040	-0.614720	-0.995392	1.1
0.3	0.3	-0.709780	-0.765963	-0.995426	1.1
0.4	0.1	-0.271440	-0.915974	-0.997420	1.1
0.3	0.4	-0.827745	-0.550199	-0.998110	1.0

Une première observation importante à souligner est que tous les coefficients de corrélation sont, en valeur absolue, supérieurs à 0.875, et certains d'entre eux sont même très proches de 1. L'ajustement s'avère plus précis lorsque la constante m se rapproche de 1, bien que l'ajustement reste très satisfaisant lorsque $m = 1,8$.

Nous sommes satisfaits d'avoir découvert une expression relativement simple, bien qu'approximative, permettant « d'expliquer » le biais en fonction du nombre de générations observées.

Ainsi, pour les distributions où $p_0 \neq 0$, nous pouvons estimer le biais de manière approximative à l'aide de l'expression suivante :

$$B = \frac{e^\alpha}{N^{-\beta}}, \quad i = 1, \dots, 5$$

Il est essentiel de déterminer les paramètres dont α et β dépendent. Le lecteur peut envisager que la moyenne, la variance, voire l'écart-type de la distribution initiale, peuvent jouer un rôle. Dans notre tentative d'explication de α et β , nous avons exploré des relations avec la moyenne et la variance, ainsi qu'avec la moyenne et l'écart-type. Nous avons finalement retenu cette dernière approche. Les calculs des matrices de corrélation ont conduit aux résultats présentés dans le tableau 4.2 :

TABLEAU 4.2 – Corrélation entre les différents paramètres

	α	β	m	σ
α	1	0,113341	-0,166585	0,979901
β	0,113341	1	-0,956967	0,197323
m	-0,166585	-0,956967	1	-0,266336
σ	0,979901	0,197323	-0,266336	1

L'examen du tableau suggère une forte corrélation entre le terme α et l'écart-type de la distribution *a priori*, tandis que la liaison de β est plus marquée avec la moyenne m .

Pour évaluer la relation potentielle entre α et β , nous allons neutraliser l'effet linéaire des paramètres m et σ en utilisant le coefficient de corrélation partielle, défini comme :

$$r_{ij.k} = \frac{r_{ij} - r_{ik}r_{jk}}{\sqrt{(1 - r_{ik}^2)(1 - r_{jk}^2)}}$$

où k représente la variable dont nous souhaitons neutraliser l'effet, et r_{ij} désigne la corrélation entre i et j . En éliminant l'effet linéaire de σ , nous obtenons le tableau 4.3.

La corrélation linéaire entre β et m a augmenté, de même que celle entre α et β .

TABLEAU 4.3 – Corrélation partielle entre les différents paramètres après avoir éliminé l'effet de σ

	α	β	m
α	1	-0.409159	0.490942
β	-0.409159	1	-0.957122
m	0.490942	-0.957122	1

Par ailleurs, en éliminant également l'effet linéaire de m , la corrélation partielle entre α et β s'élève à 0,2406. Cela signifie que, avec m et σ constants, la corrélation entre α et β n'est pas très forte.

En éliminant uniquement l'effet linéaire de m , les corrélations obtenues sont présentées dans le tableau 4.4 :

TABLEAU 4.4 – Corrélation partielle entre les différents paramètres après avoir éliminé l'effet de m

	α	β	σ
α	1	-0.161023	0.984345
β	-0.161023	1	-0.205751
σ	0.984345	-0.205751	1

Enfin, lorsque l'on considère les tableaux des corrélations partielles on voit une fois de plus que les termes α et β semblent être pleinement expliqués respectivement par l'écart-type et la moyenne de la distribution.

Cependant, nous ne jugeons pas utile de donner les résultats de la régression de (α, β) sur (m, σ) , car les coefficients obtenus n'ont pas vraiment pu éclairer les relations que nous avons constatées.

Deux résultats importants

§5.1 INTRODUCTION

Nous établirons la relation entre les estimateurs 8 et 9. Le lien entre les deux méthodes d'estimation offre un moyen clair de réduire significativement les calculs nécessaires pour déterminer l'estimateur à vraisemblance maximale de la distribution initiale.

Nous présentons ici la démarche qui nous a conduit à ce résultat. Par la suite, nous exposerons une conjecture que nous avons cru vraie pendant un certain temps. Cependant, nous fournirons un contre-exemple pour démontrer qu'elle est fausse. En conclusion de ce chapitre, nous présenterons une propriété dont jouit l'estimation à entropie maximale.

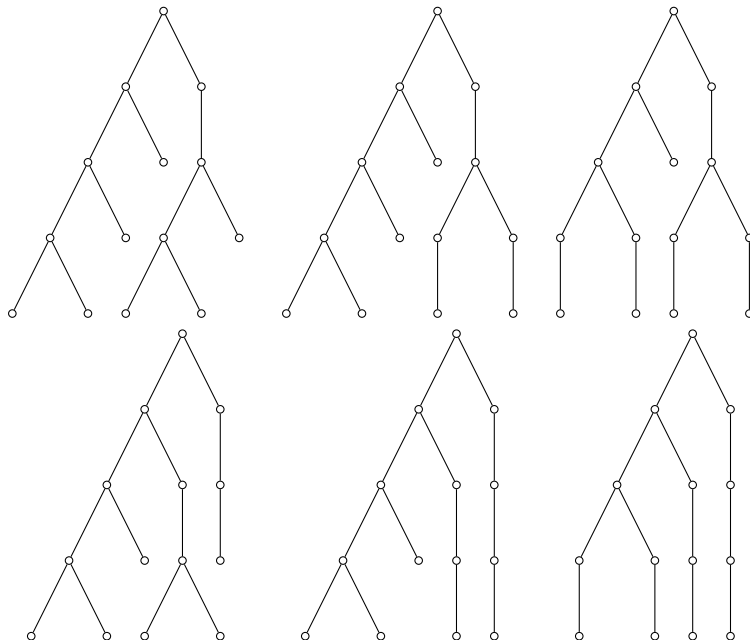
§5.2 LE LIEN ENTRE LE HUITIÈME ET LE NEUVIÈME ESTIMATEUR

Avant de démarrer la démonstration du résultat principal de cette section, nous aborderons le problème à travers deux exemples.

TABLEAU 5.1 – Énumération des partages pour la réalisation $\langle 1, 2, 3, 4, 4 \rangle$

Réalisation			Partages admissibles		
i	z_i	z_{i+1}	$\lambda_{0,k}^{i+1}$	$\lambda_{1,k}^{i+1}$	$\lambda_{2,k}^{i+1}$
0	1	2	0	0	1
1	2	3	0	1	1
2	3	4	1 0	0 2	2 1
3	4	4	2 1 0	0 2 4	2 1 0

Exemple 5.2.1 Considérons la réalisation $\langle 1, 2, 3, 4, 4 \rangle$. Le tableau 5.1 présente les partages admissibles pour le passage de z_i à z_{i+1} , $i = 0, \dots, 3$. En utilisant le tableau 5.1, on observe qu'il existe essentiellement 6 arbres possibles. En réalité, tout arbre qui aurait pu donner lieu à cette observation du processus α , à toutes fins pratiques, la structure de l'un des six arbres suivants :



Une estimation effectuée au sein de chacun des arbres nous conduit aux distributions suivantes :

$$\begin{array}{ccc} \langle 0.3, 0.1, 0.6 \rangle & \langle 0.2, 0.3, 0.5 \rangle & \langle 0.1, 0.5, 0.4 \rangle \\ \langle 0.2, 0.3, 0.5 \rangle & \langle 0.1, 0.5, 0.4 \rangle & \langle 0.0, 0.7, 0.3 \rangle \end{array}$$

Or tous les points obtenus sont sur un même segment de droite.

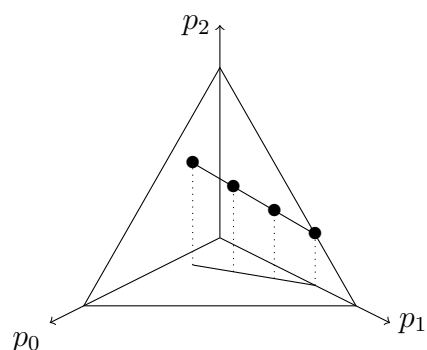


FIGURE 5.1 – Différentes estimations pour la réalisation $\langle 1, 2, 3, 4, 4 \rangle$

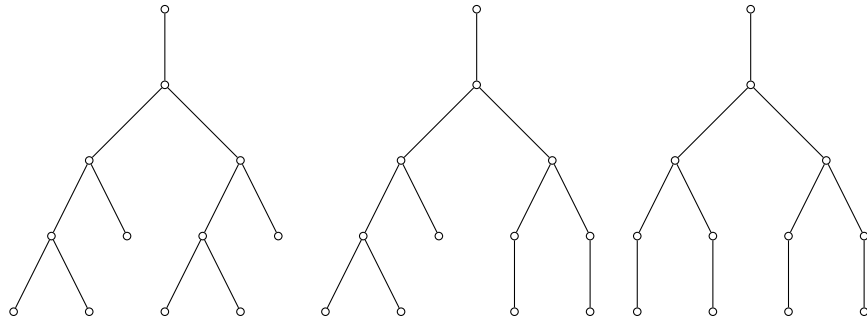
De plus, l'estimation à vraisemblance maximale pour la réalisation observée du processus de Galton-Watson est $\langle 0.0, 0.7, 0.3 \rangle$. Il est évident que ce point se situe sur le segment mentionné précédemment. \square

Exemple 5.2.2 Considérons la réalisation $\langle 1, 1, 2, 4, 4 \rangle$. Le tableau 5.2 énumère les partages admissibles.

TABLEAU 5.2 – Énumération des partages pour la réalisation $\langle 1, 1, 2, 4, 4 \rangle$

Réalisation			Partages admissibles		
i	z_i	z_{i+1}	$\lambda_{0,k}^{i+1}$	$\lambda_{1,k}^{i+1}$	$\lambda_{2,k}^{i+1}$
0	1	1	0	1	0
1	1	2	0	0	1
2	2	4	0	0	2
3	4	4	2	0	2
			1	2	1
			0	4	0

A l'aide du tableau 5.2, on voit qu'il y a 3 arbres à considérer.



Une estimation au sein de chacun des arbres donne respectivement les distributions suivantes :

$$\langle 1/4, 1/8, 5/8 \rangle \quad \langle 1/8, 3/8, 1/2 \rangle \quad \langle 0, 5/8, 3/8 \rangle$$

Et encore une fois, ces trois points sont alignés.

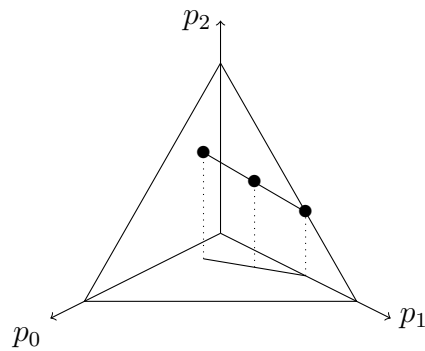


FIGURE 5.2 – Différentes estimations pour la réalisation $\langle 1, 1, 2, 4, 4 \rangle$

On vérifie facilement que l'équation de la droite est :

$$d : (0, 5/8, 3/8) + t(1/8, -2/8, 1/8)$$

En fixant $t = 1.09864$, le point résultant est $(0.13733, 0.35034, 0.51233)$, qui correspond à l'estimateur à vraisemblance maximale de la distribution initiale du processus de Galton-Watson (voir annexe B). \square

Ces résultats ne sont pas des cas particuliers. Nous démontrerons que dans le cas où le nombre maximal de descendants est 2, cet alignement se produit systématiquement.

Dans l'exemple 5.2.1, nous avons affirmé qu'il existait essentiellement 6 arbres. Cependant, certains d'entre eux conduisent à la même estimation. Pour éviter cette

« redondance, » il suffirait de considérer l'ensemble \mathcal{A}/V , tel que défini au chapitre 3. En réalité, certains d'entre eux étaient V -équivalents.

Notons que si la cardinalité de \mathcal{A}/V est égale à 1, alors les ensembles \mathcal{P}_i intervenant dans la formule de la fonction de vraisemblance du processus sont tous de cardinalité 1. En effet, supposons le contraire. Il serait alors possible de construire deux arbres conformes à l'observation du processus, mais ayant des fonctions de vraisemblance distinctes. Ainsi, si la cardinalité de \mathcal{A}/V est 1, la fonction de vraisemblance du processus est, à une constante près, celle d'un arbre représentant \mathcal{A}/V .

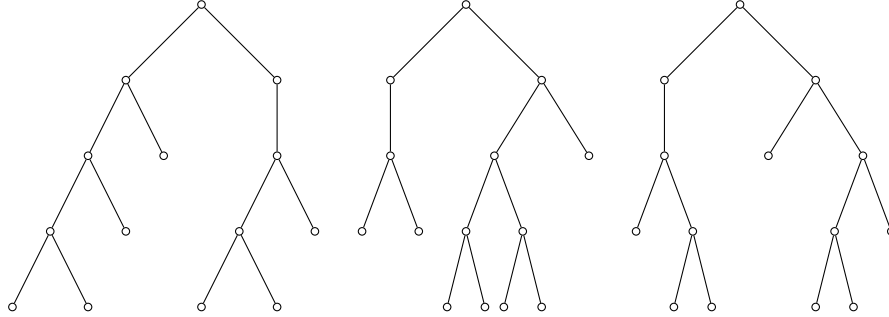
En effet, si tous les ensembles \mathcal{P}_i sont de cardinalité 1, on a λ^i comme seul partage à considérer lors de la transition de Z_{i-1} à Z_i , et la fonction de vraisemblance du processus est :

$$\begin{aligned} L(z_0, \dots, z_n) &= \prod_{i=1}^n C(\lambda^i) p^{\lambda^i} \\ &= \prod_{i=1}^n C(\lambda^i) \left(p_0^{\sum_{j=0}^n \lambda_j^i} p_1^{\sum_{j=1}^n \lambda_j^i} p_2^{\sum_{j=2}^n \lambda_j^i} \right) \end{aligned}$$

Or, $\sum_{j=0}^n \lambda_j^i$ représente dans ce cas le nombre de fois où un individu a donné naissance à j ($j = 0, 1, 2$) descendants dans l'arbre. De plus, l'expression entre parenthèses est la fonction de vraisemblance de l'arbre. Ainsi, le maximum est le même pour les deux. Dans ce cas particulier, les estimateurs 8 et 9 correspondent. Conséquemment, le résultat que nous avons annoncé trouve son véritable intérêt lorsque \mathcal{A}/V contient au moins deux classes. Dans notre premier exemple, nous avons constaté que des arbres ayant une structure différente pouvaient avoir la même vraisemblance. Nous préférons définir une équivalence structurale.

Définition 5.2.1 (Équivalence structurale) Deux arbres sont structurellement équivalents si, et seulement si, pour $i = 1, \dots, n$, la transition de Z_{i-1} à Z_i fait intervenir le même partage de Z_i en au plus Z_i parts. \square

Exemple 5.2.3 Trois arbres structurellement équivalents :



□

Considérons la relation d'équivalence \mathcal{S} de la définition 5.2.1. Il suffit de démontrer le résultat en travaillant sur l'ensemble \mathcal{A}/\mathcal{S} , car la partition induite sur \mathcal{A} par la relation V est plus grossière que celle induite par \mathcal{S} . De plus, le choix d'un représentant est beaucoup plus facile au niveau de \mathcal{A}/\mathcal{S} . En effet, pour $i = 1, \dots, n$, on choisit un élément dans \mathcal{P}_i , disons λ^i . La suite des λ^i détermine complètement la structure de l'arbre. Évidemment, deux arbres \mathcal{S} -équivalents donnent la même estimation à vraisemblance maximale des p_j , car le nombre de fois où dans l'arbre un individu donne naissance à j ($j = 0, 1, 2$) descendants est le même pour tous les arbres de la même classe.

L'équation de la droite

L'équation de la droite fait intervenir les valeurs :

$$k_0 = z_0 - z_n + \sum_{i=1}^n \lfloor z_i/2 \rfloor$$

$$k_1 = \sum_{i=1}^n z_i - 2 \sum_{i=1}^n \lfloor z_i/2 \rfloor$$

$$k_2 = \sum_{i=1}^n \lfloor z_i/2 \rfloor$$

$$k = k_0 + k_1 + k_2 \quad (\text{nombre de pères})$$

Nous nous fixons comme but de démontrer qu'une estimation à vraisemblance maximale associée à un élément de \mathcal{A}/\mathcal{S} est la droite :

$$d : \frac{1}{k}(k_0, k_1, k_2) + \frac{t}{k}(1, -2, 1)$$

L'estimation associée à une structure d'arbre est tout simplement obtenue en posant :

$$r_j = \sum_{i=1}^n \lambda_j^i, \quad \text{et} \quad \hat{p}_j = \frac{r_j}{k} \quad \text{pour } j = 0, 1, 2,$$

car r_j compte le nombre de fois où un individu a donné naissance à j descendants au sein de la structure considérée. Il suffirait donc, dans un premier temps, de démontrer la proposition 5.2.1.

Proposition 5.2.1 *Le triplet (r_0, r_1, r_2) est sur la droite dont l'équation paramétrique est*

$$d : (k_0, k_1, k_2) + t(1, -2, 1)$$

Preuve Posons $t = r_2 - k_2$. Alors,

i) D'une part

$$k_2 + t = r_2$$

ii) D'autre part

$$\begin{aligned} k_1 - 2t &= k_1 - 2r_2 + 2k_2 \\ &= \sum_{i=1}^n z_i - 2 \sum_{i=1}^n [z_i/2] - 2 \sum_{i=1}^n \lambda_2^i + 2 \sum_{i=1}^n [z_i/2] \\ &= \sum_{i=1}^n (\lambda_1^i + 2\lambda_2^i) - 2 \sum_{i=1}^n \lambda_2^i \quad \text{car } z_i = \lambda_1^i + 2\lambda_2^i \\ &= \sum_{i=1}^n \lambda_1^i = r_1 \end{aligned}$$

ii) De plus

$$\begin{aligned} k_0 + t &= k_0 + r_2 - k_2 \\ &= \sum_{i=1}^n [z_i/2] + z_0 - z_n + \sum_{i=1}^n \lambda_2^i - \sum_{i=1}^n [z_i/2] \\ &= z_0 - z_n + \sum_{i=1}^n \lambda_2^i \end{aligned}$$

Puisque $z_i = 0\lambda_0^i + 1\lambda_1^i + 2\lambda_2^i$, on a :

$$\sum_{i=1}^n z_i = \sum_{i=1}^n \lambda_1^i + 2 \sum_{i=1}^n \lambda_2^i$$

et puisque $z_{i-1} = \lambda_0^i + \lambda_1^i + \lambda_2^i$, on obtient :

$$\sum_{i=1}^n z_{i-1} = \sum_{i=1}^n \lambda_0^i + \sum_{i=1}^n \lambda_1^i + \sum_{i=1}^n \lambda_2^i$$

En soustrayant la deuxième équation de la première, on trouve

$$z_0 - z_n = \sum_{i=1}^n \lambda_0^i - \sum_{i=1}^n \lambda_2^i$$

d'où

$$k_0 + t = \sum_{i=1}^n \lambda_0^i = r_0$$

Ce qui termine la démonstration. \square

Une conséquence directe de cette proposition est que le domaine sur lequel on maximisait la fonction de vraisemblance pour aboutir à l'estimateur 8 était formé d'un nombre fini de points situés sur un segment de droite.

Montrons maintenant que l'estimateur à vraisemblance maximale (pour le processus de Galton-Watson) est également sur la droite. Supposons que le point soit à l'intérieur du simplexe. On a vu au chapitre 1 que ce point doit satisfaire la relation suivante :

$$p_j = \left[\sum_{i=0}^{n-1} z_i \right]^{-1} \sum_{i=1}^n \frac{1}{\ell_{z_{i-1}z_i}} \left(\sum_{\lambda^i \in \mathcal{P}_i} \lambda_j^i C(\lambda^i) p^{\lambda^i} \right)$$

pour $j = 0, 1, 2$. Posons $r_j = p_j k$. On a la proposition 5.2.2.

Proposition 5.2.2 *Le triplet (r_0, r_1, r_2) est sur la droite dont l'équation paramétrique est*

$$(k_0, k_1, k_2) + t(1, -2, 1).$$

Preuve Posons $t = r_2 - k_2$.

i) Évidemment, $k_2 + t = r_2$

ii) On a

$$\begin{aligned}
k_1 - 2t &= k_1 - 2p_2k + 2k_2 \\
&= \sum_{i=1}^n z_i - 2 \sum_{i=1}^n \frac{1}{\ell_{z_{i-1}z_i}} \left(\sum_{\lambda^i \in P_i} \lambda_2^i C(\lambda^i) \mathbf{p}^{\lambda^i} \right) \\
&= \sum_{i=1}^n z_i - \sum_{i=1}^n \frac{1}{\ell_{z_{i-1}z_i}} \left(\sum_{\lambda^i \in P_i} 2\lambda_2^i C(\lambda^i) \mathbf{p}^{\lambda^i} \right) \\
&= \sum_{i=1}^n z_i - \sum_{i=1}^n \frac{1}{\ell_{z_{i-1}z_i}} \sum_{\lambda^i \in P_i} (z_i - \lambda_1^i) C(\lambda^i) \mathbf{p}^{\lambda^i} \\
&= \sum_{i=1}^n z_i - \sum_{i=1}^n z_i + \sum_{i=1}^n \frac{1}{\ell_{z_{i-1}z_i}} \sum_{\lambda^i \in P_i} \lambda_1^i C(\lambda^i) \mathbf{p}^{\lambda^i} \\
&= r_1
\end{aligned}$$

iii) De plus,

$$\begin{aligned}
k_0 + t &= k - k_1 - k_2 + t \\
&= k - k_1 + 2t - k_2 - t \\
&= k - r_1 - r_2 \\
&= k(1 - p_1 - p_2) \\
&= kp_0 \\
&= r_0
\end{aligned}$$

Ceci termine la démonstration. □

La dernière démonstration suppose que le point est à l'intérieur de S . On peut aussi donner une preuve lorsque le point est sur la frontière, c'est-à-dire qu'une ou deux composantes sont nulles. La preuve est très simple et nous ne croyons pas utile de la présenter ici.

Il existe une explication très simple du phénomène. Si (a, b, c) est un point sur la droite et sur le simplexe, on peut toujours considérer ce point comme une fonction de masse du nombre de descendants générés par un individu et on peut calculer son espérance mathématique :

$$\begin{aligned}
0a + 1b + 2c &= \frac{1}{k}(k_1 - 2t + 2k_2 + 2t) \\
&= \frac{1}{k}(k_1 + 2k_2)
\end{aligned}$$

$$\begin{aligned}
&= \left(\sum_{i=0}^{n-1} z_i \right)^{-1} \sum_{i=1}^n z_i \\
&= \hat{m}
\end{aligned}$$

Malheureusement ce point important nous avait échappé au début de l'étude. Nous avons donc perdu un certain temps à écrire des programmes relativement complexes pour trouver le maximum, sous certaines contraintes, d'une fonction à deux variables. Il suffisait de maximiser une fonction à une seule variable sur un petit segment de droite. Nous aurions pu, dès le départ, restreindre notre recherche du maximum sur le segment $p_1 = (2 - p_0) + \hat{m}$, ($p_0 \geq 0, p_1 \geq 0$). On imagine facilement que le temps de calcul en aurait été considérablement réduit.

Ce constat ouvre la voie à l'espoir d'examiner davantage de générations successives. Il est évident qu'une telle étude ne pourrait être menée à bien en utilisant des distributions exactes, comme démontré au chapitre 3. Cependant, la simulation pourrait prendre le relais. Ainsi, il serait peut-être possible d'acquérir une meilleure compréhension du comportement des estimateurs pertinents.

Il serait même possible d'envisager l'étude de cas où le nombre de descendants par individu serait limité à trois. Les programmes déjà implantés peuvent être à nouveau utilisés à condition de changer la procédure calculant le gradient. Il suffirait de considérer une nouvelle fonction L^* définie comme suit :

$$L^*(p_0, p_1) = L(p_0, p_1, 3 - 3p_0 - 2p_1 - \hat{m}, 2p_0 + p_1 + \hat{m} - 2)$$

et le gradient de L^* serait donné par :

$$\begin{aligned}
\frac{\partial L^*}{\partial p_0} &= \frac{\partial L}{\partial p_0} - 3 \frac{\partial L}{\partial p_2} + 2 \frac{\partial L}{\partial p_3} \\
\frac{\partial L^*}{\partial p_1} &= \frac{\partial L}{\partial p_1} - 2 \frac{\partial L}{\partial p_2} + \frac{\partial L}{\partial p_3}
\end{aligned}$$

Bien sûr, il faudrait aussi faire d'autres changements mineurs, comme par exemple, donner de nouvelles dimensions aux vecteurs.

§5.3 L'ARBRE LE PLUS VRAISEMBLABLE

Lorsque nous voulons déterminer le huitième estimateur, nous devons générer un arbre de chacune des classes d'équivalence. Nous faisons ensuite une estimation

à vraisemblance maximale au sein de l'arbre. De toutes les estimations obtenues de cette façon, nous retenons celle qui donne la plus grande vraisemblance à la réalisation observée du processus de Galton-Watson. Or, l'arbre donnant lieu à cette estimation n'est pas nécessairement l'arbre ayant la plus grande vraisemblance,

Avant d'illustrer ce fait par un exemple concret, nous allons préalablement expliquer le critère qui nous a permis de l'identifier. Soit k , le nombre maximum de descendants par individu au sein de ce processus. Soit z_0, \dots, z_n une réalisation. Soit n_i le nombre de fois où, dans un arbre donner un individu a donné naissance à i descendants ($i = 0, 1, \dots, k$). Introduisons les variables p et d respectivement comme le nombre de pères et le nombre total de descendants présents dans l'arbre. Évidemment :

$$\begin{aligned} \sum_{i=0}^k i n_i &= d = \sum_{i=1}^n z_i \\ \sum_{i=0}^k n_i &= p = \sum_{i=1}^n z_{i-1} \end{aligned}$$

Or, ces deux équations montrent que :

$$\begin{aligned} 0 &\leq n_0 = p - d + \sum_{i=2}^k (i-1)n_i \\ 0 &\leq n_1 = d - \sum_{i=2}^k i n_i \end{aligned}$$

et la vraisemblance de l'arbre est :

$$\begin{aligned} L(A) &= \prod_{i=0}^k p_i^{n_i} \\ &= p_0^{p-d+\sum_{i=2}^k (i-1)n_i} p_1^{d-\sum_{i=2}^k i n_i} \prod_{i=2}^k p_i^{n_i} \\ &= p_0^{p-d} p_1^d \prod_{i=2}^k (p_0^{i-1} p_1^{-i} p_i)^{n_i} \end{aligned}$$

Évidemment,

$$\sum_{i=2}^k i n_i \leq d \quad \text{et} \quad \sum_{i=2}^k (i-1)n_i \leq p - d.$$

De plus, si $k = 2$ on a :

$$L(A) = p_0^{p-d} p_1^d \left(\frac{p_0 p_2}{p_1} \right)^{n_2}$$

Ainsi, lorsqu'on recherche la structure d'arbre présentant la plus grande probabilité pour une fonction de masse donnée, on évalue la valeur de $\Delta = p_0 p_2 / p_1$. Si $\Delta \leq 1$, on sélectionne le plus petit nombre possible pour n_2 , tandis que si $\Delta \geq 1$, on opte pour le plus grand nombre possible. L'application de ce critère nous a conduit à identifier l'exemple suivant.

Exemple 5.3.1 *Considérons la réalisation $\langle 1, 1, 2, 4, 4 \rangle$. Dans ce cas, il y a trois classes d'équivalence. Les fonctions de vraisemblance sont :*

$$\begin{aligned} L_1(p_0, p_1, p_2) &= p_0^2 p_1^1 p_2^5 \\ L_2(p_0, p_1, p_2) &= p_0^1 p_1^3 p_2^4 \\ L_3(p_0, p_1, p_2) &= p_0^0 p_1^5 p_2^3 \end{aligned}$$

En faisant une estimation à vraisemblance maximale au sein de chacune des classes, on trouve les distributions $(1/4, 1/8, 5/8)$, $(1/8, 3/8, 1/2)$, $(0, 5/8, 3/8)$. La distribution donnant la plus grande vraisemblance au processus est la deuxième. Cependant, puisque $\Delta = 1/6$, l'arbre le plus vraisemblable est celui où intervient la plus petite puissance de 2 dans la fonction de vraisemblance. Il s'agit donc de l'arbre de la troisième classe.

Toutefois, le lecteur peut vérifier qu'il y a 6 arbres dans la première classe, 12 dans la seconde et 1 dans la troisième. Aussi, sous la distribution $(1/8, 3/8, 1/2)$, la deuxième classe est celle qui a le plus gros poids. N'y aurait-il pas lieu de croire que la classe correspondant au huitième estimateur serait celle qui aurait la plus grande probabilité sous cette distribution ? Nous n'avons pas tenté de vérifier cette conjecture. \square

§5.4 L'ENTROPIE MAXIMALE

Nous avons intégré la notion d'entropie maximale dans la formulation d'un de nos estimateurs en raison de la découverte d'une particularité fort intéressante. Soit \hat{m} le nombre moyen estimé de descendants. L'entropie de la fonction de masse du nombre de descendants par individu est définie comme suit :

$$H(p_0, p_1, \dots) = - \sum_{i \geq 0} p_i \ln p_i$$

Nous ne posons aucune contrainte sur le nombre maximal de descendants par individu. Supposons que les p_i sont inconnus et cherchons à maximiser la fonction H sous les contraintes suivantes :

$$\begin{aligned} \text{i)} \quad & \sum_{i \geq 0} ip_i = \hat{m}; \\ \text{ii)} \quad & \sum_{i \geq 0} p_i = 1. \end{aligned}$$

Évidemment, nous ne considérons que les solutions positives, c'est-à-dire $p_i \geq 0$. En introduisant des multiplicateurs de Lagrange, nous dérivons la fonction suivante :

$$f(p_0, p_1, \dots) = - \sum_{i \geq 0} p_i \ln(p_i) + \gamma_1 \left(\sum_{i \geq 0} ip_i - \hat{m} \right) + \gamma_2 \left(\sum_{i \geq 0} p_i - 1 \right)$$

On a

$$\frac{\partial f}{\partial p_i} = -\ln(p_i) - 1 + \gamma_1 i + \gamma_2$$

En posant la dérivée égale à zéro, on trouve :

$$\ln(p_i) = \gamma_1 i + \gamma_2 - 1$$

d'où

$$p_i = e^{\gamma_1 i + \gamma_2 - 1} \tag{5.1}$$

En sommant sur les valeurs de i , on obtient :

$$\begin{aligned} 1 &= \sum_{i \geq 0} p_i = \sum_{i \geq 0} e^{\gamma_1 i + \gamma_2 - 1} \\ \hat{m} &= \sum_{i \geq 0} ip_i = \sum_{i \geq 0} ie^{\gamma_1 i + \gamma_2 - 1} \end{aligned}$$

On voit alors que :

$$\begin{aligned} e^{\gamma_2 - 1} \sum_{i \geq 0} e^{\gamma_1 i} &= 1 \\ e^{\gamma_2 - 1} \times \frac{1}{1 - e^{\gamma_1}} &= 1 \\ e^{\gamma_2 - 1} &= 1 - e^{\gamma_1} \end{aligned}$$

et que :

$$e^{\gamma_2 - 1} \sum_{i \geq 0} ie^{\gamma_1 i} = \hat{m}$$

$$e^{\gamma_2-1} \frac{e^{\gamma_1}}{(1-e^{\gamma_1})^2} = \widehat{m}$$

$$(1-e^{\gamma_1}) \frac{e^{\gamma_1}}{(1-e^{\gamma_1})^2} = \widehat{m}$$

Finalement, on trouve :

$$\widehat{m} = e^{\gamma_1} (1-e^{\gamma_1})^{-1}$$

$$\widehat{m} (1-e^{\gamma_1}) = e^{\gamma_1}$$

$$\widehat{m} = (1+\widehat{m})e^{\gamma_1}$$

$$\frac{\widehat{m}}{1+\widehat{m}} = e^{\gamma_1}$$

De plus, puisque

$$e^{\gamma_2-1} = 1 - e^{\gamma_1},$$

on a que

$$e^{\gamma_2-1} = 1 - \frac{\widehat{m}}{1+\widehat{m}}$$

et donc

$$e^{\gamma_2-1} = \frac{1}{1+\widehat{m}}.$$

En remplaçant dans (5.1) on obtient

$$\hat{p}_i = e^{\gamma_1 i \gamma_2 - 1} = \frac{1}{1+\widehat{m}} \left(\frac{\widehat{m}}{1+\widehat{m}} \right)^i \quad (5.2)$$

Maintenant sachant que

$$\frac{\partial^2 f}{\partial p_i^2} = \frac{-1}{p_i} \text{ et } \frac{\partial^2 f}{\partial p_i \partial p_j} = 0, \quad i \neq j,$$

on voit que la matrice Hessienne est définie négative. Donc, la fonction H est à son maximum au point $(\hat{p}_0, \hat{p}_1, \dots)$ et les contraintes sont respectées. On remarque que la distribution maximisant l'entropie, dont la moyenne est \widehat{m} , est une distribution géométrique. Étudions maintenant la vraisemblance du processus sous cette distribution. Afin de simplifier l'écriture posons $\alpha = (1 + \widehat{m})^{-1}$ et $\beta = \widehat{m}/(1 + \widehat{m})$. On a :

$$\zeta(t) = \sum_{i \geq 0} \alpha \beta^i t^i = \frac{\alpha}{1 - \beta}$$

$$\Lambda(s, t) = \frac{1}{1 - s\zeta(t)}$$

$$\begin{aligned}
&= \sum_{i \geq 0} s^i \zeta^i(t) \\
&= \sum_{i \geq 0} s^i \alpha^i (1 - \beta t)^{-i} \\
&= \sum_{i \geq 0} s^i \alpha^i \left(\sum_{n \geq 0} \binom{-i}{n} (-\beta)^n t^n \right) \\
&= \sum_{i \geq 0} \left(\sum_{j \geq 0} \alpha^i \binom{-i}{j} (-\beta)^j s^i t^j \right)
\end{aligned}$$

Or, $A(s, t)$ est la fonction génératrice des $\ell_{m, n}$. Donc,

$$\begin{aligned}
\ell_{m, n} &= \alpha^m \binom{-m}{n} (-\beta)^n \\
&= \alpha^m \beta^n \frac{m(m+1) \dots (m+n-1)}{n!}
\end{aligned}$$

Ainsi

$$\begin{aligned}
\prod_{i=0}^{n-1} \ell_{z_i z_{i+1}} &= \left(\prod_{i=0}^{n-1} \binom{-z_i}{z_{i+1}} \right) \alpha^{\sum_{i=0}^{n-1} z_i} (-\beta)^{\sum_{i=1}^n z_i} \\
&= C \alpha^{\sum_{i=1}^n z_{i-1}} \beta^{\sum_{i=1}^n z_i}
\end{aligned}$$

où C est une constante. Nous venons donc de calculer la vraisemblance du processus observé sous la distribution géométrique trouvée précédemment. Il est possible de démontrer que, parmi toutes les distributions géométriques, celle-ci maximise la vraisemblance de la réalisation du processus.

Considérons une distribution géométrique générique, exprimée comme suit :

$$P[Z_1 = k] = p(1 - p)^k$$

où $0 < p < 1$. Nous connaissons la forme de la probabilité de la réalisation observée :

$$C p^{\sum_{i=1}^n z_{i-1}} (1 - p)^{\sum_{i=1}^n z_i}$$

Nous allons donc maximiser la fonction suivante :

$$f(p) = \left(\sum_{i=1}^n z_{i-1} \right) \ln(p) + \left(\sum_{i=1}^n z_i \right) \ln(1 - p)$$

Or,

$$\frac{df}{dp} = \frac{1}{p} \sum_{i=1}^n z_{i-1} - \frac{1}{1-p} \sum_{i=1}^n z_i$$

En posant cette dérivée égale à zéro, on obtient :

$$\begin{aligned} p &= \left(1 + \left(\sum_{i=1}^n z_i \right) \left(\sum_{i=1}^n z_{i-1} \right)^{-1} \right)^{-1} \\ &= \frac{1}{1 + \widehat{m}} \end{aligned}$$

La dérivée seconde est $-1/p^2 - 1/(1-p)^2$. Elle est donc négative en $p = 1/(1 + \widehat{m})$ et il s'agit d'un point maximum. Ceci termine la démonstration.

Conclusion

Comme expliqué dans le chapitre 1, l'expression particulière de la fonction de vraisemblance que nous avons développée a permis d'obtenir des fonctions génératrices qui se sont révélées extrêmement utiles dans nos calculs.

L'utilisation de cet outil combinatoire ouvre probablement la voie à de nouvelles découvertes. Il va sans dire que, dans la plupart des cas, des démonstrations plus conventionnelles peuvent être fournies.

Cependant, les fonctions génératrices que nous avons utilisées à titre exploratoire ont largement orienté nos recherches. De plus, étant donné les résultats énoncés au chapitre 5, nous pensons que des études futures sur l'estimation à vraisemblance maximale pourraient s'avérer très enrichissantes. L'exploration de la relation entre l'estimation à vraisemblance maximale et l'estimation à entropie maximale mérite également une attention particulière.

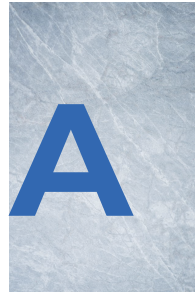
Liste des ouvrages cités

- BENDER, E. A. et J. R. GOLDMAN. 1971, « Enumerative uses of generating functions », *Indiana Univ. Math. J.*, vol. 20, n° 8, p. 753–765.
- BERGE, C. 1968, *Principes de combinatoire*, Dunod, Paris, vi+149 p..
- DION, J.-P. 1972, *Estimation des probabilités initiales et de la moyenne d'un processus de Galton-Watson*, thèse de doctorat, Université de Montréal.
- DION, J.-P. et N. KEIDING. 1978, « Statistical inference in branching processes », dans *Branching processes (Conf., Saint Hippolyte, Que., 1976)*, Dekker, New York, p. 105–140.
- DOUBILET, P., G.-C. ROTA et J. STEIN. 1974, « on the foundations of combinatorial theory. ix. combinatorial methods in invariant theory », p. 185–216.
- FEIGIN, P. D. 1977, « A note on maximum likelihood estimation for simple branching processes », *Austral. J. Statist.*, vol. 19, n° 2, p. 152–154.
- GIRI, N. C. 1977, *Multivariate statistical inference*, Academic Press [Harcourt Brace Jovanovich Publishers], New York, xvi+319 p.. Probability and Mathematical Statistics.
- HARRIS, T. E. 1948, « Branching processes », *Ann. Math. Statistics*, vol. 19, p. 474–494.
- HEYDE, C. C. et E. SENETA. 1972, « Studies in the history of probability and statistics. XXXI. The simple branching process, a turning point test and a fundamental inequality : a historical note on I. J. Bienaymé », *Biometrika*, vol. 59, p. 680–683.
- KARLIN, S. et H. M. TAYLOR. 1975, *A first course in stochastic processes*, 2^e éd., Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, xvii+557 p..

KNUTH, D. E. 1969, *The art of computer programming. Vol. 1 : Fundamental algorithms*, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont, xxi+634 p..

RIORDAN, J. 1958, *An introduction to combinatorial analysis*, John Wiley & Sons Inc., New York, xi+244 p..

ANNEXE



Programmes utilisés

L'annexe A se veut un complément au chapitre 3. Nous fournissons au lecteur les différentes procédures programmées permettant de trouver l'estimateur à vraisemblance maximale. Tous les programmes ont été écrits en FORTRAN et nous avons utilisé le compilateur FTN.

§A.1 PROCÉDURE MAXLIK

```

1      SUBROUTINE MAXLIK(LIKE,DLIKE,NSIG,MAXFN,X,F,IER,XPRIME,IT)
2  C      FONCTION          - UN ALGORITHME UTILISANT LA METHODE DU GRADIENT
3  C                          POUR TROUVER LE MAXIMUM D'UNE FONCTION DE
4  C                          VRAISEMBLANCE D'UN PROCESSUS DE RAMIFICA-
5  C                          TION OU PO>=0, P1>=0, P2>=0, MAIS PI=0 POUR
6  C                          I > 3.
7      EXTERNAL LIKE,DLIKE
8      INTEGER AIG
9      DOUBLE PRECISION EPS,XNSIG,DELTA,XNORME,F
10     DOUBLE PRECISION X(2),XPRIME(2),XO(2),X1(2), VAL(2)
11     DOUBLE PRECISION SUP(2)
12     EPS = 0.5D-10
13     IER = 0
14     ICO = 0
15     XNSIG = (10.**(-NSIG-1))*0.05
16 4     CALL GRAD(X,XPRIME)
17     XO(1) = X(1)
18     XO(2) = X(2)
19     ICO = ICO + 1
20     IF(ICO.GT.MAXFN)GO TO 7
21     AIG = 0
22     ICOME = 1
23     CALL CONTRA(,XPRIME,AIG)
24     IF(AIG.EQ.0)GO TO 1
25     CALL LIKE(,VAL(1))
26     X1(1) = XO(1) + XPRIME(1)
27     X1(2) = XO(2) + XPRIME(2)
28     DELTA = VAL(2) - VAL(1)
29 3     CALL LIKE(X1,VAL(2))
30     DELTA = VAL(2) - VAL(1)
31     IF(DELTA.LE.0.0D+00)GO TO 2
32     ICOME = 2
33     XO(1) = X1(1)
34     XO(2) = X1(2)
35     VAL(1) = VAL(2)
36     X1(1) = XO(1)+XPRIME(1)
37     X1(2) = XO(2)+XPRIME(2)
38     CALL CONTRA(X1,XPRIME,AIG)
39     IF (AIG.NE.0)GO TO 3
40  C
41  C      SI AIG = 0, NOUS NOUS SOMMES ARRETES SUR LA CONTRAINTE SPECIFIEE
42  C      PAR AIG AU DEPART DE CETTE ITERATION.
43  C
44     XO(1) = X1(1)
45     XO(2) = X1(2)
46     CALL CONTRA(XO,XPRIME,AIG)
47     IF(AIG.EQ.0)GO TO 8
48 6     IF(DABS(XO(1)-X(1)).GT.XNSIG.OR.DABS(XO(2)-X(2)).GT
49     +.XNSIG)GOTO 5
50     GO TO 8
51 5     X(1) = XO(1)
52     X(2) = XO(2)
53     GO TO 4
54 7     IER = 1
55     X(1) = XO(1)
56     X(2) = XO(2)
57     CALL LIKE(X,F)
58     WRITE (6,101)
59 101  FORMAT(//,1X,14H*** EREUR ***,20X,6HMAXLIK,20X,9H(IER =1),

```

```

60     +10X,6H*****,//)
61     RETURN
62     2 IF(ICOME.EQ.1)GO TO 10
63     XNORME = DSQRT(XPRIME(1)**2+XPRIME(2)**2)
64     SUP(1) = XPRIME(1)/XNORME
65     SUP(2) = XPRIME(2)/XNORME
66     CALL DLIKE(,VAL(1),SUP)
67     IF(VAL(1).GE.0.D+00)GO TO 10
68     X1(1) = X0(1)
69     X1(2) = X0(2)
70     (1) = X1(1)-XPRIME(1)
71     (2) = X1(2)-XPRIME(2)
72     10 CALL NEWTON(,X1,EPS,XPRIME,IER)
73     IF(IER.EQ.0)GO TO 6
74     X(1) = X1(1)
75     X(2) = X1(2)
76     CALL GRAD(X,XPRIME)
77     CALL LIKE(X,F)
78     IT = ICO
79     RETURN
80     8 X(1) = X0(1)
81     X(2) = X0(2)
82     1 CALL GRAD(X,XPRIME)
83     CALL LIKE(X,F)
84     IT = ICO
85     RETURN
86     END

```

Procédure A.1 – MAXLIK

§A.2 PROCÉDURE PART

```

1     SUBROUTINE PART(MBR,P,M,PARTA,K,NO,XMULT)
2     C
3     C     CETTE PROCEDURE ENGENDRE TOUS LES PARTAGES DU NOMBRE MBR,
4     C     EN AU PLUS P PARTS, AUCUNE N'ETANT PLUS GRANDE QUE M.
5     C
6     C     ON CALCULE AUSSI LE COEFFICIENT MULTINOMIAL ASSOCIE EN
7     C     FAISANT UN APPEL A LA PROCEDURE MULT. LES PARTAGES SONT
8     C     PLACES DANS LA MATRICE PARTA ET LES COEFFICIENTS MUL-
9     C     TINOMIAUX DANS LE VECTEUR XMULT. LES PARAMETRES K ET NO
10    C     SONT UTILISES PAR LE PROGRAMME PRINCIPAL ET SERVENT DE
11    C     COMPTEURS.
12    C
13    C     ELLE FAIT APPEL AUX PROCEDURES INIT ET SUCC.
14    C     INIT DONNE LE PARTAGE LE PLUS GROSSIER ET SUCC
15    C     GENERE LE PARTAGE SUIVANT.
16    C
17    DOUBLE PRECISION XMULT(100)
18    LOGICAL VAL
19    COMMON /NBR/NBR/VEC/VEC(21)
20    NBR = MBR
21    K = 0
22    CALL INIT(NBR,P,M,M+1)
23    1 CONTINUE
24    K = K + 1
25    NO = NO + 1
26    DO 100 NIALA=1,3
27    PARTA(NO,NIALA) = VEC(NIALA)
28    100 CONTINUE

```

```

29      CALL MULT(XMULT, PARTA, NO)
30      CALL SUCC(VAL)
31      IF(VAL) GO TO 1
32      RETURN
33      END

```

Procédure A.2 – PART

§A.3 PROCÉDURE INIT

```

1      SUBROUTINE INIT(N,P,M,POS)
2      COMMON/VEC/VEC (21)
3      INTEGER VEC,P,POS,Q,R
4      DO 1 I = 1,POS
5 1    VEC(I) = 0
6      Q = N/M
7      VEC(M+1) = Q
8      R = N-M*Q
9      VEC(1) = P-Q
10     IF(R.EQ. 0)RETURN
11     VEC(R+1) = 1
12     VEC(1) = P-Q-1
13     RETURN
14     END

```

Procédure A.3 – INIT

§A.4 PROCÉDURE SUCC

```

1      SUBROUTINE SUCC(VAL)
2      COMMON /NBR/NBR/VEC/VEC(21)
3      INTEGER VEC,SOMME,P
4      LOGICAL VAL
5      VAL = .FALSE.
6      SOMME = 0
7      P = 0
8      J = 0
9 1    J = J + 1
10     IF (VEC(J).EQ.0) GO TO 1
11     MIN = J
12 4    P = P + VEC(J)
13     SOMME = SOMME + (J-1)*VEC(J)
14     IF(SOMME.GE.NBR)RETURN
15 2    J = J + 1
16     IF(VEC(J).EG.0) GO TO 2
17     IF(J-1.LE.MIN)GO TO 4
18     VAL = .TRUE.
19     P = P+1
20     SOMME = SOMME + J - 1
21     VEC(J) = VEC(J) - 1
22     CALL INIT(SOMME,P,J-2,J-1)
23     RETURN
24     END

```

Procédure A.4 – SUCC

§A.5 PROCÉDURE MULT

```

1      SUBROUTINE MULT(XMULT, PARTA.K)
2  C    CETE PROCEDURE CALCULE DES COEFFICIENTS MULTINOMIAUX.
3  C
4  C    ELLE FAIT APPEL A LA PROCEDURE TRI DONT LE ROLE
5  C    EST DE TRIER UN VECTEUR DANS L'ORDRE ASCENDANT.
6
7      DOUBLE PRECISION XMULT(100), XLOG(84), XMOLT
8      DIMENSION LLL(3), NUM(100), NDEN(100)
9      INTEGER PARTA(100,3)
10     IF (KLUX.EQ.1)GO TO 99
11     DO 100 NIALA = 1,84
12 100 XLOG(NIALA) = DLOG(DBLE(FLOAT(NIALA)))
13     KLUX = 1
14     DO 9 I= 1,3
15     LLL(I) = PARTA(K,I)
16     9 CONTINUE
17     CALL TRI(3,LLL)
18     XMOLT = 0.0D+00
19     IF(LLL(1).NE.0)GO TO 5
20     GO TO 6
21     5 KL = LLL(1) + LLL(2)
22     DO 1 II = 1,KL
23     1 NUM(II) =LLL(3) + II
24     KL = LLL(1)
25     DO 2 II = 1,KL
26     L =2*II-1
27     NDEN(L) = II
28     2 NDEN(L+1) = II
29     L = 2*LLL(1) + 1
30     L1 = LLL(1) + 1
31     KL = LLL(1) + LLL(2)
32     DO 3 II = L,KL
33     NDEN(II) = L1
34     3 L1 = L1 + 1
35     DO 4 II = 1,KL
36     4 XMOLT = XMOLT + XLOG(NUM(II))-XLOG(NDEN(II))
37     XMULT(K) = DEXP(XMOLT)
38     RETURN
39     6 IF(LLL(2).NE.0)GO TO 7
40     XMULT(K)=1.0D+00
41     RETURN
42     7 IK = LLL(2)
43     DO 8 II = 1,IK
44     LINE = LLL(3) + II
45     8 XMOLT = XMOLT + XLOG(LINE)-XLOG(II)
46     XMULT(K) = DEXP(XMOLT)
47     RETURN
48     END

```

Procédure A.5 – MULT

§A.6 PROCÉDURE TRI

```

1      SUBROUTINE TRI(N, M)
2  C    CETTE PROCEDURE TRIANT PAR ORDRE CROISSANT LES
3  C    ELEMENTS D'UN VECTEUR "M" EST NECESSAIRE A L'EXECUTION
4  C    DE LA PROCEDURE MULT.

```

```

5 C
6   DIMENSION M(N)
7   DO 1 I = 2,N
8     IF(M(I).GE.M(I-1))GO TO 1
9     N1 = I-1
10    CALL RECH(J,N1,M,M(I))
11    K = M(I)
12    IK = I-J
13    DO 2 JJ = 1,IK
14      2 M(I-JJ+1) = M(I-JJ)
15      M(J) = K
16    1 CONTINUE
17    RETURN
18    END

```

Procédure A.6 – TRI

§A.7 PROCÉDURE RECH

```

1   SUBROUTINE RECH(IX,N,VEC,IT)
2 C
3 C   CETTE PROCEDURE EST NECESSAIRE A LA PROCEDURE TRI.
4 C   IL S'AGIT D'UNE FOUILLE BINAIRE DETERMINANT L'ENDROIT
5 C   OU DOIT ETRE INSERE UN ELEMENT QUI N'A PAS ENCORE ETE
6 C   TRIE.
7 C
8   INTEGER VEC(N)
9   MIN = 1
10  MAX = N
11  1 IF(MAX-MIN.EQ.0)GO TO 3
12  K = (MAX-MIN)/2
13  IF(IT.LE.VEC(MIN+K))GO TO 2
14  MIN = MIN + K + 1
15  GO TO 1
16  2 MAX = MIN + K
17  GO TO 1
18  3 IX = MIN
19  RETURN
20  END

```

Procédure A.7 – RECH

§A.8 PROCÉDURE NEWTON

```

1   SUBROUTINE NEWTON (XL, XR, EPS, SIP, IER)
2 C
3 C   METHODE REGULA FALSI
4 C
5 C   NOUS UTILISONS CETTE METHODE POUR ANNULER LA DERIVEE DANS LA
6 C   DIRECTION DE SUP. XL EST UN POINT INITIAL ET XP EST UN POINT
7 C   FINAL, C'EST-A-DIRE QUE SUR LE SEGMENT DETERMINE PAR CES DEUX
8 C   POINTS, LA DERIVEE S'ANNULE.
9 C
10  C   EPS EST L'ERREUR QUE NOUS TOLERONS SUR L'EVALUATION DE CETTE
11  C   DERIVEE DIRECTIONNELLE.
12  C
13  C   DOUBLE PRECISION SIP(2),FXAPP,FXL,FXR,XNORME
14  C   DOUBLE PRECISION EPS,P,T
15  C   DOUBLE PRECISION XL(2),XR(2),XAPP(2),SUP(2)

```

```

15     XNORME = DSQRT(SIP(1)**2+SIP(2)**2)
16     SUP(1) = SIP(1)/XNORME
17     SUP(2) = SIP(2)/XNORME
18     P1 = 0
19 C
20 C     ON CALCULE LES VALEURS DE DLIKE AUX DEUX POINTS
21 C
22     CALL DLIKE(XL,FXL,SUP)
23     CALL DLIKE(XR,FXR,SUP)
24 C
25 C     SI LE PRODUIT DE FXL ET FXR EST NEGATIF, LA VRAIE VALEUR SE SI-
26 C     TUE DANS L'INTERVALLE ENTRE XL ET XR. SI C'EST EGAL A ZERO, NOUS
27 C     AVONS LE POINT QUE NOUS CHERCHIONS.
28 C
29     IF (FXL*FXR) 31,23,24
30 24 WRITE (6,11)
31 11 FORMAT (1X, 10H*** NEWTON, 20X,11HIER = 2 ***)
32     IER = 2
33     RETURN
34 23 IF(FXL.EQ.0.0D+00)GO TO 25
35     XL(1) = XR(1)
36     XL(2) = XR(2)
37 25 IER = 0
38     RETURN
39 31 XAPP(1) = XL(1) + (XR(1)-XL(1))/2.
40     XAPP(2) = XL(2) + (XR(2)-XL(2))/2.
41     CALL DLIKE(XAPP,FXAPP,SUP)
42 C
43 C     SI LA VALEUR ABSOLUE DE FXAPP EST MOINDRE QUE LA TOLERANCE
44 C     SPECIFIEE, ARRETER LES CALCULS.
45 C
46     IF(DABS(FXAPP).GT.EPS)GO TO 27
47     XL(1) = XAPP(1)
48     XL(2) = XAPP(2)
49     IER = 0
50     RETURN
51 C
52 C     LES INSTRUCTIONS SUIVANTES NOUS PERMETTENT DE TROUVER UN NOU-
53 C     VEL INTERVALLE.
54 C
55 27 IF (FXAPP.LT.0.0D+00) GO TO 28
56     XL(1) = XAPP(1)
57     XL(2) = XAPP(2)
58     GO TO 31
59 28 XR(1) = XAPP(1)
60     XR(2) = XAPP(2)
61     GO TO 31
62     END

```

Procédure A.8 – NEWTON

§A.9 PROCÉDURE EVAL

```

1     SUBROUTINE EVAL(MIN,F,P,MAX,NM)
2     DOUBLE PRECISION P(NM),A(2,100),F,XP
3     IF(M.EQ.0.OR.M.NE.0)GO TO 98
4     F = 1.0D+00
5     RETURN
6 98 IF(M.GT.1)GO TO 99
7     F = P(N+1)

```

```

8      RETURN
9      99 K2 = N+MAX+1
10     DO 1 I = 1,K2
11     DO 1 J = 1,2
12     1 A(J,I) = 0.0D+00
13     K1 = MAX + 1
14     A(1,K1) = 1.0D+00
15     K2 = 2*MAX+1
16     DO 2 I = K1,K2
17     2 A(2,I) = P(I-MAX)
18     NINF = N/MAX + 1
19     IF (N.LE.MAX)NINF=1
20     NSUP = M - NINF + 1
21     IHAUT = 2
22     IBAS = 1
23     ISHIFT = N - MAX*(N/MAX)
24     K = 3
25     MAXABS = N + MAX + 1
26     C
27     C      AVONS-NOUS FINI DE CALCULER LES LIGNES SUPERIEURES?
28     C
29     IF(NSUP-2)12,13,3
30     C
31     C      SI NSUP < 2, ALORS NSUP=-1. NOUS AURIONS DONC QUE LA PRO-
32     C      BABILITE CHERCHEE SE TROUVE SUR LA DEUXIEME LIGNE CE QUE
33     C      NOUS AURIONS DECELE PLUS HAUT, OU QUE LA PROBABILITE QUE
34     C      NOUS CHERCHONS NE NECESSITE PAS LA CONNAISSANCE DE TOUTE
35     C      LA LIGNE 2.
36     C
37     12 IF(NINF.EQ.0)GO TO 15
38     NSUP= M+1
39     K1 = ISHIFT + MAX
40     GO TO 7
41     13 IF(NINF.NE.0)GO TO 6
42     15 IBAS = 2
43     IHAUT = 1
44     GO TO 11
45     3 IF(K.GT.NSUP)GO TO 6
46     K2 = K*MAX+1
47     IF(K2.GT.MAXABS)K2 = MAXABS
48     DO 5 I = K1,K2
49     XP = 0.0D+00
50     DO 4 J = 1,K1
51     4 XP = XP + P(J)*A(IHAUT,I-J+1)
52     5 A(IBAS,I) = XP
53     ICOPIE = IBAS
54     IBAS = IHAUT
55     IHAUT = ICOPIE
56     K = K+ 1
57     GO TO 3
58     6 NSUP = M + 1
59     K1 = ISHIFT + 1
60     7 IF(K.GT.NSUP)GO TO 11
61     K1 = K1 + MAX
62     K2 = K*MAX+1
63     IF(K2.GT.MAXABS)K2 = MAXABS
64     DO 9 I = K1,K2
65     XP = 0.0D+00
66     K3 = 1
67     K4 = MAX + 1
68     DO 8 J1 = K3,K4
69     K5 = I-J1+1

```



```

70     XP = XP + P(J1)*A(IHAUT,K5)
71     8 CONTINUE
72     9 A(IBAS,I) = XP
73     ICOPIE = IBAS
74     IBAS = IHAUT
75     IHAUT = ICOPIE
76     K = K+ 1
77     GO TO 7
78     11 F = A(IHAUT,N+MAX+1)
79     RETURN
80     END

```

Procédure A.9 – EVAL

§A.10 PROCÉDURE LIKE

```

1     SUBROUTINE LIKE(X, F)
2     COMMON /BLOC1/IX(11),NGEN
3     DOUBLE PRECISION X(2),XP(3),H,F,G
4     MAX = 2
5     NM = 3
6     P(1) = X(1)
7     P(2) = X(2)
8     P(3) = 1.0D+00 - P(1) - P(2)
9     H = 1.0D+00
10    DO 3 I = 2, NGEN
11    CALL EVAL(IX(I-1), IX(I), G, P, MAX, NM)
12    3 H = H*G
13    F = H
14    RETURN
15    END

```

Procédure A.10 – LIKE

§A.11 PROCÉDURE GRAD

```

1     SUBROUTINE GRAD(X, XPRIME)
2     COMMON /BLOC1/IX(11),NGEN
3     DOUBLE PRECISION SOMME,G,L(10),P
4     DOUBLE PRECISION X(2),XP(3),XPRIME(2),YPRIME(3)
5     XP(1) = X(1)
6     XP(2) = X(2)
7     XP(3) = 1.0D+00 - X(1) - X(2)
8     ITEST = 0
9     P = 1.D+00
10    DO 1 I = 2, NGEN
11    CALL EVAL(IX(I-1), IX(I), L(I-1), XP, 2, 3)
12    IF(L(I-1).NE.0)GO TO 5
13    ITEST = ITEST+1
14    IF (ITEST.EQ.2)GO TO 4
15    IND = I-1
16    GO TO 1
17    5 P = P*L(I-1)
18    1 CONTINUE
19    IF (ITEST.EQ.0)GO TO 6
20    DO 7 J = 1,3
21    II = IX(IND) - 1
22    JJ = IX(IND+1) - J+1

```

```

23     IF(II.LT.0.OR.JJ.LT.0.OR.2*II.LT.JJ)GO TO 8
24     CALL EVAL(II,JJ,G,XP,2,3)
25     YPRIME(J) = P*G
26     GO TO 7
27     8 YPRIME(J) = 0.0D+00
28     7 CONTINUE
29     GO TO 9
30     6 DO 3 J = 1, 3
31     SOMME = 0.D+00
32     MANIA = NGEN-1
33     DO 2 I = 1,MANIA
34     II = IX(I) -1
35     JJ = IX(I+1)-J+1
36     IF(II.LT.0.OR.JJ.LT.0)GO TO 2
37     IF(2*II.LT.JJ)GO TO 2
38     CALL EVAL(II,JJ,G,XP,2,3)
39     SOMME = SOMME + IX(I)*G/L(I)
40     2 CONTINUE
41     YPRIME(J) = SOMME
42     3 CONTINUE
43     9 XPRIME(1) = (YPRIME(1)-YPRIME(3))*P
44     XPRIME(2) = (YPRIME(2)-YPRIME(3))*P
45     IF(DABS(XPRIME(1)).LT..5D-15)XPRIME(1)= 0.0D+00
46     IF(DABS(XPRIME(2)).LT..5D-15)XPRIME(2)= 0.0D+00
47     RETURN
48     4 XPRIME(1) = 0.0D+00
49     XPRIME(2) = 0.0D+00
50     RETURN
51     END

```

Procédure A.11 – GRAD

§A.12 PROCÉDURE DLIKE

```

1     SUBROUTINE DLIKE (X,FX,SUP)
2     DOUBLE PRECISION X(2),XPRIME(2),SUP(2),FX
3     CALL GRAD(X,XPRIME)
4     FX = XPRIME(1)*SUP(1) + XPRIME(2)*SUP(2)
5     RETURN
6     END

```

Procédure A.12 – DLIKE

§A.13 PROCÉDURE CONTRA

```

1     SUBROUTINE CONTRA(X,XPRIME,AIG)
2     INTEGER AIG
3     LOGICAL TIC
4     DOUBLE PRECISION XPRIME(2),X(2),Z(2),TX(2),T,VEC(3,2),XNORME
5     DIMENSION TETA(3)
6     IF(AIG.GT.0)GO TO 1
7     VEC(1,1) = -X(1)
8     VEC(1,2) = -X(2)
9     VEC(2,1) = VEC(1,1)
10    VEC(2,2) = VEC(1,2) + 1.D+00
11    VEC(3,1) = VEC(1,1) + 1.D+00
12    VEC(3,2) = VEC(1,2)
13    CALL ANGLE(XPRIME,ALPHA)

```

```

14     IF (ALPHA.NE.7.)GO TO 14
15 23 XPRIME(1) = 0.
16     XPRIME(2) * 0.
17     AIG = 0
18     RETURN
19 14 IF (ITO.EQ.1)GO TO 61
20     TX(1) = 1.
21     TX(2) = 1.
22     CALL ANGLE(TX,PISU4)
23     TX(1) = 0.
24     TX(2) = 1.
25     CALL ANGLE(TX,PISU2)
26     TX(1) = -1.
27     TX(2) = 1.
28     CALL ANGLE(TX, PI3SU4)
29     TX(1) = -1.
30     TX(2) = 0.
31     CALL ANGLE(TX,PI)
32     TX(1) = -1.
33     TX(2) = -1.
34     CALL ANGLE(TX,PI5SU4)
35     TX(1) = 0.
36     TX(2) = -1.
37     CALL ANGLE(TX,PI3SU2)
38     TX(1) = 1.
39     TX(2) = -1.
40     CALL ANGLE(TX, PI7SU4)
41     ITO = 1
42 61 DO 2 I = 1,3
43     DO 3 J = 1,2
44     3 TX(J) = VEC(I,J)
45     CALL ANGLE(TX,TETA(I))
46     IF (TETA(I).EQ.7.)GO TO 10
47     2 CONTINUE
48     IF (TETA(3).EQ.0.)GO TO 11
49     IF (TETA(2).EQ.PISU2)GO TO 12
50     IF (TETA(2).EQ.PI3SU4)GO TO 13
51 C
52 C     LE POINT EST A L'INTERIEUR
53 C
54     IF (ALPHA.LE.TETA(1).AND.ALPHA.GE.TETA(2))GO TO 4
55     IF (ALPHA.LE.TETA(3).AND.ALPHA.GE.TETA(1))GO TO 5
56 26 AIG = 3
57     IF (XPRIME(1))15,16,15
58 16 T * 1.D+00 -X(1)-X(2)
59     Z(1) = X(1)
60     Z(2) = 1.00D+00 - X(1)
61 40 XNORME = DSQRT(XPRIME(1)**2+XPRIME(2)**2)
62     T = T/(10.D+00*XNORME)
63     XPRIME(1) = XPRIME(1)*T
64     XPRIME(2) = XPRIME(2)*T
65     RETURN
66 15 T = XPRIME(2)/XPRIME(1)
67     Z(1) = (1.D+00-(X(2)-T*X(1)))/(T+1.D+00)
68     Z(2) = 1.D+00 - Z(1)
69 41 TX(1) = Z(1) - X(1)
70     TX(2) = Z(2) - X(2)
71     T = DSQRT(TX(1)**2+TX(2)**2)
72     GO TO 40
73 C
74 C     ON POURRAIT VIOLER Y >= 0
75 C

```

```

76     5 AIG = 2
77     IF(XPRIME(1))19,20,19
78   20 T = X(2)
79     Z(1) = X(1)
80     Z(2) = 0.D+00
81     GO TO 40
82   19 T = XPRIME(2)/XPRIME(1)
83     Z(1) = X(1) - X(2)/T
84     Z(2) = 0.D+00
85     GO TO 41
86   C
87   C     ON POURRAIT VIOLER X >= 0
88   C
89     4 AIG = 1
90     IF (XPRIME(1))21,22,21
91   22 T = X(1)
92     Z(1) = 0.D+00
93     Z(2) = X(2)
94     GO TO 40
95   21 T = XPRIME(2)/XPRIME(1)
96     Z(1) = 0.D+00
97     Z(2) = X(2) - T*X(1)
98     GO TO 41
99   C
100  C     ON EST A L'INTERIEUR D'UN SEGMENT DE X+Y=1
101  C
102   13 IF(ALPHA.LE.TETA(1).AND.ALPHA.GE.TETA(2))GO TO 4
103     IF(ALPHA.LE.TETA(3).AND.ALPHA.GE.TETA(1))GO TO 5
104     IF(ALPHA-PISU4)24,23,25
105   24 AIG = 2
106     XPRIME(1) = X(2)/10.D+00
107     XPRIME(2) = -XPRIME(1)
108     Z(1) = 1.D+00
109     Z(2) = 0.D+00
110     RETURN
111   25 IF(ALPHA.GE.PI7SU4)GO TO 24
112     AIG = 1
113     XPRIME(1) = -X(1)/10.D+00
114     XPRIME(2) = -XPRIME(1)
115     Z(1) = 0.D+00
116     Z(2) = 1.D+00
117     RETURN
118  C
119  C     ON EST SUR X = 0
120  C
121   12 IF(ALPHA.LE.TETA(3).AND.ALPHA.GE.PI3SU2)GO TO 5
122     IF(ALPHA.LE.PISU2.OR.ALPHA.GE.TETA(3))GO TO 26
123     IF (ALPHA-PI)27,23,29
124   27 AIG = 3
125     Z(1) = 0.D+00
126     Z(2) = 1.D+00
127     T = 1.D+00-X(2)
128     XPRIME(1) = 0.D+00
129     XPRIME(2) = T/10.D+00
130     RETURN
131   29 AIG = 2
132     Z(1) = 0.D+00
133     Z(2) = 0.D+00
134     T = X(2)
135     XPRIME(1) = 0.D+00
136     XPRIME(2) = -T/10.D+00
137     RETURN

```

```

138 C
139 C     ON EST SUR X = 0
140 C
141 11 IF (ALPHA .LE. PI .AND. ALPHA .GE. TETA(2)) GO TO 4
142 IF (ALPHA .LE. TETA(2) .AND. ALPHA .GE. 0.000) GO TO 26
143 IF (ALPHA - PI3SU2) 30, 23, 31
144 30 AIG = 1
145 Z(1) = 0.D+00
146 Z(2) = 0.D+00
147 T = X(1)
148 XPRIME(1) = -T/10.D+00
149 XPRIME(2) = 0.D+00
150 RETURN
151 31 AIG = 3
152 Z(1) = 1.D+00
153 Z(2) = 0.0D+00
154 T = 1.D+00 - X(1)
155 XPRIME(1) = T/10.D+00
156 XPRIME(2) = 0.D+00
157 RETURN
158 C
159 C     ON EST SUR UN SOMMET
160 C
161 10 GOTO(32, 33, 34) I
162 C
163 C     ON EST A L'ORIGINE
164 C
165 32 IF (XPRIME(1) .LE. 0.D+00 .AND. XPRIME(2) .LE. 0.D+00) GO TO 23
166 IF (XPRIME(1) .GE. 0.D+00 .AND. XPRIME(2) .GE. 0.D+00) GO TO 26
167 IF (XPRIME(1)) 27, 27, 31
168 C
169 C     ON EST AU POINT (1,0)
170 C
171 34 IF (ALPHA .LE. PISU4 .OR. ALPHA .GE. PI3SU2) GO TO 23
172 IF (ALPHA .LE. PI3SU4) GO TO 25
173 IF (ALPHA .GE. PI .AND. ALPHA .LE. PI3SU2) GO TO 30
174 GO TO 4
175 C
176 C     ON EST AU POINT (0,1)
177 C
178 33 IF (ALPHA .GE. PISU4 .AND. ALPHA .LE. PI) GO TO 23
179 IF (ALPHA .LE. PISU4 .OR. ALPHA .GE. PI7SU4) GO TO 24
180 IF (ALPHA .GE. PI .AND. ALPHA .LE. PI3SU2) GO TO 29
181 GO TO 5
182 C
183 C
184 C
185 1 TIC = X(1) .GE. 0.D+00 .AND. X(2) .GE. 0.D+00 .AND. X(1)+X(2) .LE. 1.D+00
186 IF (TIC) RETURN
187 AIG = 0
188 X(1) = Z(1)
189 X(2) = Z(2)
190 CALL GRAD(X, XPRIME)
191 RETURN
192 END

```

§A.14 PROCÉDURE ANGLE

```

1      SUBROUTINE ANGLE(X,R)
2      C
3      C      CETTE PROCEDURE CALCULE L'ANGLE ENTRE L'AXE DES X ET LE VECTEUR 'X'.
4      C      IL S'AGIT DE L'ANGLE PRIS DANS LE SENS ANTI-HORAIRE.
5      C
6      DOUBLE PRECISION X(2)
7      DATA PI/3.141592653589793238462643/
8      IF(X(1))5,1,6
9      1 IF(X(2))2,3,4
10     2 R = 4.712388980384689857693965
11     RETURN
12     C
13     C      SI LE VECTEUR EST UN VECTEUR NUL (0,0),
14     C      ON ATTRIBUE UNE VALEUR DE 7 A L'ANGLE R.
15     C
16     3 R = 7
17     RETURN
18     4 R = 1.570796326794896619231322
19     RETURN
20     5 R = X(2)/X(1)
21     R = ATAN(R)
22     R = R + PI
23     RETURN
24     6 R = X(2)/X(1)
25     R = ATAN(R)
26     IF(X(2).LT.0.D+00)R = R + 2.*PI
27     RETURN
28     END

```

Procédure A.14 – ANGLE

ANNEXE



Exemples d'estimation

Le lecteur trouvera dans cette annexe quelques exemples d'estimations faites au cours de notre travail, Il s'agit de cas auxquels nous faisons référence dans le texte.

TABLEAU B.1 – Énumération des partages pour la réalisation $\langle 1, 2, 2, 2, 2 \rangle$

Réalisation			Partages admissibles			Coefficient multinomial
i	z_i	z_{i+1}	$\lambda_{0,k}^{i+1}$	$\lambda_{1,k}^{i+1}$	$\lambda_{2,k}^{i+1}$	
0	1	2	0	0	1	1
1	2	2	1	0	1	2
			0	2	0	1
2	2	2	1	0	1	2
			0	2	0	1
3	2	2	1	0	1	2
			0	2	0	1

TABLEAU B.2 – Valeurs des estimateurs pour la réalisation $\langle 1, 2, 2, 2, 2 \rangle$

No de l'estimateur	\hat{p}_0	\hat{p}_1	\hat{p}_2
1	.23077	.46154	.30769
2	.24000	.48000	.28000
3	.31579	.31579	.36842
4	.32432	.32432	.35135
5	.15000	.60000	.25000
6	.15789	.63158	.21053
7	.28571	.28571	.42857
8	.42857	.00000	.57143
9	.42857	.00000	.57143
10	.21429	.42857	.35714
11	.30000	.30000	.40000
Estimation de la moyenne, \hat{m} :			1.1429

TABLEAU B.3 – Énumération des partages pour la réalisation $\langle 1, 2, 4, 5, 7 \rangle$

Réalisation			Partages admissibles			Coefficient multinomial
i	z_i	z_{i+1}	$\lambda_{0,k}^{i+1}$	$\lambda_{1,k}^{i+1}$	$\lambda_{2,k}^{i+1}$	
0	1	2	0	0	1	1
1	2	4	0	0	2	1
2	4	5	1 0	1 3	2 1	12 4
3	5	7	1 0	1 3	3 2	20 10

TABLEAU B.4 – Valeurs des estimateurs pour la réalisation $\langle 1, 2, 4, 5, 7 \rangle$

No de l'estimateur	\hat{p}_0	\hat{p}_1	\hat{p}_2
1	.09524	.38095	.52381
2	.10345	.41379	.48276
3	.14747	.34101	.51152
4	.14639	.34224	.51137
5	.02623	.23279	.74098
6	.04142	.36095	.59763
7	.08333	.33333	.58333
8	.00000	.50000	.50000
9	.00000	.50000	.50000
10	.08333	.33333	.58333
11	.14545	.33636	.51818
Estimation de la moyenne, \hat{m} :			1.5000

TABLEAU B.5 – Énumération des partages pour la réalisation $\langle 1, 2, 3, 4, 4 \rangle$

Réalisation			Partages admissibles			Coefficient multinomial
i	z_i	z_{i+1}	$\lambda_{0,k}^{i+1}$	$\lambda_{1,k}^{i+1}$	$\lambda_{2,k}^{i+1}$	
0	1	2	0	0	1	1
1	2	3	0	1	1	2
2	3	4	1	0	2	3
			0	2	1	3
3	4	4	2	0	2	6
			1	2	1	12
			0	4	0	1

TABLEAU B.6 – Valeurs des estimateurs pour la réalisation $\langle 1, 2, 3, 4, 4 \rangle$

No de l'estimateur	\hat{p}_0	\hat{p}_1	\hat{p}_2
1	.19048	.42857	.38095
2	.21127	.45070	.33803
3	.27273	.36364	.36364
4	.28610	.36512	.34877
5	.08333	.59259	.32407
6	.09195	.67816	.22989
7	.20000	.30000	.50000
8	.00000	.70000	.30000
9	.00000	.70000	.30000
10	.15000	.40000	.45000
11	.24153	.36441	.39407
Estimation de la moyenne, \hat{m} :			1.3000

TABLEAU B.7 – Énumération des partages pour la réalisation $\langle 1, 1, 2, 4, 4 \rangle$

Réalisation			Partages admissibles			Coefficient multinomial
i	z_i	z_{i+1}	$\lambda_{0,k}^{i+1}$	$\lambda_{1,k}^{i+1}$	$\lambda_{2,k}^{i+1}$	
0	1	1	0	1	0	1
1	1	2	0	0	1	1
2	2	4	0	0	2	1
3	4	4	2	0	2	6
			1	2	1	12
			0	4	0	1

TABLEAU B.8 – Valeurs des estimateurs pour la réalisation $\langle 1, 1, 2, 4, 4 \rangle$

No de l'estimateur	\hat{p}_0	\hat{p}_1	\hat{p}_2
1	.18750	.43750	.37500
2	.22222	.46296	.31481
3	.30000	.36250	.33750
4	.30968	.36452	.32581
5	.04630	.57407	.37963
6	.06410	.67949	.25641
7	.12500	.37500	.50000
8	.12500	.37500	.50000
9	.13733	.35034	.51233
10	.12500	.37500	.50000
11	.27273	.35227	.37500
Estimation de la moyenne, \hat{m} :			1.3750



Tableaux et graphiques des biais

§C.1 OBJECTIFS

Cette annexe est divisée en trois parties. La première nous donne les distances entre l'espérance des différents estimateurs et les paramètres pour chacune des vingt distributions étudiées. Nous présentons ensuite les graphiques qui ont permis de faire la discussion détaillée au chapitre 5.

La seconde partie de cette section contient vingt graphiques illustrant les courbes du biais versus le nombre de générations observées. Nous n'avons pas tracé toutes les courbes.

Finalement, on trouve les graphiques du biais versus le nombre de générations observées pour les estimateurs 1, 9 et 8. Remarquons que pour les estimateurs 8 et 9, nous aurions pu tracer qu'une courbe par distribution puisque leurs courbes respectives sont presque toujours confondues.

§C.2 TABLEAUX DES BIAIS

TABLEAU C.1 – Biais des estimateurs pour $p_0 = 0.0, p_1 = 0.1, p_2 = 0.9$

Estim. 1	Estim. 2	Estim. 3	Estim. 4
0.0000×10^0	0.0000×10^0	0.0000×10^0	0.0000×10^0
1.3530×10^{-2}	2.4820×10^{-2}	2.3692×10^{-2}	3.1185×10^{-2}
1.1765×10^{-2}	1.6061×10^{-2}	4.0095×10^{-2}	3.6204×10^{-2}
7.7755×10^{-3}	6.5503×10^{-3}	4.5872×10^{-2}	3.2875×10^{-2}
8.4384×10^{-3}	1.2897×10^{-2}	3.5501×10^{-2}	2.3218×10^{-2}
Estim. 5	Estim. 6	Estim. 7	Estim. 8
0.0000×10^0	0.0000×10^0	0.0000×10^0	0.0000×10^0
7.9550×10^{-4}	1.4800×10^{-2}	1.2772×10^{-2}	1.0607×10^{-2}
2.0820×10^{-2}	1.3374×10^{-2}	1.2185×10^{-2}	5.5216×10^{-3}
3.9937×10^{-2}	3.9157×10^{-2}	5.5774×10^{-3}	5.4515×10^{-3}
5.2585×10^{-2}	5.5966×10^{-2}	4.7743×10^{-3}	7.6653×10^{-3}
Estim. 9	Estim. 10	Estim. 11	\hat{m}
0.0000×10^0	0.0000×10^0	0.0000×10^0	0.0000×10^0
1.0607×10^{-2}	1.2772×10^{-2}	2.3202×10^{-2}	1.5000×10^{-2}
5.5527×10^{-3}	8.7714×10^{-3}	3.8608×10^{-2}	1.4717×10^{-2}
5.2173×10^{-3}	3.8612×10^{-3}	4.3309×10^{-2}	1.1422×10^{-2}
7.4157×10^{-3}	1.3426×10^{-1}	1.7500×10^{-1}	8.0169×10^{-3}

TABLEAU C.2 – Biais des estimateurs pour $p_0 = 0.0, p_1 = 0.3, p_2 = 0.7$

Estim. 1	Estim. 2	Estim. 3	Estim. 4
0.0000×10^0	0.0000×10^0	0.0000×10^0	0.0000×10^0
2.5297×10^{-2}	4.8112×10^{-2}	4.0610×10^{-2}	5.5595×10^{-2}
2.2601×10^{-2}	3.1195×10^{-2}	4.8789×10^{-2}	4.5471×10^{-2}
2.4100×10^{-2}	2.7602×10^{-2}	3.7878×10^{-2}	3.1069×10^{-2}
3.8546×10^{-2}	4.9373×10^{-2}	3.2889×10^{-2}	3.7187×10^{-2}
Estim. 5	Estim. 6	Estim. 7	Estim. 8
0.0000×10^0	0.0000×10^0	0.0000×10^0	0.0000×10^0
5.5685×10^{-3}	3.3801×10^{-2}	1.9953×10^{-2}	1.1068×10^{-2}
2.6236×10^{-2}	1.2708×10^{-2}	1.2964×10^{-2}	1.3008×10^{-2}
5.9369×10^{-2}	5.3772×10^{-2}	4.4574×10^{-2}	1.2093×10^{-2}
8.5740×10^{-2}	9.4134×10^{-2}	6.1260×10^{-2}	1.1651×10^{-2}
Estim. 9	Estim. 10	Estim. 11	\hat{m}
0.0000×10^0	0.0000×10^0	0.0000×10^0	0.0000×10^0
1.1068×10^{-2}	1.9953×10^{-2}	3.7071×10^{-2}	3.5000×10^{-2}
1.2790×10^{-2}	1.1954×10^{-2}	4.1905×10^{-2}	3.7660×10^{-2}
1.1634×10^{-2}	2.6661×10^{-2}	3.0739×10^{-2}	3.3261×10^{-2}
1.2086×10^{-2}	4.3229×10^{-2}	3.2488×10^{-2}	2.7031×10^{-2}

TABLEAU C.3 – Biais des estimateurs pour $p_0 = 0.0, p_1 = 0.5, p_2 = 0.5$

Estim. 1	Estim. 2	Estim. 3	Estim. 4
0.0000×10^0	0.0000×10^0	0.0000×10^0	0.0000×10^0
2.4296×10^{-2}	4.6564×10^{-2}	3.3693×10^{-2}	4.7839×10^{-2}
2.7437×10^{-2}	3.5731×10^{-2}	3.8417×10^{-2}	4.1200×10^{-2}
4.1786×10^{-2}	4.5786×10^{-2}	5.8581×10^{-2}	6.4892×10^{-2}
6.3907×10^{-2}	7.2482×10^{-2}	8.6280×10^{-2}	9.5171×10^{-2}
Estim. 5	Estim. 6	Estim. 7	Estim. 8
0.0000×10^0	0.0000×10^0	0.0000×10^0	0.0000×10^0
1.1049×10^{-2}	3.9707×10^{-2}	1.4731×10^{-2}	2.9463×10^{-2}
1.7762×10^{-2}	2.7360×10^{-2}	5.2632×10^{-2}	1.6071×10^{-2}
3.8604×10^{-2}	3.0460×10^{-2}	9.3658×10^{-2}	1.4955×10^{-2}
6.1795×10^{-2}	5.8103×10^{-2}	1.2150×10^{-1}	1.3663×10^{-2}
Estim. 9	Estim. 10	Estim. 11	\hat{m}
0.0000×10^0	0.0000×10^0	0.0000×10^0	0.0000×10^0
2.9463×10^{-2}	1.4731×10^{-2}	2.6882×10^{-2}	4.1667×10^{-2}
1.5868×10^{-2}	2.3878×10^{-2}	3.1091×10^{-2}	4.8214×10^{-2}
1.4928×10^{-2}	4.8660×10^{-2}	5.4838×10^{-2}	4.6828×10^{-2}
1.3597×10^{-2}	7.3845×10^{-2}	8.4428×10^{-2}	4.2597×10^{-2}

TABLEAU C.4 – Biais des estimateurs pour $p_0 = 0.0, p_1 = 0.7, p_2 = 0.3$

Estim. 1	Estim. 2	Estim. 3	Estim. 4
0.0000×10^0	0.0000×10^0	0.0000×10^0	0.0000×10^0
1.8065×10^{-2}	3.1613×10^{-2}	2.2618×10^{-2}	2.9439×10^{-2}
3.0071×10^{-2}	3.2846×10^{-2}	4.9573×10^{-2}	5.1153×10^{-2}
5.1893×10^{-2}	5.3048×10^{-2}	8.7748×10^{-2}	9.3018×10^{-2}
7.5731×10^{-2}	7.9696×10^{-2}	1.2400×10^{-1}	1.3214×10^{-1}
Estim. 5	Estim. 6	Estim. 7	Estim. 8
0.0000×10^0	0.0000×10^0	0.0000×10^0	0.0000×10^0
1.2993×10^{-2}	3.3204×10^{-2}	1.2374×10^{-2}	4.5634×10^{-2}
1.7714×10^{-2}	3.3088×10^{-2}	7.3581×10^{-2}	2.1069×10^{-2}
2.4003×10^{-2}	3.0543×10^{-2}	1.1321×10^{-1}	1.4113×10^{-2}
3.1610×10^{-2}	3.1984×10^{-2}	1.5499×10^{-1}	2.4134×10^{-2}
Estim. 9	Estim. 10	Estim. 11	\hat{m}
0.0000×10^0	0.0000×10^0	0.0000×10^0	0.0000×10^0
4.5634×10^{-2}	1.2374×10^{-2}	1.8382×10^{-2}	3.5000×10^{-2}
2.0884×10^{-2}	3.3697×10^{-2}	4.7766×10^{-2}	4.2860×10^{-2}
1.4105×10^{-2}	5.8667×10^{-2}	8.4160×10^{-2}	4.4592×10^{-2}
2.4327×10^{-2}	8.2562×10^{-2}	1.1879×10^{-1}	4.3962×10^{-2}

TABLEAU C.5 – Biais des estimateurs pour $p_0 = 0.0, p_1 = 0.8, p_2 = 0.2$

Estim. 1	Estim. 2	Estim. 3	Estim. 4
0.0000×10^0	0.0000×10^0	0.0000×10^0	0.0000×10^0
1.3867×10^{-2}	2.1944×10^{-2}	1.8291×10^{-2}	2.1080×10^{-2}
2.7822×10^{-2}	2.8067×10^{-2}	4.8410×10^{-2}	4.8944×10^{-2}
4.9095×10^{-2}	4.9023×10^{-2}	8.4981×10^{-2}	8.8776×10^{-2}
7.1053×10^{-2}	7.3208×10^{-2}	1.1973×10^{-1}	1.2648×10^{-1}
Estim. 5	Estim. 6	Estim. 7	Estim. 8
0.0000×10^0	0.0000×10^0	0.0000×10^0	0.0000×10^0
1.1314×10^{-2}	2.5327×10^{-2}	1.1926×10^{-2}	4.2999×10^{-2}
1.5219×10^{-2}	2.6207×10^{-2}	6.5088×10^{-2}	2.6430×10^{-2}
2.0625×10^{-2}	2.5806×10^{-2}	1.0199×10^{-1}	1.4547×10^{-2}
2.6647×10^{-2}	2.7895×10^{-2}	1.4261×10^{-1}	2.6959×10^{-2}
Estim. 9	Estim. 10	Estim. 11	\hat{m}
0.0000×10^0	0.0000×10^0	0.0000×10^0	0.0000×10^0
4.2999×10^{-2}	1.1926×10^{-2}	1.7233×10^{-2}	2.6667×10^{-2}
2.6364×10^{-2}	3.2309×10^{-2}	4.7976×10^{-2}	3.3417×10^{-2}
1.4586×10^{-2}	5.4213×10^{-2}	8.1681×10^{-2}	3.5690×10^{-2}
2.7116×10^{-2}	7.5098×10^{-2}	1.1376×10^{-1}	3.6242×10^{-2}

TABLEAU C.6 – Biais des estimateurs pour $p_0 = 0.0, p_1 = 0.9, p_2 = 0.1$

Estim. 1	Estim. 2	Estim. 3	Estim. 4
0.0000×10^0	0.0000×10^0	0.0000×10^0	0.0000×10^0
8.2377×10^{-3}	1.1526×10^{-2}	1.1976×10^{-2}	1.2467×10^{-2}
1.9480×10^{-2}	1.8743×10^{-2}	3.4052×10^{-2}	3.4127×10^{-2}
3.5164×10^{-2}	3.4784×10^{-2}	5.9836×10^{-2}	6.1988×10^{-2}
5.1250×10^{-2}	5.2379×10^{-2}	8.5247×10^{-2}	8.9631×10^{-2}
Estim. 5	Estim. 6	Estim. 7	Estim. 8
0.0000×10^0	0.0000×10^0	0.0000×10^0	0.0000×10^0
7.1595×10^{-3}	1.4295×10^{-2}	8.5513×10^{-3}	2.8853×10^{-2}
9.4980×10^{-3}	1.4130×10^{-2}	4.1257×10^{-2}	2.4691×10^{-2}
1.4792×10^{-2}	1.5349×10^{-2}	6.7674×10^{-2}	1.8675×10^{-2}
2.1107×10^{-2}	1.9741×10^{-2}	9.6363×10^{-2}	2.0007×10^{-2}
Estim. 9	Estim. 10	Estim. 11	\hat{m}
0.0000×10^0	0.0000×10^0	0.0000×10^0	0.0000×10^0
2.8853×10^{-2}	8.5513×10^{-3}	1.2556×10^{-2}	1.5000×10^{-2}
2.4684×10^{-2}	2.2715×10^{-2}	3.4259×10^{-2}	1.9174×10^{-2}
1.8680×10^{-2}	3.7772×10^{-2}	5.7626×10^{-2}	2.0922×10^{-2}
2.0031×10^{-2}	5.2419×10^{-2}	8.0498×10^{-2}	2.1745×10^{-2}

TABLEAU C.7 – Biais des estimateurs pour $p_0 = 0.1, p_1 = 0.0, p_2 = 0.9$

Estim. 1	Estim. 2	Estim. 3	Estim. 4
7.0711×10^{-2}	7.0711×10^{-2}	7.0711×10^{-2}	7.0711×10^{-2}
5.1826×10^{-2}	5.9076×10^{-2}	3.6836×10^{-2}	4.0442×10^{-2}
9.8482×10^{-2}	1.1582×10^{-1}	1.1445×10^{-1}	1.2161×10^{-1}
1.2156×10^{-1}	1.2907×10^{-1}	1.7526×10^{-1}	1.7143×10^{-1}
1.2565×10^{-1}	1.2214×10^{-1}	1.8703×10^{-1}	1.7746×10^{-1}
Estim. 5	Estim. 6	Estim. 7	Estim. 8
7.0711×10^{-2}	7.0711×10^{-2}	7.0711×10^{-2}	7.0711×10^{-2}
6.5002×10^{-2}	7.6442×10^{-2}	4.3334×10^{-2}	2.7856×10^{-2}
7.9715×10^{-2}	1.0345×10^{-1}	1.1753×10^{-1}	7.4799×10^{-3}
6.3357×10^{-2}	7.2258×10^{-2}	1.0870×10^{-1}	6.5572×10^{-4}
4.9247×10^{-2}	4.6365×10^{-2}	1.1114×10^{-1}	3.7627×10^{-3}
Estim. 9	Estim. 10	Estim. 11	\hat{m}
7.0711×10^{-2}	7.0711×10^{-2}	7.0711×10^{-2}	2.0000×10^{-1}
2.7856×10^{-2}	4.3334×10^{-2}	3.2778×10^{-2}	7.8788×10^{-2}
7.4799×10^{-3}	6.7017×10^{-2}	9.5370×10^{-2}	2.1156×10^{-2}
6.5572×10^{-4}	7.9394×10^{-2}	1.5502×10^{-1}	1.8547×10^{-3}
3.7627×10^{-3}	1.5962×10^{-1}	2.4405×10^{-1}	9.5288×10^{-3}

TABLEAU C.8 – Biais des estimateurs pour $p_0 = 0.3, p_1 = 0.0, p_2 = 0.7$

Estim. 1	Estim. 2	Estim. 3	Estim. 4
2.1213×10^{-1}	2.1213×10^{-1}	2.1213×10^{-1}	2.1213×10^{-1}
1.3156×10^{-1}	1.4515×10^{-1}	9.6706×10^{-2}	1.0105×10^{-1}
2.0665×10^{-1}	2.3832×10^{-1}	2.0335×10^{-1}	2.1643×10^{-1}
2.4690×10^{-1}	2.7007×10^{-1}	2.6172×10^{-1}	2.6542×10^{-1}
2.6586×10^{-1}	2.8176×10^{-1}	2.7891×10^{-1}	2.7969×10^{-1}
Estim. 5	Estim. 6	Estim. 7	Estim. 8
2.1213×10^{-1}	2.1213×10^{-1}	2.1213×10^{-1}	2.1213×10^{-1}
1.6338×10^{-1}	1.8767×10^{-1}	1.1929×10^{-1}	1.0335×10^{-1}
2.1292×10^{-1}	2.6163×10^{-1}	2.1622×10^{-1}	5.4580×10^{-2}
2.3401×10^{-1}	2.8103×10^{-1}	1.8747×10^{-1}	2.9855×10^{-2}
2.4670×10^{-1}	2.9090×10^{-1}	2.0663×10^{-1}	2.3781×10^{-2}
Estim. 9	Estim. 10	Estim. 11	\hat{m}
2.1213×10^{-1}	2.1213×10^{-1}	2.1213×10^{-1}	6.0000×10^{-1}
1.0335×10^{-1}	1.1929×10^{-1}	9.5149×10^{-2}	2.9231×10^{-1}
5.4580×10^{-2}	1.5025×10^{-1}	1.7009×10^{-1}	1.5438×10^{-1}
2.9855×10^{-2}	1.7722×10^{-1}	2.3482×10^{-1}	8.4443×10^{-2}
2.3781×10^{-2}	1.9711×10^{-1}	2.6280×10^{-1}	4.4232×10^{-2}

TABLEAU C.9 – Biais des estimateurs pour $p_0 = 0.4, p_1 = 0.0, p_2 = 0.6$

Estim. 1	Estim. 2	Estim. 3	Estim. 4
2.8284×10^{-1}	2.8289×10^{-1}	2.8284×10^{-1}	2.8284×10^{-1}
1.6660×10^{-1}	1.8009×10^{-1}	1.2660×10^{-1}	1.2894×10^{-1}
2.2921×10^{-1}	2.6107×10^{-1}	2.0928×10^{-1}	2.2270×10^{-1}
2.6253×10^{-1}	2.8835×10^{-1}	2.5632×10^{-1}	2.6355×10^{-1}
2.8022×10^{-1}	3.0106×10^{-1}	2.7618×10^{-1}	2.8108×10^{-1}
Estim. 5	Estim. 6	Estim. 7	Estim. 8
2.8289×10^{-1}	2.8284×10^{-1}	2.8284×10^{-1}	2.8284×10^{-1}
2.0429×10^{-1}	2.3096×10^{-1}	1.5707×10^{-1}	1.4816×10^{-1}
2.5290×10^{-1}	3.0298×10^{-1}	2.3404×10^{-1}	9.1770×10^{-2}
2.8391×10^{-1}	3.3543×10^{-1}	1.9706×10^{-1}	6.2495×10^{-2}
3.0995×10^{-1}	3.6362×10^{-1}	2.1939×10^{-1}	5.3894×10^{-2}
Estim. 9	Estim. 10	Estim. 11	\hat{m}
2.8284×10^{-1}	2.8284×10^{-1}	2.8284×10^{-1}	8.0000×10^{-1}
1.4816×10^{-1}	1.5707×10^{-1}	1.2936×10^{-1}	4.1905×10^{-1}
9.1770×10^{-2}	1.7629×10^{-1}	1.7754×10^{-1}	2.5957×10^{-1}
6.2495×10^{-2}	1.9802×10^{-1}	2.3006×10^{-1}	1.7676×10^{-1}
5.3894×10^{-2}	2.1479×10^{-1}	2.5793×10^{-1}	1.2603×10^{-1}

TABLEAU C.10 – Biais des estimateurs pour $p_0 = 0.2, p_1 = 0.1, p_2 = 0.7$

Estim. 1	Estim. 2	Estim. 3	Estim. 4
1.2500×10^{-1}	1.2500×10^{-1}	1.2500×10^{-1}	1.2500×10^{-1}
8.7729×10^{-2}	1.0496×10^{-1}	7.3566×10^{-2}	8.3856×10^{-2}
1.3738×10^{-1}	1.6467×10^{-1}	1.4928×10^{-1}	1.5782×10^{-1}
1.6940×10^{-1}	1.8957×10^{-1}	1.8982×10^{-1}	1.9061×10^{-1}
1.8348×10^{-1}	1.9536×10^{-1}	1.9652×10^{-1}	1.9440×10^{-1}
Estim. 5	Estim. 6	Estim. 7	Estim. 8
1.2500×10^{-1}	1.2500×10^{-1}	1.2500×10^{-1}	1.2500×10^{-1}
9.7717×10^{-2}	1.2108×10^{-1}	7.9153×10^{-2}	5.8976×10^{-2}
1.1654×10^{-1}	1.5647×10^{-1}	1.1316×10^{-1}	3.8641×10^{-2}
1.2107×10^{-1}	1.5957×10^{-1}	1.0532×10^{-1}	5.8184×10^{-2}
1.1964×10^{-1}	1.5283×10^{-1}	1.1973×10^{-1}	7.3960×10^{-2}
Estim. 9	Estim. 10	Estim. 11	\hat{m}
1.2500×10^{-1}	1.2500×10^{-1}	1.2500×10^{-1}	3.7500×10^{-1}
5.8976×10^{-2}	7.9153×10^{-2}	6.8714×10^{-2}	1.7952×10^{-1}
3.9516×10^{-2}	9.9715×10^{-2}	1.2926×10^{-1}	8.4764×10^{-2}
6.0310×10^{-2}	1.1769×10^{-1}	1.7182×10^{-1}	3.5994×10^{-2}
7.5001×10^{-2}	1.3177×10^{-1}	1.8675×10^{-1}	9.4240×10^{-3}

TABLEAU C.11 – Biais des estimateurs pour $p_0 = 0.1, p_1 = 0.2, p_2 = 0.7$

Estim. 1	Estim. 2	Estim. 3	Estim. 4
5.7198×10^{-2}	5.7198×10^{-2}	5.7198×10^{-2}	5.7198×10^{-2}
4.9372×10^{-2}	7.0813×10^{-2}	5.2228×10^{-2}	6.7047×10^{-2}
7.4349×10^{-2}	9.4450×10^{-2}	9.6483×10^{-2}	9.9491×10^{-2}
8.7464×10^{-2}	9.8498×10^{-2}	1.1241×10^{-1}	1.0743×10^{-1}
8.6985×10^{-2}	8.8837×10^{-2}	1.0404×10^{-1}	9.7216×10^{-2}
Estim. 5	Estim. 6	Estim. 7	Estim. 8
5.7198×10^{-2}	5.7198×10^{-2}	5.7198×10^{-2}	5.7198×10^{-2}
4.4823×10^{-2}	6.8840×10^{-2}	4.2359×10^{-2}	2.2507×10^{-2}
4.3403×10^{-2}	7.2700×10^{-2}	4.3896×10^{-2}	2.7138×10^{-2}
3.4518×10^{-2}	5.5718×10^{-2}	2.8136×10^{-2}	4.5447×10^{-2}
2.8385×10^{-2}	3.3532×10^{-2}	3.0725×10^{-2}	5.4616×10^{-2}
Estim. 9	Estim. 10	Estim. 11	\hat{m}
5.7198×10^{-2}	5.7198×10^{-2}	5.7198×10^{-2}	1.7778×10^{-1}
2.2507×10^{-2}	4.2359×10^{-2}	4.6768×10^{-2}	7.0867×10^{-2}
2.8170×10^{-2}	4.9069×10^{-2}	8.4122×10^{-2}	2.0920×10^{-2}
4.6965×10^{-2}	5.0371×10^{-2}	9.9578×10^{-2}	3.2635×10^{-3}
5.4136×10^{-2}	5.1822×10^{-2}	9.8671×10^{-2}	1.4227×10^{-2}

TABLEAU C.12 – Biais des estimateurs pour $p_0 = 0.1, p_1 = 0.7, p_2 = 0.2$

Estim. 1	Estim. 2	Estim. 3	Estim. 4
5.7198×10^{-2}	5.7198×10^{-2}	5.7198×10^{-2}	5.7198×10^{-2}
4.3401×10^{-2}	5.3368×10^{-2}	3.7905×10^{-2}	4.4325×10^{-2}
1.6936×10^{-2}	2.2644×10^{-2}	6.6796×10^{-3}	4.5205×10^{-3}
1.5507×10^{-2}	9.9308×10^{-3}	4.6922×10^{-2}	4.9342×10^{-2}
3.7923×10^{-2}	3.7522×10^{-2}	8.2648×10^{-2}	8.8641×10^{-2}
Estim. 5	Estim. 6	Estim. 7	Estim. 8
5.7198×10^{-2}	5.7198×10^{-2}	5.7198×10^{-2}	5.7198×10^{-2}
4.6976×10^{-2}	5.9599×10^{-2}	3.8744×10^{-2}	2.6386×10^{-2}
3.3559×10^{-2}	4.5493×10^{-2}	3.8435×10^{-2}	2.1530×10^{-2}
2.2420×10^{-2}	3.1093×10^{-2}	7.1794×10^{-2}	2.5374×10^{-2}
1.5244×10^{-2}	2.0307×10^{-2}	1.1248×10^{-1}	4.2684×10^{-2}
Estim. 9	Estim. 10	Estim. 11	\hat{m}
5.7198×10^{-2}	5.7198×10^{-2}	5.7198×10^{-2}	1.2222×10^{-1}
2.6386×10^{-2}	3.8744×10^{-2}	3.4809×10^{-2}	8.1965×10^{-2}
2.1805×10^{-2}	2.0528×10^{-2}	1.2816×10^{-2}	6.4706×10^{-2}
2.4868×10^{-2}	2.9022×10^{-2}	4.5136×10^{-2}	5.3390×10^{-2}
3.9847×10^{-2}	4.7567×10^{-2}	7.7180×10^{-2}	4.4834×10^{-2}

TABLEAU C.13 – Biais des estimateurs pour $p_0 = 0.1, p_1 = 0.6, p_2 = 0.3$

Estim. 1	Estim. 2	Estim. 3	Estim. 4
5.2705×10^{-2}	5.2705×10^{-2}	5.2705×10^{-2}	5.2705×10^{-2}
3.9031×10^{-2}	5.4580×10^{-2}	3.5638×10^{-2}	4.6468×10^{-2}
1.2974×10^{-2}	2.7336×10^{-2}	9.4409×10^{-3}	1.6050×10^{-2}
1.1749×10^{-2}	1.7524×10^{-2}	4.7020×10^{-2}	5.1910×10^{-2}
3.2423×10^{-2}	3.5709×10^{-2}	7.8848×10^{-2}	8.5759×10^{-2}
Estim. 5	Estim. 6	Estim. 7	Estim. 8
5.2705×10^{-2}	5.2705×10^{-2}	5.2705×10^{-2}	5.2705×10^{-2}
4.0848×10^{-2}	5.8504×10^{-2}	3.3742×10^{-2}	2.8858×10^{-2}
2.8384×10^{-2}	4.7527×10^{-2}	4.6114×10^{-2}	1.8182×10^{-2}
1.9218×10^{-2}	3.8087×10^{-2}	7.9522×10^{-2}	2.5416×10^{-2}
1.4128×10^{-2}	3.4248×10^{-2}	1.1575×10^{-1}	3.7709×10^{-2}
Estim. 9	Estim. 10	Estim. 11	\hat{m}
5.2705×10^{-2}	5.2705×10^{-2}	5.2705×10^{-2}	1.3333×10^{-1}
2.8858×10^{-2}	3.3742×10^{-2}	3.1183×10^{-2}	7.5986×10^{-2}
1.8404×10^{-2}	1.6258×10^{-2}	8.5275×10^{-3}	5.0756×10^{-2}
2.5331×10^{-2}	2.8285×10^{-2}	4.4014×10^{-2}	3.4641×10^{-2}
3.5078×10^{-2}	4.5876×10^{-2}	7.3895×10^{-2}	2.3075×10^{-2}

TABLEAU C.14 – Biais des estimateurs pour $p_0 = 0.3, p_1 = 0.4, p_2 = 0.3$

Estim. 1	Estim. 2	Estim. 3	Estim. 4
1.5152×10^{-1}	1.5152×10^{-1}	1.5152×10^{-1}	1.5152×10^{-1}
1.2011×10^{-1}	1.3109×10^{-1}	1.1307×10^{-1}	1.2090×10^{-1}
9.2436×10^{-2}	9.9572×10^{-2}	6.9469×10^{-2}	7.1900×10^{-2}
7.3081×10^{-2}	7.6951×10^{-2}	3.7829×10^{-2}	3.8447×10^{-2}
6.0029×10^{-2}	6.2650×10^{-2}	1.6958×10^{-2}	2.1344×10^{-2}
Estim. 5	Estim. 6	Estim. 7	Estim. 8
1.5152×10^{-1}	1.5152×10^{-1}	1.5152×10^{-1}	1.5152×10^{-1}
1.2656×10^{-1}	1.3735×10^{-1}	1.1770×10^{-1}	9.8412×10^{-2}
1.1289×10^{-1}	1.2310×10^{-1}	7.7533×10^{-2}	9.0342×10^{-2}
1.0673×10^{-1}	1.1760×10^{-1}	6.5341×10^{-2}	9.1369×10^{-2}
1.0558×10^{-1}	1.1904×10^{-1}	6.4791×10^{-2}	9.3548×10^{-2}
Estim. 9	Estim. 10	Estim. 11	\hat{m}
1.5152×10^{-1}	1.5152×10^{-1}	1.5152×10^{-1}	4.2857×10^{-1}
9.8412×10^{-2}	1.1770×10^{-1}	1.1140×10^{-1}	3.0561×10^{-1}
9.1386×10^{-2}	9.1401×10^{-2}	7.1857×10^{-2}	2.4338×10^{-1}
9.1884×10^{-2}	7.4067×10^{-2}	4.2703×10^{-2}	2.0387×10^{-1}
8.9931×10^{-2}	6.2336×10^{-2}	2.2248×10^{-2}	1.7611×10^{-1}

TABLEAU C.15 – Biais des estimateurs pour $p_0 = 0.2, p_1 = 0.5, p_2 = 0.3$

Estim. 1	Estim. 2	Estim. 3	Estim. 4
1.0308×10^{-1}	1.0308×10^{-1}	1.0308×10^{-1}	1.0308×10^{-1}
8.0174×10^{-2}	9.2853×10^{-2}	7.5326×10^{-2}	8.4252×10^{-2}
5.2505×10^{-2}	6.0942×10^{-2}	3.1281×10^{-2}	3.4773×10^{-2}
3.0483×10^{-2}	3.4697×10^{-2}	5.1694×10^{-3}	1.3909×10^{-2}
1.4740×10^{-2}	1.7886×10^{-2}	3.1802×10^{-2}	3.8820×10^{-2}
Estim. 5	Estim. 6	Estim. 7	Estim. 8
1.0308×10^{-1}	1.0308×10^{-1}	1.0308×10^{-1}	1.0308×10^{-1}
8.4013×10^{-2}	9.7342×10^{-2}	7.6851×10^{-2}	6.0115×10^{-2}
7.0321×10^{-2}	8.3072×10^{-2}	4.8848×10^{-2}	5.5172×10^{-2}
6.2453×10^{-2}	7.4865×10^{-2}	5.8260×10^{-2}	5.8056×10^{-2}
5.9423×10^{-2}	7.3830×10^{-2}	8.0908×10^{-2}	6.3213×10^{-2}
Estim. 9	Estim. 10	Estim. 11	\hat{m}
1.0308×10^{-1}	1.0308×10^{-1}	1.0308×10^{-1}	2.7500×10^{-1}
6.0115×10^{-2}	7.6851×10^{-2}	7.2812×10^{-2}	1.8924×10^{-1}
5.5734×10^{-2}	5.2142×10^{-2}	3.3719×10^{-2}	1.4605×10^{-1}
5.8161×10^{-2}	3.7507×10^{-2}	1.0238×10^{-2}	1.1749×10^{-1}
5.9638×10^{-2}	3.1820×10^{-2}	2.7781×10^{-2}	9.6688×10^{-2}

TABLEAU C.16 – Biais des estimateurs pour $p_0 = 0.4, p_1 = 0.1, p_2 = 0.5$

Estim. 1	Estim. 2	Estim. 3	Estim. 4
2.4037×10^{-1}	2.4037×10^{-1}	2.4037×10^{-1}	2.4037×10^{-1}
1.5657×10^{-1}	1.6981×10^{-1}	1.3126×10^{-1}	1.3628×10^{-1}
1.7997×10^{-1}	2.0743×10^{-1}	1.6065×10^{-1}	1.7215×10^{-1}
2.0013×10^{-1}	2.2603×10^{-1}	1.8508×10^{-1}	1.9448×10^{-1}
2.1340×10^{-1}	2.3626×10^{-1}	1.9819×10^{-1}	2.0562×10^{-1}
Estim. 5	Estim. 6	Estim. 7	Estim. 8
2.4037×10^{-1}	2.4037×10^{-1}	2.4037×10^{-1}	2.4037×10^{-1}
1.7747×10^{-1}	1.9888×10^{-1}	1.5297×10^{-1}	1.3976×10^{-1}
1.9380×10^{-1}	2.3296×10^{-1}	1.5156×10^{-1}	9.3753×10^{-2}
2.1264×10^{-1}	2.5738×10^{-1}	1.4008×10^{-1}	8.6409×10^{-2}
2.3046×10^{-1}	2.7957×10^{-1}	1.4944×10^{-1}	1.0520×10^{-1}
Estim. 9	Estim. 10	Estim. 11	\hat{m}
2.4037×10^{-1}	2.4037×10^{-1}	2.4037×10^{-1}	7.3333×10^{-1}
1.3976×10^{-1}	1.5297×10^{-1}	1.3228×10^{-1}	4.2431×10^{-1}
9.4195×10^{-2}	1.5160×10^{-1}	1.4553×10^{-1}	2.8832×10^{-1}
8.9241×10^{-2}	1.6036×10^{-1}	1.7068×10^{-1}	2.1408×10^{-1}
1.0727×10^{-1}	1.6947×10^{-1}	1.8672×10^{-1}	1.6698×10^{-1}

TABLEAU C.17 – Biais des estimateurs pour $p_0 = 0.3, p_1 = 0.3, p_2 = 0.4$

Estim. 1	Estim. 2	Estim. 3	Estim. 4
1.5152×10^{-1}	1.5152×10^{-1}	1.5152×10^{-1}	1.5152×10^{-1}
1.1516×10^{-1}	1.2835×10^{-1}	1.0575×10^{-1}	1.1495×10^{-1}
1.0187×10^{-1}	1.1882×10^{-1}	8.3751×10^{-2}	9.1545×10^{-2}
9.7676×10^{-2}	1.1494×10^{-1}	7.4373×10^{-2}	8.2203×10^{-2}
9.7255×10^{-2}	1.1397×10^{-1}	7.1501×10^{-2}	7.9148×10^{-2}
Estim. 5	Estim. 6	Estim. 7	Estim. 8
1.5152×10^{-1}	1.5152×10^{-1}	1.5152×10^{-1}	1.5152×10^{-1}
1.2353×10^{-1}	1.3682×10^{-1}	1.1287×10^{-1}	9.6203×10^{-2}
1.1701×10^{-1}	1.3839×10^{-1}	7.4400×10^{-2}	8.4380×10^{-2}
1.2030×10^{-1}	1.4776×10^{-1}	5.6945×10^{-2}	8.8984×10^{-2}
1.2882×10^{-1}	1.6205×10^{-1}	4.4513×10^{-2}	9.4676×10^{-2}
Estim. 9	Estim. 10	Estim. 11	\hat{m}
1.5152×10^{-1}	1.5152×10^{-1}	1.5152×10^{-1}	4.7143×10^{-1}
9.6203×10^{-2}	1.1287×10^{-1}	1.0405×10^{-1}	3.0418×10^{-1}
8.5336×10^{-2}	9.3225×10^{-2}	8.0692×10^{-2}	2.2095×10^{-1}
9.1110×10^{-2}	8.2829×10^{-2}	7.0089×10^{-2}	1.7037×10^{-1}
9.2736×10^{-2}	7.7541×10^{-2}	6.6506×10^{-2}	1.3612×10^{-1}

TABLEAU C.18 – Biais des estimateurs pour $p_0 = 0.3, p_1 = 0.2, p_2 = 0.5$

Estim. 1	Estim. 2	Estim. 3	Estim. 4
1.6320×10^{-1}	1.6320×10^{-1}	1.6320×10^{-1}	1.6320×10^{-1}
1.1620×10^{-1}	1.3159×10^{-1}	1.0281×10^{-1}	1.1237×10^{-1}
1.2765×10^{-1}	1.5268×10^{-1}	1.1754×10^{-1}	1.2774×10^{-1}
1.4256×10^{-1}	1.6652×10^{-1}	1.3179×10^{-1}	1.3985×10^{-1}
1.5356×10^{-1}	1.7472×10^{-1}	1.3924×10^{-1}	1.4534×10^{-1}
Estim. 5	Estim. 6	Estim. 7	Estim. 8
1.6320×10^{-1}	1.6320×10^{-1}	1.6320×10^{-1}	1.6320×10^{-1}
1.2664×10^{-1}	1.4516×10^{-1}	1.1261×10^{-1}	9.6535×10^{-2}
1.3217×10^{-1}	1.6580×10^{-1}	8.9218×10^{-2}	7.6998×10^{-2}
1.4516×10^{-1}	1.8601×10^{-1}	7.8171×10^{-2}	8.5991×10^{-2}
1.6031×10^{-1}	2.0702×10^{-1}	7.8457×10^{-2}	9.7110×10^{-2}
Estim. 9	Estim. 10	Estim. 11	\hat{m}
1.6320×10^{-1}	1.6320×10^{-1}	1.6320×10^{-1}	5.1429×10^{-1}
9.6535×10^{-2}	1.1261×10^{-1}	1.0082×10^{-1}	3.0140×10^{-1}
7.7972×10^{-2}	1.0615×10^{-1}	1.0686×10^{-1}	1.9826×10^{-1}
8.9200×10^{-2}	1.0942×10^{-1}	1.2086×10^{-1}	1.3875×10^{-1}
9.7312×10^{-2}	1.1491×10^{-1}	1.3005×10^{-1}	1.0019×10^{-1}

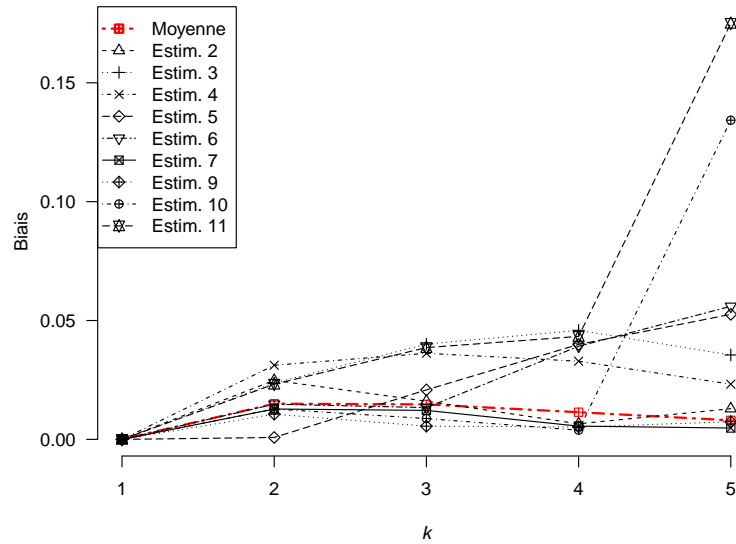
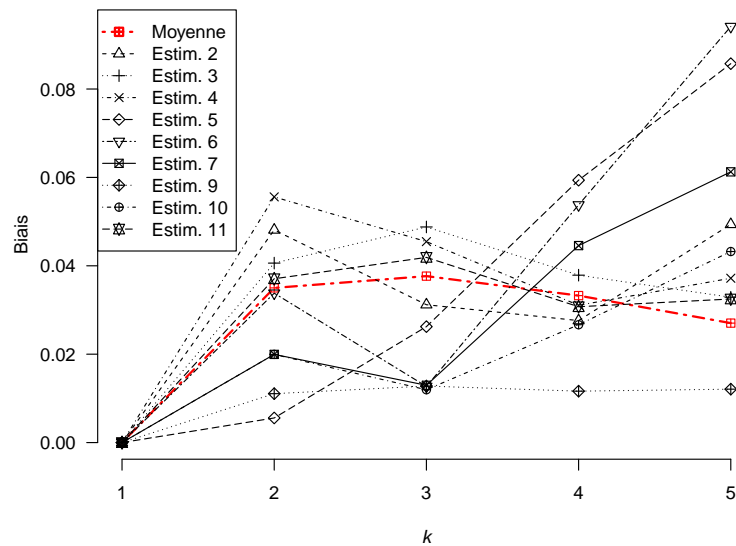
TABLEAU C.19 – Biais des estimateurs pour $p_0 = 0.2, p_1 = 0.3, p_2 = 0.5$

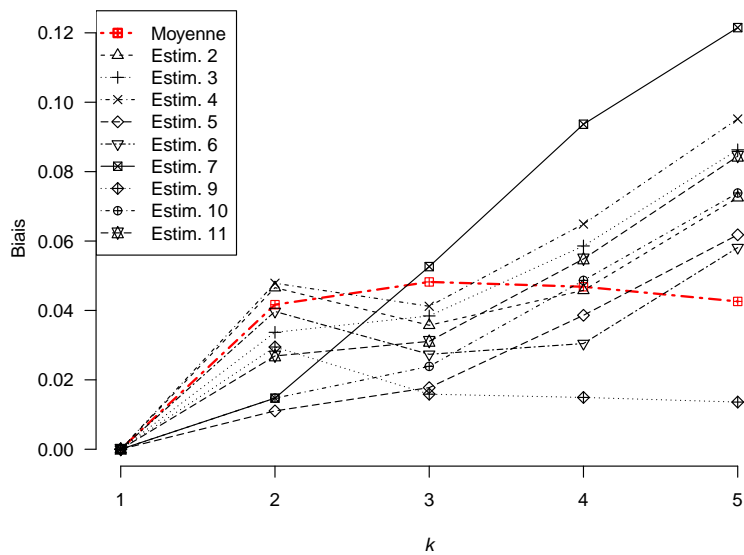
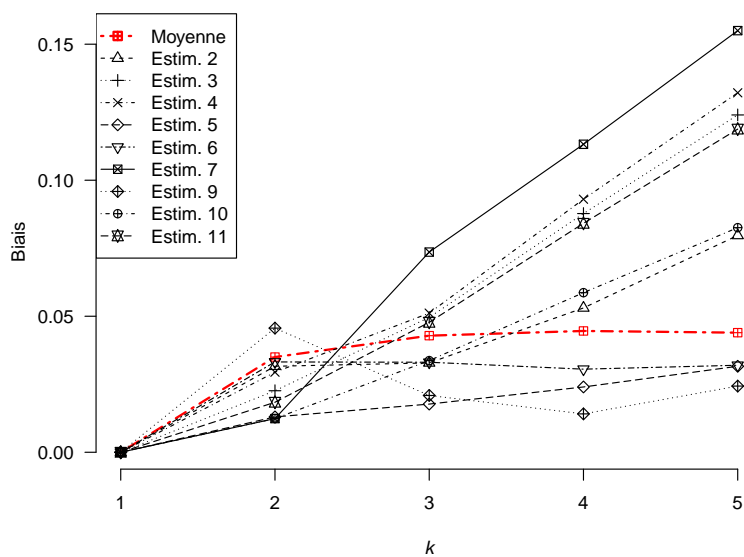
Estim. 1	Estim. 2	Estim. 3	Estim. 4
1.0308×10^{-1}	1.0308×10^{-1}	1.0308×10^{-1}	1.0308×10^{-1}
7.7660×10^{-2}	9.6169×10^{-2}	7.3209×10^{-2}	8.6528×10^{-2}
8.0363×10^{-2}	1.0285×10^{-1}	7.6892×10^{-2}	8.5492×10^{-2}
8.7069×10^{-2}	1.0698×10^{-1}	8.0726×10^{-2}	8.6275×10^{-2}
9.1758×10^{-2}	1.0818×10^{-1}	8.1162×10^{-2}	8.4783×10^{-2}
Estim. 5	Estim. 6	Estim. 7	Estim. 8
1.0308×10^{-1}	1.0308×10^{-1}	1.0308×10^{-1}	1.0308×10^{-1}
8.1198×10^{-2}	1.0004×10^{-1}	7.3428×10^{-2}	5.8697×10^{-2}
7.9797×10^{-2}	1.0961×10^{-1}	4.1296×10^{-2}	5.3171×10^{-2}
8.5296×10^{-2}	1.2054×10^{-1}	2.4378×10^{-2}	6.5964×10^{-2}
9.3929×10^{-2}	1.3353×10^{-1}	1.5050×10^{-2}	7.4436×10^{-2}
Estim. 9	Estim. 10	Estim. 11	\hat{m}
1.0308×10^{-1}	1.0308×10^{-1}	1.0308×10^{-1}	3.2500×10^{-1}
5.8697×10^{-2}	7.3428×10^{-2}	6.9485×10^{-2}	1.8380×10^{-1}
5.3881×10^{-2}	6.2175×10^{-2}	6.7503×10^{-2}	1.1266×10^{-1}
6.7778×10^{-2}	5.7797×10^{-2}	7.0178×10^{-2}	7.0148×10^{-2}
7.3267×10^{-2}	5.6602×10^{-2}	7.1761×10^{-2}	4.2569×10^{-2}

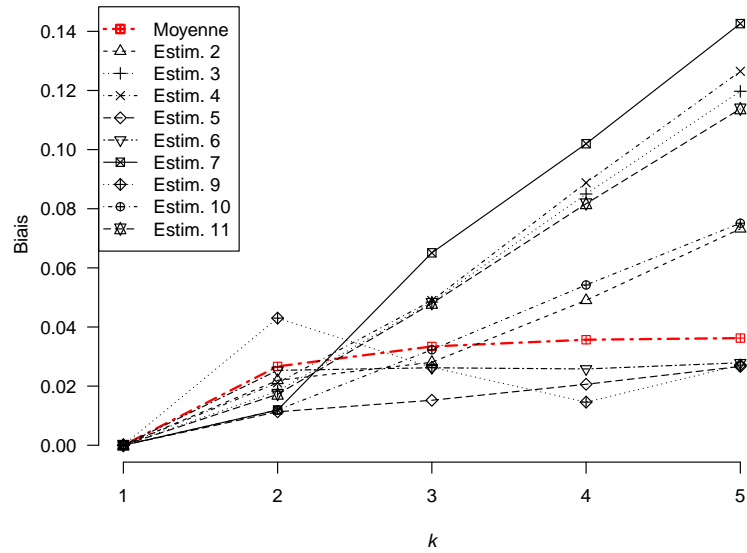
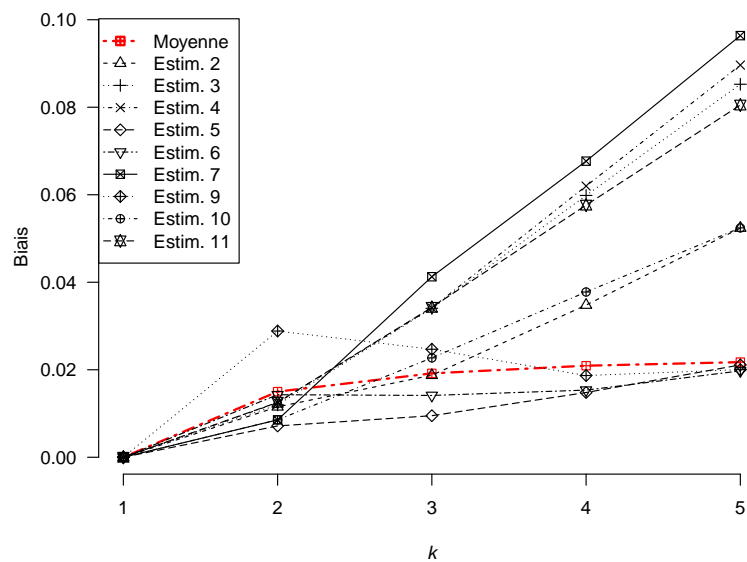
TABLEAU C.20 – Biais des estimateurs pour $p_0 = 0.1, p_1 = 0.4, p_2 = 0.5$

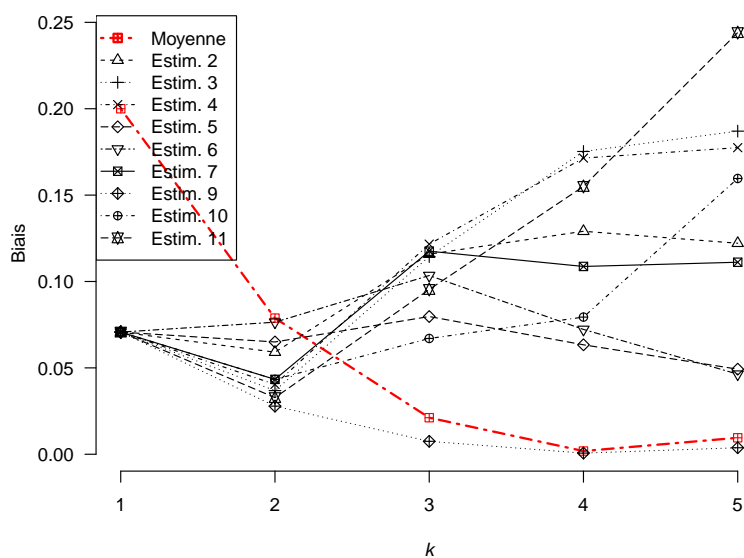
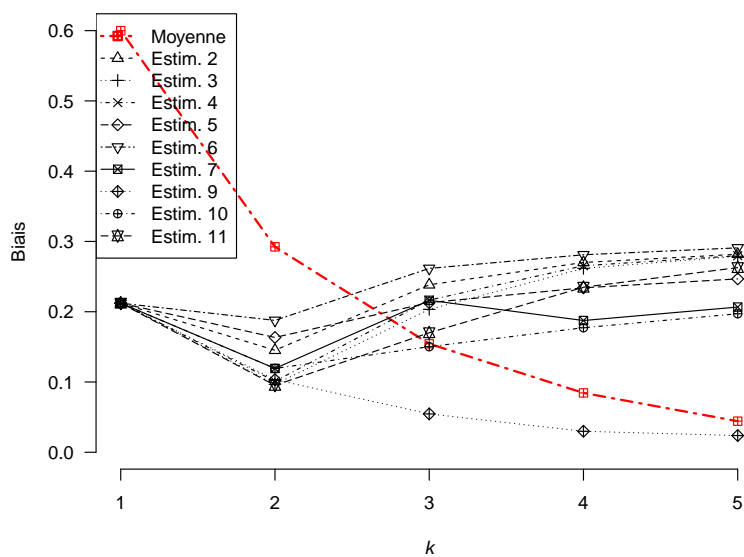
Estim. 1	Estim. 2	Estim. 3	Estim. 4
5.0308×10^{-2}	5.0308×10^{-2}	5.0308×10^{-2}	5.0308×10^{-2}
4.1473×10^{-2}	6.4717×10^{-2}	4.5060×10^{-2}	6.1594×10^{-2}
3.8177×10^{-2}	5.8515×10^{-2}	4.2931×10^{-2}	4.9562×10^{-2}
3.5765×10^{-2}	5.0413×10^{-2}	4.2248×10^{-2}	4.5590×10^{-2}
3.2723×10^{-2}	4.2744×10^{-2}	4.4119×10^{-2}	4.7097×10^{-2}
Estim. 5	Estim. 6	Estim. 7	Estim. 8
5.0308×10^{-2}	5.0308×10^{-2}	5.0308×10^{-2}	5.0308×10^{-2}
3.8017×10^{-2}	6.1748×10^{-2}	3.5014×10^{-2}	2.5554×10^{-2}
3.0694×10^{-2}	5.9372×10^{-2}	1.8510×10^{-2}	2.2740×10^{-2}
2.5573×10^{-2}	5.3948×10^{-2}	4.1867×10^{-2}	3.6044×10^{-2}
2.2388×10^{-2}	4.8652×10^{-2}	5.6995×10^{-2}	4.2924×10^{-2}
Estim. 9	Estim. 10	Estim. 11	\hat{m}
5.0308×10^{-2}	5.0308×10^{-2}	5.0308×10^{-2}	1.5556×10^{-1}
2.5554×10^{-2}	3.5014×10^{-2}	3.9251×10^{-2}	6.9786×10^{-2}
2.3175×10^{-2}	1.8994×10^{-2}	3.2047×10^{-2}	3.0639×10^{-2}
3.6514×10^{-2}	5.2353×10^{-3}	3.0149×10^{-2}	8.1931×10^{-3}
4.1513×10^{-2}	6.0647×10^{-3}	3.5037×10^{-2}	5.2876×10^{-3}

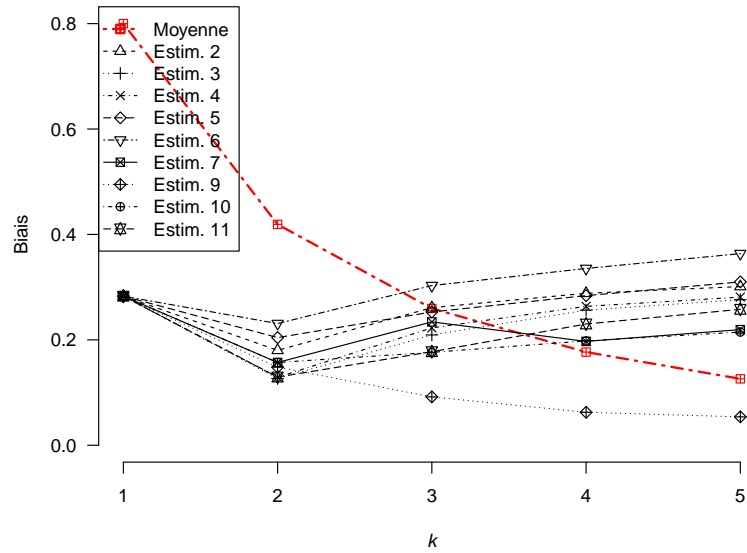
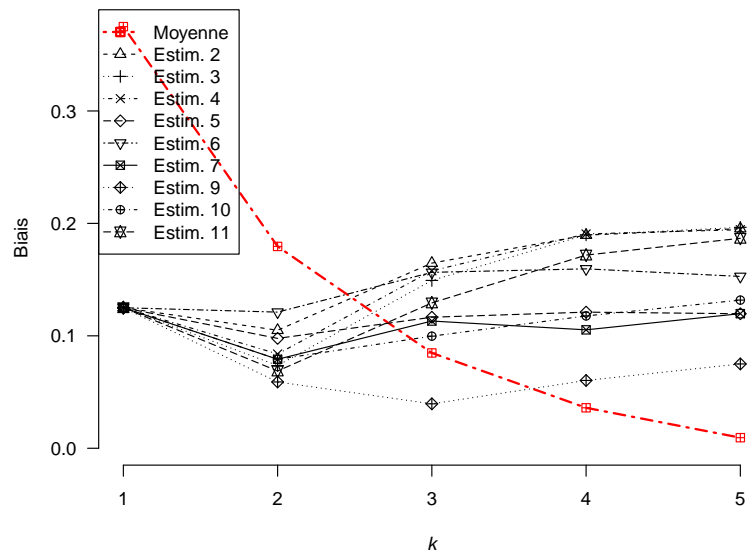
§C.3 GRAPHIQUES DES BIAIS

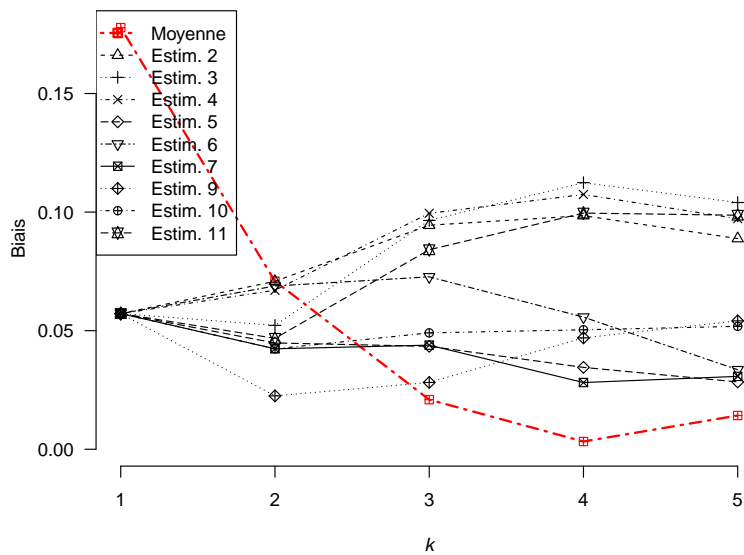
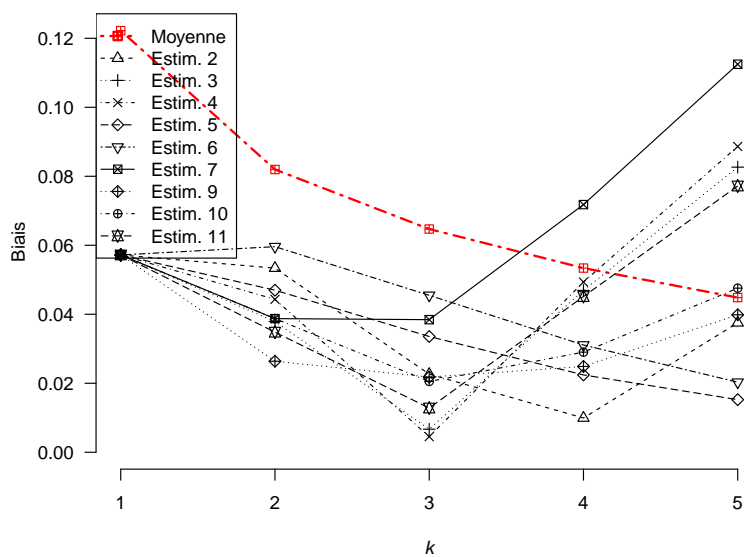
FIGURE C.1 – Biais en fonction de la taille pour $\mathbf{p} = (0.0, 0.1, 0.9)$ FIGURE C.2 – Biais en fonction de la taille pour $\mathbf{p} = (0.0, 0.3, 0.7)$

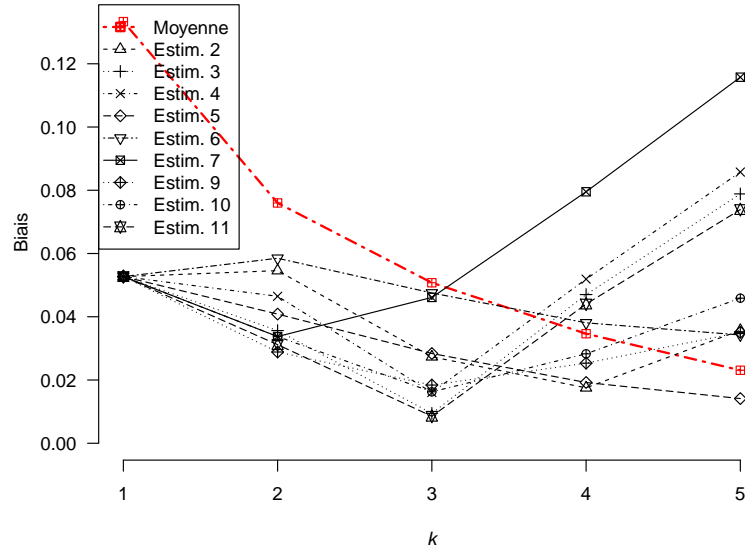
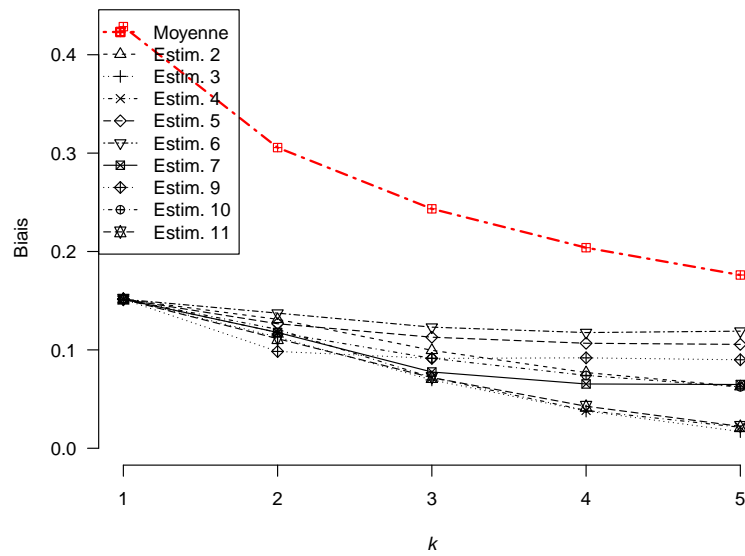
FIGURE C.3 – Biais en fonction de la taille pour $\mathbf{p} = (0.0, 0.5, 0.5)$ FIGURE C.4 – Biais en fonction de la taille pour $\mathbf{p} = (0.0, 0.7, 0.3)$

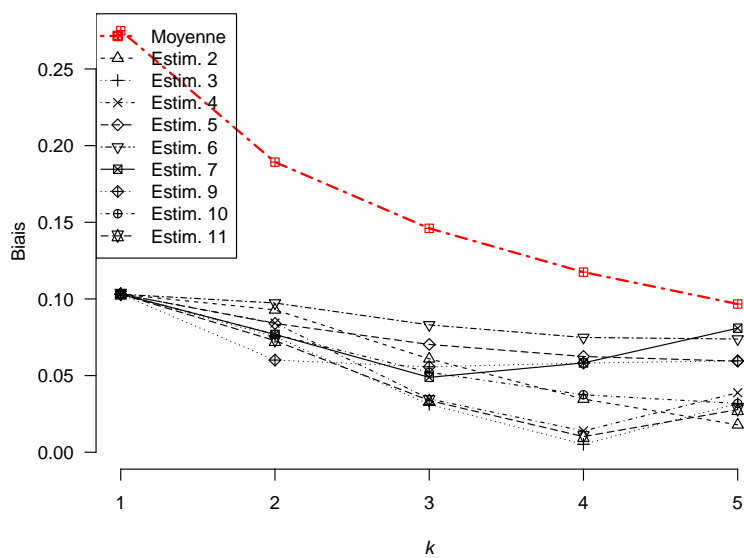
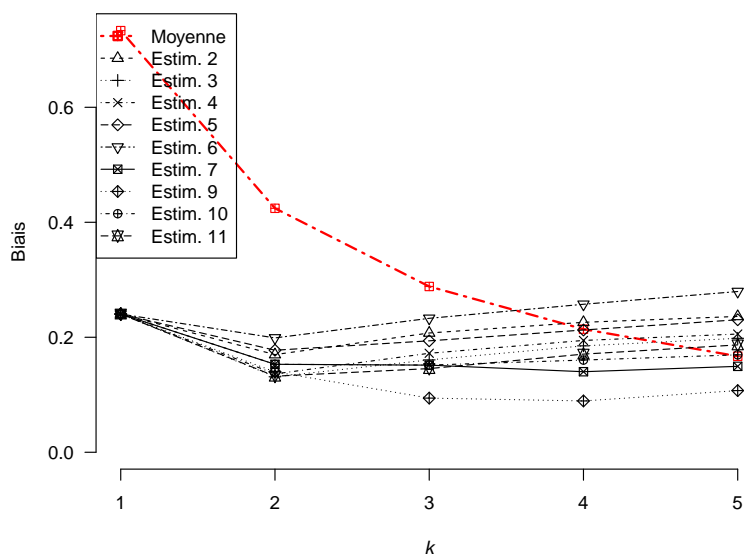
FIGURE C.5 – Biais en fonction de la taille pour $\mathbf{p} = (0.0, 0.8, 0.2)$ FIGURE C.6 – Biais en fonction de la taille pour $\mathbf{p} = (0.0, 0.9, 0.1)$

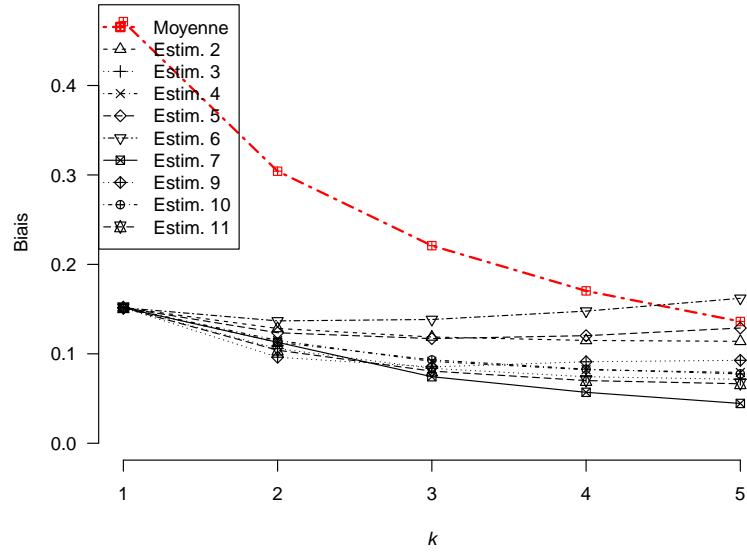
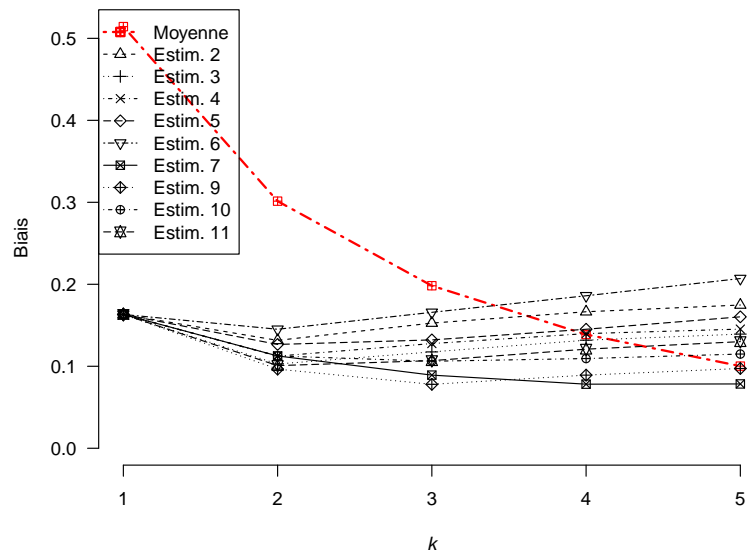
FIGURE C.7 – Biais en fonction de la taille pour $\mathbf{p} = (0.1, 0.0, 0.9)$ FIGURE C.8 – Biais en fonction de la taille pour $\mathbf{p} = (0.3, 0.0, 0.7)$

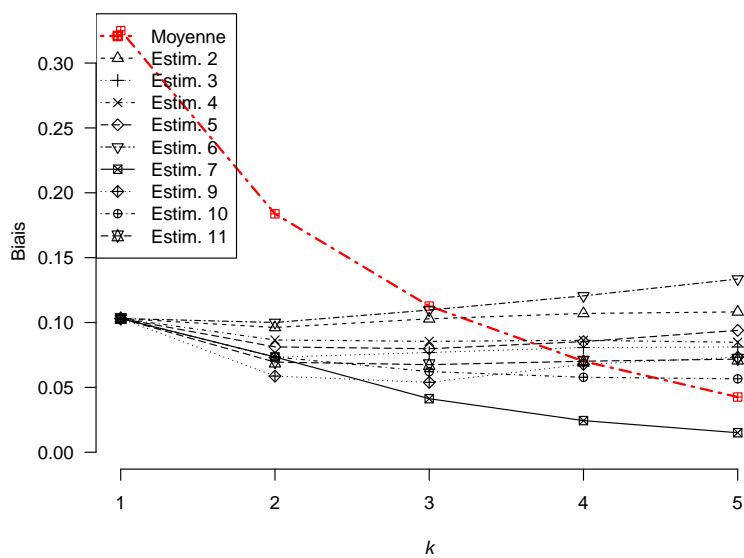
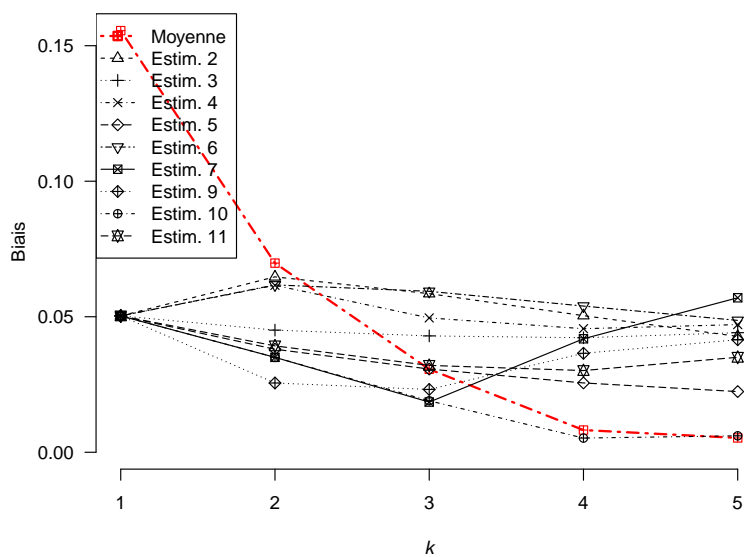
FIGURE C.9 – Biais en fonction de la taille pour $\mathbf{p} = (0.4, 0.0, 0.6)$ FIGURE C.10 – Biais en fonction de la taille pour $\mathbf{p} = (0.2, 0.1, 0.7)$

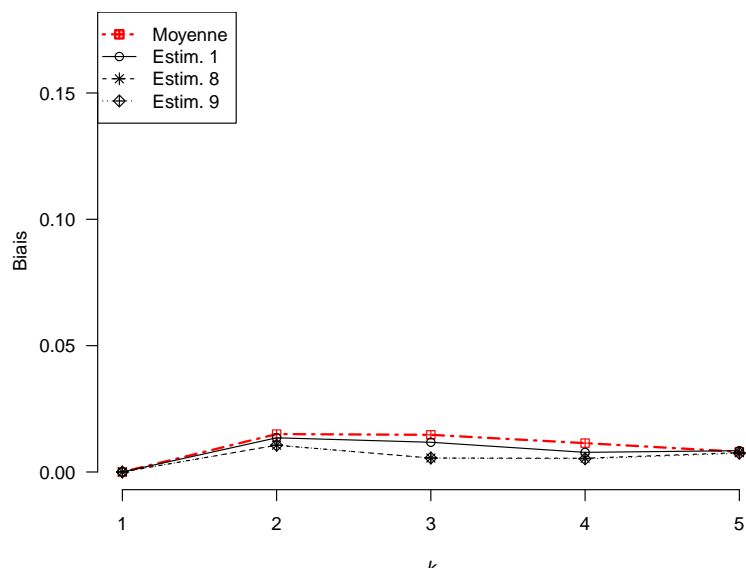
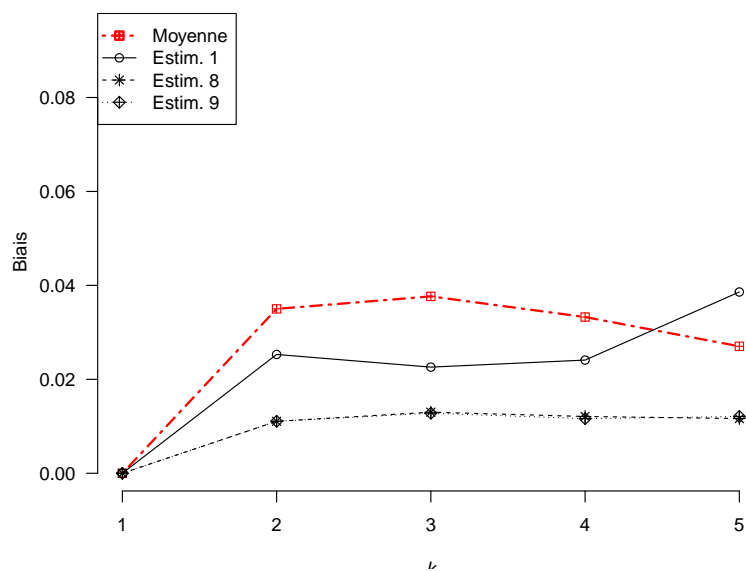
FIGURE C.11 – Biais en fonction de la taille pour $\mathbf{p} = (0.1, 0.2, 0.7)$ FIGURE C.12 – Biais en fonction de la taille pour $\mathbf{p} = (0.1, 0.7, 0.2)$

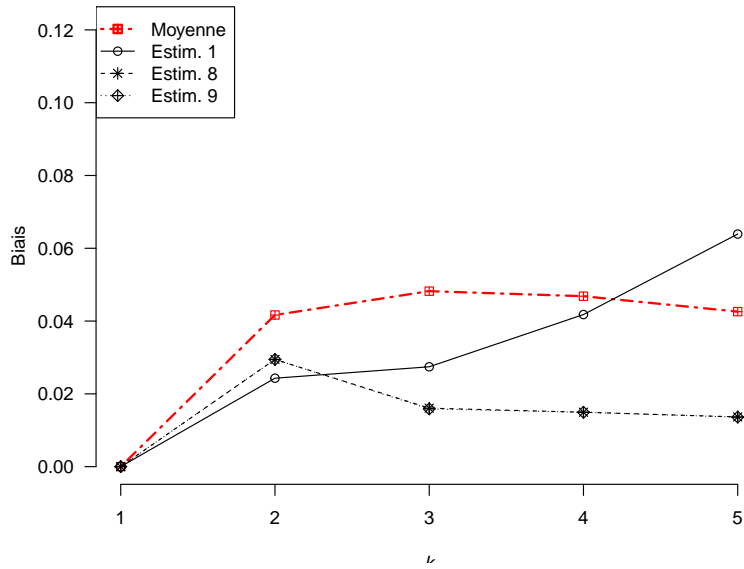
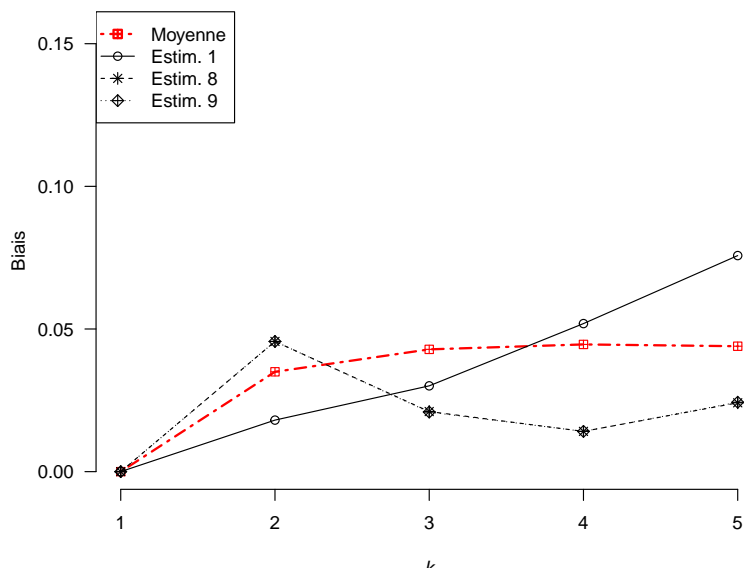
FIGURE C.13 – Biais en fonction de la taille pour $\mathbf{p} = (0.1, 0.6, 0.3)$ FIGURE C.14 – Biais en fonction de la taille pour $\mathbf{p} = (0.3, 0.4, 0.3)$

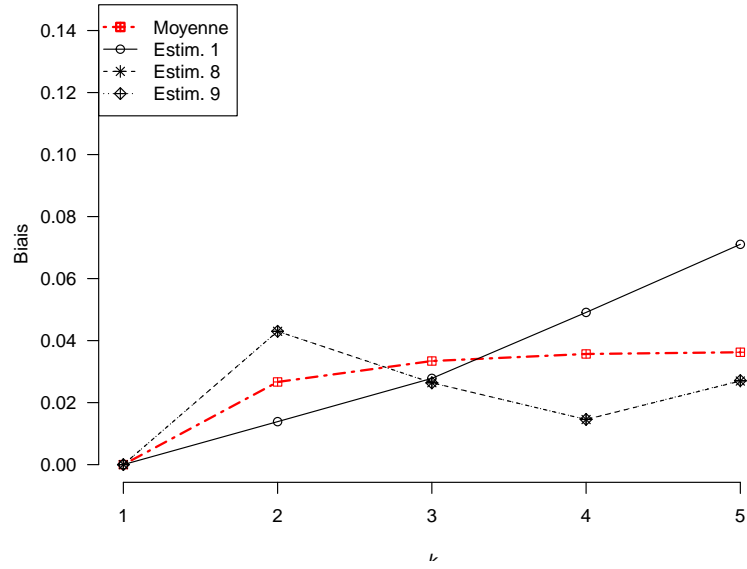
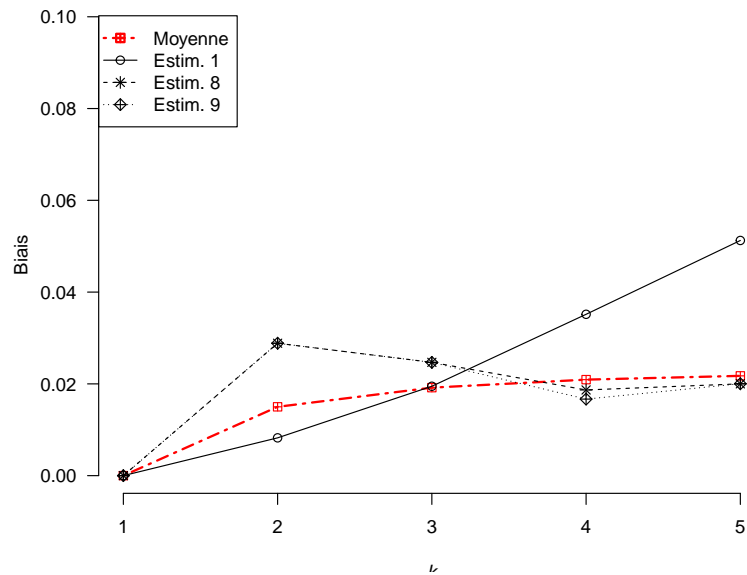
FIGURE C.15 – Biais en fonction de la taille pour $p = (0.2, 0.5, 0.3)$ FIGURE C.16 – Biais en fonction de la taille pour $p = (0.4, 0.1, 0.5)$

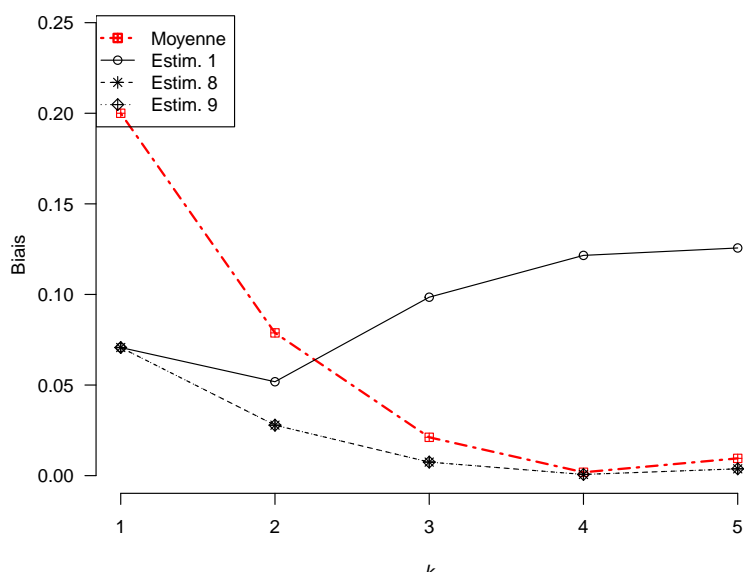
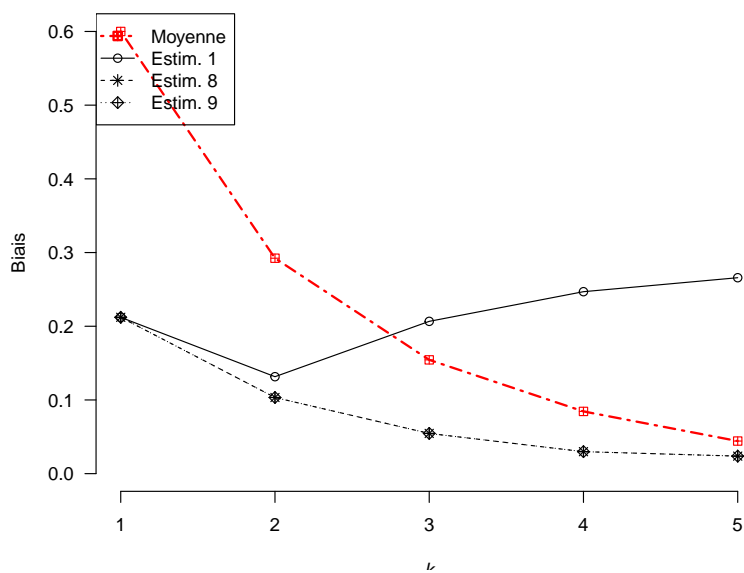
FIGURE C.17 – Biases en fonction de la taille pour $\mathbf{p} = (0.3, 0.3, 0.4)$ FIGURE C.18 – Biases en fonction de la taille pour $\mathbf{p} = (0.3, 0.2, 0.5)$

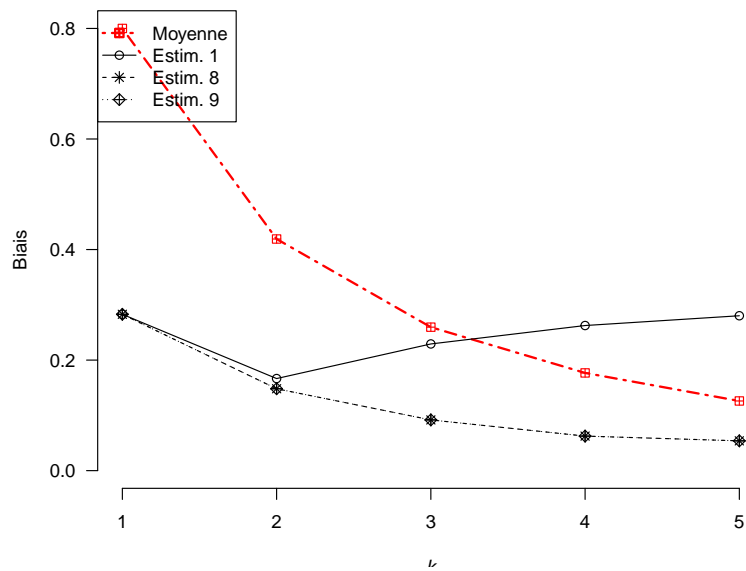
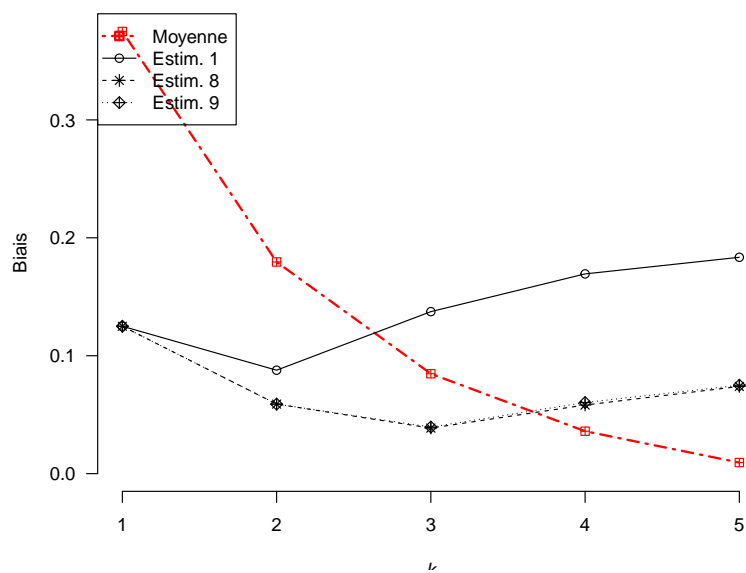
FIGURE C.19 – Biais en fonction de la taille pour $p = (0.2, 0.3, 0.5)$ FIGURE C.20 – Biais en fonction de la taille pour $p = (0.1, 0.4, 0.5)$

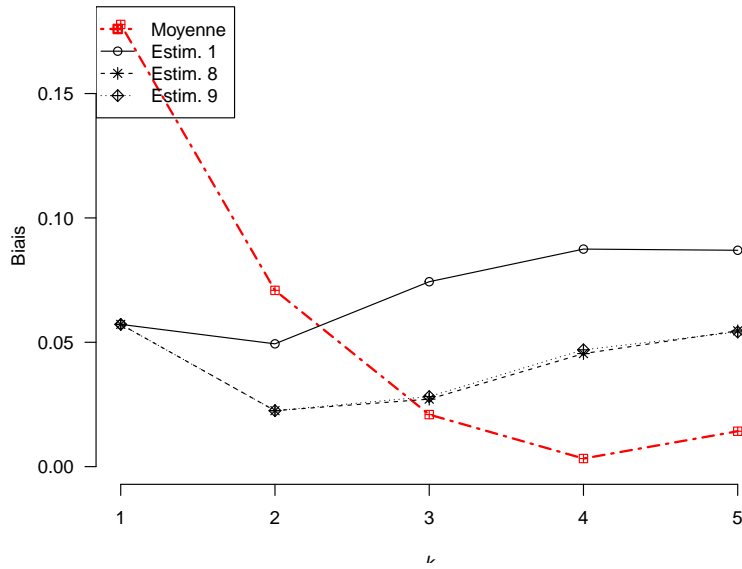
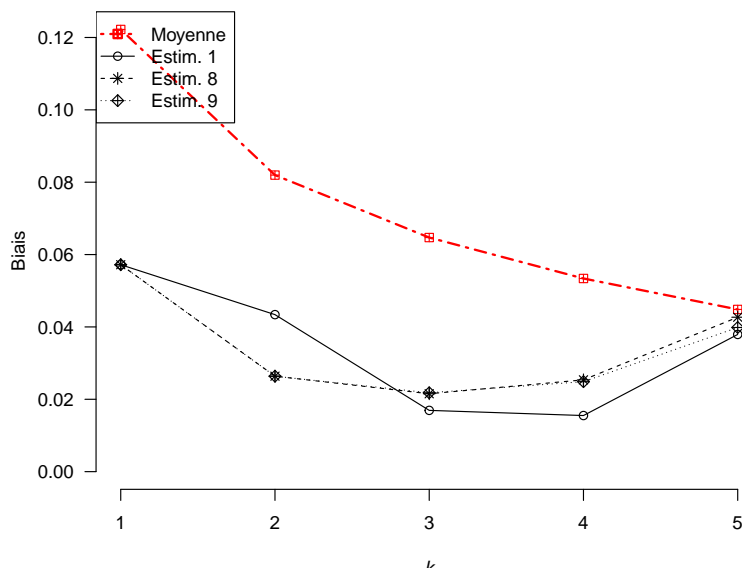
FIGURE C.21 – Biais en fonction de la taille pour $\mathbf{p} = (0.0, 0.1, 0.9)$ FIGURE C.22 – Biais en fonction de la taille pour $\mathbf{p} = (0.0, 0.3, 0.7)$

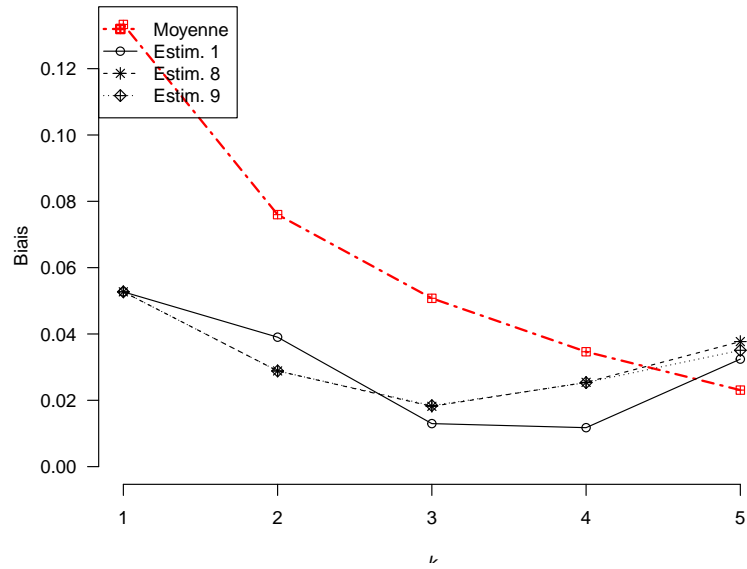
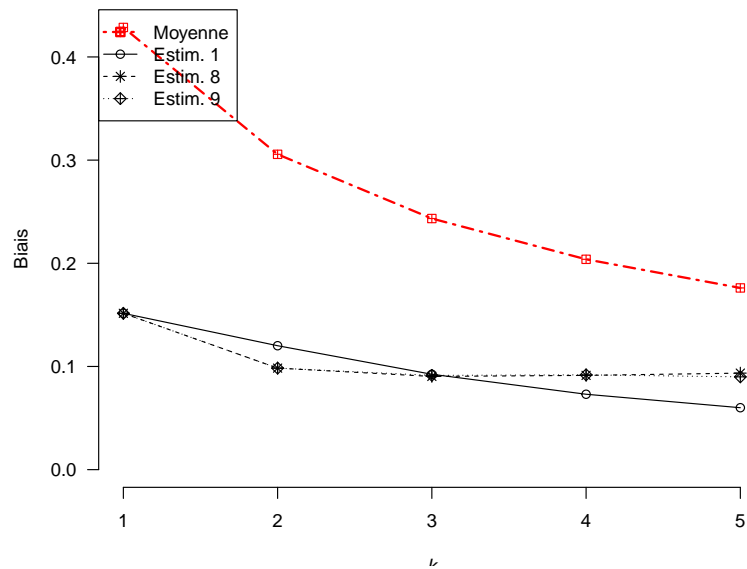
FIGURE C.23 – Biais en fonction de la taille pour $p = (0.0, 0.5, 0.5)$ FIGURE C.24 – Biais en fonction de la taille pour $p = (0.0, 0.7, 0.3)$

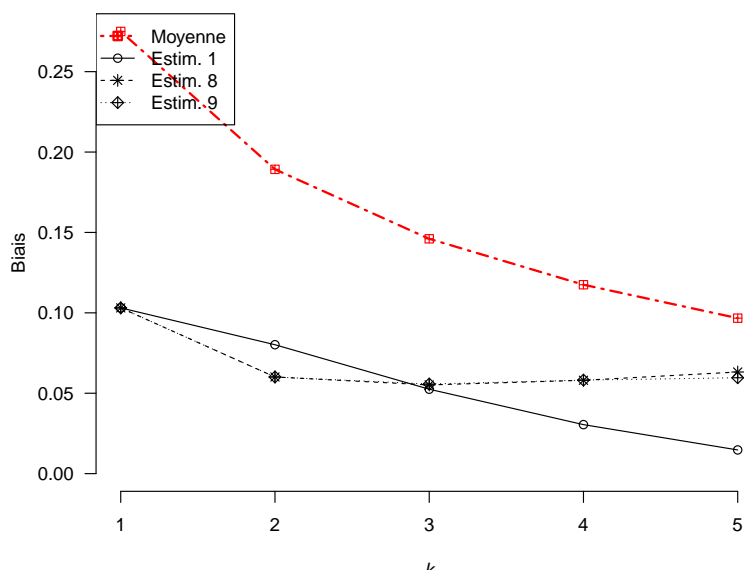
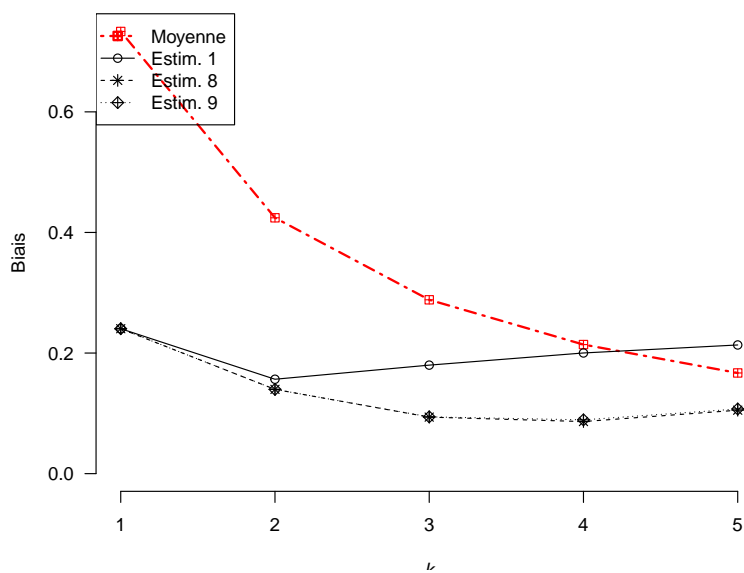
FIGURE C.25 – Biais en fonction de la taille pour $p = (0.0, 0.8, 0.2)$ FIGURE C.26 – Biais en fonction de la taille pour $p = (0.0, 0.9, 0.1)$

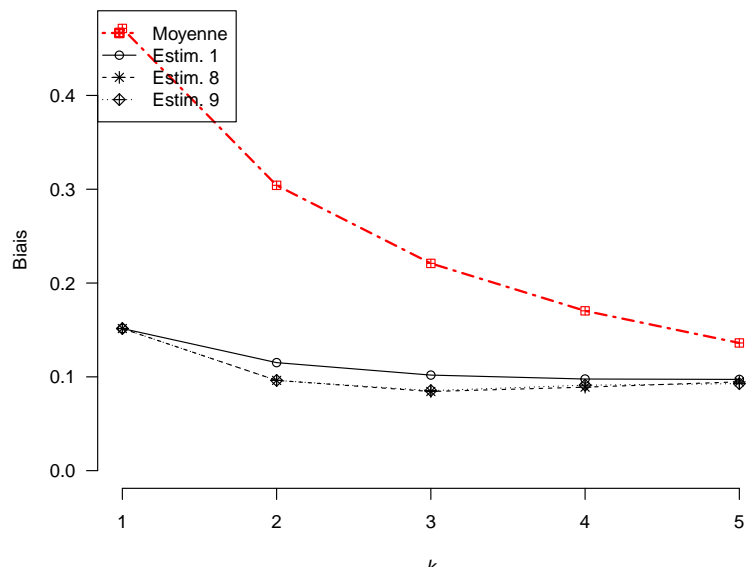
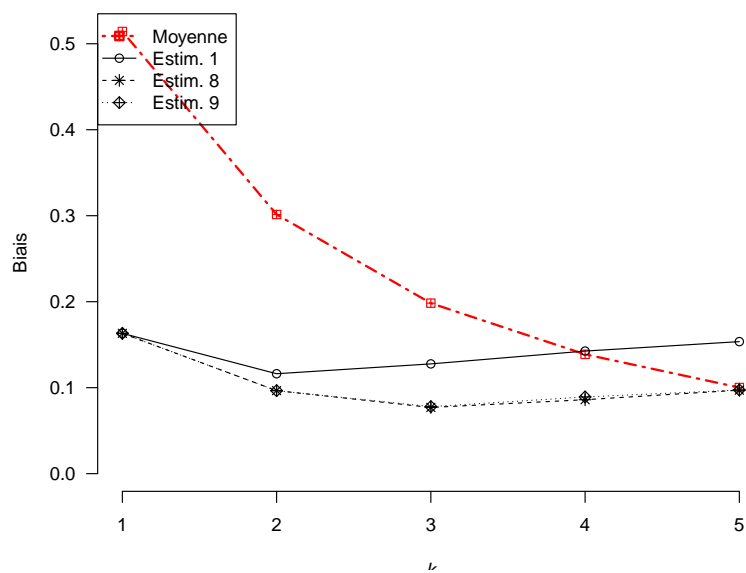
FIGURE C.27 – Biais en fonction de la taille pour $\mathbf{p} = (0.1, 0.0, 0.9)$ FIGURE C.28 – Biais en fonction de la taille pour $\mathbf{p} = (0.3, 0.0, 0.7)$

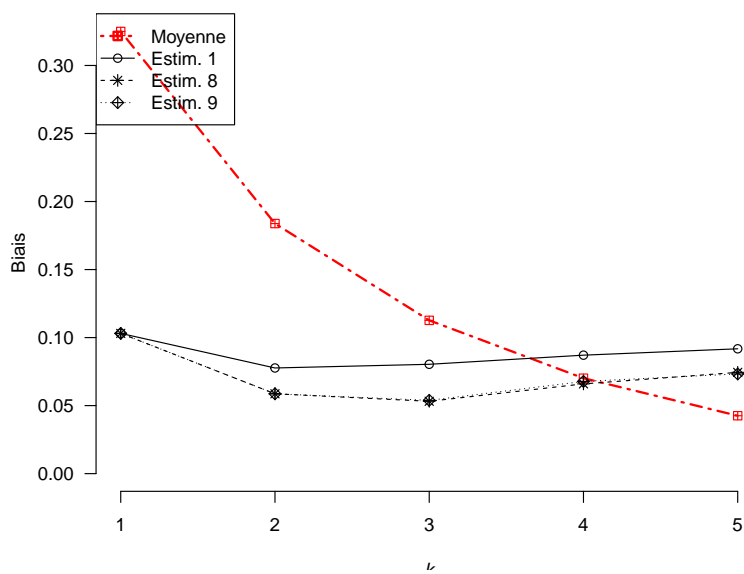
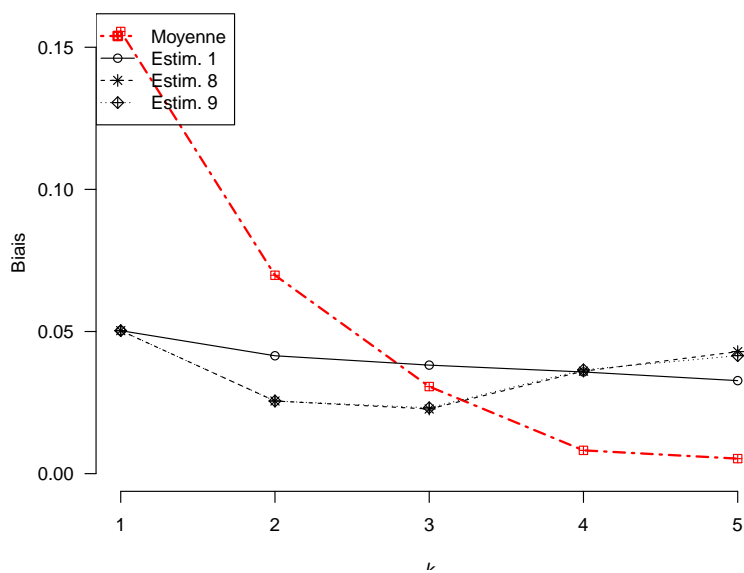
FIGURE C.29 – Biais en fonction de la taille pour $p = (0.4, 0.0, 0.6)$ FIGURE C.30 – Biais en fonction de la taille pour $p = (0.2, 0.1, 0.7)$

FIGURE C.31 – Biais en fonction de la taille pour $p = (0.1, 0.2, 0.7)$ FIGURE C.32 – Biais en fonction de la taille pour $p = (0.1, 0.7, 0.2)$

FIGURE C.33 – Biais en fonction de la taille pour $\mathbf{p} = (0.1, 0.6, 0.3)$ FIGURE C.34 – Biais en fonction de la taille pour $\mathbf{p} = (0.3, 0.4, 0.3)$

FIGURE C.35 – Biais en fonction de la taille pour $p = (0.2, 0.5, 0.3)$ FIGURE C.36 – Biais en fonction de la taille pour $p = (0.4, 0.1, 0.5)$

FIGURE C.37 – Biases en fonction de la taille pour $p = (0.3, 0.3, 0.4)$ FIGURE C.38 – Biases en fonction de la taille pour $p = (0.3, 0.2, 0.5)$

FIGURE C.39 – Biais en fonction de la taille pour $p = (0.2, 0.3, 0.5)$ FIGURE C.40 – Biais en fonction de la taille pour $p = (0.1, 0.4, 0.5)$