

# Mesures de la Biodiversité

Eric Marcon

21/12/2023



Ce document est réalisé de façon dynamique et reproductible grâce à:

- L<sup>A</sup>T<sub>E</sub>X, dans sa distribution Miktex (<http://miktex.org/>) et la classe memoir (<http://www.ctan.org/pkg/memoir>).
- R (<http://www.r-project.org/>) et RStudio (<http://www.rstudio.com/>)
- bookdown (<http://bookdown.org/>)

Son code source est sur GitHub: <https://github.com/EricMarcon/MesuresBioDiv2/>.  
Le texte mis à jour en continu peut être lu sur <https://ericmarcon.github.io/MesuresBioDiv2/>. Les versions d'étape sont déposées sur HAL: <https://hal-agroparistech.archives-ouvertes.fr/cel-01205813/>.

# Contents

<b>Contents</b>	<b>iii</b>
<b>Motivation</b>	<b>v</b>
<b>Calculs et données</b>	<b>vii</b>
<b>Notations</b>	<b>ix</b>
<b>I — Notions</b>	<b>1</b>
<b>1 Notions de Diversité</b>	<b>3</b>
1.1 Composantes . . . . .	3
Richesse . . . . .	3
Équitabilité . . . . .	4
Disparité . . . . .	4
Agrégation . . . . .	5
1.2 Niveaux de l'étude . . . . .	5
Diversité $\alpha$ , $\beta$ et $\gamma$ . . . . .	5
Décomposition . . . . .	6
1.3 Le problème de l'espèce . . . . .	7
<b>2 Outils</b>	<b>9</b>
2.1 Courbes d'accumulation . . . . .	9
2.2 Diversité asymptotique . . . . .	10
2.3 Taux de couverture . . . . .	11
Formule des fréquences de Good-Turing . . . . .	11
Taux de couverture et déficit de couverture . . . . .	12
Complétude . . . . .	14
2.4 Distribution de l'abondance des espèces (SAD) . . . . .	15
La distribution en log-séries . . . . .	16
La distribution Broken Stick . . . . .	16
La distribution log-normale . . . . .	17
La distribution géométrique . . . . .	18
Synthèse . . . . .	18
<b>Bibliography</b>	<b>23</b>



# Motivation

Le terme *biodiversity* est attribué<sup>1</sup> à Walter Rosen, un membre du *National Research Council* américain, qui a commencé à contracter les termes *biological diversity* pendant la préparation d'un colloque dont les actes seront publiés sous le titre "Biodiversity".<sup>2</sup> La question de la diversité biologique intéressait les écologues bien avant l'invention de la biodiversité, mais le néologisme a connu un succès fulgurant<sup>3</sup> en même temps qu'il devenait une notion floue, dans lequel chacun peut placer ce qu'il souhaite y trouver, au point de lui retirer son caractère scientifique.<sup>4</sup> Une cause de ce glissement est que la biodiversité a été nommée pour attirer l'attention sur son érosion, en lien avec la biologie de la conservation. Cette érosion concernant potentiellement de nombreux aspects du monde vivant, la définition de la biodiversité fluctue selon les besoins: DeLong<sup>5</sup> en recense 85 dans les dix premières années de littérature. Les indicateurs de la biodiversité peuvent englober bien d'autres choses que la diversité du vivant: le nombre d'espèces menacées (par exemple la liste rouge de l'IUCN), la taille des populations ou la surface des écosystèmes préservés, la dégradation des habitats, la menace pesant sur des espèces emblématiques... Une mesure rigoureuse et cohérente de la diversité peut pourtant être construite pour clarifier beaucoup (mais pas tous) des concepts qui constituent la biodiversité.

Dans l'introduction du premier chapitre des actes de ce qui était devenu le "Forum sur la Biodiversité", Wilson utilise le mot dans le sens étroit de nombres d'espèces. L'élargissement de la notion aux "systèmes naturels" et à l'opposé à la diversité génétique intraspécifique est venu du monde de la conservation.<sup>6</sup> La déclaration de Michel Loreau, président du comité scientifique de la conférence de Paris en 2005<sup>7</sup> en donne une définition aboutie:

La Terre abrite une extraordinaire diversité biologique, qui inclut non seulement les espèces qui habitent notre planète, mais aussi la diversité de leurs gènes, la multitude des interactions écologiques entre elles et avec leur environnement physique, et la variété des écosystèmes complexes

<sup>1</sup>C. Meine et al. "A Mission-Driven Discipline": The Growth of Conservation Biology." In: *Conservation Biology* 20.3 (2006), pp. 631–651. DOI: [10.1111/j.1523-1739.2006.00449.x](https://doi.org/10.1111/j.1523-1739.2006.00449.x).

<sup>2</sup>E. O. Wilson and F. M. Peter, eds. *Biodiversity*. Washington, D.C.: The National Academies Press, 1988.

<sup>3</sup>P. Blandin. "La Diversité Du Vivant Avant (et Après) La Biodiversité : Repères Historiques et Épistémologiques." In: *La Biodiversité En Question. Enjeux Philosophiques, Éthiques et Scientifiques*. Ed. by E. Casetta and J. Delord. Paris: Editions Matériologiques, 2014, pp. 31–68.

<sup>4</sup>J. Delord. "La Biodiversité : Imposture Scientifique Ou Ruse Épistémologique ?" In: *La Biodiversité En Question. Enjeux Philosophiques, Éthiques et Scientifiques*. Ed. by E. Casetta and J. Delord. Paris: Editions Matériologiques, 2014, pp. 83–118. DOI: [10.3917/edmat.delord.2014.01.0083](https://doi.org/10.3917/edmat.delord.2014.01.0083).

<sup>5</sup>D. C. J. DeLong. "Defining Biodiversity." In: *Wildlife Society Bulletin* 24.4 (1996), pp. 738–749. JS-TOR: [3783168](https://doi.org/10.3783168).

<sup>6</sup>J. G. Speth et al. "Foreword." In: *Global Biodiversity Strategy*. Ed. by K. Courrier. Washington, D.C.: WRI, IUCN, UNEP, 1992, pp. v–vi.

<sup>7</sup>M. Loreau. "Discours de Clôture." In: *Actes de La Conférence Internationale Biodiversité Science et Gouvernance*. Ed. by R. Barbault and J.-P. Le Duc. Paris, France.: IRD Editions, 2005, pp. 254–256.

qu'elles constituent. Cette biodiversité, qui est le produit de plus de 3 milliards d'années d'évolution, constitue un patrimoine naturel et une ressource vitale dont l'humanité dépend de multiples façons.

Aujourd'hui encore, le terme *biodiversité* concerne le plus souvent la richesse en espèces d'un écosystème. Pour simplifier la présentation, le niveau d'étude dans ce document sera en général celui des espèces.<sup>8</sup> La prise en compte de la totalité des êtres vivants est généralement hors de portée. La mesure de diversité est alors limitée à un taxocène, c'est-à-dire un sous-ensemble des espèces d'une communauté reliées taxonomiquement: les papillons, les mammifères, les arbres (la délimitation du sous-ensemble n'est pas forcément strictement taxonomique)...

Un objet privilégié des études sur la biodiversité est, depuis le Forum, la forêt tropicale parce qu'elle est très diverse et un enjeu pour la conservation. La plupart des exemples concerneront ici les arbres de la forêt tropicale, qui ont l'avantage d'être clairement définis en tant qu'individus (donc simples à compter) et posent des problèmes méthodologiques considérables pour l'estimation de leur diversité à partir de données réelles.

On peut bien évidemment s'intéresser à d'autres niveaux et d'autres objets, par exemple la diversité génétique (en termes d'allèles différents pour certains gènes ou marqueurs) à l'intérieur d'une population, ou même la diversité des interactions entre espèces d'une communauté.<sup>9</sup> On gardera toujours à l'esprit que la prise en compte de la diversité spécifique n'est pas la seule approche, les méthodes présentées ici s'appliquent à la mesure de la diversité en général, pas même nécessairement biologique.

L'objectif de ce document est de traiter la mesure de la biodiversité, pas son importance en tant que telle. On se référera par exemple à Chapin et al.<sup>10</sup> pour une revue sur cette question, Cardinale et al.<sup>11</sup> pour les conséquences de l'érosion de la biodiversité sur les services écosystémiques ou Ceballos et al.<sup>12</sup> pour les propriétés autocatalytiques de la biodiversité.

La mesure de la diversité est un sujet important en tant que tel,<sup>13</sup> pour permettre de formaliser les concepts et de les appliquer à la réalité. La question est loin d'être épuisée, et fait toujours l'objet d'une recherche active et de controverses.<sup>14</sup>

<sup>8</sup>autre concept flou, J. Hey. "The Mind of the Species Problem." In: *Trends in Ecology & Evolution* 16.7 (2001), pp. 326–329. DOI: [10.1016/S0169-5347\(01\)02145-0](https://doi.org/10.1016/S0169-5347(01)02145-0).

<sup>9</sup>Z. Jizhong et al. "An Index of Ecosystem Diversity." In: *Ecological Modelling* 59 (1991), pp. 151–163. DOI: [10.1016/0304-3800\(91\)90176-2](https://doi.org/10.1016/0304-3800(91)90176-2).

<sup>10</sup>F. S. I. Chapin et al. "Consequences of Changing Biodiversity." In: *Nature* 405.6783 (2000), pp. 234–242. DOI: [10.1038/35012241](https://doi.org/10.1038/35012241).

<sup>11</sup>B. J. Cardinale et al. "Biodiversity Loss and Its Impact on Humanity." In: *Nature* 486.7401 (2012), pp. 59–67. DOI: [10.1038/nature11148](https://doi.org/10.1038/nature11148).

<sup>12</sup>G. Ceballos et al. "Biological Annihilation via the Ongoing Sixth Mass Extinction Signaled by Vertebrate Population Losses and Declines." In: *Proceedings of the National Academy of Sciences* (2017), p. 201704949. DOI: [10.1073/pnas.1704949114](https://doi.org/10.1073/pnas.1704949114).

<sup>13</sup>A. Purvis and A. Hector. "Getting the Measure of Biodiversity." In: *Nature* 405.6783 (2000), pp. 212–9. DOI: [10.1038/35012221](https://doi.org/10.1038/35012221).

<sup>14</sup>C. Ricotta. "Through the Jungle of Biological Diversity." In: *Acta Biotheoretica* 53.1 (2005), pp. 29–38. DOI: [10.1007/s10441-005-7001-6](https://doi.org/10.1007/s10441-005-7001-6).

# Calculs et données

La présentation des mesures de diversité est donnée avec un usage intensif du formalisme mathématique. La liste des notations est fournie ci-dessous, on s’y référera autant que nécessaire.

Les calculs sont réalisés dans R,<sup>15</sup> essentiellement avec le package *entropart*.<sup>16</sup> L’ensemble du code est disponible sur GitHub<sup>17</sup> où se trouvent les mises à jour de ce document<sup>18</sup>.

Les données sont souvent celles de la parcelle 6 de la forêt de Paracou (figure 1 en Guyane française,<sup>19</sup> d’une surface de 6.25 ha. Tous les arbres de plus de 10 cm de diamètre à hauteur de poitrine (DBH: *Diameter at Breast Height*) y ont été inventoriés en 2016. La position de chaque arbre, son espèce et sa surface terrière sont fournis dans le package *SpatDiv*.

D’autres exemples utilisent la parcelle forestière permanente de Barro Colorado Island, souvent abrégée BCI:<sup>20</sup> 50 ha de forêt tropicale dont les arbres de plus de 1 cm de diamètre à hauteur de poitrine (DBH: *Diameter at Breast Height*) ont été inventoriés. Le jeu de données utilisé pour les exemples est une version réduite aux arbres de plus de 10 cm disponible dans le package *vegan*,<sup>21</sup> soient 21457 arbres dans 225 espèces.

Enfin, les données fournies par le package *entropart* sous le nom “Paracou618” sont deux inventaires de l’année 2012 limités à un hectare chacun dans les parcelles 6 et 18 de la forêt de Paracou, soient 1124 arbres dans 229 espèces.

Le code R pour réaliser la figure est le suivant:

```
library("dbmss")
library("SpatDiv")
autoplot(Paracou6, labelSize = expression(paste("Surface terrière (",
  cm^2, ")")), labelColor = "Espèce")
```

<sup>15</sup>R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2022.

<sup>16</sup>E. Marcon and B. Hérault. “Entropart, an R Package to Measure and Partition Diversity.” In: *Journal of Statistical Software* 67.8 (2015), pp. 1–26. DOI: [10.18637/jss.v067.i08](https://doi.org/10.18637/jss.v067.i08).

<sup>17</sup><https://github.com/EricMarcon/MesuresBioDiv2/>

<sup>18</sup><https://ericmarcon.github.io/MesuresBioDiv2/>

<sup>19</sup>S. Gourlet-Fleury et al. *Ecology & Management of a Neotropical Rainforest. Lessons Drawn from Paracou, a Long-Term Experimental Research Site in French Guiana*. Paris: Elsevier, 2004.

<sup>20</sup>R. Condit et al. “Thirty Years of Forest Census at Barro Colorado and the Importance of Immigration in Maintaining Diversity.” In: *PLoS ONE* 7.11 (2012), e49826. DOI: [10.1371/journal.pone.0049826](https://doi.org/10.1371/journal.pone.0049826).

<sup>21</sup>J. Oksanen et al. “Vegan: Community Ecology Package.” In: (2012).

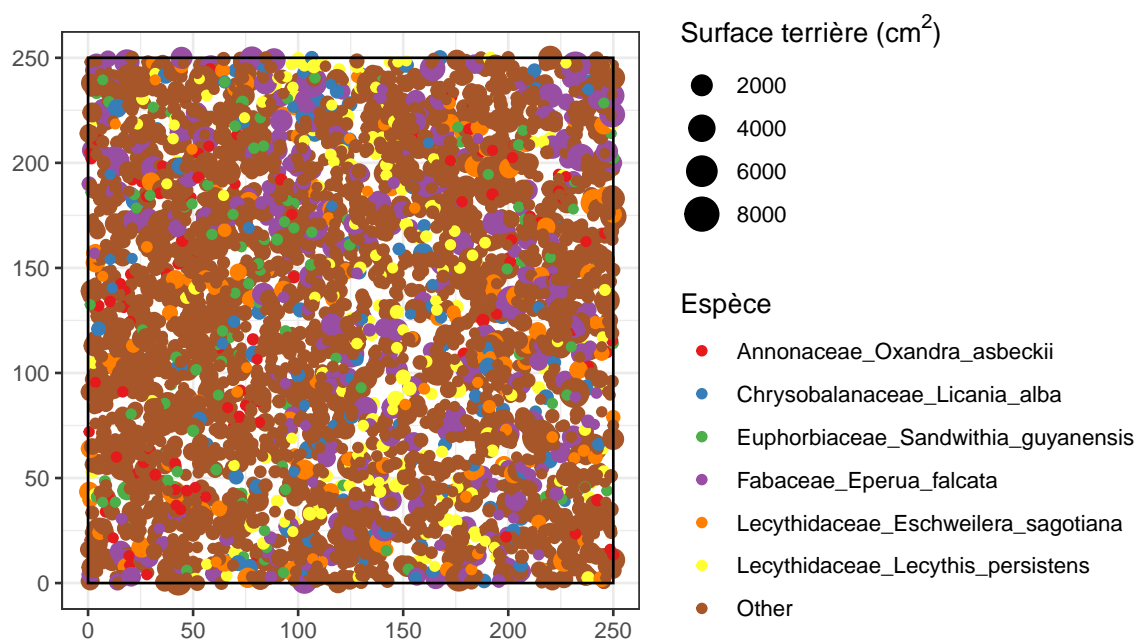


Figure 1: Carte de la parcelle 6 de Paracou. Les points représentent les arbres. Leur taille est proportionnelle à leur surface terrière. Seules les espèces les plus fréquentes sont identifiées sur la carte.



# Notations

Les notations mathématiques peuvent différer de celles de la littérature citée pour l'homogénéité de ce document.

Les matrices sont notées en caractères gras et majuscules: **X**. Les éléments de la matrice **X** sont notés  $x_{i,j}$ .

Les vecteurs sont notés en gras minuscule: **p**. Les nombres sont notés en minuscules,  $n$ , et les variables aléatoires en majuscules:  $N$ . Les valeurs maximales des énumérations font exception: elles sont notées en majuscules pour les distinguer des indices:  $\sum_{s=1}^S p_s = 1$ .

Le produit matriciel de **X** et **Y** est noté **XY**. Dans les scripts R, l'opérateur est `%*%`. Le produit de Hadamard (terme à terme) est noté **X**  $\circ$  **Y** (opérateur `*` dans R). De même **X** <sup>$n$</sup>  indique la puissance  $n$  au sens du produit matriciel d'une matrice carrée (opérateur `%^%` du package *expm*), alors que **X** <sup>$\circ n$</sup>  est la matrice dont chaque terme est celui de **X** à la puissance  $n$  (opérateur `^` de R). La matrice transposée de **X** est notée **X'**.

Les notations sont les suivantes:

**1**( $\cdot$ ): la fonction indicatrice, qui vaut 1 si la condition dans la parenthèse est vraie, 0 sinon.

**1** <sub>$s$</sub> : le vecteur de longueur  $s$  composé uniquement de 1. **1** <sub>$s$</sub> **1** <sub>$s$</sub> ' = **J** <sub>$s$</sub>  où **J** <sub>$s$</sub>  est la matrice carrée de taille  $s$  ne contenant que des 1.

$A$ : l'aire d'étude, et, selon le contexte, sa surface.

$\alpha_\nu$ : la probabilité moyenne des espèces représentées par  $\nu$  individus.

$C$ : le taux de couverture de l'échantillon, c'est-à-dire la probabilité qu'un individu de la communauté appartienne à une des espèces échantillonnées.  $C^n$  est le taux de couverture correspondant à un échantillon de taille  $n$ .

${}^qD$ : la diversité vraie (nombre de Hill pour les diversités  $\alpha$  et  $\gamma$ ), nombre équivalent de communautés pour la diversité  $\beta$ .  ${}_i^qD_\alpha$  est la diversité  $\alpha$  mesurée dans la communauté  $i$ .  ${}^q\bar{D}(T)$  est la diversité phylogénétique.

$\Delta$ : la matrice de dissimilarité dont les éléments sont  $\delta_{s,t}$ , la dissimilarité entre l'espèce  $s$  et l'espèce  $t$ .

$\mathbb{E}(X)$ : l'espérance de la variable aléatoire  $X$ .

${}^qH$ : l'entropie de Tsallis (ou HCDT).  ${}_iH_\alpha$  est l'entropie  $\alpha$  mesurée dans la communauté  $i$ . Si nécessaire, le vecteur des probabilités servant au calcul est précisé sous la forme  ${}^qH(\mathbf{p})$ .  ${}^q\bar{H}(T)$  est l'entropie phylogénétique.

$I$ : le nombre de communautés qui constituent une partition de la méta-communauté dans le cadre de la décomposition de la diversité. Les communautés sont indexées par  $i$ .

$I(p_s)$ : l'information apportée par l'observation d'un événement de probabilité  $p_s$ .  $I(q_s, p_s)$  est le gain d'information apporté par l'expérience ( $q_s$  est observé) par rapport aux probabilités  $p_s$  attendues.

$\mathbf{I}_s$ : la matrice identité de rang  $s$ : matrice carrée de taille  $s \times s$  dont la diagonale ne comporte que des 1 et les autres éléments sont nuls.

$N$ : le nombre (aléatoire) d'individus se trouvant dans l'aire d'étude.  $N_s$  est la même variable aléatoire, mais restreinte aux individus de l'espèce  $s$ .

$n$ : le nombre d'individus échantillonnés.  $n_{s,i}$  est le nombre d'individus de l'espèce  $s$  dans la communauté  $i$ . Les effectifs totaux sont  $n_{s+}$  (pour l'espèce  $s$ ),  $n_{+i}$  pour la communauté  $i$  et  $n$  le total général. S'il n'y a qu'une communauté, le nombre d'individus par espèce est  $n_s$ .

$p_s$ : la probabilité qu'un individu tiré au hasard appartienne à l'espèce  $s$ . Son estimateur,  $\hat{p}_s$  est la fréquence observée.  $p_{s|i}$  est la même probabilité dans la communauté  $i$ .

$\mathbf{p} = (p_1, p_2, \dots, p_s, \dots, p_S)$ : le vecteur décrivant la distribution des probabilités  $p_s$ , appelé simplexe en référence à sa représentation dans l'espace à  $S$  dimensions.

$\pi_\nu$ : la probabilité qu'une espèce choisie au hasard soit représentée par  $\nu$  individus,  $\sum_{\nu=1}^n \pi_\nu = 1$ . Si l'espèce est choisie explicitement, la probabilité est notée  $\pi_{n_s}$ .

${}^qR$ : l'entropie de Rényi d'ordre  $q$ .

$S$ : le nombre d'espèces, considéré comme une variable aléatoire, estimé par  $\hat{S}$ .

$S_\nu^n$ : le nombre d'espèces, considéré comme une variable aléatoire, observées  $\nu$  fois dans l'échantillonnage. L'indice est le nombre de fois où l'espèce est détectée: par exemple  $S_1$  ou  $S_{\neq 0}$ . L'exposant est la taille de l'échantillon:  $S^A$  pour la surface  $A$  ou  $\hat{S}^n$  pour un échantillon de  $n$  individus.  $S_0^A$  est le nombre d'espèces non rencontrées dans la surface  $A$ . Pour alléger les notations, s'il n'y a pas d'ambiguïté, l'indice est omis pour les espèces présentes:  $S_{\neq 0}^A$  est noté  $S^A$ . Si l'exposant n'est pas noté, l'échantillon n'est pas précisé et peut être aussi bien un nombre d'individus qu'une surface.

$s_\nu^n$ : le nombre d'espèces observées, avec les mêmes notations

que ci-dessus.  $s_\nu^n$  peut être considéré comme une réalisation de  $S_\nu^n$ .

$t_{1-\alpha/2}^n$ : le quantile d'une loi de Student à  $n$  degrés de liberté au seuil de risque  $\alpha$ , classiquement 1,96 pour  $n$  grand et  $\alpha = 5\%$ .

**Z**: la matrice de similarité entre espèces dont les éléments sont  $z_{s,t}$ , la similarité entre l'espèce  $s$  et l'espèce  $t$ .

$(\cdot)$ : la fonction gamma.

$(\cdot)$ : la fonction digamma.

$\binom{n}{k}$ : le nombre de combinaisons de  $k$  éléments parmi  $n$ :

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

.



# Part I

## Notions



# CHAPTER 1

## Notions de Diversité

### 1.1 Composantes

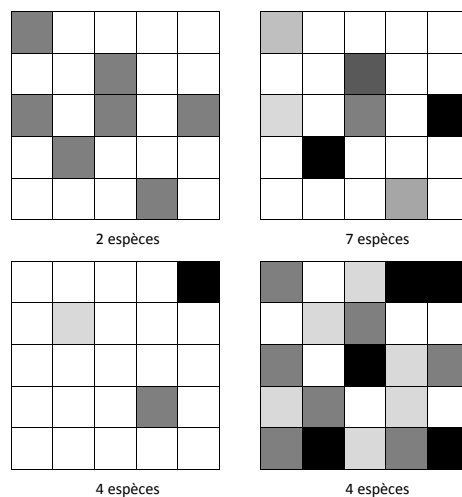


Figure 1.1: Importance de la richesse (en haut) et de l'équitabilité (en bas) pour la définition de la diversité. Ligne du haut: toutes choses égales par ailleurs, une communauté de 7 espèces semble plus diverse qu'une communauté de 2 espèces. Ligne du bas: à richesse égale, une communauté moins équitable (à gauche) semble moins diverse. Colonne de gauche: une communauté moins riche (en haut) peut sembler plus diverse si elle est plus équitable. Colonne de droite: idem pour la communauté du bas.

Une communauté comprenant beaucoup d'espèces mais avec une espèce dominante n'est pas perçue intuitivement comme plus diverse qu'une communauté avec moins d'espèces, mais dont les effectifs sont proches (figure 1.1, colonne de gauche). La prise en compte de deux composantes de la diversité, appelées richesse et équitabilité, est nécessaire.<sup>1</sup>

#### Richesse

La richesse<sup>2</sup> est le nombre (ou une fonction croissante du nombre) de classes différentes présentes dans le système étudié, par exemple le nombre d'espèces d'arbres dans une forêt.

Un certain nombre d'hypothèses sont assumées plus ou moins explicitement:

- Les classes sont bien connues: compter le nombre d'espèces a peu de sens si la taxonomie n'est pas bien

<sup>1</sup>R. H. Whittaker. "Dominance and Diversity in Land Plant Communities." In: *Science* 147.3655 (1965), pp. 250–260. DOI: [10.1126/science.147.3655.250](https://doi.org/10.1126/science.147.3655.250).

<sup>2</sup>terme introduit par R. P. McIntosh. "An Index of Diversity and the Relation of Certain Concepts to Diversity." In: *Ecology* 48.3 (1967), pp. 392–404. DOI: [10.2307/1932674](https://doi.org/10.2307/1932674).

établie. C'est parfois une difficulté majeure quand on travaille sur les microorganismes;

- Les classes sont équidistantes: la richesse augmente d'une unité quand on rajoute une espèce, que cette espèce soit proche des précédentes ou extrêmement originale.

L'indice de richesse le plus simple et le plus utilisé est tout simplement le nombre d'espèces  $S$ .

## Équitabilité

La régularité de la distribution des espèces (équitabilité en Français, *evenness* ou *equitability* en anglais) est un élément important de la diversité. Une espèce représentée abondamment ou par un seul individu n'apporte pas la même contribution à l'écosystème. Sur la figure 1.1, la ligne du bas présente deux communautés de 4 espèces, mais celle de droite est beaucoup plus équitable de celle de gauche et semble intuitivement plus diverse. À nombre d'espèces égal, la présence d'espèces très dominantes entraîne mathématiquement la rareté de certaines autres: on comprend donc assez intuitivement que le maximum de diversité sera atteint quand les espèces auront une répartition très régulière.

Un indice d'équitabilité est indépendant du nombre d'espèces (donc de la richesse).

La plupart des indices courants, comme ceux de Simpson ou de Shannon, évaluent à la fois la richesse et l'équitabilité.

## Disparité

Les mesures classiques de la diversité, dites mesures de diversité neutre (*species-neutral diversity*) ou taxonomique ne prennent pas en compte une quelconque distance entre classes. Pourtant, deux espèces du même genre sont de toute évidence plus proches que deux espèces de familles différentes. Les mesures de diversité non neutres (chapitre ??) prennent en compte cette notion, qui nécessite quelques définitions supplémentaires.<sup>3</sup>

La mesure de la différence entre deux classes est souvent une distance, mais parfois une mesure qui n'a pas toutes les propriétés d'une distance: une dissimilarité. Les mesures de *divergence*<sup>4</sup> sont construites à partir de la dissimilarité entre les classes, avec ou sans pondération par la fréquence.

Si la divergence entre espèces est une distance évolutive comme l'âge du plus récent ancêtre commun, la diversité sera dite phylogénétique. Si c'est une distance fonctionnelle, définie par exemple dans l'espace des traits fonctionnels, la diversité sera dite fonctionnelle.

<sup>3</sup>D. Mouillot et al. "Niche Overlap Estimates Based on Quantitative Functional Traits: A New Family of Non-Parametric Indices." In: *Oecologia* 145.3 (2005), pp. 345–353. DOI: [10.1007/s00442-005-0151-z](https://doi.org/10.1007/s00442-005-0151-z); C. Ricotta. "A Semantic Taxonomy for Diversity Measures." In: *Acta Biotheoretica* 55.1 (2007), pp. 23–33. DOI: [10.1007/s10441-007-9008-7](https://doi.org/10.1007/s10441-007-9008-7).

<sup>4</sup>S. Pavoine and M. B. Bonsall. "Measuring Biodiversity to Explain Community Assembly: A Unified Approach." In: *Biological Reviews* 86.4 (2011), pp. 792–812. DOI: [10.1111/j.1469-185X.2010.00171.x](https://doi.org/10.1111/j.1469-185X.2010.00171.x).



La disparité,<sup>5</sup> divergence moyenne entre deux espèces (indépendamment des fréquences), ou de façon équivalente la longueur totale des branches d'un arbre phylogénétique, est la composante qui décrit à quel point les espèces sont différentes les unes des autres.

Les mesures de *régularité* décrivent la façon dont les espèces occupent l'espace des niches (régularité fonctionnelle) ou la régularité dans le temps et entre les clades des événements de spéciation représentés par un arbre phylogénétique. Ce concept complète celui d'équitabilité dans les mesures classiques: la diversité augmente avec la richesse, la divergence entre espèces, et la régularité (qui se réduit à l'équitabilité quand toutes les espèces sont également divergentes entre elles).

## Agrégation

À partir d'une large revue de la littérature dans plusieurs disciplines scientifiques s'intéressant à la diversité (au-delà de la biodiversité), Stirling<sup>6</sup> estime que les trois composantes, qu'il nomme *variété* (richesse), *équilibre* (équitabilité) et *disparité*, recouvrent tous les aspects de la diversité.

Stirling définit la propriété d'*agrégation* comme la capacité d'une mesure de diversité à combiner explicitement les trois composantes précédentes. Cela ne signifie pas que les composantes contribuent indépendamment les unes des autres à la diversité.<sup>7</sup>

## 1.2 Niveaux de l'étude

La diversité est classiquement estimée à plusieurs niveaux emboîtés, nommés  $\alpha$ ,  $\beta$  et  $\gamma$  par Whittaker<sup>8</sup> qui a nommé  $\alpha$  la diversité locale qu'il mesurait avec l'indice  $\alpha$  de Fisher (voir le chapitre ??) et a utilisé les lettres suivantes selon ses besoins.

### Diversité $\alpha$ , $\beta$ et $\gamma$

La diversité  $\alpha$  est la diversité locale, mesurée à l'intérieur d'un système délimité. Plus précisément, il s'agit de la diversité dans un habitat uniforme de taille fixe.

De façon générale,<sup>9</sup> la richesse spécifique diminue avec la latitude (la diversité est plus grande dans les zones tropicales, et au sein de celles-ci, quand on se rapproche de l'équateur), voir figure 1.2.<sup>10</sup> La tendance est la même pour la diversité génétique intraspécifique.<sup>11</sup> La richesse diminue avec l'altitude. Elle est généralement plus faible sur les îles, où elle décroît avec la distance au continent, source de migrations.

<sup>5</sup>B. Runnegar. "Rates and Modes of Evolution in the Mollusca." In: *Rates of Evolution*. Ed. by M. Campbell and M. F. Day. London: Allen & Unwin, 1987, pp. 39–60.

<sup>6</sup>A. Stirling. "A General Framework for Analysing Diversity in Science, Technology and Society." In: *Journal of the Royal Society, Interface* 4.15 (2007), pp. 707–719. DOI: [10.1098/rsif.2007.0213](https://doi.org/10.1098/rsif.2007.0213).

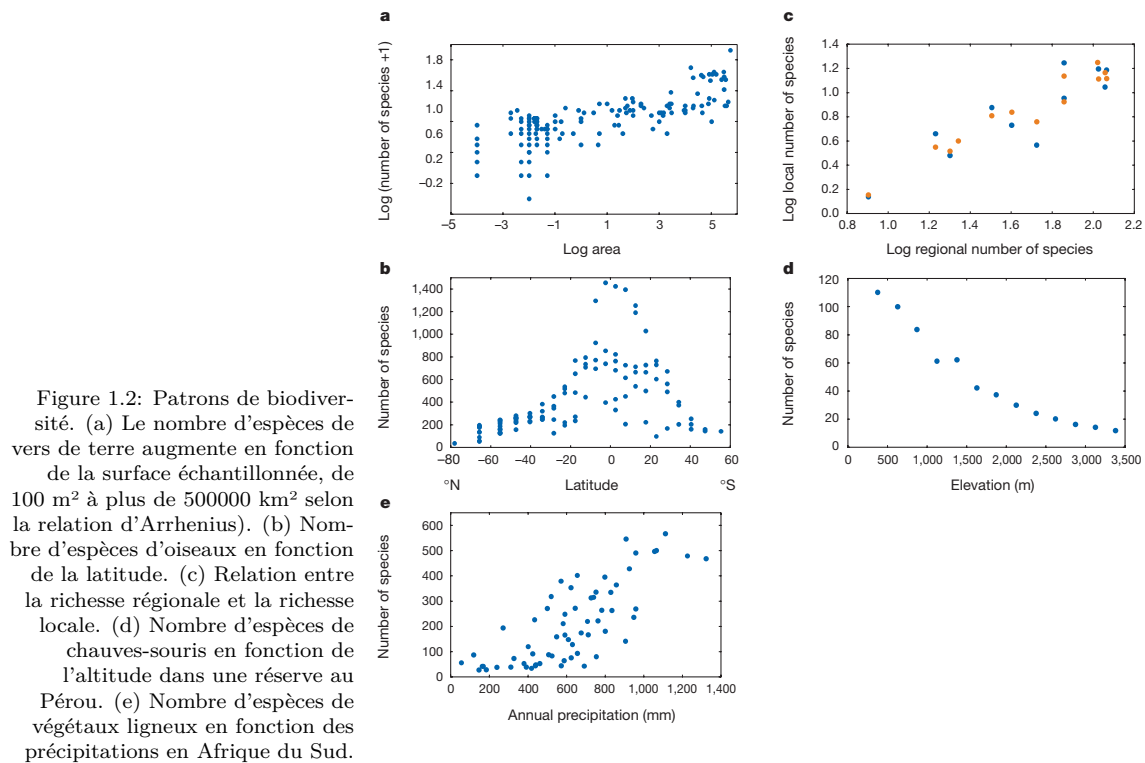
<sup>7</sup>L. Jost. "The Relation between Evenness and Diversity." In: *Diversity* 2.2 (2010), pp. 207–232. DOI: [10.3390/d2020207](https://doi.org/10.3390/d2020207).

<sup>8</sup>R. H. Whittaker. "Vegetation of the Siskiyou Mountains, Oregon and California." In: *Ecological Monographs* 30.3 (1960), pp. 279–338. DOI: [10.2307/1943563](https://doi.org/10.2307/1943563). JSTOR: [1943563](https://www.jstor.org/stable/1943563), page 320.

<sup>9</sup>K. J. Gaston. "Global Patterns in Biodiversity." In: *Nature* 405.6783 (2000), pp. 220–227. DOI: [10.1038/35012228](https://doi.org/10.1038/35012228).

<sup>10</sup>*Ibid.*, figure 1.

<sup>11</sup>A. Miraldo et al. "An Anthropocene Map of Genetic Diversity." In: *Science* 353.6307 (2016), pp. 1532–1535. DOI: [10.1126/science.aaf4381](https://doi.org/10.1126/science.aaf4381).



<sup>12</sup>C. E. Moreno and P. Rodríguez. "A Consistent Terminology for Quantifying Species Diversity?" In: *Oecologia* 163.2 (2010), pp. 279–82. DOI: [10.1007/s00442-010-1591-7](https://doi.org/10.1007/s00442-010-1591-7).

La diversité  $\beta$  mesure à quel point les systèmes locaux sont différents. Cette définition assez vague fait toujours l'objet de débats.<sup>12</sup>

Enfin, la diversité  $\gamma$  est similaire à la diversité  $\alpha$ , prise en compte sur l'ensemble du système étudié. Les diversités  $\alpha$  et  $\gamma$  se mesurent donc de la même façon.

## Décomposition

<sup>13</sup>R. H. Whittaker. "Evolution of Species Diversity in Land Communities." In: *Evolutionary Biology* 10 (1977). Ed. by M. K. Hecht et al., pp. 1–67.

Whittaker<sup>13</sup> a proposé sans succès une normalisation des échelles d'évaluation de la biodiversité, en introduisant la diversité régionale  $\varepsilon$  ( $\gamma$  étant réservé au paysage et  $\alpha$  à l'habitat) et la diversité  $\delta$  entre les paysages. Seuls les trois niveaux originaux ont été conservés par la littérature, sans définition stricte des échelles d'observation.

La distinction entre les diversités  $\alpha$  et  $\beta$  dépend de la finesse de la définition de l'habitat. La distinction de nombreux habitats diminue la diversité  $\alpha$  au profit de la  $\beta$ . Il est donc important de définir une mesure qui ne dépende pas de ce découpage, donc une mesure cumulative (additive ou multiplicative) décrivant la diversité totale, décomposable en la somme ou le produit convenablement pondérés de toutes les diversités  $\alpha$  des habitats (diversité intra) et de la diversité  $\beta$  inter-habitat.

Nous appellerons *communauté* le niveau de découpage concernant la diversité  $\alpha$  et *méta-communauté* le niveau de re-

groupement pour l'estimation de la diversité  $\gamma$ .

### 1.3 Le problème de l'espèce

Évaluer la richesse spécifique suppose que les espèces soient définies clairement, ce qui n'est de toute évidence pas le cas.<sup>14</sup> Le premier aspect du problème concerne la nature des espèces: réalité naturelle ou seulement représentation simplificatrice. Une analyse historique et philosophique en est faite par Richards.<sup>15</sup> L'autre aspect, avec des conséquences pratiques plus immédiates, concerne leur délimitation. Mayden<sup>16</sup> recense vingt-deux définitions différentes du concept d'espèce. Wilkins<sup>17</sup> estime qu'il n'y a qu'un seul concept d'espèce mais sept définitions, c'est-à-dire sept façons de les identifier, et vingt-sept variations ou mélanges de ces définitions.

La définition historique est celle de *morphoespèce*, qui classe les espèces selon leurs formes, supposées d'abord immuables. La définition moderne la plus répandue est celle d'espèce *biologique*:<sup>18</sup> un "groupe de populations naturelles isolées reproductivement les unes des autres".<sup>19</sup> Lorsque les populations d'une espèce sont isolées géographiquement, leur capacité à se reproduire ensemble est toute théorique (et rarement vérifiée expérimentalement). Des populations allopatriques n'ont pas de flux de gènes réels entre elles et peuvent être considérées comme des espèces distinctes selon la définition d'espèce *phylogénétique*: "le plus petit groupe identifiable d'individus avec un pattern commun d'ancêtres et de descendants".<sup>20</sup> C'est l'unité génétique détectée par la méthode du coalescent pour la délimitation des espèces.<sup>21</sup> Le nombre d'espèces phylogénétiques est bien supérieur au nombre d'espèces biologiques. Enfin, Van Valen<sup>22</sup> définit les espèces par la niche écologique qu'elles occupent (à partir de l'exemple des chênes blancs européens) plutôt que par les flux de gènes (permanents entre les espèces distinctes): la définition *écologique* d'espèce est proche du concept de complexe d'espèces.<sup>23</sup>

Le choix de la définition modifie considérablement sur la quantification de la richesse.<sup>24</sup> Des problèmes méthodologiques s'ajoutent aux problèmes conceptuels:<sup>25</sup> la séparation ou le regroupement de plusieurs populations ou morphotypes en un nombre plus ou moins grand d'espèces est un choix qui reflète les connaissances du moment et peut évoluer.<sup>26</sup>

L'impact du problème de l'espèce sur la mesure de la diversité reste sans solution à ce stade, si ce n'est d'utiliser les mêmes définitions si des communautés différentes doivent être comparées. L'approche phylogénétique (chapitre ??) permet de contourner le problème: si deux taxons très semblables ap-

<sup>14</sup>E. Casetta. "Évaluer et Conserver La Biodiversité Face Au Problème Des Espèces." In: *La Biodiversité En Question. Enjeux Philosophiques, Éthiques et Scientifiques*. Ed. by E. Casetta and J. Delord. Paris: Editions Matériologiques, 2014, pp. 139–154.

<sup>15</sup>R. A. Richards. *The Species Problem. A Philosophical Analysis*. Cambridge: Cambridge University Press, 2010.

<sup>16</sup>R. L. Mayden. "A Hierarchy of Species Concepts: The Denouement in the Saga of the Species Problem." In: *Species. The Units of Biodiversity*. Ed. by M. F. Claridge et al. London: Chapman and Hall, 1997, pp. 381–424.

<sup>17</sup>J. S. Wilkins. "Philosophically Speaking, How Many Species Concepts Are There?" In: *Zootaxa* 2765.1 (Feb. 15, 2011), pp. 58–60. DOI: [10.11646/zootaxa.2765.1.5](https://doi.org/10.11646/zootaxa.2765.1.5).

<sup>18</sup>T. Dobzhansky. *Genetics and the Origin of Species*. New York: Columbia University Press, 1937.

<sup>19</sup>E. Mayr. *Systematics and the Origin of Species from the Viewpoint of a Zoologist*. New York: Columbia University Press, 1942.

<sup>20</sup>J. Cracraft. "Species Concepts and Speciation Analysis." In: *Current Ornithology Volume 1*. Ed. by R. F. Johnston. Vol. 1. Current Ornithology. Springer US, 1983, pp. 159–187. DOI: [10.1007/978-1-4615-6781-3\\_6](https://doi.org/10.1007/978-1-4615-6781-3_6).

<sup>21</sup>J. Sukumaran and L. L. Knowles. "Multispecies Coalescent Delimits Structure, Not Species." In: *Proceedings of the National Academy of Sciences of the United States of America* in press (2017). DOI: [10.1073/PNAS.1607921114](https://doi.org/10.1073/PNAS.1607921114).

<sup>22</sup>L. Van Valen. "Ecological Species, Multispecies, and Oaks." In: *Taxon* 25.2/3 (1976), pp. 233–239. DOI: [10.2307/1219444](https://doi.org/10.2307/1219444).

<sup>23</sup>ensemble d'espèces voisines échangeant des gènes, J. Pernès, ed. *Gestion Des Ressources Génétiques Des Plantes. Tome 2 : Manuel*. Paris: Agence de Coopération culturelle et technique, 1984.

<sup>24</sup>P. M. Agapow et al. "The Impact of Species Concept on Biodiversity Studies." In: *The Quarterly Review of Biology* 79.2 (2004), pp. 161–179. DOI: [10.1086/383542](https://doi.org/10.1086/383542).

<sup>25</sup>Hey, see n. 8.

<sup>26</sup>A. Barberousse and S. Samadi. "La Taxonomie et Les Collections d'histoire Naturelle à l'heure de La Sixième Extinction." In: *La Biodiversité En Question. Enjeux Philosophiques, Éthiques et Scientifiques*. Ed. by E. Casetta and J. Delord. Paris: Editions Matériologiques, 2014, pp. 155–182.

portent à peine plus de diversité qu'un seul taxon, le choix de les distinguer ou non n'est pas critique.

## CHAPTER 2

# Outils



### L'essentiel

La diversité peut être décrite localement par une courbe d'accumulation (SAC) qui représente le nombre d'espèces échantillonnées en fonction de l'effort. À une échelle plus large, cette courbe s'appelle relation aire-espèces (SAR).

La distribution des abondances des espèces (SAD) est représentée par une histogramme des fréquences ou un diagramme rang-abondance.

Le taux de couverture est la somme des probabilités des espèces observées étant donné l'effort d'échantillonnage. Il peut être estimé précisément à partir des données d'inventaire. Le taux de complétude est la proportion (en nombre) des espèces observées.

Quelques outils sont nécessaires avant d'entrer dans le coeur du sujet. Les relations décrivant le nombre d'espèces en fonction de la taille de l'échantillon (relations aire-espèces) et la distribution de l'abondance des espèces sont importants pour les écologues. La mesure de l'exhaustivité de l'échantillonnage par le taux de couverture sera la base de l'estimation de la diversité à partir de données réelles.

### 2.1 Courbes d'accumulation

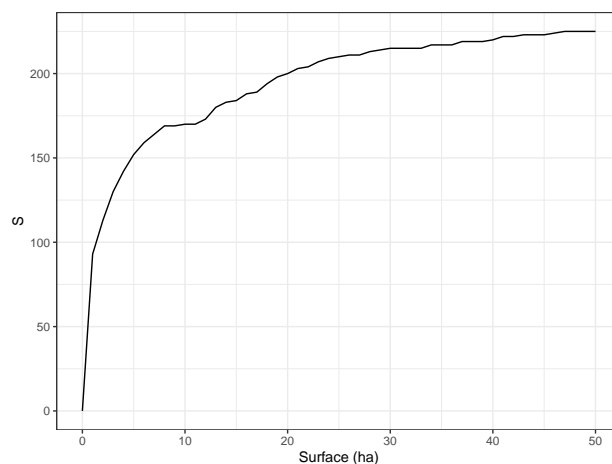


Figure 2.1: Courbe d'accumulation des espèces d'arbres du dispositif de Barro Colorado Island. Le nombre d'espèces est cumulé dans l'ordre des carrés d'un hectare du dispositif.

Evaluer la diversité d'une communauté nécessite en pratique de l'inventorier. Le nombre d'espèces découvertes en fonction de l'effort d'échantillonnage permet de tracer une courbe d'accumulation (SAC: *Species Accumulation Curve*). Une courbe de raréfaction (*Rarefaction Curve*) peut être calculée en réduisant par des outils statistiques l'effort d'échantillonnage réel pour obtenir une SAC théorique, libérée des aléas de l'ordre de prise en compte des données.

La figure 2.1 montre l'accumulation des espèces pour les données de BCI. Une SAC peut être tracée en fonction de la surface, du nombre d'individus ou du nombre de placettes d'échantillonnage, selon les besoins.

Code R pour réaliser la figure 2.1:

```
library("vegan")
data(BCI)
Cumul <- apply(BCI, 2, cumsum)
Richesse <- apply(Cumul, 1, function(x) sum(x > 0))
SARplot <- ggplot(data.frame(A = 0:50,
                             S = c(0, Richesse))) +
  aes(A, S) +
  geom_line() +
  labs(x = "Surface (ha)")
SARplot
```

Les courbes d'accumulation peuvent aussi concerner la diversité (voir le chapitre ??), mesurée au-delà du nombre d'espèces.

Plus généralement, une courbe aire-espèces (SAR: *Species Area Relationship*) représente le nombre d'espèces observées en fonction de la surface échantillonnée (figure ??). Il existe plusieurs façons de prendre en compte cette relation,<sup>1</sup> classables en deux grandes familles:<sup>2</sup>

- Dans une SAR au sens strict, chaque point représente une communauté. La question traitée est la relation entre le nombre d'espèces et la taille de chaque communauté, en lien avec des processus écologiques;
- Une courbe d'accumulation (SAC) ne représente que l'effet statistique de l'échantillonnage. Pour éviter toute confusion, le terme SAR ne doit pas être utilisé pour décrire une SAC.

<sup>1</sup>S. M. Scheiner. "Six Types of Species-Area Curves." In: *Global Ecology and Biogeography* 12.6 (2003), pp. 441–447. DOI: [10.1046/j.1466-822X.2003.00061.x](https://doi.org/10.1046/j.1466-822X.2003.00061.x).

<sup>2</sup>J. Dengler. "Which Function Describes the Species-Area Relationship Best? A Review and Empirical Evaluation." In: *Journal of Biogeography* 36.4 (2009), pp. 728–744. DOI: [10.1111/j.1365-2699.2008.02038.x](https://doi.org/10.1111/j.1365-2699.2008.02038.x).

## 2.2 Diversité asymptotique

Augmenter l'effort d'échantillonnage peut permettre d'atteindre un stade où la diversité n'augmente plus: sa valeur est appelée *diversité asymptotique*. Dans des communautés très diverses comme les forêts tropicales, la diversité asymptotique ne peut en général pas être observée sur le terrain à cause de la variabilité de l'environnement:

l'augmentation de la surface inventoriée amène à échantillonner dans des communautés différentes avant d'atteindre la diversité asymptotique de la communauté étudiée. La diversité asymptotique est donc celle d'une communauté théorique qui n'existe généralement pas. En d'autres termes, c'est la diversité d'une communauté dont l'inventaire disponible serait un échantillon représentatif.

Evaluer la diversité asymptotique nécessite d'utiliser des estimateurs de diversité, dont la précision dépend de l'exhaustivité de l'échantillonnage. La diversité peut aussi être estimée pour un effort donné: un hectare de forêt ou 5000 arbres par exemple, ou encore un taux de couverture choisi, qui décrit mieux la qualité de l'échantillonnage.

## 2.3 Taux de couverture

### Formule des fréquences de Good-Turing

La relation fondamentale entre les fréquences des espèces dans un échantillon est due à Turing et a été publiée par Good.<sup>3</sup> En absence de toute information sur la loi de distribution des espèces, en supposant seulement que les individus sont tirés indépendamment les uns des autres selon une loi multinomiale respectant ces probabilités, la formule de Good-Turing relie la probabilité moyenne  $\alpha_\nu$  d'une espèce représentée  $\nu$  fois (c'est-à-dire par  $\nu$  individus) au rapport entre les nombres d'espèces représentées  $\nu + 1$  fois et  $\nu$  fois:

$$\alpha_\nu \approx \frac{(\nu + 1)}{n} \frac{s_{\nu+1}^n}{s_\nu^n}. \quad (2.1)$$

Les singletons ( $s_1^n$ : le nombre d'espèces observées une seule fois) et les doubletons ( $s_2^n$ : le nombre d'espèces observées deux fois) ont une importance centrale. Pour  $\nu = 1$ , on a:  $\alpha_1 = 2s_2^n/(ns_1^n)$ : la fréquence d'une espèce typiquement représentée par un singleton est proportionnelle au rapport entre le nombre des doubletons et des singletons. Pour  $\nu = 0$ , l'ignorance du nombre d'espèces non échantillonnées  $s_0^n$  pose problème mais le produit  $\alpha_0 \times s_0^n = \pi_0$ , la probabilité totale des espèces non représentées, peut être estimée par  $s_1^n/n$ . Ces relations sont le fondement des estimateurs de richesse de Chao présentées ci-dessous.

La relation a été précisée<sup>4</sup> en limitant les approximations dans les calculs. La seule nécessaire est que les probabilités des espèces représentées le même nombre de fois  $\nu$  varient peu et puissent être considérées toutes égales à  $\alpha_\nu$ . Alors,  $\alpha_\nu$  est estimé par

$$\hat{\alpha}_\nu = \frac{(\nu + 1) s_{\nu+1}^n}{(n - \nu) s_\nu^n + (\nu + 1) s_{\nu+1}^n}. \quad (2.2)$$

<sup>3</sup>I. J. Good. "The Population Frequency of Species and the Estimation of Population Parameters." In: *Biometrika* 40.3/4 (1953), pp. 237–264. DOI: [10.1093/biomet/40.3-4.237](https://doi.org/10.1093/biomet/40.3-4.237).

<sup>4</sup>C.-H. Chiu et al. "An Improved Nonparametric Lower Bound of Species Richness via a Modified Good-Turing Frequency Formula." In: *Biometrics* 70.3 (2014), pp. 671–682. DOI: [10.1111/biom.12200](https://doi.org/10.1111/biom.12200). PMID: [24945937](https://pubmed.ncbi.nlm.nih.gov/24945937/), eq. 6 et 7a.

Ce nouvel estimateur est à la base de l'estimateur de Chao amélioré et des estimateurs d'entropie de Chao et Jost (sections ?? et ??).

### Taux de couverture et déficit de couverture

<sup>5</sup>Good, "The Population Frequency of Species and the Estimation of Population Parameters," see n. 3, p. 11.

Good<sup>5</sup> définit le taux de couverture (*sample coverage*) de l'échantillonnage comme la proportion des espèces découvertes,

$$C = \sum_{s=1}^S \mathbf{1}(n_s > 0) p_s, \quad (2.3)$$

où  $\mathbf{1}(\cdot)$  est la fonction indicatrice. Son complément à 1 est appelé déficit de couverture (*coverage deficit*).

Le taux de couverture augmente avec l'effort d'échantillonnage. Plus il est élevé, meilleures seront les estimations de la diversité. Pour comparer deux communautés par des courbes de raréfaction, Chao and Jost<sup>6</sup> montrent que le taux de couverture plutôt que la taille de l'échantillon doit être identique. Les estimateurs de la diversité développés plus loin reposent largement sur cette notion pour la correction du biais d'échantillonnage<sup>7</sup> (la sous-estimation systématique de la diversité due aux espèces non observées, un des éléments du biais d'estimation).

L'estimateur du taux de couverture, que Good attribue à Turing, est selon la relation des fréquences vues plus haut:

$$\hat{C} = 1 - \frac{s_1^n}{n}. \quad (2.4)$$

Cet estimateur est biaisé.<sup>8</sup> En réalité,

$$C = 1 - \frac{\mathbb{E}(S_1^n) - \pi_1}{n}. \quad (2.5)$$

L'estimateur de Good néglige le terme  $\pi_1$ , la somme des probabilités des espèces observées une fois. Ce terme peut être estimé avec un biais plus petit. Chao et al.<sup>9</sup> puis Z. Zhang and Huang<sup>10</sup> proposent l'estimateur suivant, qui utilise toute l'information disponible et a le plus petit biais possible:

$$\hat{C} = 1 - \sum_{\nu=1}^n (-1)^{\nu+1} \binom{n}{\nu}^{-1} s_{\nu}^n. \quad (2.6)$$

Les termes de la somme décroissent très vite avec  $\nu$ . En se limitant à  $\nu = 1$ , l'estimateur se réduit à celui de Good.

Esty,<sup>11</sup> complété par C.-H. Zhang and Z. Zhang,<sup>12</sup> a montré que l'estimateur était asymptotiquement normal et a calculé l'intervalle de confiance de  $\hat{C}$ :

<sup>6</sup>A. Chao and L. Jost. "Coverage-Based Rarefaction and Extrapolation: Standardizing Samples by Completeness Rather than Size." In: *Ecology* 93.12 (2012), pp. 2533–2547. DOI: [10.1890/11-1952.1](https://doi.org/10.1890/11-1952.1).

<sup>7</sup>G. Dauby and O. J. Hardy. "Sampled-Based Estimation of Diversity Sensus Stricto by Transforming Hurlbert Diversities into Effective Number of Species." In: *Ecography* 35.7 (2012), pp. 661–672. DOI: [10.1111/j.1600-0587.2011.06860.x](https://doi.org/10.1111/j.1600-0587.2011.06860.x).

<sup>8</sup>Z. Zhang and H. Huang. "Turing's Formula Revisited." In: *Journal of Quantitative Linguistics* 14.2-3 (2007), pp. 222–241. DOI: [10.1080/09296170701514189](https://doi.org/10.1080/09296170701514189).

<sup>9</sup>A. Chao et al. "A Generalized Good's Nonparametric Coverage Estimator." In: *Chinese Journal of Mathematics* 16 (1988), pp. 189–199. JSTOR: [43836340](https://www.jstor.org/stable/43836340).

<sup>10</sup>Z. Zhang and Huang, see n. 8.

<sup>11</sup>W. W. Esty. "A Normal Limit Law for a Nonparametric Estimator of the Coverage of a Random Sample." In: *The Annals of Statistics* 11.3 (1983), pp. 905–912. DOI: [10.2307/2240652](https://doi.org/10.2307/2240652). JSTOR: [2240652](https://www.jstor.org/stable/2240652).

<sup>12</sup>C.-H. Zhang and Z. Zhang. "Asymptotic Normality of a Nonparametric Estimator of Sample Coverage." In: *Annals of Statistics* 37 (5A 2009), pp. 2582–2595. DOI: [10.1214/08-aos658](https://doi.org/10.1214/08-aos658).



$$C = \hat{C} \pm t_{1-\alpha/2}^n \frac{\sqrt{s_1^n \left(1 - \frac{s_1^n}{n}\right) + 2s_2^n}}{n}. \quad (2.7)$$

Où  $t_{1-\alpha/2}^n$  est le quantile d'une loi de Student à  $n$  degrés de liberté au seuil de risque  $\alpha$ , classiquement 1,96 pour  $n$  grand et  $\alpha = 5\%$ .

Un autre estimateur est utilisé dans le logiciel SPADE<sup>13</sup> et son portage sous R, le package *spadeR*.<sup>14</sup> Il est la base des estimateurs d'entropie de Chao et Jost (section ??). L'estimation de l'équation (2.5) donne la relation

$$\hat{C} = 1 - \frac{s_1^n - \hat{\pi}_1}{n}. \quad (2.8)$$

Or,  $\hat{\pi}_1 = s_1^n \hat{\alpha}_1$ .  $\alpha_1$  peut être estimé par la relation de Good-Turing (2.2), en remplaçant  $s_0^n$  par l'estimateur Chao1 (??). Alors:

$$\hat{C} = 1 - \frac{s_1^n}{n} (1 - \hat{\alpha}_1) = 1 - \frac{s_1^n}{n} \left[ \frac{(n-1) s_1^n}{(n-1) s_1^n + 2s_2^n} \right]. \quad (2.9)$$

Dans le package *entropart*, la fonction `Coverage` calcule les trois estimateurs (celui de Zhang et Huang par défaut):

```
library("entropart")
Ns <- colSums(BCI)
Coverage(Ns)
```

```
## ZhangHuang
## 0.9991146
```

Le taux de couverture de BCI est proche de 1 parce que l'inventaire couvre 50 ha. Il est moindre sur les 6.25 ha de la parcelle 6 de Paracou:

```
library("SpatDiv")
Coverage(as.AbdVector(Paracou6))
```

```
## ZhangHuang
## 0.9723258
```

Chao and Jost<sup>15</sup> montrent que la pente de la courbe d'accumulation donnant l'espérance du nombre d'espèces en fonction du nombre d'individus (courbe de raréfaction de la figure ??) est égale au déficit de couverture,

$$1 - \mathbb{E}(C^n) = \mathbb{E}(S^{n+1}) - \mathbb{E}(S^n), \quad (2.10)$$

où  $C^n$  est le taux de couverture d'un échantillon de taille  $n$  et  $S^n$  le nombre d'espèces découvertes dans cet échantillon.

<sup>13</sup>A. Chao and T.-J. Shen. "Program SPADE: Species Prediction and Diversity Estimation. Program and User's Guide." In: (2010).

<sup>14</sup>A. Chao et al. "SpadeR: Species Prediction and Diversity Estimation with R." in: (2016).

<sup>15</sup>Chao and Jost, see n. 6.

<sup>16</sup>A. Chao and C.-W. Lin. “Non-parametric Lower Bounds for Species Richness and Shared Species Richness under Sampling without Replacement.” In: *Biometrics* 68.3 (2012), pp. 912–921. doi: [10.1111/j.1541-0420.2011.01739.x](https://doi.org/10.1111/j.1541-0420.2011.01739.x).

<sup>17</sup>W.-H. Hwang et al. “Good-Turing Frequency Estimation in a Finite Population.” In: *Biometrical journal* 57.2 (2014), pp. 321–339. doi: [10.1002/bimj.201300168](https://doi.org/10.1002/bimj.201300168).

Les estimateurs présentés ici supposent une population de taille infinie (de façon équivalente, les individus sont tirés avec remise). Le cas des populations de taille finie est traité par Chao and Lin<sup>16</sup> et Hwang et al.<sup>17</sup>

## Complétude

La complétude de l’échantillonnage est la proportion du nombre d’espèces observées:  $s_{\neq 0}^n/S$ . Elle compte simplement les espèces et ne doit pas être confondue avec la couverture qui somme leurs probabilités: le taux de complétude est toujours très inférieur au taux de couverture parce que les espèces non échantillonnées sont les plus rares.

La complétude du même échantillon d’arbres de forêt tropicale que dans l’exemple précédent peut être estimée en divisant le nombre d’espèces observées par le nombre d’espèces estimées (voir section ??). À BCI:

```
# Espèces observées
(Obs <- Richness(Ns, Correction = "None"))

## None
## 225

# Richesse estimée
(Est <- Richness(Ns, Correction = "Jackknife"))

## Jackknife 1
## 244

# Complétude
as.numeric(Obs/Est)

## [1] 0.9221311
```

À Paracou:

```
# Espèces observées
(Obs <- Richness(Paracou6, Correction = "None"))

## None
## 334

# Richesse estimée
(Est <- Richness(Paracou6, Correction = "Jackknife"))

## Jackknife 2
## 471

# Complétude
as.numeric(Obs/Est)

## [1] 0.7091295
```

## 2.4 Distribution de l'abondance des espèces (SAD)

La distribution de l'abondance des espèces (SAD: *Species Abundance Distribution*) est la loi statistique qui donne l'abondance attendue de chaque espèce d'une communauté. Les espèces ne sont pas identifiées individuellement, mais par le nombre d'individus leur appartenant.

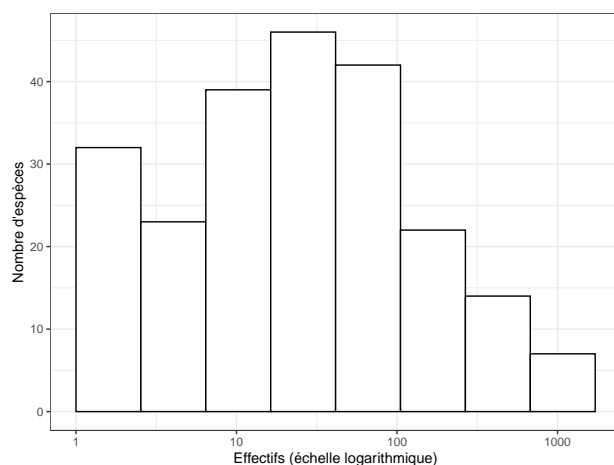


Figure 2.2: Histogramme des fréquences (diagramme de Preston) des arbres du dispositif de Barro Colorado Island. En abscisse: le nombre d'arbres de chaque espèce (en logarithme); en ordonnée: le nombre d'espèces.

Elle peut être représentée sous la forme d'un histogramme des fréquences (diagramme de Preston,<sup>18</sup> figure 2.2) ou bien d'un diagramme rang-abondance (RAC: *Rank Abundance Curve* ou diagramme de Whittaker,<sup>19</sup> figure 2.3). Le RAC est souvent utilisé pour reconnaître des distributions connues. Izsák and Pavoine<sup>20</sup> ont étudié les propriétés des RAC pour les principales SAD.

<sup>18</sup>F. W. Preston. "The Commonness, and Rarity, of Species." In: *Ecology* 29.3 (1948), pp. 254–283. DOI: [10.2307/1930989](https://doi.org/10.2307/1930989).

<sup>19</sup>Whittaker, "Dominance and Diversity in Land Plant Communities," see n. 1, p. 3.

<sup>20</sup>J. Izsák and S. Pavoine. "Links between the Species Abundance Distribution and the Shape of the Corresponding Rank Abundance Curve." In: *Ecological Indicators* 14.1 (2012), pp. 1–6. DOI: [10.1016/j.ecolind.2011.06.030](https://doi.org/10.1016/j.ecolind.2011.06.030).

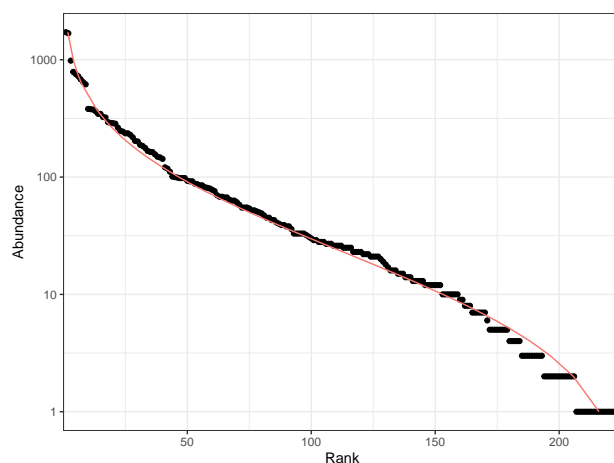


Figure 2.3: Diagramme rang-abondance (diagramme de Whittaker) des arbres du dispositif de Barro Colorado Island. Les points sont les données: en abscisse: le rang de l'espèce, à partir de la plus abondante; en ordonnée: l'abondance de l'espèce. La courbe est l'ajustement d'une distribution log-normale.

Code de la figure 2.2:

```
Ns <- sort(colSums(BCI), decreasing = TRUE)
N <- sum(Ns)
SADhist <- ggplot(data.frame(Ns), aes(Ns)) +
```

```
geom_histogram(bins=nclass.Sturges(log(Ns)),
               color="black", fill="white", boundary = 0) +
scale_x_log10() +
labs(x="Effectifs (échelle logarithmique)",
     y="Nombre d'espèces")
SADhist
```

Code de la figure 2.3:

```
library("entropart")
autoplot(as.AbdVector(Ns), Distribution = "lnorm")
```

<sup>21</sup>A. E. Magurran. *Ecological Diversity and Its Measurement*. Princeton, NJ: Princeton University Press, 1988.

<sup>22</sup>B. J. McGill et al. "Species Abundance Distributions: Moving beyond Single Prediction Theories to Integration within an Ecological Framework." In: *Ecology Letters* 10.10 (2007), pp. 995–1015. DOI: [10.1111/j.1461-0248.2007.01094.x](https://doi.org/10.1111/j.1461-0248.2007.01094.x).

<sup>23</sup>R. A. Fisher et al. "The Relation between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population." In: *Journal of Animal Ecology* 12 (1943), pp. 42–58. DOI: [10.2307/1411](https://doi.org/10.2307/1411).

<sup>24</sup>I. Motomura. "On the statistical treatment of communities." In: *Zoological Magazine* 44 (1932), pp. 379–383; R. H. Whittaker. "Evolution and Measurement of Species Diversity." In: *Taxon* 21.2/3 (1972), pp. 213–251. DOI: [10.2307/1218190](https://doi.org/10.2307/1218190).

<sup>25</sup>Preston, "The Commonness, and Rarity, of Species," see n. 18, p. 15.

<sup>26</sup>R. H. MacArthur. "On the Relative Abundance of Bird Species." In: *Proceedings of the National Academy of Sciences of the United States of America* 43.3 (1957), pp. 293–295. DOI: [10.1073/pnas.43.3.293](https://doi.org/10.1073/pnas.43.3.293). JSTOR: 89566.

<sup>27</sup>S. P. Hubbell. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, 2001.

<sup>28</sup>MacArthur, see n. 26.

Les SAD sont traitées en détail par Magurran<sup>21</sup> ou McGill et al.<sup>22</sup> Les principales distributions, nécessaires à la compréhension de la suite sont présentées ici:

- La distribution en log-séries de Fisher et al.,<sup>23</sup>
- La distribution géométrique;<sup>24</sup>
- La distribution log-normale;<sup>25</sup>
- Le modèle Broken Stick.<sup>26</sup>

Formellement, la distribution des probabilités des espèces, notées  $p_s$ , est à établir.

## La distribution en log-séries

Cette distribution est traitée en détail dans le chapitre ??.

Le nombre d'espèces est lié au nombre d'individus par la relation  $\mathbb{E}(S^n) = \alpha \ln(1 + n/\alpha)$  où  $S^n$  indique le nombre d'espèces observées dans un échantillon de  $n$  individus.  $\alpha$  est le paramètre qui fixe la pente de la partie linéaire de la relation, valide dès que  $n \gg \alpha$ , où le nombre d'espèces  $S^n$  augmente proportionnellement au logarithme du nombre d'individus  $\ln(n)$ .

La distribution a été obtenue à partir d'inventaires de communautés de papillons en Malaisie et en Angleterre. Le modèle est tombé en désuétude faute de confirmation empirique à l'échelle de la communauté, avant d'être remis en valeur par la théorie neutre<sup>27</sup> dans lequel la distribution de la *méta-communauté* est en log-séries.

## La distribution Broken Stick

Si les espèces se partagent les ressources ou l'espace des niches, représentées par un bâton, par un processus de cassure aléatoire et simultanée (précisément, les  $S - 1$  cassures du bâton sont distribuées uniformément sur sa longueur) et que leur abondance est proportionnelle à la quantité de ressources ou d'espace de niche obtenus, alors leur distribution suit le modèle Broken Stick de MacArthur.<sup>28</sup>

Parmi les distributions classiques, c'est la plus équitable: la distribution uniforme des probabilités ( $p_s = 1/S$  pour tout  $s$ ) n'est jamais approchée.

Elle est peu observée empiriquement.

### La distribution log-normale

Dans une distribution log-normale, le logarithme des probabilités des espèces (notées  $p_s$  pour l'espèce  $s$ ) suit une loi normale. L'écart-type  $\sigma$  de cette distribution règle l'équité de la distribution. Son espérance est obtenue à partir du nombre d'espèces et de  $\sigma$ , pour que la somme des probabilités égale 1.

May<sup>29</sup> explique cette distribution par le théorème de la limite centrale: la variable aléatoire valant 1 si un individu est de l'espèce  $s$  et 0 sinon est le produit de nombreuses variables de loi inconnues valant 1 en cas de succès (germination d'une graine, survie à de nombreux événements...). Le logarithme de ce produit est une somme de variables aléatoires dont la loi est forcément normale par application du théorème de la limite centrale.

La distribution est aussi le résultat d'un algorithme de bâton brisé (*broken stick*) hiérarchique:<sup>30</sup>

- Si les ressources (représentées par un bâton) sont partagées une première fois aléatoirement, selon une loi quelconque,
- Si chaque bâton obtenu est partagé à nouveau selon le même procédé, et que l'opération est répétée un assez grand nombre de fois,
- Si l'abondance de chaque espèce est proportionnelle aux ressources dont elle dispose,
- Alors la distribution des espèces est log-normale.

Ce mécanisme décrit assez bien un mécanisme de partage successif des ressources, par exemple entre groupes d'espèces de plus en plus petits, correspondant à des niches écologiques de plus en plus étroites.

D'autres arguments existent dans la littérature. Par exemple, Engen and Lande<sup>31</sup> obtiennent une distribution normale à partir d'un modèle de dynamique des populations.

La distribution log-normale décrit assez bien (mais pas exactement) une communauté locale dans le cadre de la théorie neutre<sup>32</sup> comme le montre la figure 2.3. Le nombre d'espèces rares est un peu surestimé. La distribution exacte est donnée par Volkov et al.<sup>33</sup>

<sup>29</sup>R. M. May. "Patterns of Species Abundance and Diversity." In: *Ecology and Evolution of Communities*. Ed. by M. L. Cody and J. M. Diamond. Harvard University Press, 1975, pp. 81–120.

<sup>30</sup>M. G. Bulmer. "On Fitting the Poisson Lognormal Distribution to Species-Abundance Data." In: *Biometrics* 30.1 (1974), pp. 101–110. DOI: [10.2307/1939021](https://doi.org/10.2307/1939021).

<sup>31</sup>S. Engen and R. Lande. "Population Dynamic Models Generating the Lognormal Species Abundance Distribution." In: *Mathematical Biosciences* 132.2 (1996), pp. 169–183. DOI: [10.1016/0025-5564\(95\)00054-2](https://doi.org/10.1016/0025-5564(95)00054-2).

<sup>32</sup>Hubbell, see n. 27.

<sup>33</sup>I. Volkov et al. "Neutral Theory and Relative Species Abundance in Ecology." In: *Nature* 424.6952 (2003), pp. 1035–1037. DOI: [10.1038/nature01883](https://doi.org/10.1038/nature01883).

## La distribution géométrique

Si les espèces se partagent les ressources selon un algorithme *broken stick* séquentiel (comme dans la distribution log-normale) mais de proportion fixe  $0 < k < 1$ , alors la distribution obtenue est géométrique. Les abondances successives sont proportionnelles à  $k, k(1-k), k(1-k)^2, \dots, k(1-k)^s, \dots, k(1-k)^S$ .

Ce modèle a été établi par Motomura<sup>34</sup> cité par May.<sup>35</sup> Ses propriétés ont été étudiées par Whittaker.<sup>36</sup>

C'est la distribution qui traduit l'absence de relation entre la taille de l'échantillon et l'abondance des espèces:<sup>37</sup> la distribution du logarithme de ses probabilités est uniforme. En d'autres termes, l'ordre de grandeur de l'abondance d'une espèce est uniformément distribué.

La distribution est observée dans les communautés pionnières,<sup>38</sup> peu diverses, ou les communautés microbiennes.<sup>39</sup>

<sup>34</sup>Motomura, "On the statistical treatment of communities," see n. 24, p. 16.

<sup>35</sup>May, "Patterns of Species Abundance and Diversity," see n. 29, p. 17.

<sup>36</sup>Whittaker, "Evolution and Measurement of Species Diversity," see n. 24, p. 16.

<sup>37</sup>S. Pueyo et al. "The Maximum Entropy Formalism and the Idiosyncratic Theory of Biodiversity." In: *Ecology letters* 10.11 (2007), pp. 1017–28. DOI: [10.1111/j.1461-0248.2007.01096.x](https://doi.org/10.1111/j.1461-0248.2007.01096.x).

<sup>38</sup>F. A. Bazzaz. "Plant Species Diversity in Old-Field Successional Ecosystems in Southern Illinois." In: *Ecology* 56.2 (Mar. 1975), pp. 485–488. DOI: [10.2307/1934981](https://doi.org/10.2307/1934981).

<sup>39</sup>B. Haegeman et al. "Robust Estimation of Microbial Diversity in Theory and in Practice." In: *The ISME journal* 7.6 (2013), pp. 1092–101. DOI: [10.1038/ismej.2013.10](https://doi.org/10.1038/ismej.2013.10).

## Synthèse

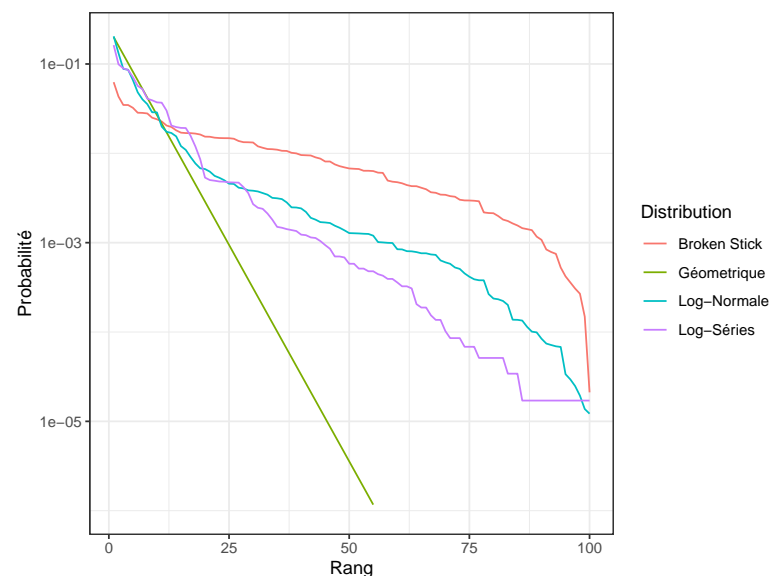


Figure 2.4: Diagramme rang-fréquence des distributions de probabilité classiques. Toutes les distributions sont de 100 espèces. Les probabilités inférieures à  $10^{-6}$  ne sont pas affichées. Les paramètres choisis sont  $\alpha = 11$  pour la distribution log-séries,  $k = 0,2$  pour la distribution géométrique et  $\sigma = 2$  pour la distribution log-normale.

<sup>40</sup>Magurran, *Ecological Diversity and Its Measurement*, see n. 21, p. 16.

La figure 2.4 est inspirée de la figure très connue de Magurran.<sup>40</sup> Elle montre bien une gradation en termes de décroissance de probabilité entre des distributions de même richesse: de la plus équitable (broken stick) à la plus inéquitable (géométrique). Elle doit être nuancée: la proportion  $k$  de la distribution géométrique fixe la pente de la droite qui la représente sur la figure.  $k = 10\%$  sur la figure: une valeur plus faible diminuerait la pente. De même, l'écart-type de la distribution log-normale décrit sa dispersion. Une valeur supérieure augmenterait sa décroissance.

Le code utilisé pour produire la figure 2.4 est le suivant:

```
library("entropart")
# Tirage d'une communauté en log-séries
lseries <- rCommunity(1, size=1E5, Distribution="lseries", alpha=11)
# Nombre d'espèces de référence
S <- length(lseries)
# Part des ressources accaparées dans la distribution géométrique
prob=0.2
# Calcul des probabilités de la distribution géométrique
geom <- prob/(1-(1-prob)^S)*(1-prob)^(0:(S-1))
# Dispersion de la loi lognormale
sd <- 2
# Tirage de S valeurs dans une loi lognormale
Nslnorm <- rlnorm(S, 0, sd)
# Normalisation pour obtenir des probabilités
lnorm <- Nslnorm/sum(Nslnorm)
# Tirage des probabilités de la distribution broken stick
bstick <- c(cuts <- sort(stats::runif(S-1)), 1)- c(0, cuts)
# Graphique
tibble(Rang=1:S,
  `Log-Séries`=sort(lseries/sum(lseries), decreasing=TRUE),
  `Géométrique`=sort(geom, decreasing=TRUE),
  `Log-Normale`=sort(lnorm, decreasing=TRUE),
  `Broken Stick`=sort(bstick, decreasing=TRUE)) %>%
  pivot_longer(cols=-Rang) %>%
  ggplot() +
  geom_line(aes(x=Rang, y=value, color=name)) +
  scale_y_log10(limits=c(1E-6, NA)) +
  labs(y="Probabilité", color="Distribution")
```

La simulation de ces quatre distributions peut être réalisée par la fonction `rCommunity()` du package *entropart*, où l'argument `Distribution` peut valoir “bstick”, “lnorm”, “geom” ou “lseries”. La simulation des communautés autres que log-séries passe par le tirage des probabilités des espèces (le calcul est déterministe dans le cas de la distribution géométrique) puis le tirage d'un nombre d'individus dans une loi multinomiale respectant ces probabilités et l'effectif total.

Les fonction `RACbstick()`, `RAClnorm()`, `RACgeom()` et `RAClseries()` permettent d'inférer à partir d'un vecteur d'abondance les paramètres d'une distribution. La distribution correspondant au modèle estimé peut être affichée sur la figure Rang-Abondance (figure 2.2).

Le package *sads* fournit les fonctions classiques de R (densité de probabilité, cumulative, quantile, simulation) pour les distributions utiles en écologie, au-delà de celles présentées ici. La distribution de Volkov notamment peut être simulée. Les fonctions `fitxxx()` complètent les fonctions `RACxxx()` d'*entropart*.

Le code suivant montre comment ajuster une distribution log-normale aux données de BCI avec *entropart* ou *sads*.

```
# entropart
library("entropart")
fit_entropart_lnorm <- RAClnorm(Ns)
# Affichage des paramètres estimés
fit_entropart_lnorm$mu
```

```
## [1] 3.142695
```

```
fit_entropart_lnorm$sigma
```

```
## [1] 1.791177
```

```
# sads
library("sads")
# Estimation. Les données sont tronquées: les espèces
# observées 0 fois ne sont pas comptées.
fit_sads_lnorm <- fitlnorm(Ns, trunc = 0)
fit_sads_lnorm@fullcoef
```

```
## meanlog    sdlog
## 3.142695 1.787195
```

L'ajustement du modèle de Volkov peut être comparé à celui d'une distribution log-normale.

```
# Ajustement du modèle de Volkov
fit_volkov <- fitvolkov(Ns)
fit_volkov@fullcoef
```

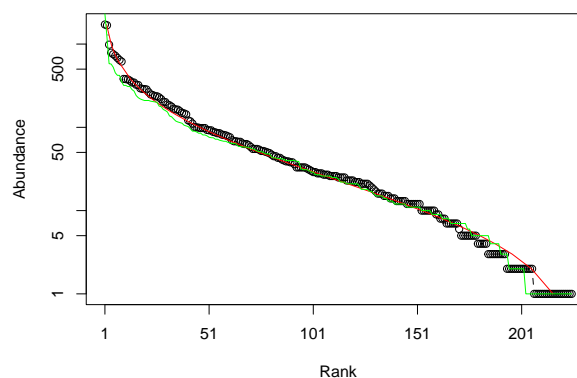
```
##      theta      m      J
## 4.767129e+01 9.238201e-02 2.145700e+04
```

Graphiquement, l'ajustement est très proche mais la distribution de Volkov prévoit explicitement des effectifs égaux parce qu'entiers.

```
# Comparaison graphique des deux modèles. Log-normal en
# rouge.
plot(as.AbdVector(Ns), Distribution = "lnorm")
```

```
## $mu
## [1] 3.142695
##
## $sigma
## [1] 1.791177
```

```
# Volkov en vert
lines(sort(rvolkov(length(Ns), fit_volkov@fullcoef[1], fit_volkov@fullcoef[2],
  fit_volkov@fullcoef[3]), decreasing = TRUE), col = "green")
```





Les vraisemblances sont proches.

```
# Comparaison des vraisemblances  
fit_sads_lnorm@min
```

```
## [1] 1157.013
```

```
fit_volkov@min
```

```
## [1] 1150.182
```

Les paramètres du modèle de communauté locale de la théorie neutre sont  $\theta$ , le “nombre fondamental de la biodiversité” égal à deux fois le nombre d’espèces apparaissant par pas de temps dans la méta-communauté,  $m$ , le taux de migration, et  $J$ , la taille de la communauté locale (qui n’est pas à proprement parler un paramètre mais une statistique décrivant les données).

La différence entre les logarithmes de vraisemblance des deux modèles en faveur du modèle de Volkov, alors que le nombre de paramètres est le même. L’ajustement est donc meilleur mais la différence est petite et la complexité du modèle et des calculs pour l’estimer ne se justifient pas en général: le modèle de Volkov est très peu utilisé en pratique.



# Bibliography

- Agapow, P. M., O. R. P. Binindal-Emonds, K. A. Crandall, J. L. Gittleman, G. M. Mace, J. C. Marshall, and A. Purvis. "The Impact of Species Concept on Biodiversity Studies." In: *The Quarterly Review of Biology* 79.2 (2004), pp. 161–179. DOI: [10.1086/383542](#) (cit. on p. 7).
- Barberousse, A. and S. Samadi. "La Taxonomie et Les Collections d'histoire Naturelle à l'heure de La Sixième Extinction." In: *La Biodiversité En Question. Enjeux Philosophiques, Éthiques et Scientifiques*. Ed. by E. Casetta and J. Delord. Paris: Editions Matériologiques, 2014, pp. 155–182 (cit. on p. 7).
- Bazzaz, F. A. "Plant Species Diversity in Old-Field Successional Ecosystems in Southern Illinois." In: *Ecology* 56.2 (Mar. 1975), pp. 485–488. DOI: [10.2307/1934981](#) (cit. on p. 18).
- Blandin, P. "La Diversité Du Vivant Avant (et Après) La Biodiversité : Repères Historiques et Épistémologiques." In: *La Biodiversité En Question. Enjeux Philosophiques, Éthiques et Scientifiques*. Ed. by E. Casetta and J. Delord. Paris: Editions Matériologiques, 2014, pp. 31–68 (cit. on p. v).
- Bulmer, M. G. "On Fitting the Poisson Lognormal Distribution to Species-Abundance Data." In: *Biometrics* 30.1 (1974), pp. 101–110. DOI: [10.2307/1939021](#) (cit. on p. 17).
- Cardinale, B. J., J. E. Duffy, A. Gonzalez, D. U. Hooper, C. Perrings, P. Venail, A. Narwani, G. M. Mace, D. Tilman, D. A. Wardle, A. P. Kinzig, G. C. Daily, M. Loreau, J. B. Grace, A. Larigauderie, D. S. Srivastava, and S. Naeem. "Biodiversity Loss and Its Impact on Humanity." In: *Nature* 486.7401 (2012), pp. 59–67. DOI: [10.1038/nature11148](#) (cit. on p. vi).
- Casetta, E. "Évaluer et Conserver La Biodiversité Face Au Problème Des Espèces." In: *La Biodiversité En Question. Enjeux Philosophiques, Éthiques et Scientifiques*. Ed. by E. Casetta and J. Delord. Paris: Editions Matériologiques, 2014, pp. 139–154 (cit. on p. 7).
- Ceballos, G., P. R. Ehrlich, and R. Dirzo. "Biological Annihilation via the Ongoing Sixth Mass Extinction Signaled by Vertebrate Population Losses and Declines." In: *Proceedings of the National Academy of Sciences* (2017), p. 201704949. DOI: [10.1073/pnas.1704949114](#) (cit. on p. vi).
- Chao, A. and L. Jost. "Coverage-Based Rarefaction and Extrapolation: Standardizing Samples by Completeness Rather than Size." In: *Ecology* 93.12 (2012), pp. 2533–2547. DOI: [10.1890/11-1952.1](#) (cit. on pp. 12, 13).
- Chao, A., S.-M. Lee, and T.-C. Chen. "A Generalized Good's Nonparametric Coverage Estimator." In: *Chinese Journal of Mathematics* 16 (1988), pp. 189–199. JSTOR: [43836340](#) (cit. on p. 12).
- Chao, A. and C.-W. Lin. "Nonparametric Lower Bounds for Species Richness and Shared Species Richness under Sampling without Replacement." In: *Biometrics* 68.3 (2012), pp. 912–921. DOI: [10.1111/j.1541-0420.2011.01739.x](#) (cit. on p. 14).
- Chao, A., K. H. Ma, T. C. Hsieh, and C.-H. Chiu. "SpadeR: Species Prediction and Diversity Estimation with R." In: (2016) (cit. on p. 13).
- Chao, A. and T.-J. Shen. "Program SPADE: Species Prediction and Diversity Estimation. Program and User's Guide." In: (2010) (cit. on p. 13).
- Chapin, F. S. I., E. S. Zavaleta, V. T. Eviner, R. L. Naylor, P. M. Vitousek, H. L. Reynolds, D. U. Hooper, S. Lavorel, O. E. Sala, S. E. Hobbie, M. C. Mack, and S. Díaz. "Consequences of Changing Biodiversity." In: *Nature* 405.6783 (2000), pp. 234–242. DOI: [10.1038/35012241](#) (cit. on p. vi).
- Chiu, C.-H., Y.-T. Wang, B. A. Walther, and A. Chao. "An Improved Nonparametric Lower Bound of Species Richness via a Modified Good-Turing Frequency Formula." In: *Biometrics* 70.3 (2014), pp. 671–682. DOI: [10.1111/biom.12200](#). PMID: [24945937](#) (cit. on p. 11).
- Condit, R., R. A. Chisholm, and S. P. Hubbell. "Thirty Years of Forest Census at Barro Colorado and the Importance of Immigration in Maintaining Diversity." In: *PLoS ONE* 7.11 (2012), e49826. DOI: [10.1371/journal.pone.0049826](#) (cit. on p. vii).
- Cracraft, J. "Species Concepts and Speciation Analysis." In: *Current Ornithology Volume 1*. Ed. by R. F. Johnston. Vol. 1. Current Ornithology. Springer US, 1983, pp. 159–187. DOI: [10.1007/978-1-4615-6781-3\\_6](#) (cit. on p. 7).
- Dauby, G. and O. J. Hardy. "Sampled-Based Estimation of Diversity Sensu Stricto by Transforming Hurlbert Diversities into Effective Number of Species." In: *Ecography* 35.7 (2012), pp. 661–672. DOI: [10.1111/j.1600-0587.2011.06860.x](#) (cit. on p. 12).
- DeLong, D. C. J. "Defining Biodiversity." In: *Wildlife Society Bulletin* 24.4 (1996), pp. 738–749. JSTOR: [3783168](#) (cit. on p. v).
- Delord, J. "La Biodiversité : Imposture Scientifique Ou Ruse Épistémologique ?" In: *La Biodiversité En Question. Enjeux Philosophiques, Éthiques et Scientifiques*. Ed. by E. Casetta and J. Delord. Paris: Editions Matériologiques, 2014, pp. 83–118. DOI: [10.3917/edmat.delord.2014.01.0083](#) (cit. on p. v).

- Dengler, J. "Which Function Describes the Species-Area Relationship Best? A Review and Empirical Evaluation." In: *Journal of Biogeography* 36.4 (2009), pp. 728–744. DOI: [10.1111/j.1365-2699.2008.02038.x](https://doi.org/10.1111/j.1365-2699.2008.02038.x) (cit. on p. 10).
- Dobzhansky, T. *Genetics and the Origin of Species*. New York: Columbia University Press, 1937 (cit. on p. 7).
- Engen, S. and R. Lande. "Population Dynamic Models Generating the Lognormal Species Abundance Distribution." In: *Mathematical Biosciences* 132.2 (1996), pp. 169–183. DOI: [10.1016/0025-5564\(95\)00054-2](https://doi.org/10.1016/0025-5564(95)00054-2) (cit. on p. 17).
- Esty, W. W. "A Normal Limit Law for a Nonparametric Estimator of the Coverage of a Random Sample." In: *The Annals of Statistics* 11.3 (1983), pp. 905–912. DOI: [10.2307/2240652](https://doi.org/10.2307/2240652). JSTOR: [2240652](https://www.jstor.org/stable/2240652) (cit. on p. 12).
- Fisher, R. A., A. S. Corbet, and C. B. Williams. "The Relation between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population." In: *Journal of Animal Ecology* 12 (1943), pp. 42–58. DOI: [10.2307/1411](https://doi.org/10.2307/1411) (cit. on p. 16).
- Gaston, K. J. "Global Patterns in Biodiversity." In: *Nature* 405.6783 (2000), pp. 220–227. DOI: [10.1038/35012228](https://doi.org/10.1038/35012228) (cit. on p. 5).
- Good, I. J. "The Population Frequency of Species and the Estimation of Population Parameters." In: *Biometrika* 40.3/4 (1953), pp. 237–264. DOI: [10.1093/biomet/40.3-4.237](https://doi.org/10.1093/biomet/40.3-4.237) (cit. on pp. 11, 12).
- Gourlet-Fleury, S., J. M. Guehl, and O. Laroussinie. *Ecology & Management of a Neotropical Rainforest. Lessons Drawn from Paracou, a Long-Term Experimental Research Site in French Guiana*. Paris: Elsevier, 2004 (cit. on p. vii).
- Hageman, B., J. Hamelin, J. Moriarty, P. Neal, J. Dushoff, and J. S. Weitz. "Robust Estimation of Microbial Diversity in Theory and in Practice." In: *The ISME journal* 7.6 (2013), pp. 1092–101. DOI: [10.1038/ismej.2013.10](https://doi.org/10.1038/ismej.2013.10) (cit. on p. 18).
- Hey, J. "The Mind of the Species Problem." In: *Trends in Ecology & Evolution* 16.7 (2001), pp. 326–329. DOI: [10.1016/S0169-5347\(01\)02145-0](https://doi.org/10.1016/S0169-5347(01)02145-0) (cit. on pp. vi, 7).
- Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, 2001 (cit. on pp. 16, 17).
- Hwang, W.-H., C.-W. Lin, and T.-J. Shen. "Good-Turing Frequency Estimation in a Finite Population." In: *Biometrical journal* 57.2 (2014), pp. 321–339. DOI: [10.1002/bimj.201300168](https://doi.org/10.1002/bimj.201300168) (cit. on p. 14).
- Izsák, J. and S. Pavoine. "Links between the Species Abundance Distribution and the Shape of the Corresponding Rank Abundance Curve." In: *Ecological Indicators* 14.1 (2012), pp. 1–6. DOI: [10.1016/j.ecolind.2011.06.030](https://doi.org/10.1016/j.ecolind.2011.06.030) (cit. on p. 15).
- Jizhong, Z., M. Shijun, and C. Changming. "An Index of Ecosystem Diversity." In: *Ecological Modelling* 59 (1991), pp. 151–163. DOI: [10.1016/0304-3800\(91\)90176-2](https://doi.org/10.1016/0304-3800(91)90176-2) (cit. on p. vi).
- Jost, L. "The Relation between Evenness and Diversity." In: *Diversity* 2.2 (2010), pp. 207–232. DOI: [10.3390/d2020207](https://doi.org/10.3390/d2020207) (cit. on p. 5).
- Loreau, M. "Discours de Clôture." In: *Actes de La Conférence Internationale Biodiversité Science et Gouvernance*. Ed. by R. Barbault and J.-P. Le Duc. Paris, France.: IRD Editions, 2005, pp. 254–256 (cit. on p. v).
- MacArthur, R. H. "On the Relative Abundance of Bird Species." In: *Proceedings of the National Academy of Sciences of the United States of America* 43.3 (1957), pp. 293–295. DOI: [10.1073/pnas.43.3.293](https://doi.org/10.1073/pnas.43.3.293). JSTOR: [89566](https://www.jstor.org/stable/89566) (cit. on p. 16).
- Magurran, A. E. *Ecological Diversity and Its Measurement*. Princeton, NJ: Princeton University Press, 1988 (cit. on pp. 16, 18).
- Marcon, E. and B. Hérault. "Entropart, an R Package to Measure and Partition Diversity." In: *Journal of Statistical Software* 67.8 (2015), pp. 1–26. DOI: [10.18637/jss.v067.i08](https://doi.org/10.18637/jss.v067.i08) (cit. on p. vii).
- May, R. M. "Patterns of Species Abundance and Diversity." In: *Ecology and Evolution of Communities*. Ed. by M. L. Cody and J. M. Diamond. Harvard University Press, 1975, pp. 81–120 (cit. on pp. 17, 18).
- Mayden, R. L. "A Hierarchy of Species Concepts: The Denouement in the Saga of the Species Problem." In: *Species. The Units of Biodiversity*. Ed. by M. F. Claridge, H. A. Dawah, and M. R. Wilson. London: Chapman and Hall, 1997, pp. 381–424 (cit. on p. 7).
- Mayr, E. *Systematics and the Origin of Species from the Viewpoint of a Zoologist*. New York: Columbia University Press, 1942 (cit. on p. 7).
- McGill, B. J., R. S. Etienne, J. S. Gray, D. Alonso, M. J. Anderson, H. K. Bence, M. Dornelas, B. J. Enquist, J. L. Green, F. He, A. H. Hurlbert, A. E. Magurran, P. A. Marquet, B. A. Maurer, A. Ostling, C. U. Soykan, K. I. Ugland, and E. P. White. "Species Abundance Distributions: Moving beyond Single Prediction Theories to Integration within an Ecological Framework." In: *Ecology Letters* 10.10 (2007), pp. 995–1015. DOI: [10.1111/j.1461-0248.2007.01094.x](https://doi.org/10.1111/j.1461-0248.2007.01094.x) (cit. on p. 16).
- Mcintosh, R. P. "An Index of Diversity and the Relation of Certain Concepts to Diversity." In: *Ecology* 48.3 (1967), pp. 392–404. DOI: [10.2307/1932674](https://doi.org/10.2307/1932674) (cit. on p. 3).
- Meine, C., M. Soulé, and R. F. Noss. "A Mission-Driven Discipline": The Growth of Conservation Biology." In: *Conservation Biology* 20.3 (2006), pp. 631–651. DOI: [10.1111/j.1523-1739.2006.00449.x](https://doi.org/10.1111/j.1523-1739.2006.00449.x) (cit. on p. v).
- Miraldo, A., S. Li, M. K. Borregaard, A. Florez-Rodriguez, S. Gopalakrishnan, M. Rizvanovic, Z. Wang, C. Rahbek, K. A. Marske, and D. Nogues-Bravo. "An Anthropocene Map of Genetic Diversity." In: *Science* 353.6307 (2016), pp. 1532–1535. DOI: [10.1126/science.aaf4381](https://doi.org/10.1126/science.aaf4381) (cit. on p. 5).
- Moreno, C. E. and P. Rodríguez. "A Consistent Terminology for Quantifying Species Diversity?" In: *Oecologia* 163.2 (2010), pp. 279–82. DOI: [10.1007/s00442-010-1591-7](https://doi.org/10.1007/s00442-010-1591-7) (cit. on p. 6).
- Motomura, I. "On the statistical treatment of communities." In: *Zoological Magazine* 44 (1932), pp. 379–383 (cit. on pp. 16, 18).
- Mouillot, D., W. Stubbs, M. Faure, O. Dumay, J.-A. Tomasini, J. B. Wilson, and T. D. Chi. "Niche Overlap Estimates Based on Quantitative Functional Traits: A New Family of Non-Parametric Indices." In: *Oecologia* 145.3 (2005), pp. 345–353. DOI: [10.1007/s00442-005-0151-z](https://doi.org/10.1007/s00442-005-0151-z) (cit. on p. 4).

- Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. "Vegan: Community Ecology Package." In: (2012) (cit. on p. vii).
- Pavoine, S. and M. B. Bonsall. "Measuring Biodiversity to Explain Community Assembly: A Unified Approach." In: *Biological Reviews* 86.4 (2011), pp. 792–812. DOI: [10.1111/j.1469-185X.2010.00171.x](https://doi.org/10.1111/j.1469-185X.2010.00171.x) (cit. on p. 4).
- Pernès, J., ed. *Gestion Des Ressources Génétiques Des Plantes. Tome 2 : Manuel*. Paris: Agence de Coopération culturelle et technique, 1984 (cit. on p. 7).
- Preston, F. W. "The Commonness, and Rarity, of Species." In: *Ecology* 29.3 (1948), pp. 254–283. DOI: [10.2307/1930989](https://doi.org/10.2307/1930989) (cit. on pp. 15, 16).
- Pueyo, S., F. He, and T. Zillio. "The Maximum Entropy Formalism and the Idiosyncratic Theory of Biodiversity." In: *Ecology letters* 10.11 (2007), pp. 1017–28. DOI: [10.1111/j.1461-0248.2007.01096.x](https://doi.org/10.1111/j.1461-0248.2007.01096.x) (cit. on p. 18).
- Purvis, A. and A. Hector. "Getting the Measure of Biodiversity." In: *Nature* 405.6783 (2000), pp. 212–9. DOI: [10.1038/35012221](https://doi.org/10.1038/35012221) (cit. on p. vi).
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2022 (cit. on p. vii).
- Richards, R. A. *The Species Problem. A Philosophical Analysis*. Cambridge: Cambridge University Press, 2010 (cit. on p. 7).
- Ricotta, C. "Through the Jungle of Biological Diversity." In: *Acta Biotheoretica* 53.1 (2005), pp. 29–38. DOI: [10.1007/s10441-005-7001-6](https://doi.org/10.1007/s10441-005-7001-6) (cit. on p. vi).
- "A Semantic Taxonomy for Diversity Measures." In: *Acta Biotheoretica* 55.1 (2007), pp. 23–33. DOI: [10.1007/s10441-007-9008-7](https://doi.org/10.1007/s10441-007-9008-7) (cit. on p. 4).
- Runnegar, B. "Rates and Modes of Evolution in the Mollusca." In: *Rates of Evolution*. Ed. by M. Campbell and M. F. Day. London: Allen & Unwin, 1987, pp. 39–60 (cit. on p. 5).
- Scheiner, S. M. "Six Types of Species-Area Curves." In: *Global Ecology and Biogeography* 12.6 (2003), pp. 441–447. DOI: [10.1046/j.1466-822X.2003.00061.x](https://doi.org/10.1046/j.1466-822X.2003.00061.x) (cit. on p. 10).
- Speth, J. G., M. W. Holdgate, and M. K. Tolba. "Foreword." In: *Global Biodiversity Strategy*. Ed. by K. Courrier. Washington, D.C.: WRI, IUCN, UNEP, 1992, pp. v–vi (cit. on p. v).
- Stirling, A. "A General Framework for Analysing Diversity in Science, Technology and Society." In: *Journal of the Royal Society, Interface* 4.15 (2007), pp. 707–719. DOI: [10.1098/rsif.2007.0213](https://doi.org/10.1098/rsif.2007.0213) (cit. on p. 5).
- Sukumaran, J. and L. L. Knowles. "Multispecies Coalescent Delimits Structure, Not Species." In: *Proceedings of the National Academy of Sciences of the United States of America* in press (2017). DOI: [10.1073/PNAS.1607921114](https://doi.org/10.1073/PNAS.1607921114) (cit. on p. 7).
- Van Valen, L. "Ecological Species, Multispecies, and Oaks." In: *Taxon* 25.2/3 (1976), pp. 233–239. DOI: [10.2307/1219444](https://doi.org/10.2307/1219444) (cit. on p. 7).
- Volkov, I., J. R. Banavar, S. P. Hubbell, and A. Maritan. "Neutral Theory and Relative Species Abundance in Ecology." In: *Nature* 424.6952 (2003), pp. 1035–1037. DOI: [10.1038/nature01883](https://doi.org/10.1038/nature01883) (cit. on p. 17).
- Whittaker, R. H. "Vegetation of the Siskiyou Mountains, Oregon and California." In: *Ecological Monographs* 30.3 (1960), pp. 279–338. DOI: [10.2307/1943563](https://doi.org/10.2307/1943563). JSTOR: [1943563](https://www.jstor.org/stable/1943563) (cit. on p. 5).
- "Dominance and Diversity in Land Plant Communities." In: *Science* 147.3655 (1965), pp. 250–260. DOI: [10.1126/science.147.3655.250](https://doi.org/10.1126/science.147.3655.250) (cit. on pp. 3, 15).
- "Evolution and Measurement of Species Diversity." In: *Taxon* 21.2/3 (1972), pp. 213–251. DOI: [10.2307/1218190](https://doi.org/10.2307/1218190) (cit. on pp. 16, 18).
- "Evolution of Species Diversity in Land Communities." In: *Evolutionary Biology* 10 (1977). Ed. by M. K. Hecht, W. C. Steere, and B. Wallace, pp. 1–67 (cit. on p. 6).
- Wilkins, J. S. "Philosophically Speaking, How Many Species Concepts Are There?" In: *Zootaxa* 2765.1 (Feb. 15, 2011), pp. 58–60. DOI: [10.11646/zootaxa.2765.1.5](https://doi.org/10.11646/zootaxa.2765.1.5) (cit. on p. 7).
- Wilson, E. O. and F. M. Peter, eds. *Biodiversity*. Washington, D.C.: The National Academies Press, 1988 (cit. on p. v).
- Zhang, C.-H. and Z. Zhang. "Asymptotic Normality of a Nonparametric Estimator of Sample Coverage." In: *Annals of Statistics* 37 (5A 2009), pp. 2582–2595. DOI: [10.1214/08-aos658](https://doi.org/10.1214/08-aos658) (cit. on p. 12).
- Zhang, Z. and H. Huang. "Turing's Formula Revisited." In: *Journal of Quantitative Linguistics* 14.2-3 (2007), pp. 222–241. DOI: [10.1080/09296170701514189](https://doi.org/10.1080/09296170701514189) (cit. on p. 12).

**Résumé** La biodiversité peut être mesurée de nombreuses façons.

La dualité entropie-diversité fournit un cadre clair et rigoureux pour le faire. L'entropie est la surprise moyenne fournie par les individus d'une communauté. Le choix de la fonction d'information qui mesure cette surprise à partir des probabilités d'occurrence des espèces (ou d'autres catégories) permet de définir les mesures de diversités neutres, fonctionnelles ou phylogénétique présentées ici. L'entropie est transformée en diversité au sens strict par une fonction croissante (l'exponentielle déformée), ce qui simplifie son interprétation en tant que nombre équivalent d'espèces.

L'entropie phylogénétique généralise les indices de diversité classique, intègre si nécessaire la distance entre espèces, peut être écomposée et corrigée des biais d'estimation. Sa transformation en diversité au sens strict permet d'interpréter les valeurs sous une forme unique : un nombre équivalent d'espèces et un nombre équivalent de communautés. La diversité de Leinster et Cobbold généralise à son tour la diversité phylogénétique et permet d'autres définitions de la distance entre espèces. Le paramétrage des mesures (l'ordre de la diversité) permet de donner plus ou moins d'importance aux espèces rares et de tracer des profils de diversité.

La construction de ce cadre méthodologique est présentée en détail ainsi que plusieurs approches différentes, qui constituent l'état de l'art de la mesure de la biodiversité.

