The Pennsylvania State University

The Graduate School

Department of Statistics

The Mathematical Statistics of Diversity and Abundance

A Thesis in

Statistics

by

Charles Taillie Jr.

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 1977

Date of Signature:

G. P. Patil
Professor of Mathematical Statistics
Chairman of Committee
Thesis Adviser

William L. Harkness
Head of the Department of Statistics

ACKNOWLEDGMENTS

I wish to thank Professor G. P. Patil for his enthusiastic support and for his many valuable insights, and Professor W. L. Harkness for allowing me the time that made this research possible. I have also had the benefit of several illuminating discussions with Brian Dennis and with Professor M. T. Boswell. The approach to size-biased sampling in Chapter 3 was suggested by Professor F. Haight.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

APPENDIX

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF FIGURES (Continued)

CHAPTER 1

DIVERSITY AS A CONCEPT AND ITS MEASUREMENT

1.1 Introduction and Summary

Diversity is an important concept in community ecology. Under various names it also appears in several of the biological, social, and management sciences. Despite an extensive literature on diversity related issues, formal definition and logical development of diversity as a concept and its measurement have been largely lacking. Chapter 1 attempts to fill this gap.

Section 1.2 puts forth the view that diversity is an average property of a community and identifies that property as species rarity. Introducing the concepts of dichotomy and ranking in this context, corresponding measures of species rarity are defined. It is shown that in either case the rarity measure R is uniquely determined by the diversity index Δ.

Like any other concept, diversity remains elusive until it can be quantified with some underlying unity. It is shown in Section 1.3 that the three standard indices called Species Count, Shannon, and Simpson can be interpreted in terms of a single formulation of intraspecific encounters. Further, the rarity measures associated with these indices satisfy what we call the proportionality equation. But other solutions of this equation exist and enable us to embed the three standard indices in a one-parameter family Δ_β of indices. Defining the MacArthur transformation to species richness scale, the log numbers equivalent of Δ_β turns out to be the generalized entropy. Section 1.3 also develops a general technique for associating species rarity measures with a given discrete probability distribution. Both the Hurlbert-Smith index Δ_{ω}^{H-S} and the index Δ_{β} are obtained as special cases.

Section 1.4 identifies two instances in which it can be argued on intuitive grounds that one community is more diverse than another. These cases are formalized into two operations called introducing a species and transferring abundance. Intuitively, introducing a species increases the "species richness" component of diversity while transferring abundance increases the "evenness" component. Criterion C1, which requires dichotomous rarity measures R to be decreasing functions, is shown to imply that introducing a species increases the diversity measure Δ of a community. However, an example is given to show that this criterion is not sufficient to ensure that transferring abundance increases Δ. This leads to Criterion C2 which requires that Δ increase under both operations. It is shown that Δ_{β} satisfies Criterion C1 for all β while Criterion C2 is satisfied for $\beta \geq -1$. Thus, the Species Count, the Shannon index, and the Simpson index satisfy both criteria. The Hurlbert-Smith index also satisfies both criteria.

Different indices may inconsistently order a given pair of communities with respect to their diversity. In view of these inconsistencies, it becomes of interest to define an intrinsic diversity ordering without reference to indices. We introduce such an ordering in Section 1.5 and observe that it is a partial order, i.e., two communities may not be comparable. The choice of a particular index effectively smooths this partial order into a linear order. The inconsistencies arise because different indices result in different smoothings. After reformulating the intrinsic diversity ordering in terms of majorization, we discover that Criterion C2 can be restated as the requirement that

Δ be Schur concave. This suggests the stronger Criterion C3 which requires that Δ be a concave function.

Section 1.6 develops indices based on ranking. The simplest example is provided by taking the rarity of the ith ranked species to be $R(i) = i - 1$, which leads to the Average Rank index. A proportionality equation for ranks allows us to embed this index into a one-parameter family of rank type indices $\Delta_\rho^{(\text{rank})}$. For the general measure of rarity based on ranks, the analogue of Criterion C1 requires $R(i)$ to be an increasing function of i. In contrast with the case of dichotomous indices, this monotonicity is sufficient for Criterion C2 and Criterion C3.

The choice of an index implicitly involves a decision regarding the diversity ordering of communities which are not intrinsically comparable. With such a wealth of indices available, there is a need for a theory of index sensitivity to assist in this choice. In Section 1.7, we give a definition of the sensitivity of an index to rare species and relate it to the question of index inconsistency. The term "sensitive to rare species" occurs frequently in the diversity literature but to our knowledge no precise definition has previously appeared. The sensitivity of each of the three families of indices Δ_β, Δ_ω^{H-S}, and $\Delta_\rho^{(\text{rank})}$ is shown to be a monotone function of their parameters. A revealing crossing point theorem is also established. Section 1.7 concludes with a theory of response to perturbations. The indices Δ_β and Δ_ω^{H-S} are characterized by the form of their response functions.

Section 1.8 is concerned with diversity decomposition. Sometimes it is possible to decompose a community in some natural way and, as in the analysis of variance, apportion the total diversity among and between the various components. We consider two types of decompositions: two-way classifications and mixtures. As measured by the Shannon index, the total diversity of a two-way classification decomposes into the sum of two terms: the diversity of the row marginals and the average diversity of the normalized rows. We replace this last average by a "deflated" average and extend the decomposition to a wider class of indices through what we call a deflated ANOVA formula. The indices Δ_β are then characterized as the only indices to satisfy this formula. Section 1.8 also develops a mixture decomposition for indices satisfying Criterion C3. The between-community component of this decomposition is identified as the average rarity gain analogous to the information gain. Section 1.8 concludes with an application to genetics. A law of increasing entropy is established for the approach to Hardy-Weinberg equilibrium. Modifications of the argument lead to a general proof of the Hardy-Weinberg law.

In recent years, there has been considerable interest in the use of diversity indices as indicators of environmental quality. A problem associated with such an application is the time and level of professional expertise required for a complete taxonomic classification of the sample. A sequential comparison method has sometimes been employed to alleviate this difficulty. After a bias correction, the method is shown to provide an unbiased estimator of Simpson's index. Section 1.9 also establishes the asymptotic normality of this estimator via an extension of the Noether central limit theorem.

Estimation of the number of species in the community is one of the more interesting and intriguing problems facing the ecologist. Section

1.10 discusses several methods for estimating species richness and proposes a new estimator based on the number of singletons in the sample.

The previous sections have emphasized ecological diversity. Section 1.11 surveys several additional areas in which the diversity concept has found application.

The last section discusses income inequality in some depth. It is first noted that inequality is not the opposite of diversity; it is the opposite of evenness. After developing the notion of relative inequality, we give an intrinsic inequality ordering, similar to the intrinsic diversity ordering, and observe that it is equivalent to the Lorenz ordering familiar in economics. Finally, we propose a general method of constructing a measure of inequality from a diversity index and pair up the Gini coefficient with the Average rank, the Theil index with the Shannon index, and the coefficient of variation with the Simpson index.

1.1.1 Notation and Conventions. The terms "increasing," "decreasing," "more," and "less" always have their loose interpretations unless qualified by "strictly." Except as otherwise indicated, logarithms are to the base e. When $X = 0$, the expression $X \cdot Y$ is taken to be zero even if Y is infinite or undefined. All vectors are finite dimensional but are to be padded on the right with an infinite string of zeros. Thus (1, 2, 3) is identified with (1, 2, 3, 0, 0, 0,...) and expressions such as (1, 2, 3) + (1, 2, 3, 4) are well-defined. Except briefly in Section 1.11.2, there is no need to distinguish row and column vectors. The mathematical expectation of a random varaible X is denoted E[X].

1.2 Diversity as an Average Property of a Community

We view diversity as an average property of a community. But the average of what? To an outside observer, variety is a most striking feature of a diverse community. Alfred Russel Wallace's [1, p. 65] description of a tropical forest is a vivid illustration:

> If the traveller notices a particular species and wishes to find more like it, he may turn his eyes in vain in any direction. Trees of varied forms, dimensions and colors are around him, but he rarely sees any one of them repeated. Time after time he goes towards a tree which looks like the one he seeks, but a closer examination proves it to be distinct. He may at length, perhaps, meet with a second specimen half a mile off, or may fail altogether, till on another occasion he stumbles on one by accident.

In a diverse community, such as that described by Wallace, the typical species is relatively rare. Consequently, we propose that diversity be defined as the average rarity within a community. To make this idea precise, the concepts of "community" and "rarity" must be formalized.

Consider some quantity distributed among a countable set of categories, labeled $i = 1,2,3,\ldots$, with π_i as the proportionate share received by category i. This quantity may be continuous (income, biomass, etc.) or it may be discrete (word occurrences in a text, biological organisms, etc.). For concreteness, we will usually speak of biological organisms grouped into species and call $\underset{\sim}{\pi} = (\pi_1, \pi_2, \pi_3,\ldots)$ the <u>species abundance vector</u>; arranging the components of $\underset{\sim}{\pi}$ in descending order gives the <u>ranked abundance vector</u>, $\underset{\sim}{\pi}^* = (\pi_1^*, \pi_2^*, \pi_3^*,\ldots)$ where $\pi_1^* \geq \pi_2^* \geq \pi_3^* \geq \ldots$. For our purposes, a community may be identified with the pair $C = (s, \underset{\sim}{\pi})$ where s is the number of nonzero components of $\underset{\sim}{\pi}$, i.e., the number of species that are physically present; s is assumed to be finite. A community is said to be completely even when $\pi_1^* = \pi_2^* = \ldots = \pi_s^* = 1/s$.

Given the community, a numerical measure of rarity is to be associated with each species; denote the rarity of species i by $R(i; \underset{\sim}{\pi})$.

Definition 1.1: The diversity measure of a community $C = (s, \underset{\sim}{\pi})$ is its average rarity and is given by $\Delta(C) = \Sigma\pi_i R(i; \underset{\sim}{\pi})$ where Δ is the diversity index associated with the measure of rarity R. Sometimes we write $\Delta(\underset{\sim}{\pi})$ instead of $\Delta(C)$.

The measures of rarity considered here will be based on one of the following:

Dichotomy: The rarity of species i depends only on the numerical value of π_i. For notational simplicity, write $R(i; \underset{\sim}{\pi})$ as $R(\pi_i)$.

Ranking: The rarity of a species depends only on its (descending) rank and not explicitly on the numerical values of the components of $\underset{\sim}{\pi}$. Denoting the rarity of the ith ranked species by R(i), the index takes the form $\Delta = \Sigma R(i)\pi_i^*$. The ranks may be assigned arbitrarily within tied sets without affecting the value of Δ.

We show that in either of these two cases the measure of rarity is uniquely determined by the diversity index.

Theorem 1.1: a) Assume that the functions Q and R are defined and measurable on the interval (0,1] and that $\Sigma\pi_i Q(\pi_i) = \Sigma\pi_i R(\pi_i)$ for all probability vectors $\underset{\sim}{\pi}$. Then $Q(x) = R(x)$ for all $x \in (0,1]$.

b) Assume that Q(i) and R(i) are defined for each positive integer i and that $\Sigma Q(i)\pi_i^* = \Sigma R(i)\pi_i^*$ for all probability vectors $\underset{\sim}{\pi}$. Then $Q(i) = R(i)$ for all i.

Proof: a) Let $L(x) = x[Q(x) - R(x)]$ so that L is measurable, $L(0) = 0 = L(1)$ and $\Sigma L(\pi_i) = 0$ for all probability vectors $\underset{\sim}{\pi}$. Take $x, y \in [0,1]$ with $x + y \leq 1$; comparing the two equations

$$L(x) + L(y) + L(1 - x - y) = 0$$
$$L(x + y) + L(1 - x - y) = 0,$$

the function L is seen to be a measurable solution of Cauchy's equation on the unit interval so that ([2, p. 35]) $L(x) = cx$ for some constant c. But $c = 0$ since $L(1) = 0$. Thus $L \equiv 0$ and $Q(x) = R(x)$ for all $x \in (0,1]$. The proof of b) is trivial.

Remark 1.1: The assumption that both Q and R are measurable is essential. For suppose $\Delta = \Sigma\pi_i R(\pi_i)$ with R measurable and take $Q(x) = R(x) + L(x)/x$ where $L(x)$ is any nonmeasurable solution of Cauchy's equation with $L(1) = 0$. Then Q is nonmeasurable but $\Delta = \Sigma\pi_i Q(\pi_i)$; in particular, measurability of the index does not imply measurability of the rarity measure.

A given rarity measure R may always be replaced by a positive affine transformation having the form $R^* = bR + a$ where a and b are real numbers with $b > 0$. The index is then transformed in the same way: $\Delta^* = b\Delta + a$. For a dichotomous index $R(\pi)$, the values of a and b can usually be fixed by imposing the standardizing requirements:

$$R(1) = 0, \tag{1.1a}$$
$$R'(1) = -1. \tag{1.1b}$$

The first requirement has the effect of assigning diversity zero to a single-species community. The second is available only if $R(\pi)$ has a nonzero derivative at $\pi = 1$. The corresponding requirements for a rank type rarity measure are:

$$R(1) = 0, \tag{1.2a}$$

$$R(2) = 1. \tag{1.2b}$$

Here, (1.2b) is available only when $R(1) \neq R(2)$.

1.3 The Interpretation of Diversity Indices With Examples.

Three widely used indices of ecological diversity are the Species Count, the Shannon index, and Simpson index. These will be denoted as follows:

Species Count: $\Delta_{-1} = s - 1.$

Shannon index: $\Delta_0 = -\Sigma\pi_i \log(\pi_i).$

Simpson index: $\Delta_1 = 1 - \Sigma\pi_i^2\ .$

All three assign diversity zero to a single-species community.

It is shown in Section 1.3.1 that these indices can all be interpreted in terms of the single notion of interspecific encounters and in Section 1.3.2 the three indices are embedded in a one-parameter family $\{\Delta_\beta : \beta \text{ real}\}$ of indices. Section 1.3.3 describes a general scheme for obtaining rarity measures from a given discrete probability distribution. This scheme yields Δ_β as well as the Hurlbert-Smith index as special cases.

For the purpose of comparing the diversity of two communities, any strictly increasing transformation of an index is effectively equivalent to the original index. After transformation, the interpretation as average community rarity will, in general, be lost. But Section 1.3.4 shows that this interpretation, in an appropriate generalized sense, is retained under a particular transformation introduced by MacArthur [4].

1.3.1 Interspecific Encounters. As Simpson [3] has observed, Δ_1 is the probability that two randomly selected members of the community belong to different species. When rewritten in the form $\Delta_1 = \Sigma\pi_i(1 - \pi_i) = \Sigma\pi_i R(\pi_i)$, the Simpson index may also be interpreted as average community rarity, with the understanding that species rarity is measured by $R(\pi) = 1 - \pi$. Now contemplate Wallace's traveler who first comes upon a member of, say, the ith species. As his journey continues, the traveler encounters other organisms, sometimes of this species and sometimes not. The rarer the ith species, the more likely are the interspecific encounters; but $R(\pi_i) = 1 - \pi_i$ is precisely the probability that a given encounter is interspecific. In what follows, this concept of inter- versus intraspecific encounters is explored further. Three different schemes are considered.

Waiting time for an intraspecific encounter. Again consider the traveler in search of the ith species as, for example, in Figure 1.1. With $Y + 1$ equal to the number of encounters up to and including the first intraspecific one, we have,

$$E[Y|\pi_i] = (1 - \pi_i)/\pi_i, \; E[Y + 1|\pi_i] = 1/\pi_i,$$
$$E[1/(Y + 1)|\pi_i] = -\pi_i \log(\pi_i)/(1 - \pi_i),$$

when

$$P(Y = y|\pi_i) = \pi_i(1 - \pi_i)^y, \; y = 0,1,2,\ldots .$$

Since large Y are associated with rare species, both Y and $Y/(Y + 1)$ are reasonable measures of rarity. But these are random variables

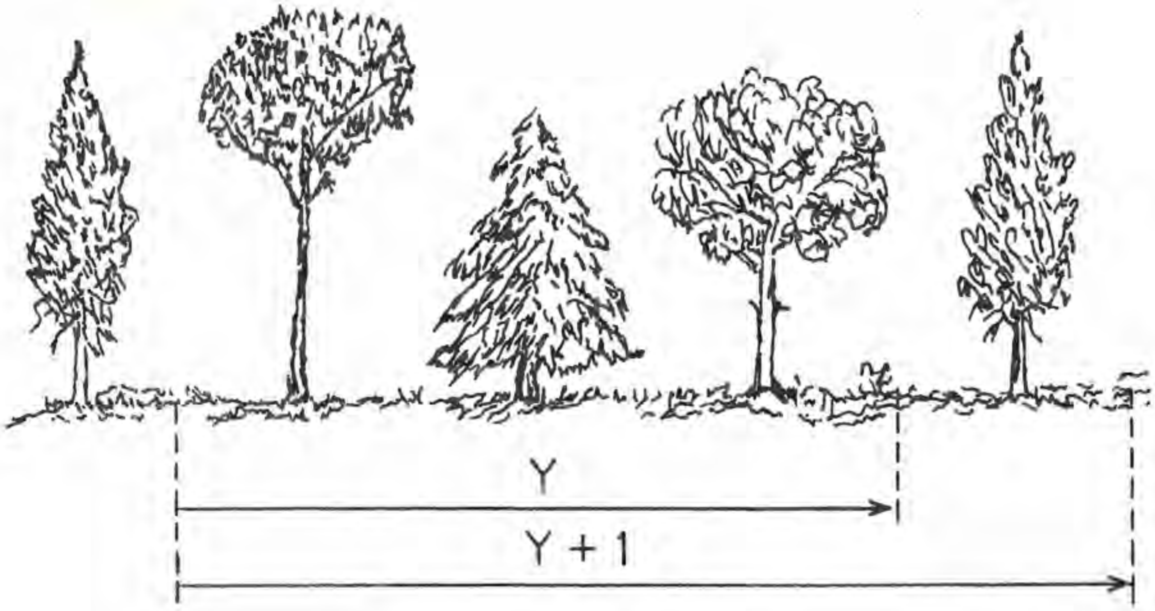


Figure 1.1 Waiting Time for an Intraspecific Encounter

and should be replaced by average quantities. There are several ways to interpret the "average" of a ratio; each gives rise to a different index.

1. Species Count:

$$R(\pi_i) = E[Y|\pi_i] = (1 - \pi_i)/\pi_i,$$
$$\Sigma\pi_i R(\pi_i) = s - 1.$$

2. Simpson index:

$$R(\pi_i) = E[Y|\pi_i]/E[Y + 1|\pi_i] = 1 - \pi_i,$$
$$\Sigma\pi_i R(\pi_i) = \Sigma\pi_i(1 - \pi_i).$$

3. Shannon index:

$$R(\pi_i) = E[1/(Y + 1)|\pi_i]\cdot E[Y|\pi_i] = -\log(\pi_i),$$
$$\Sigma\pi_i R(\pi_i) = -\Sigma\pi_i \log(\pi_i).$$

4. An unfamiliar index:

$$R(\pi_i) = E[Y/(Y + 1)|\pi_i] = 1 + \pi_i \log(\pi_i)/(1 - \pi_i),$$
$$\Sigma\pi_i R(\pi_i) = 1 + \Sigma\pi_i^2 \log(\pi_i)/(1 - \pi_i).$$

Waiting time for an interspecific encounter. Here we suppose the traveler to be in search of a new species and put $Z + 1$ equal to the number of encounters up to and including the first interspecific one. See Figure 1.2. Small Z are associated with rare species and the variables of interest are $1/Z$ and $1/(Z + 1)$.

1. Species Count:

$$R(\pi_i) = 1/E[Z|\pi_i] = (1 - \pi_i)/\pi_i.$$

2. Simpson index:

$$R(\pi_i) = 1/E[Z + 1|\pi_i] = 1 - \pi_i.$$

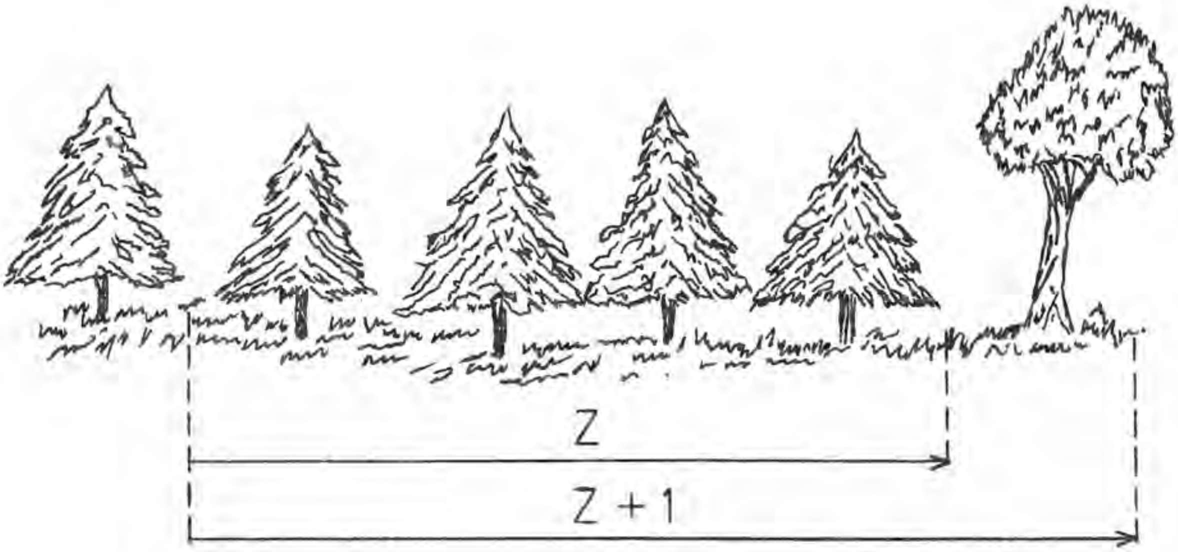


Figure 1.2 Waiting Time for an Interspecific Encounter

3. A second unfamiliar index:

$$R(\pi_i) = E[1/(Z+1)|\pi_i] = -(1-\pi_i)\log(1-\pi_i)/\pi_i.$$

The index itself is $\Sigma\pi_i R(\pi_i) = -\Sigma(1-\pi_i)\log(1-\pi_i)$.

Note that $E[1/Z|\pi_i] = \infty$ so that this case does not lead to a useful measure of rarity.

The two unfamiliar indices can be related by a duality principle. Let $R(\pi)$ be a dichotomous rarity measure which has a finite limit as π approaches zero from above, and define the dual measure by $R^\dagger(\pi) = 1 - R(1-\pi)$ for $0 < \pi < 1$ and $R^\dagger(1) = R(0+)$. It is easy to show that:

(i) $R^{\dagger\dagger} = R$,

(ii) the Simpson index is its own dual,

(iii) the two unfamiliar indices are duals of one another.

Fixed number of encounters. Here we let Y be the number of interspecific and Z the number of intraspecific encounters out of a fixed total of N encounters.

1. Species Count:

$$R(\pi_i) = E[Y|\pi_i]/E[Z|\pi_i] = (1-\pi_i)/\pi_i.$$

2. Simpson index:

$$R(\pi_i) = E[Y|\pi_i]/E[Y+Z|\pi_i] = 1-\pi_i.$$

Note that $E[Y/Z|\pi_i] = \infty$, while $E[Y/(Y+Z)|\pi_i] = 1-\pi_i$.

Some biological motivations for considering interspecific encounters have been discussed by Hurlbert [5], who concluded that if an index is to be used, Simpson's is conceptually preferable to Shannon's. While it is curious that the Shannon index arises in only the first scheme, Hurlbert's negative assessment of this index would appear to be somewhat pessimistic.

1.3.2 The Proportionality Equation. Consider a species with abundance π and a subspecies whose abundance is a fraction q of $\pi (0 < q \leq 1)$. The difference $R(q\pi) - R(\pi)$ will be compared with the corresponding change in rarity that occurs upon taking a fraction q of an entire community. For the Shannon index the two changes are equal,

$$R(q \cdot \pi) - R(\pi) = R(q \cdot 1) - R(1),$$

while for the Species Count and the Simpson index we have, respectively,

$$R(q\pi) - R(\pi) = \pi^{-1}[R(q) - R(1)]$$

and

$$R(q\pi) - R(\pi) = \pi[R(q) - R(1)].$$

These suggest consideration of the functional equation

$$R(q\pi) - R(\pi) = W(\pi)[R(q) - R(1)], \tag{1.3}$$

which will be called the proportionality equation with $W(\pi)$ as the deflation factor. Notice that (1.3) is invariant under affine transformations of R. A degenerate solution is given by

$$W(\pi) = \begin{cases} 0 & \text{if} \quad 0 < \pi < 1 \\ 1 & \text{if} \quad \pi = 1 \end{cases}$$

and, up to an affine transformation,

$$R(\pi) = \begin{cases} 1 & \text{if} \quad 0 < \pi < 1 \\ 0 & \text{if} \quad \pi = 1. \end{cases}$$

The next theorem gives essentially all other solutions.

Theorem 1.2: Suppose R and W are defined on the interval (0,1] with R measurable and nonconstant. Then R and W satisfy the proportionality equation (1.3) <=> either R and W are the degenerate solution or there is a real number β for which $W(\pi) = \pi^{\beta}$ and, up to an affine transformation,

$$R(\pi) = \begin{cases} (1 - \pi^{\beta})/\beta & \text{if } \beta \neq 0 \\ -\log(\pi) & \text{if } \beta = 0. \end{cases}$$

A similar conclusion holds if the interval (0,1] is replaced by either $(0,\infty)$ or $[1,\infty)$.

Proof: Hardy, Littlewood and Polya [6, p. 69] give a proof under more stringent regularity assumptions; the present version requires only minor technical modifications in their argument.

For any real number β, define the diversity index of order β, denoted Δ_β, as the dichotomous index whose rarity measure $R(\pi)$ is given by Theorem 1.2. Explicitly,

$$\Delta_\beta = (1 - \Sigma\pi_i^{\beta+1})/\beta,$$

where the usual limiting convention is understood when $\beta = 0$. It will be shown in Section 1.4 that a restriction ($\beta \geq -1$) must be imposed if Δ_β is to have certain desirable properties.

In Section 1.8.2 we also need the index Δ_∞ obtained from the degenerate solution of (1.3). This index assigns diversity zero to a single-species community and diversity one to all other communities.

1.3.3 The Hurlbert-Smith Index. Let ω be a nonnegative integer. The Hurlbert-Smith index of order ω is the expected number of species obtained when $\omega + 1$ individuals are randomly selected from the community.

The definition was given by Hurlbert [5] and by Smith [7] for sampling without and with replacement, respectively. Only the case of sampling with replacement will be considered here. It is also convenient to subtract one from the expected number of species so that a single-species community has diversity zero. With this change, the Hurlbert-Smith index takes the form

$$\begin{aligned}\Delta_\omega^{H-S} &= \Sigma(1 - \pi_i)[1 - (1 - \pi_i)^\omega] \\ &= s - 1 - \Sigma'(1 - \pi_i)^{\omega+1},\end{aligned}$$

where the second summation ranges over the nonzero components of $\underset{\sim}{\pi}$. Some special cases are tabulated below:

ω:	0	1	∞
Δ_ω^{H-S}:	0	Simpson	$s - 1$.

For a given community C, the plot of $\Delta_\omega^{H-S}(C)$ versus ω is the familiar species-individual curve. Empirical species-individual curves are generally concave and one may expect the same for the Hurlbert-Smith index. For the proof, it is convenient to let the parameter ω assume arbitrary nonnegative real values.

Theorem 1.3: Both $\Delta_\beta(C)$ and $\Delta_\omega^{H-S}(C)$ are identically zero when C is a single-species community. Otherwise, $\Delta_\beta(C)$ is a strictly decreasing, strictly convex function of β with $\lim \Delta_\beta(C) = 0$ when $\beta \to \infty$; and $\Delta_\omega^{H-S}(C)$ is a strictly increasing, strictly concave function of ω. In addition,

$$\omega\, \Delta_1 < \Delta_\omega^{H-S} < \Delta_1 \quad \text{if} \quad 0 < \omega < 1,$$

and

The desired result now follows by taking the 1's complement of these inequalities, multiplying by $1 - \pi_i$, and summing over the nonzero components of $\underset{\sim}{\pi}$.

Remark 1.2: The shape of the empirical species-individual curve depends on the sampling scheme as well as the underlying community structure. The preceding theorem justifies a concave shape only for multinomial sampling. A similar analysis shows that the species-area curve is strictly increasing and strictly concave for Poisson sampling. It is perhaps well to point out that the sampling considered in the definition of the Hurlbert-Smith index is conceptual. There is no requirement that it be physically possible to draw a random sample.

The Hurlbert-Smith index is of the dichotomous type with rarity measure $R(\pi) = (1 - \pi)[1 - (1 - \pi)^{\omega}]/\pi$. In this form $R(\pi)$ seems to have no obvious interpretation. However, there is a general unifying scheme for the construction of rarity measures which yields those of both Δ_{β} and Δ_{ω}^{H-S} after imposing the standardizing requirements (1.1). Consider once again Wallace's traveler who initially encounters a member of species i and subsequently encounters X additional individuals where X is a positive integer valued random variable. Define the type I rarity measure to be the probability that a new species is encountered, i.e., the probability that at least one of the X additional individuals belongs to a species different from i. A type II rarity measure, on the other hand, is the probability that each of the additional individuals belongs to species different from i.

Both types of rarity measures are readily expressed in terms of the probability generating function G(t) of X as follows:

	$R(\pi)$	
Type	Raw Form	Standardized Form (1.1)
I	$1 - G(\pi)$	$[1 - G(\pi)]/G'(1) = [1 - G(\pi)]/E[X]$
II	$G(1 - \pi)$	$G(1 - \pi)/G'(0) = G(1 - \pi)/P(X = 1)$.

Notice that, in the raw form, the two types of rarity measures are duals of one another as defined in Section 1.3.1. The type II rarity measure cannot be standardized unless $P(X = 1) > 0$, but it will be shown later that this is a necessary requirement if the associated index is to possess certain desirable properties. See Section 1.3 and Section 1.5.

Particular rarity measures may now be constructed by specializing on the distribution of X. Three cases will be considered by first letting the distribution be degenerate, then uniform discrete and, finally, log series. These give rise, respectively, to the index Δ_β, the Hurlbert-Smith index and, after a limit, the Shannon index.

Letting X be degenerate at the positive integer β, the generating function is $G(t) = t^\beta$ and we obtain the following rarity measures:

	$R(\pi)$	
Type	Raw Form	Standardized Form (1.1)
I	$1 - \pi^\beta$	$(1 - \pi^\beta)/\beta$
II	$(1 - \pi)^\beta$	Not possible unless $\beta = 1$.

Here, Δ_β is the index associated with the standardized type I measure. The raw form of the type II measure will be used to generate counter-examples in Section 1.4.

Next, let X be a discrete uniform random variable with support $\{1, 2, 3, \ldots, \omega\}$. Since the generating function in this case is $G(t) = \omega^{-1}t(1 - t^{\omega})/(1 - t)$, we obtain:

Type	$R(\pi)$ Raw Form	 Standard Form (1.1)
I	$1 - \omega^{-1}\pi(1 - \pi^{\omega})/(1 - \pi)$	$2^{-1}\omega(\omega + 1)[1 - \omega^{-1}\pi(1 - \pi^{\omega})/(1 - \pi)]$
II	$\omega^{-1}(1 - \pi)[1 - (1 - \pi)^{\omega}]/\pi$	$(1 - \pi)[1 - (1 - \pi)^{\omega}]/\pi$

In this case, the Hurlbert-Smith index is associated with the standardized type II rarity measure. Also, letting $\omega \to 0$ in the two raw forms gives the unfamiliar indices of Section 1.3.1. Conceptually ω is an integer but real values make mathematical sense.

Finally, let X have a log series distribution with parameter θ and generating function $G(t) = \log(1 - \theta t)/\log(1 - \theta)$. The standardized type II rarity measure is then $R(\pi) = -\log(1 - \theta + \theta\pi)/\theta$ which converges to the rarity measure of the Shannon index when $\theta \to 1$. Note that the number X of additional encounters converges to infinity under this limit.

Other types of rarity measures could be defined as the probability of particular events associated with the random variable X. Usually such a probability turns out to be a function of the entire vector $\underset{\sim}{\pi}$ and not merely π_i where i is the initially encountered species. The resulting rarity measure then has the general form $R(i;\underset{\sim}{\pi})$ and is not of the dichotomous type.

1.3.4 The Numbers Equivalent. MacArthur [4] has introduced a method for transforming diversity indices to a species richness scale. For a given community C and a given index Δ, he poses the question: how many species must a completely even community have in order that its diversity be $\Delta(C)$? This number is called the *equivalent number of species* and is denoted by $S_\Delta(C)$. The same transformation has appeared in the economics literature as the *numbers equivalent* [8, 9, 10]. We will use the latter term.

The numbers equivalent remains invariant under one-one transformations of the index, i.e., $S_{f(\Delta)} = S_\Delta$ whenever f is one-one. For a dichotomous index, $\Delta = \Sigma\pi_i R(\pi_i)$, the numbers equivalent is any solution of the equation $R(1/S_\Delta) = \Delta$; assuming that R is continuous and strictly monotonic, the solution is unique and may be written as

$$S_\Delta = 1/R^{-1}(\Sigma\pi_i R(\pi_i)). \tag{1.4}$$

The denominator of the right hand side of (1.4) has the form of a generalized average with R as the Kolmogorov-Nagumo function [6, chap. 3; 11]. Specializing to the index Δ_β, the numbers equivalent,

$$S_\beta = 1/(1 - \beta\Delta_\beta)^{1/\beta} = 1/(\Sigma\pi_i\cdot\pi_i^\beta)^{1/\beta}, \tag{1.5}$$

is the reciprocal of the generalized mean of order β of $\pi_1, \pi_2, \pi_3, \ldots$ with weights $\pi_1, \pi_2, \pi_3, \ldots$; in particular, the numbers equivalents of the indices Δ_{-1}, Δ_0 and Δ_1 are the respective reciprocals of the harmonic mean, the geometric mean and the arithmetic mean of $\pi_1, \pi_2, \pi_3, \ldots$ with weights $\pi_1, \pi_2, \pi_3, \ldots$.

For a given community, the indices S_β have the interesting property of being a monotone decreasing function of β. The reason for this effect is clear from (1.5): with β large, the rare species make only a negligible contribution to S_β. In the limit as β approaches ∞, S_β converges to $1/\pi_1^*$ -- the reciprocal of the abundance of the single most dominant species. Hill [12] has plotted S_β versus distance along a line transect and has found that, to a remarkable degree, the plots have the same shape irrespective of β.

In general, it is not possible to obtain simple closed form expressions for the numbers equivalent of the Hurlbert-Smith index.

The logarithmic numbers equivalent $\log(S_\Delta)$ is sometimes useful. For the index Δ_β,

$$\log(S_\beta) = -\log(1 - \beta\Delta_\beta)/\beta = -\log(\Sigma\pi_i^{\beta+1})/\beta$$

is Renyi's [11] entropy of order $\beta + 1$. Within the family of dichotomous indices, the Shannon index is characterized as being its own logarithmic numbers equivalent.

1.4 Two Criteria for Diversity Indices

Recalling the definition of dichotomous indices, $\Delta(C) = \Sigma\pi_i R(\pi_i)$, observe that $R(0)$ is inherently undefined while the value $R(1)$ is germane only to a single-species community, and, in fact, equals the diversity of such a community. $R(1) = 0$ is a natural normalizing requirement.

What else might be required of R? On intuitive grounds, $R(\pi)$ should be a decreasing function of π since rarer species correspond to smaller values of π.

Criterion C1: R is a decreasing function defined on the interval (0,1]. If the normalizing condition $R(1) = 0$ is also imposed, R will, as a consequence, be nonnegative.

This monotonicity requirement, simple and intuitive though it is, has a striking implication. Consider two communities $C = (s, \underset{\sim}{\pi})$ and $C' = (s', \underset{\sim}{\pi})$. We say that C leads to C' by introducing a species (see Figure 1.3) if $s' = s + 1$ and if there are two distinct positive integers i and j such that

$$\pi'_k = \begin{cases} \pi_k & \text{if } k \neq i,j \\ \pi_i - h & \text{if } k = i \\ h & \text{if } k = j \end{cases}$$

where $0 < h < \pi_i$. Note that $\pi_j = 0$. A possible biological interpretation is that species i shares its resources with a newly arrived competing species.

Theorem 1.4: Assume $R(\pi)$ is decreasing in π. Then introducing a species increases the diversity measure of a community.

Proof: By assumption $\pi_i > \pi_i - h > 0$ and $\pi_i > h > 0$ so that $R(\pi_i - h) \geq R(\pi_i)$ and $R(h) \geq R(\pi_i)$. But

$$\begin{aligned} \Delta(C') - \Delta(C) &= (\pi_i - h)R(\pi_i - h) + hR(h) - \pi_i R(\pi_i) \\ &= h\,[R(h) - R(\pi_i - h)] + \pi_i[R(\pi_i - h) - R(\pi_i)] \\ &\geq h\,[R(h) - R(\pi_i - h)] + h[R(\pi_i - h) - R(\pi_i)] \\ &= h[R(h) - R(\pi_i)] \\ &\geq 0. \end{aligned}$$

Any community with finitely many species can be constructed from a single-species community by successively introducing new species

Figure 1.3 Introducing a Species

(see Figure 1.4). Theorem 1.4 asserts that the diversity increases at each step. None the less, indices satisfying Criterion C1 may have undesirable properties, as illustrated by the next example.

Example 1.1: Let $R(\pi) = 1/\pi^2 - 1$ and $\Delta = \Sigma_1^s(1/\pi_i^*) - 1$. This index satisfies Criterion C1 and assigns diversity zero to a single-species community. Figure 1.5 includes a plot of the values of Δ for communities with ranked abundance vector $(1 - \pi, \pi)$, $0 \leq \pi \leq 1/2$. The point A represents a single-species community, while B, C and D represent successively more even two-species communities. In accord with Theorem 1.4, each of B, C and D is more diverse than A. But among the two-species communities, the diversity, as measured by Δ, decreases as the evenness increases.

In going from B to C to D in Example 1.1, the change in community composition may be described as a transfer of abundance from one species to another strictly less abundant species. The next definition formalizes this operation for many species communities.

Definition 1.2: Let $C = (s, \underset{\sim}{\pi})$ and $C' = (s', \underset{\sim}{\pi}')$ be two communities. C leads to C' by a transfer of abundance if $s = s'$ and if there are positive integers i and j such that $\pi_i > \pi_j > 0$ and

$$\pi'_k = \begin{cases} \pi_k & \text{if} \quad k \neq i,j \\ \pi_i - h & \text{if} \quad k = i \\ \pi_j + h & \text{if} \quad k = j \end{cases}$$

where $0 < h < \pi_i - \pi_j$.

For mathematical purposes, it is sometimes convenient to consider the operation of introducing a species as a special case of transferring abundance: in Definition 1.2, delete the requirement that $s = s'$ and

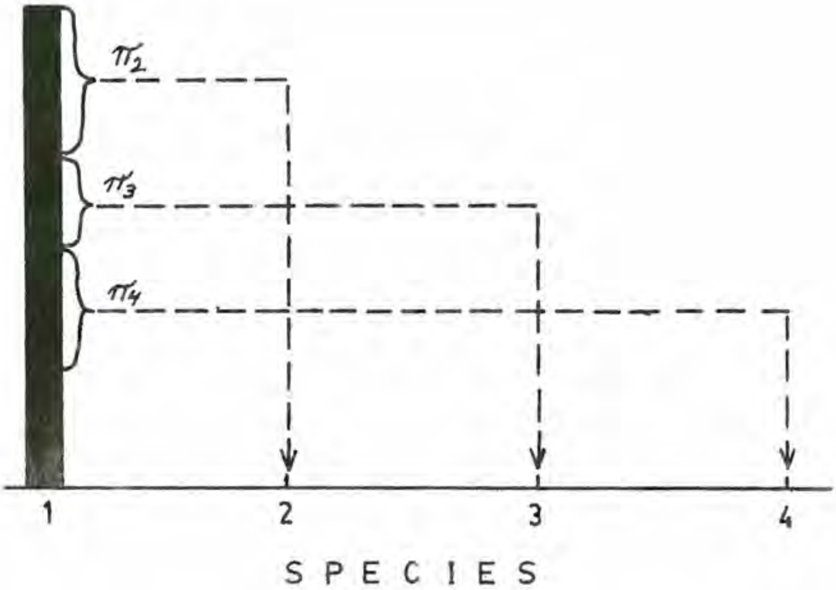


Figure 1.4 Construction of a Four-Species Community From a Single-Species Community

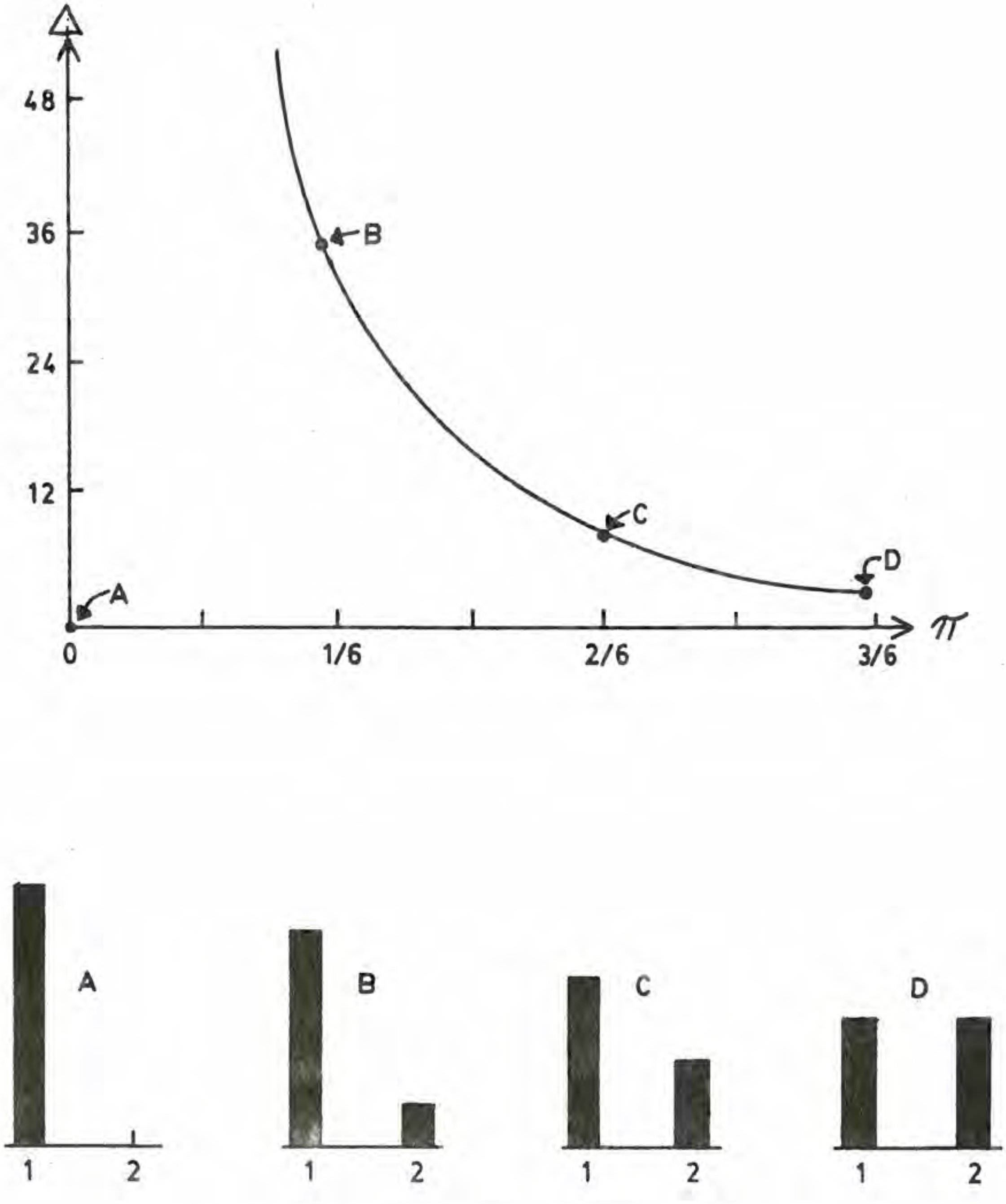


Figure 1.5 Inadequacy of Criterion C1

and permit π_j to be zero. The two operations are conceptually quite different, however; introducing a species increases the "species richness" component of diversity while transferring abundance increases the "evenness" component.

Criterion C2: $\Delta(C) \leq \Delta(C')$ whenever C leads to C' either by introducing a species or by transferring abundance.

Remark 1.3: The requirement that transferring abundance should increase the index is known in the economics literature as Dalton's [13] "principle of transfers" and was originally proposed in connection with the measurement of income inequality. We discuss income inequality in Section 1.12.

While Criterion C1 is meaningful only for dichotomous indices, Criterion C2 makes sense for any proposed index irrespective of its functional form. But Criterion C2 should then be supplemented by the requirement that $\Delta(\underset{\sim}{\pi})$ be invariant to permutations of $\underset{\sim}{\pi}$. Note that permutation invariance is automatic for dichotomous as well as rank type indices.

It is an obvious (but useful) fact that Criterion C2 is preserved under monotone transformations of the index. In particular, when an index satisfies Criterion C2 so does its numbers equivalent.

To state conditions under which the two criteria will be satisfied, it is convenient to define an auxiliary function V by

$$V(\pi) = \begin{cases} \pi R(\pi) & \text{if} \quad 0 < \pi \leq 1 \\ 0 & \text{if} \quad \pi = 0. \end{cases}$$

For the index s = 1, $V(\pi) = 1 - \pi$ for positive π, which shows that V may be discontinuous at the origin.

Theorem 1.5: Criterion C2 is satisfied <=>

$$V(\pi_j + h) - V(\pi_j) \geq V(\pi_i - h) - V(\pi_i) \qquad (1.6)$$

whenever $\pi_i > \pi_i - h \geq \pi_j + h > \pi_j \geq 0$ and $\pi_i + \pi_j \leq 1$. Assuming differentiability of V, (1.6) may be replaced by $V'(\pi_j) \geq V'(\pi_i)$ whenever $\pi_i > \pi_j \geq 0$ and $\pi_i + \pi_j \leq 1$.

Proof: Straightforward.

Example 1.2: The condition (1.6) has a simple geometric interpretation, as illustrated by Figure 1.6 for communities with abundance vectors (π_1, π_2, π_3). Points on the vertices, edges, and interior of the triangle represent respectively single-species, two-species, and three-species communities. The centroid of the triangle represents the completely even three-species community. Condition (1.6) requires that the diversity increase as a point moves toward the center of the triangle along any line segment parallel to an edge. The arrows on the segment AB in the figure indicate the direction of increasing diversity.

Theorem 1.6: Criterion C1 and Criterion C2 are both satisfied if V is concave on the closed unit interval [0,1].

Proof: Criterion C1: Let $0 < x < y \leq 1$. Observe that $V(x)/x$ is the slope of the line from the origin to the point $(x, V(x))$. The concavity of V implies that this slope is a decreasing function of x. Hence $V(x)/x \geq V(y)/y$ and $R(x) \geq R(y)$. Criterion C2: Referring to Figure 1.7, the condition (1.6) requires that the line L should have greater slope than the line L'. But this is a well known consequence of concavity.

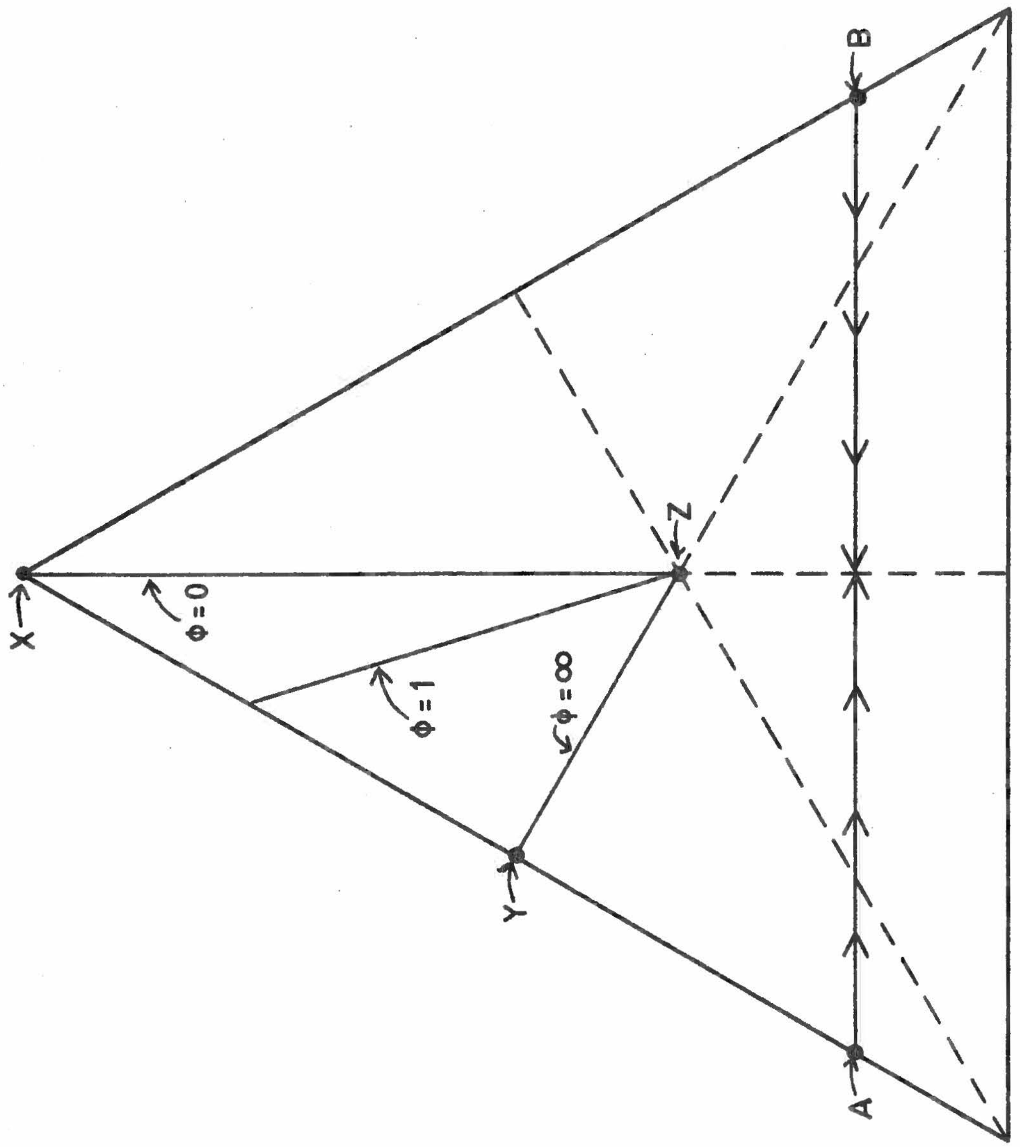


Figure 1.6 Species Abundance Simplex

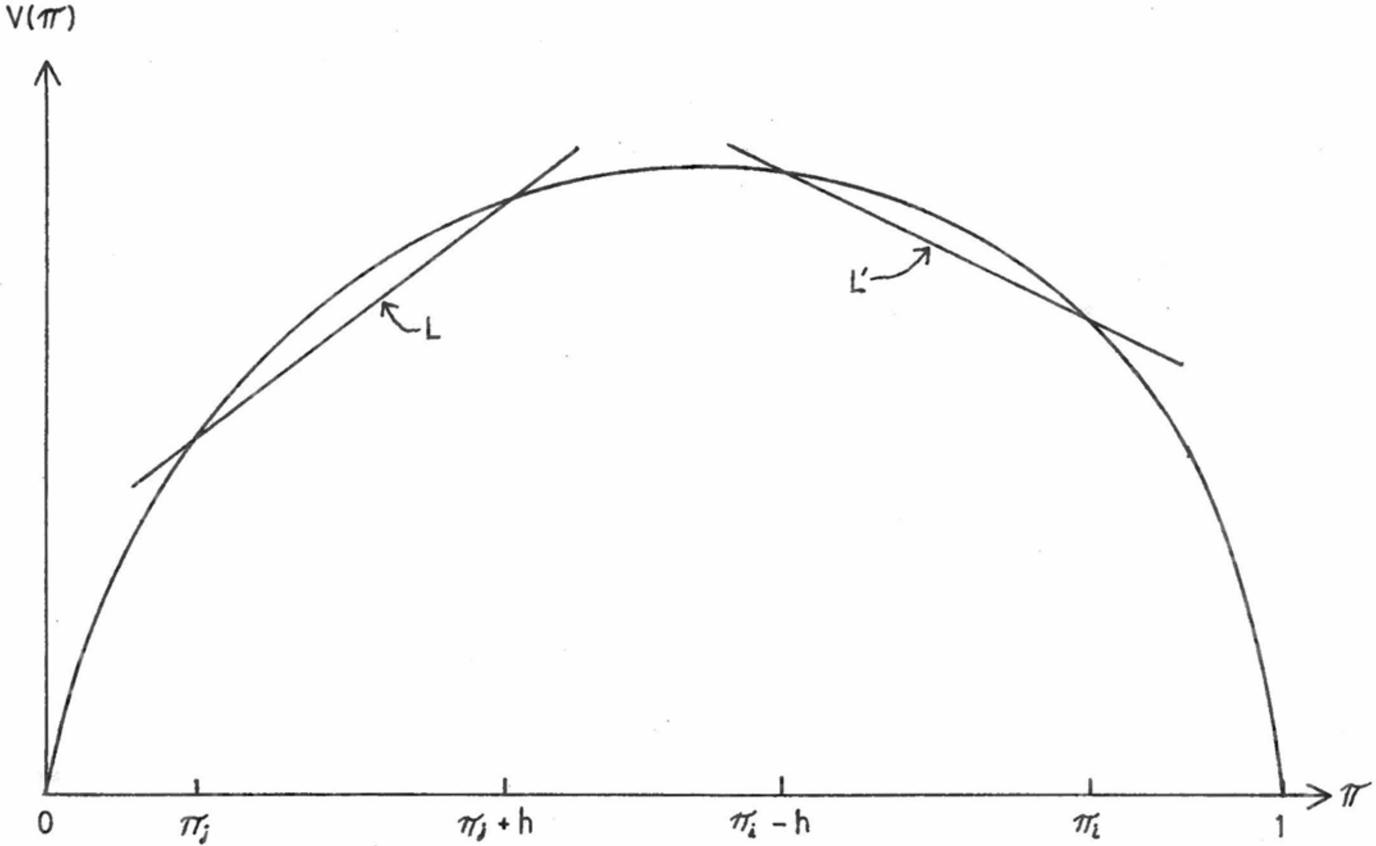


Figure 1.7 A Concave Auxiliary Function and Two Secant Lines L and L'

Remark 1.4: Because of the constraint $\pi_i + \pi_j \leq 1$, the converse of Theorem 1.6 is not quite true. For a counterexample take $R(\pi) = (1 - \pi)^{\beta}$ with $\beta = 2$ or 3. However, Criterion C2 does require that $V(\pi)$ be concave for $0 \leq \pi \leq 1/2$.

Corollary 1.1: The index Δ_{β} satisfies Criterion C1 for all real β. Criterion C2 is satisfied <=> V is concave <=> $\beta \geq -1$. In particular, the Species Count, the Shannon index, and the Simpson index satisfy both criteria.

Corollary 1.2: When $\omega \geq 0$, the Hulbert-Smith index Δ_{ω}^{H-S} has a concave auxiliary function and satisfies both Criterion C1 and Criterion C2.

Corollary 1.3: Any index based on a type I rarity measure has a concave auxiliary function and satisfies both Criterion C1 and Criterion C2.

Proof: Apply Theorem 1.6 using the fact that a probability generating function is increasing and convex.

Corollary 1.4: An index based on a type II rarity measure satisfies Criterion C1 but not necessarily C2. A sufficient condition for the auxiliary function to be concave is that $P(X = i)$ be decreasing, i.e., $P(X = 1) \geq P(X = 2) \geq P(X = 3) \geq \ldots$ while $P(X = 1) > 0$ is a necessary condition. (The random variable X is defined in Section 1.3.3.)

Proof: Criterion C1 is satisfied since $R(\pi) = G(1 - \pi)$ is decreasing in π. For a conterexample to Criterion C2, take $R(\pi) = (1 - \pi)^{\beta}$ ($\beta = 4, 5, 6, \ldots$) and check that the auxiliary function is not concave for $0 \leq \pi \leq 1/2$. To prove the last assertion, let $V(\pi) = \pi G(1 - \pi)$ where $G(t) = \sum_{i=1}^{\infty} p_i t^i$ is the generating function of X. Now, the second derivative of $V(\pi)$ is

$$V''(\pi) = -2G'(1-\pi) + \pi G''(1-\pi)$$

$$= -\sum_{i=0}^{\infty} (i+1)(i+2)(p_{i+1} - p_{i+2})(1-\pi)^i. \quad (1.7)$$

Clearly $V''(\pi) \leq 0$ when $p_1 \geq p_2 \geq p_3 \geq \ldots$. On the other hand, if $p_1 = 0$ the first nonzero term in the expansion (1.7) has a positive coefficient and $V''(\pi)$ will be positive for π sufficiently close to one.

Example 1.3: Good [14] has proposed the family of indices

$$C(m,n) = \sum_{i=1}^{s} (\pi_i^*)^m (-\log \pi_i^*)^n \; (m,\, n = 0,1,2,\ldots) \; .$$

Criterion C2 is satisfied by only three members of this family: C(1,1) is the Shannon index; C(0,0) is the number of species; C(1,0) is identically one. The special case C(2,0) is the 1's complement of the Simpson index. Thus -C(2,0) satisfies Criterion C2. More generally, the only members of Good's family whose negatives satisfy Criterion C2 are C(m,0) (m = 1, 2, 3,...). Note that the affine transformation $[1 - C(m,0)]/(m-1)$ converts C(m, 0) into Δ_{m-1} and also gives the Shannon index as the limiting case when m = 1.

1.5 Diversity Ordering

A numerical-valued diversity index imposes a linear ordering on the diversity of ecological communities. But, as emphasized by Hurlbert [5], different indices may give inconsistent orderings. For example, the Species Count and the Simpson index order the two communities $C = (2, (.5,0,.5))$ and $C' = (3, (.1, .8, .1))$ in the opposite sense:

$$\Delta_{-1}(C) = 1 < 2 = \Delta_{-1}(C')$$
$$\Delta_1(C) = .5 > .34 = \Delta_1(C').$$

Inconsistent measures such as these are a familiar problem and should not be a cause for undue pessimism. The mean and the median are inconsistent measures of "central tendency"; the variance, the mean absolute deviation, and the range are inconsistent measures of "spread." However, in view of these inconsistencies, it becomes of interest to define an intrinsic diversity ordering without reference to indices. We propose the following:

Definition 1.3: Community C' is *intrinsically more diverse* than community C (written $C' \geq C$) provided C leads to C' by a finite sequence of the following operations:

1. Introducing a species;
2. Transferring abundance;
3. Relabeling species, i.e. permuting the components of the abundance vector.

It is easy to see that this intrinsic diversity ordering is a *partial order* in the sense that (i) $C \geq C$; (ii) $C'' \geq C$ whenever $C'' \geq C'$ and $C' \geq C$; and (iii) a given pair of communities need not be comparable. Solomon [15] has also proposed a diversity ordering based on the notion of majorization [16]. A fundamental theorem of Hardy, Littlewood, Polya, and Rado [6, 17, 18, 19, 20] shows that Solomon's definition is equivalent to ours. For completeness we sketch a proof.

Theorem 1.7: Let $C = (s, \underset{\sim}{\pi})$ and $C' = (s', \underset{\sim}{\nu})$ be two communities. Then the following statements are equivalent.

a) C' is intrinsically more diverse than C.

b) $\underset{\sim}{\nu}$ is a convex linear combination of permutations of $\underset{\sim}{\pi}$.

c) $\underset{\sim}{\nu}^*$ is stochastically greater than $\underset{\sim}{\pi}^*$, i.e.,

$$\sum_{i>k} \nu_i^* \geq \sum_{i>k} \pi_i^*, \; k = 1,2,3,\ldots .$$

d) $\underset{\sim}{\pi}^*$ majorizes $\underset{\sim}{\nu}^*$, i.e.,

$$\sum_{i\leq k} \pi_i^* \geq \sum_{i\leq k} \nu_i^*, \; k = 1,2,3,\ldots \; .$$

Proof: For the implication a) => b), we suppose, by induction, that C leads to C' by one of the three operations of Definition 1.3. If $\underset{\sim}{\nu}$ is a permutation of $\underset{\sim}{\pi}$, there is nothing to prove; otherwise there are two positive integers $i \neq j$ such that $\pi_k = \nu_k$ for $k \neq i, j$ and $\nu_i = \pi_i - h$ and $\nu_j = \pi_j + h$ where $0 < h < \pi_i - \pi_j$. Now ν_i lies between π_i and π_j and so may be written as a convex linear combination, $\nu_i = (1 - w)\pi_i + w\pi_j$. By symmetry, $\nu_j = (1 - w)\pi_j + w\pi_i$. Thus $\underset{\sim}{\nu} = (1 - w)\underset{\sim}{\pi} + w\underset{\sim}{\pi}'$ where $\underset{\sim}{\pi}'$ is the same as $\underset{\sim}{\pi}$ but with components i and j interchanged. To prove the implication b) => c), notice that

$$\sum_{i>k} \pi_{\sigma i} \geq \sum_{i>k} \pi_i^* \tag{1.8}$$

for any permutation σ. But $\underset{\sim}{\nu}$ (and hence $\underset{\sim}{\nu}^*$) is a convex linear combination of permutations of $\underset{\sim}{\pi}$,

$$\nu_i^* = \sum_{\sigma} w_\sigma \pi_{\sigma i} \; . \tag{1.9}$$

Sum (1.9) for $i > k$ and use (1.8) to conclude that $\underset{\sim}{\nu}^*$ is stochastically greater than $\underset{\sim}{\pi}^*$. Since the implication c) => d) is clear, it remains only to prove that d) => a). Suppose that $\underset{\sim}{\pi}^*$ majorizes $\underset{\sim}{\nu}^*$. If $\pi_i^* \geq \nu_i^*$ for all i, then $\underset{\sim}{\nu}^* = \underset{\sim}{\pi}^*$; otherwise, let k be the smallest positive integer for which $\nu_k^* > \pi_k^*$. By the majorization property

$$\sum_{i<k} (\pi_i^* - \nu_i^*) \geq \nu_k^* - \pi_k^*,$$

and so we may successively transfer abundance from $\pi_{k-1}^*, \pi_{k-2}^*, \ldots, \pi_1^*$ to π_k^* until π_k^* equals ν_k^*. Moreover, the inequalities $\pi_i^* \geq \nu_i^*$, $i < k$, can be maintained throughout the transfers. Proceeding by induction, it follows that $\underset{\sim}{\nu}^*$ is obtainable from $\underset{\sim}{\pi}^*$ by finitely many transfers of abundance or (when $\pi_k^* = 0$) introductions of species. Since $\underset{\sim}{\nu}$ and $\underset{\sim}{\pi}$ are permutations of $\underset{\sim}{\nu}^*$ and $\underset{\sim}{\pi}^*$, the proof is complete.

Remark 1.5: A fifth statement, namely, that $\underset{\sim}{\nu}$ be a doubly stochastic transform of $\underset{\sim}{\pi}$, can be added to the list of equivalences in Theorem 1.7. See Mirsky [18].

With the help of Theorem 1.7(b), the intrinsic diversity ordering may be conveniently represented on the abundance simplex. Figure 1.8 shows a three-species community C together with the regions of the simplex which are intrinsically more or intrinsically less diverse than C. Also shown are the contours through C of the indices Δ_β. Any community whose abundance vector lies between two of these contours would be inconsistently ordered by the corresponding indices. As β decreases toward -1, the contours become increasingly elongated toward the vertices of the simplex and sweep out a region whose boundaries are labeled $\beta = \infty$ and $\beta = -1$. (The index Δ_{-1} does not have well-defined contours. The boundary labeled $\beta = -1$ has defining equation $\pi_1 \pi_2 \pi_3 = \text{constant}$, while the boundary $\beta = \infty$ is defined by the equation $\pi_1^* = \text{constant}$.) The contours do not sweep out the entire region of incomparability and there are communities which are consistently ordered by the entire family $\{\Delta_\beta : \beta \geq -1\}$ even though they are not intrinsically comparable. The communities labeled C and C" in Figure 1.8 are an example. A family of indices will be called complete if C is intrinsically more diverse than C" $\Leftrightarrow \Delta(C) \geq \Delta(C'')$ for all indices Δ in the family. Thus $\{\Delta_\beta : \beta \geq -1\}$ is not complete.

Figure 1.8 Geometric Representation of the Intrinsic Diversity Ordering, Together with Some Contours of Δ_β and Δ_ω^{H-S} ($C = (3, (.49, .36, .15))$ and $C'' = (3, (.62, .21, .17))$)

$$\Delta_1 < \Delta_\omega^{H-S} < \omega\,\Delta_1 \quad \text{if } 1 < \omega < \infty,$$

provided C is not a single-species community.

Proof: The first statement is clear and it is easy to see that $\lim \Delta_\beta(C) = 0$ when $\beta \to \infty$. Now, the second derivative of Δ_β with respect to β is

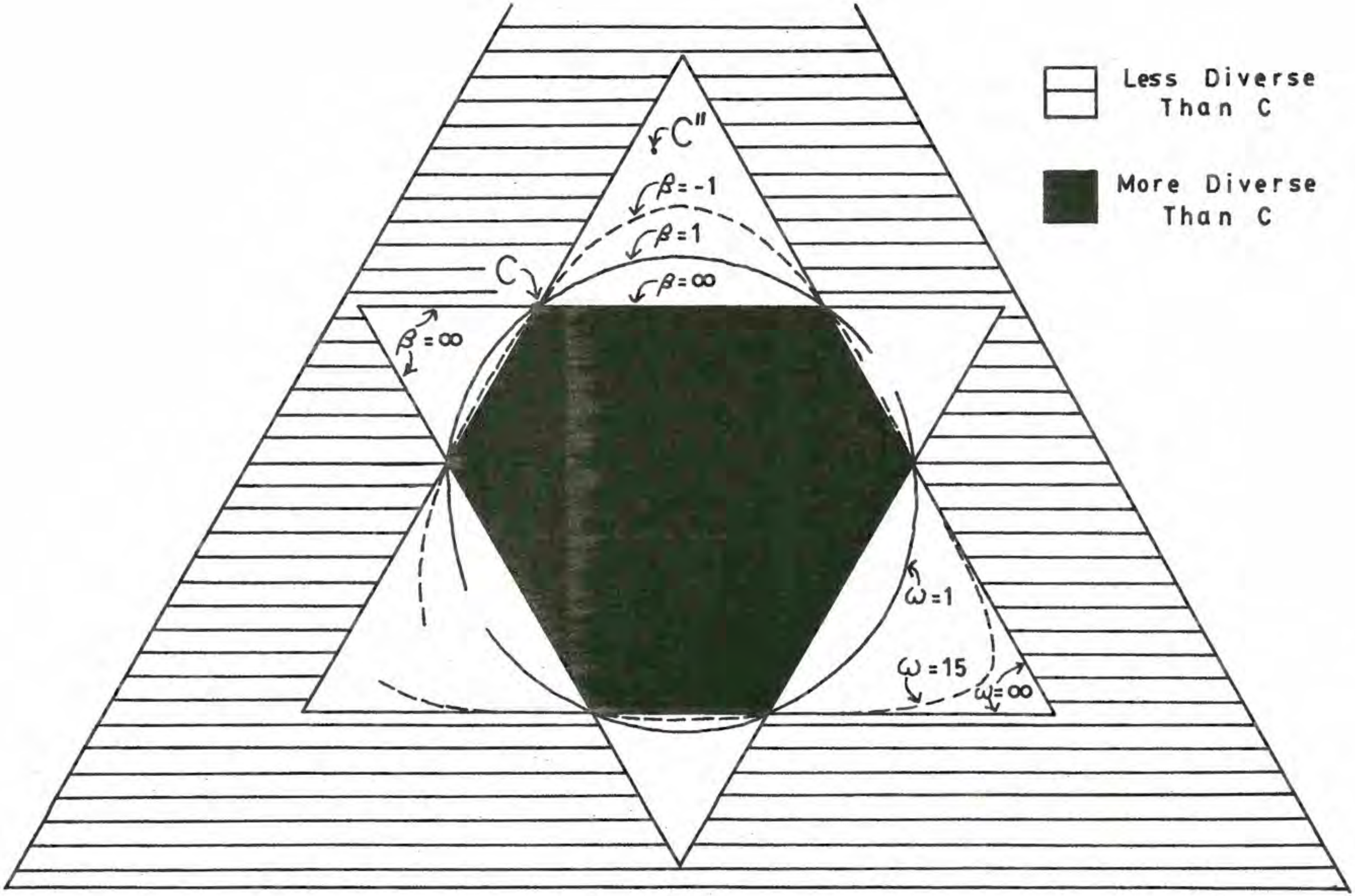
$$\beta^{-3}\Sigma\pi_i[-\beta^2\pi_i^\beta \log^2(\pi_i) + 2\beta\pi_i^\beta\log(\pi_i) - 2\pi_i^\beta + 2] =$$
$$\beta^{-3}\Sigma\pi_i[-\pi_i^\beta \log^2(\pi_i^\beta) + 2\pi_i^\beta \log(\pi_i^\beta) - 2\pi_i^\beta + 2].$$

Write $x = \log(\pi_i^\beta)$ and $f(x) = -x^2e^x + 2xe^x + 2$. It suffices to show that $f(x)$ has the same sign as β, i.e., the sign opposite to x. But $f(0) = 0$ and $f'(x) = -x^2e^x$ which is strictly negative when $x \neq 0$. This proves that Δ_β is strictly convex. But a nonnegative, strictly convex function which is asymptotically zero must be strictly decreasing. As for the Hurlbert-Smith index, the first derivative with respect to ω is $-\Sigma(1 - \pi_i)^{\omega+1}\log(1 - \pi_i) > 0$ while the second derivative is $-\Sigma(1 - \pi_i)^{\omega+1}\log^2(1 - \pi_i) < 0$. Pictorially, the last two assertions of the theorem are obvious. A formal proof may be given by noting that, as a function of π $(0 \leq \pi \leq 1)$, $(1 - \pi)^\omega$ is strictly concave when $0 < \omega < 1$ and strictly convex when $1 < \omega < \infty$. For $0 < \pi_i < 1$, this implies that

$$1 - \omega\pi_i > (1 - \pi_i)^\omega > 1 - \pi_i \text{ if } 0 < \omega < 1,$$

and

$$1 - \pi_i > (1 - \pi_i)^\omega > 1 - \omega\pi_i \text{ if } 1 < \omega < \infty.$$


Less Diverse Than C
More Diverse Than C
C''
β = -1
β = 1
β = ∞
C
β = ∞
ω = 1
ω = 15
ω = ∞

The contours of the Hurlbert-Smith index are also shown in Figure 1.8. As $\omega \to 0$ there is a limiting contour with, in this case, $\Sigma(1 - \pi_i)\log(1 - \pi_i)$ = constant as the defining equation. The limiting contour is not shown in Figure 1.8 since it is graphically indistinguishable from the contour labeled $\omega = 1$.

In view of Theorem 1.7(b), Criterion C2 may be restated as the requirement that $\Delta(\underset{\sim}{\nu}) \geq \Delta(\underset{\sim}{\pi})$ whenever $\underset{\sim}{\nu}$ is a convex linear combination of permutations of $\underset{\sim}{\pi}$. Functions which satisfy this property are sometimes called Schur Concave [16, 18]. Lewontin [21] has suggested the following somewhat stronger requirement.

Criterion C3: $\Delta(\underset{\sim}{\pi})$ is a permutation invariant concave function of the abundance vector $\underset{\sim}{\pi}$.

While we know of no intuitive motivation for this criterion, it will be essential to the mixture decomposition of Section 1.8.3. Criterion C3, unlike Criterion C2, is not preserved under increasing transformations of the index. For example, the numbers equivalent of the Simpson index, $1/\Sigma\pi_i^2$, satisfies Criterion C2 but not Criterion C3. We also note that the conclusion of Theorem 1.6 can be strengthened as follows:

Theorem 1.8: Criterion C3 is satisfied if the auxiliary function V is concave on the closed unit interval [0,1]. In particular, Δ_β satisfies Criterion C3 <=> $\beta \geq -1$. Also Δ_ω^{H-S} satisfies this criterion when $\omega \geq 0$.

Proof: Straightforward.

By Corollary 1.3, Criterion C3 is met by any index based on a type I rarity measure, but, for a type II measure, it is necessary that $P(X = 1) > 0$, which is also necessary to standardize the index.

1.6 Indices Based on Ranking

The indices discussed in Section 1.3 were based on dichotomous measures of rarity. A measure of species rarity, with a more detailed dependence upon community composition, is the number of more abundant species. For the ith ranked species, this number is i - 1 and average community rarity becomes $\Sigma(i - 1)\pi_i^* = \Sigma i\pi_i^* - 1$ = Average Rank - 1. Solomon [15], from quite a different point of view, has introduced the Average Rank as a diversity index. (The -1 has the effect of assigning diversity zero to a single-species community and appears to be a generally useful convention.)

A related index is Fager's [22] "Number of Moves" which is, in effect, the Average Rank rescaled to range between zero and one. However, Peet [23] has given persuasive arguments against rescaling diversity indices. Fager's basic idea is attractive, though. As an alternative to Fager's number of moves needed to convert a sample to an even distribution, one may consider the "work (= mass x distance)" required to construct a given community from a single-species community. This "work" is seen to be Average Rank - 1.

For the general measure of rarity based on ranks, the analogue of Criterion C1 is the requirement that R(i) be an increasing function of i. In contrast with the case of dichotomous indices, this monotonicity is sufficient for Criterion C2. In fact,

Theorem 1.9: For the rank type index of form $\Delta(\underset{\sim}{\pi}) = \Sigma R(i)\pi_i^*$, the following are equivalent:

a) R(i) is an increasing function of i.

b) Δ satisfies Criterion C2 (Shur concavity).

c) Δ satisfies Criterion C3 (concavity).

Proof: That c) implies b) is clear. To show that b) implies a), let i be a positive integer and consider any community whose ranked abundance vector satisfies $\pi_i^* > \pi_{i+1}^*$. Transfer a small amount of abundance h from the ith to the (i + 1)th ranked species. By Criterion C2 the change in the index must be nonnegative. Since the change is seen to be $h[R(i + 1) - R(i)]$, we may conclude that $R(i + 1) \geq R(i)$. For the proof that a) implies c), observe that $\Delta(\underset{\sim}{\pi})$ is a well-defined and continuous (in fact, piecewise linear) function for $\underset{\sim}{\pi} \in R^n$. We will prove that Δ is concave on all of R^n. For $\underset{\sim}{x}, \underset{\sim}{y} \in R^n$, let $\underset{\sim}{z}(t) = \underset{\sim}{x} + t\underset{\sim}{y}$, $-\infty < t < \infty$, be the straight line passing through $\underset{\sim}{x}$ with direction vector $\underset{\sim}{y}$. It suffices to show that the second order difference of the function $\Delta(z(t))$ is nonpositive at $t = 0$, i.e., that

$$\Delta(\underset{\sim}{z}(h)) - \Delta(\underset{\sim}{x}) \leq \Delta(\underset{\sim}{x}) - \Delta(\underset{\sim}{z}(-h)) \qquad (1.10)$$

for all sufficiently small positive h. Since Δ is permutation invariant, we may suppose that $\underset{\sim}{x}$ is ranked in descending order ($\underset{\sim}{x} = \underset{\sim}{x}^*$). Define a tied set to be any equivalence class for the equivalence relation $i \sim j \iff x_i = x_j$. (Example: If $x_1 = x_2 = x_3 > x_4 > x_5, \ldots$, then $\{1, 2, 3\}$ and $\{4\}$ are each tied sets.) The vector $\underset{\sim}{z}(h)$ need not be ranked, but when h is small the ranking can be disturbed only within tied sets. Define a permutation σ by $z_i^*(h) = z_{\sigma i}(h)$ for all sufficiently small positive h. Clearly σ maps each tied set into itself and, within each tied set, arranges the direction vector $\underset{\sim}{y}$ in descending order. Now, the left hand side of (1.10) will be compared with the right hand side. Using the fact that $\underset{\sim}{x}$ is ranked, a direct calculation shows that the left hand side is

$$\begin{aligned}\Delta(\underset{\sim}{z}(h)) - \Delta(\underset{\sim}{x}) &= \Sigma R(i)[z_i^*(h) - x_i^*] \\ &= \Sigma R(i)[z_{\sigma i}(h) - x_i] \\ &= \Sigma R(i)[x_{\sigma i} + hy_{\sigma i} - x_i] \\ &= h\Sigma R(i)y_{\sigma i}. \qquad (1.11)\end{aligned}$$

Similarly,

$$\Delta(\underset{\sim}{x}) - \Delta(\underset{\sim}{z}(-h)) = h\Sigma R(i)y_{\tau i} \qquad (1.12)$$

where the permutation τ maps each tied set into itself and, within each tied set, arranges y in ascending order. Now compare (1.11) with (1.12), tied set by tied set, and apply the following lemma of Hardy, Littlewood, and Polya [6, p. 261] to conclude that (1.10) is true.

Lemma 1.1: Let $\underset{\sim}{a}, \underset{\sim}{b} \in R^n$ and assume that $a_1 \leq a_2 \leq \ldots \leq a_n$. Then $f(\sigma) = \Sigma a_i b_{\sigma i}$ is maximized when the permutation σ arranges $\underset{\sim}{b}$ in ascending order and is minimized when σ arranges b in descending order.

Remark 1.6: It follows from Theorem 1.7 and Theorem 1.9 that Δ preserves stochastic ordering provided R(i) is an increasing function of i. Consequently, $\Sigma R(i)\nu_i^* \geq \Sigma R(i)\pi_i^*$ whenever $\underset{\sim}{\nu}^*$ is stochastically greater than $\underset{\sim}{\pi}^*$. This is, of course, a well known property of stochastically ordered distributions. For example, see Lehmann [24, p. 73, p. 112].

The proportionality equation (1.3) has an analogue

$$R(k + i) - R(i) = W(i)[R(k + 1) - R(1)], \; i, \; k = 1,2,3,\ldots,$$

whose solution is $W(i) = \rho^{i - 1}$ and, up to an affine transformation,

$$R(i) = \begin{cases} (1 - \rho^{i-1})/(1 - \rho) & \text{if} \quad \rho \neq 1 \\ i - 1 & \text{if} \quad \rho = 1, \end{cases} \tag{1.13}$$

where ρ is a constant. A necessary and sufficient condition for $R(i)$ to be increasing is that $\rho \geq 0$. The index based on (1.13) will be denoted $\Delta_\rho^{(\text{rank})}$, $\rho \geq 0$. Average Rank - 1 is included as the special case $\rho = 1$ and is the analogue of the Shannon index in the sense that the deflation factor W is identically one for both indices.

In Section 1.12, we need the numbers equivalent of Average Rank - 1 which may be shown to be $2(\Sigma i\pi_i^*) - 1$. In general, however, it is not possible to obtain a closed form expression for the numbers equivalent of $\Delta_\rho^{(\text{rank})}$.

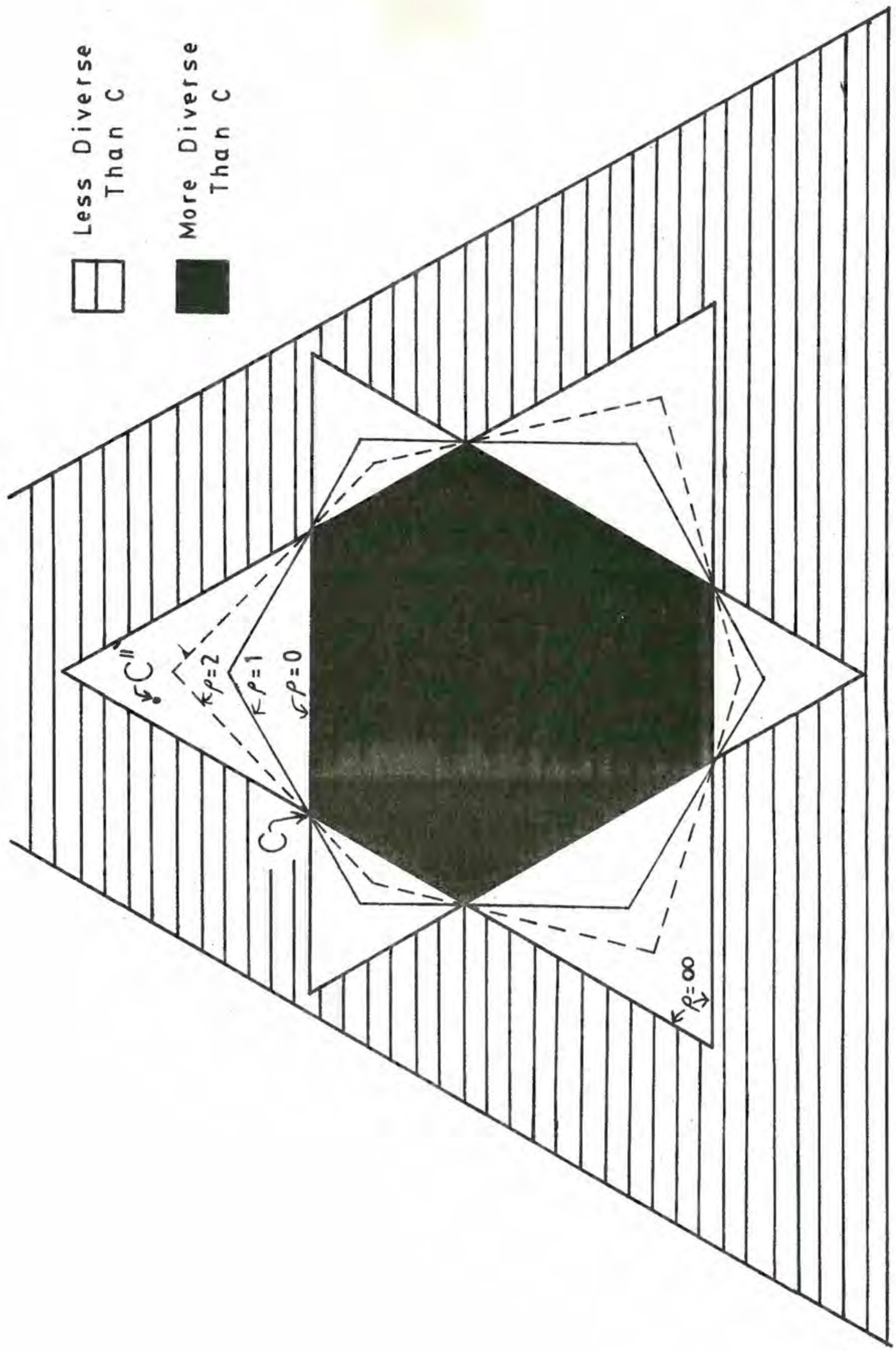
With the aid of the partial summation formula, $\Delta_\rho^{(\text{rank})}$ may be expressed as a right tail probability generating function,

$$\begin{aligned} \Delta_\rho^{(\text{rank})} &= \sum_{i \geq 1} \Big(\sum_{j > i} \pi_j^* \Big) \rho^{i-1} \qquad (1.14) \\ &= (1 - \pi_1^*) + (1 - \pi_1^* - \pi_2^*)\rho + \ldots \,. \end{aligned}$$

Clearly, then, $\Delta_\rho^{(\text{rank})}(C)$ is an increasing function of ρ and is strictly increasing when $s > 2$. This may be compared with $\Delta_\beta(C)$ which is a decreasing function of β. Typically, the two families of indices behave oppositely with respect to their parameters.

Figure 1.9 shows some contours of $\Delta_\rho^{(\text{rank})}$ for three-species communities. As is true for all rank type indices, the contours are piecewise linear. As ρ increases, the contours become increasingly elongated toward the vertices and, unlike those of Δ_β or Δ_ω^{H-S}, sweep out the entire region of imcomparability. Thus $\{\Delta_\rho^{(\text{rank})} : \rho \geq 0\}$ is

Figure 1.9 Some Contours of the Index $\Delta_{\rho}^{(\text{rank})}$


Less Diverse Than C
More Diverse Than C
C
C''
ρ=2
ρ=1
ρ=0
ρ=∞

complete for three-species communities. This fails for more than three species.

Example 1.4: Let the communities C and C" have respective ranked abundance vectors $(6,6,2,2)/16$ and $(6 + a, 5 - a, 4, 1)/16$ where $0 < a \leq 1$. Put $g(\rho) = \Delta_\rho^{(\text{rank})}(C") - \Delta_\rho^{(\text{rank})}(C) = -(\rho^2 - \rho + a)/16$. Now C and C" are not intrinsically comparable; but the polynomial $g(\rho)$ is always negative when $a > 1/4$ so that the family $\{\Delta_\rho^{(\text{rank})} : \rho \geq 0\}$ fails to detect this incomparability. When $0 < a < 1/4$, $g(\rho)$ has two positive roots and the incomparability is detected.

1.7 Index Response

The choice of an index implicitly involves a decision regarding the diversity ordering of communities which are not intrinsically comparable. Peet [25] discusses the need for a theory of index response to assist in this choice. For example, one may wish the index to be sensitive to the composition of the rare species but relatively indifferent to that of the abundant species. In Section 1.7.1, we give a definition of the sensitivity of an index and relate it to the question of index inconsistency. The term "sensitive to rare species" occurs frequently in the diversity literature but to our knowledge no precise definition has previously appeared.

Peet [25] suggests that index response be determined by examining the change in index value that results from a given change in the abundance of two equally common species. Expanding upon this idea in Section 1.7.2, we define the deflation factor of an arbitrary index and characterize Δ_β and $\Delta_\rho^{(\text{rank})}$ by the form of their deflation factors.

1.7.1 Sensitivity to Rare Species. In this section we consider communities C, C', C" with ranked abundance vectors $\underset{\sim}{\pi}^*$, $\underset{\sim}{\pi}'^*$, $\underset{\sim}{\pi}''^*$. Recall that C" is intrinsically more diverse than C if the tail sum differences,

$$T_k = T_k(C'', C) = \sum_{i \geq k} (\pi_i''^* - \pi_i^*), \; k = 1,2,3,\ldots,$$

are nonnegative. The next most complicated relation occurs when this sequence of differences has exactly one sign change. Taking the sign change to be from negative to positive, we say that C" is more diverse in the tail than C if there is a positive integer m for which $T_1, T_2, \ldots, T_m$ are less than or equal to zero but not all zero while $T_{m+1}, T_{m+2}, \ldots$ are greater than or equal to zero but not all zero.

Definition 1.4: A vector $\underset{\sim}{f} = (f_1, f_2, \ldots)$ is a transfer vector if all its tail sums are nonnegative and $f_1 + f_2 + \ldots = 0$. A pair $(\underset{\sim}{b}, \underset{\sim}{f})$ of nonzero transfer vectors is separated at m ($m = 2,3,\ldots$) if $b_i = 0$ for $m < i$ and $f_i = 0$ for $i < m$. Call $\underset{\sim}{b}$ the backward and $\underset{\sim}{f}$ the forward component.

If C" is intrinsically more diverse than C, the difference $\underset{\sim}{f} = \underset{\sim}{\pi}''^* - \underset{\sim}{\pi}^*$ is a transfer vector which codes the information needed to obtain C" from C by transfers to less abundant species (forward transfers). When C" is more diverse in the tail than C, there is a pair $(\underset{\sim}{b}, \underset{\sim}{f})$ of nonzero transfer vectors separated at m for which $\underset{\sim}{\pi}''^* = \underset{\sim}{\pi}^* + \underset{\sim}{f} - \underset{\sim}{b}$. In this case, C" may be obtained from C by a two-step operation: $\underset{\sim}{\pi}^* \to \underset{\sim}{\pi}^* + \underset{\sim}{f} \to \underset{\sim}{\pi}^* + \underset{\sim}{f} - \underset{\sim}{b}$. The first step increases the diversity in the intrinsic sense and does so by forward transfers among the species ranked $m, m+1, m+2, \ldots$. The second step decreases the diversity by backward transfers among the species ranked $1, 2, 3, \ldots, m$.

Example 1.5: We illustrate the preceding with a numerical example. Take

$$\underset{\sim}{\pi}^* = (.18,\ .18, .16, .16, .16, .16)$$

and

$$\underset{\sim}{\pi}''^* = (.50, .05, .05, .05, .05., 05, .05, .05, .05, .05, .05).$$

The sequence of tail sum differences T_k is

$$0,\ -.32,\ -.19,\ -.08,\ .03,\ .14,\ .25,\ .20,\ .15,\ .10,\ .05.$$

Since there is only one sign change, C" is more diverse in the tail than C. Putting

$$\underset{\sim}{b} = (-.32,\ .13,\ .11,\ .08,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0),$$

$$\underset{\sim}{f} = (0,\ 0,\ 0,\ -.03,\ -.11,\ -.11,\ .05,\ .05,\ .05,\ .05,\ .05),$$

the pair $(\underset{\sim}{b},\underset{\sim}{f})$ is separated at $m = 4$ and $\underset{\sim}{\pi}''^* = \underset{\sim}{\pi}^* + \underset{\sim}{f} - \underset{\sim}{b}$. Since the two communities C" and C are not intrinsically comparable, they may be inconsistently ordered by a given pair of indices. But C" is more diverse in the tail so that we might anticipate the order $\Delta(C'') \geq \Delta(C)$ when Δ is "sensitive to rare species" and the opposite ordering when Δ is "insensitive." This behavior is illustrated for the indices Δ_β, Δ_ω^{H-S} and $\Delta_\rho^{(rank)}$ in Table 1.1. Examination of the table suggests that Δ_β is sensitive to rare species for small β while Δ_ω^{H-S} and $\Delta_\rho^{(rank)}$ are sensitive for large values of ω and ρ. Figure 1.10 is a plot of $\Delta_\beta(C)$ and $\Delta_\beta(C'')$ versus β. Notice that both graphs are decreasing and have exactly one crossing point. The analogous plots for the indices

Table 1.1: Response to Increased Tail Diversity

(a)

	β			
	-1	0	1	2
$\Delta_\beta(C)$:	5	1.79	.83	.49
$\Delta_\beta(C'')$:	10	1.84	.73	.44

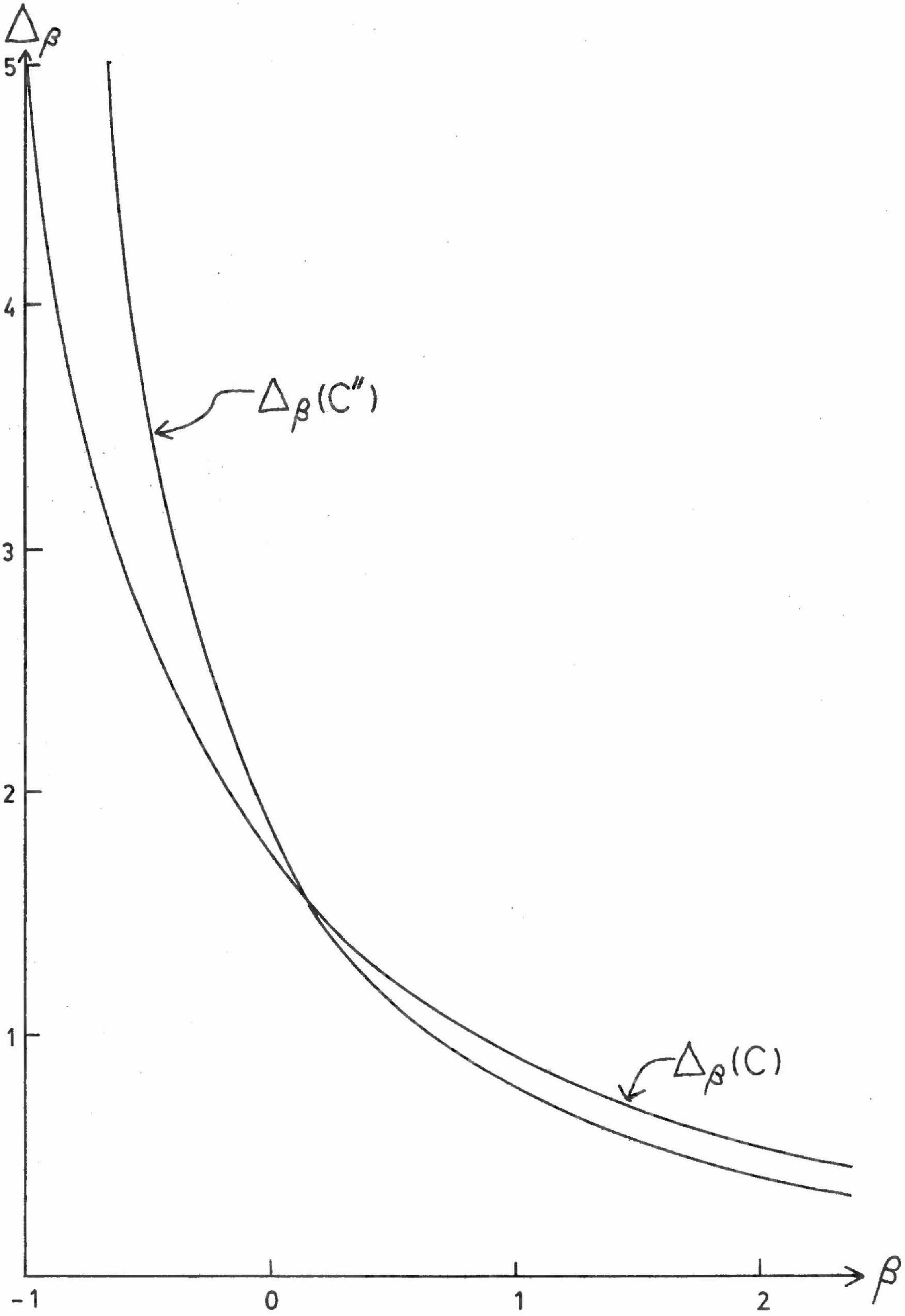
(b)

	ω			
	1	5	20	∞
$\Delta_\omega^{H-S}(C)$:	.83	2.99	4.87	5
$\Delta_\omega^{H-S}(C'')$:	.73	2.63	6.60	10

(c)

	ρ			
	0	.5	1	2
$\Delta_\rho^{(rank)}(C)$:	.82	1.31	2.42	9.14
$\Delta_\rho^{(rank)}(C'')$:	.50	.90	2.75	101.80

Figure 1.10 Illustration of the Crossing Point Theorem Δ_β


Δ_β
5
4
3
2
1
$\Delta_\beta(C'')$
$\Delta_\beta(C)$
-1
0
1
2
β

Δ_ω^{H-S} and $\Delta_\rho^{(rank)}$ are increasing in ω and ρ but again have a unique crossing point. See Figure 1.11 and Figure 1.12.

The program for the remainder of this section is to (1) give a precise definition of the sensitivity of an index to rare species, (2) show that the sensitivity of Δ_β is a strictly decreasing function of β while the sensitivities of Δ_ω^{H-S} and $\Delta_\beta^{(rank)}$ are strictly increasing functions of ω and ρ, and (3) prove the uniqueness of the crossing points described in Example 1.5.

Start with an s-species community $C = (s, \underset{\sim}{\pi}^*)$ and let $(\underset{\sim}{b}, \underset{\sim}{f})$ be a pair of nonzero transfer vectors which are separated at m $(1 < m \leq s)$. With h and k two small positive numbers, use $\underset{\sim}{b}$ and $\underset{\sim}{f}$ as direction vectors to construct the two communities

$$C' : \underset{\sim}{\pi}'^* = \underset{\sim}{\pi}^* + h\underset{\sim}{f} \quad \text{and}$$
$$C'' : \underset{\sim}{\pi}''^* = \underset{\sim}{\pi}^* + h\underset{\sim}{f} - k\underset{\sim}{b}.$$

To ensure that $\underset{\sim}{\pi}'^*$ and $\underset{\sim}{\pi}''^*$ are ranked, we assume that $f_i \geq f_{i+1}$ whenever $\pi_i^* = \pi_{i+1}^*$ and that $b_i \leq b_{i+1}$ whenever both $\pi_i^* = \pi_{i+1}^*$ and $f_i = f_{i+1}$. Now C' is intrinsically more diverse than C and $\Delta(C') \geq \Delta(C)$ for any index Δ which satisfies Criterion C2. On the other hand, $\Delta(C'')$ is a decreasing function of k and, for sufficiently small h, there is a $k = k(h)$ for which $\Delta(C'') = \Delta(C)$. We define the sensitivity of the index Δ to be $dk/dh|_{h=0}$ and denote it $\sigma(\Delta;C;\underset{\sim}{b},\underset{\sim}{f})$. Figure 1.13 shows the communities C, C', C'' along with the contour of Δ which passes through C. The sensitivity is the slope of this contour at h = 0. Comparison of Figure 1.13 with Figure 1.8 reveals that a sensitive index has contours which are elongated toward the vertices of the abundance simplex. Since contours are unaffected by strictly monotonic transformations, the sensitivity of an index is invariant to such transformations. In particular both Δ and its numbers equivalent have the same sensitivity.

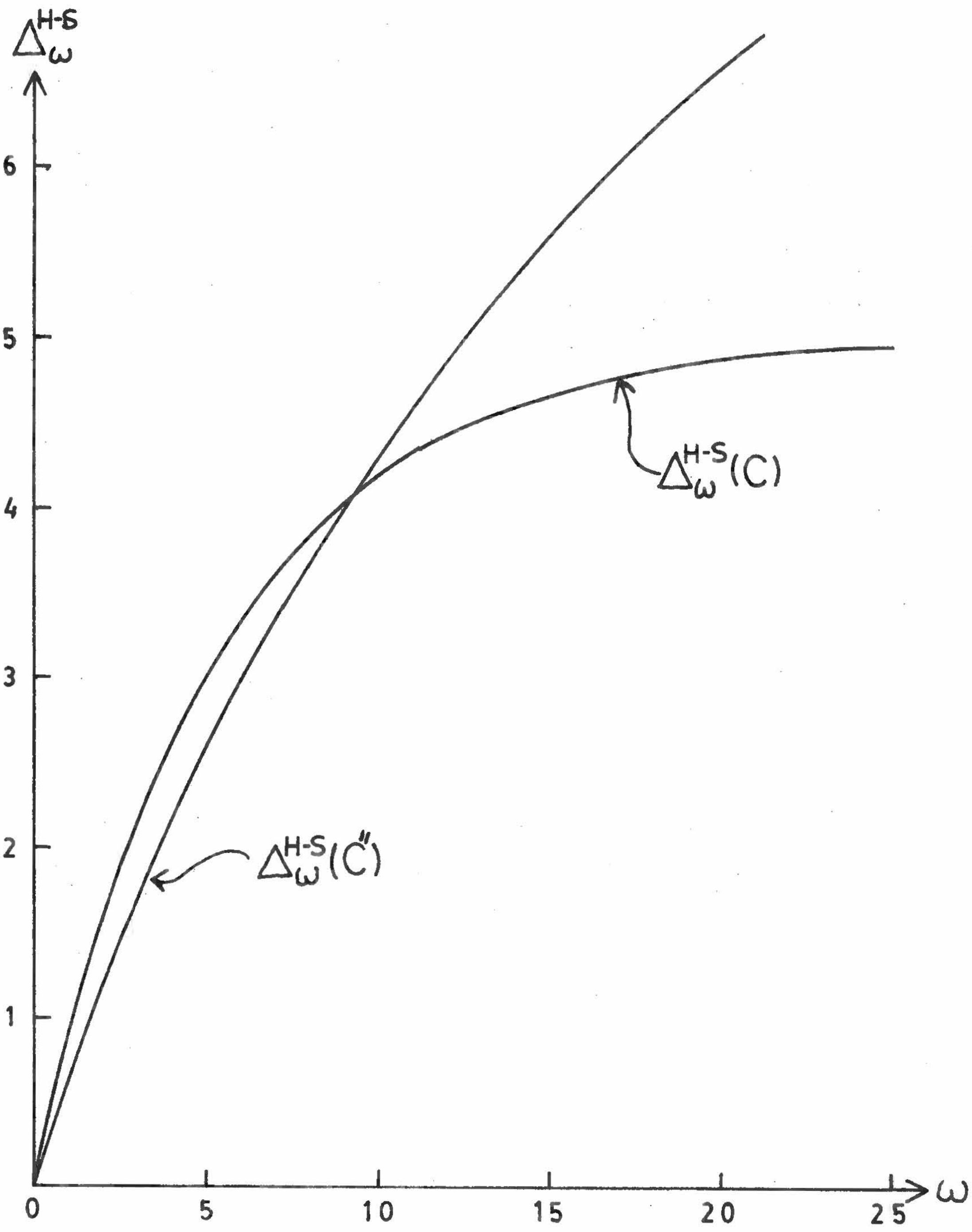


Figure 1.11 Illustration of the Crossing Point Theorem for Δ_ω^{H-S}

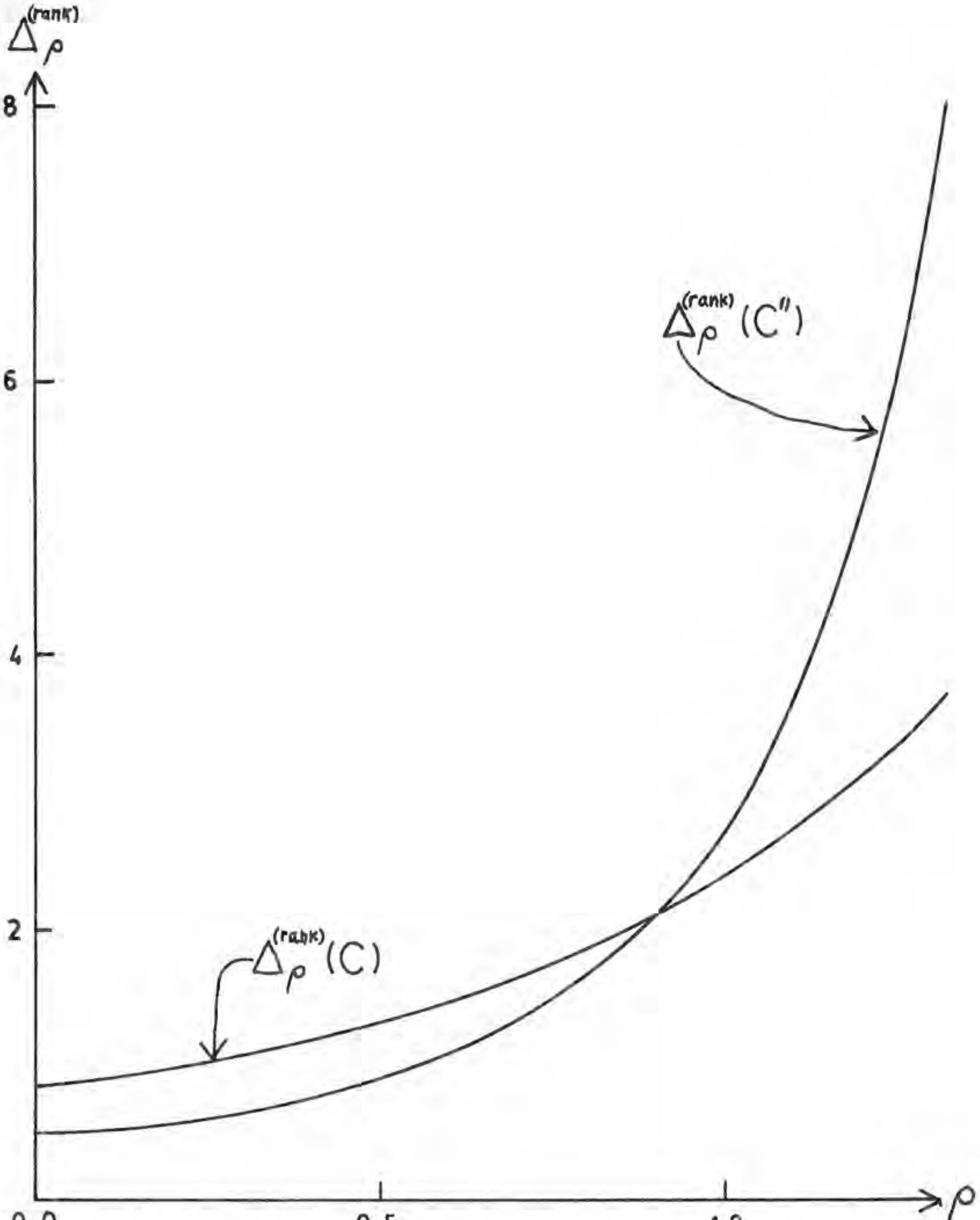


Figure 1.12 Defining the Sensitivity of $\Delta_\rho^{(\text{rank})}$

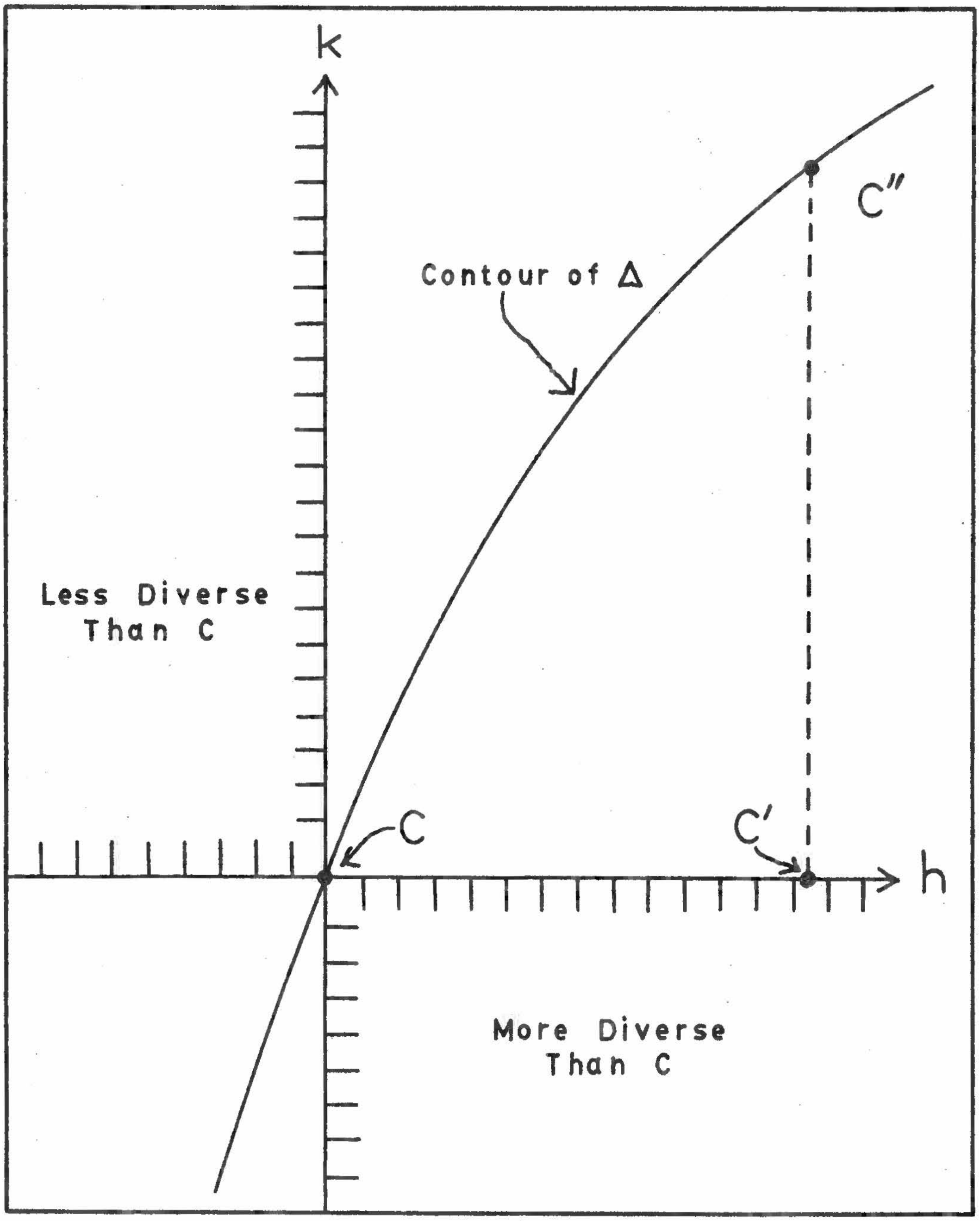


Figure 1.13 Defining the Sensitivity of Δ

Remark 1.7: Multiplying $\underset{\sim}{f}$ by $\lambda > 0$ and $\underset{\sim}{b}$ by $\mu > 0$ multiplies the sensitivity by λ/μ. A standardized sensitivity unit can be established by requiring that the positive components of $\underset{\sim}{b}$ and of $\underset{\sim}{f}$ sum to unity. Then $h(k)$ is the net abundance transferred forward (backward) and a large value of the sensitivity means that a small transfer forward to the rare species requires a large offsetting transfer back to the abundant species.

Explicit expressions for the sensitivity are obtained by implicit differentiation with respect to h of the relation $\Delta(C'') = \Delta(C) = \text{constant}$.

Theorem 1.10: The sensitivity of the rank type index $\Delta = \Sigma\pi_i^* R(i)$ does not depend on $\underset{\sim}{\pi}^*$ and is given by

$$\sigma(\Delta;\ C;\ \underset{\sim}{b},\ \underset{\sim}{f}) = \Sigma f_i R(i)/\Sigma b_j R(j).$$

In particular,

$$\sigma(\Delta_\rho^{(\text{rank})};\ C;\ \underset{\sim}{b},\ \underset{\sim}{f}) = \begin{cases} \Sigma f_i \rho^{i-1}/\Sigma b_j \rho^{j-1} & \text{if} \quad \rho > 0,\ \rho \neq 1 \\ \Sigma f_i (i-1)/\Sigma b_j (j-1) & \text{if} \quad \rho = 1. \end{cases}$$

Theorem 1.11: The dichotomous index $\Delta = \Sigma\pi_i^* R(\pi_i^*)$ with auxiliary function $V(\pi) = \pi R(\pi)$ has sensitivity

$$\sigma(\Delta;\ C;\ \underset{\sim}{b},\ \underset{\sim}{f}) = \Sigma f_i V'(\pi_i^*)/\Sigma b_j V'(\pi_j^*).$$

In particular,

$$\sigma(\Delta_\beta;\ C;\ \underset{\sim}{b},\ \underset{\sim}{f}) = \begin{cases} \Sigma f_i (\pi_i^*)^\beta/\Sigma b_j (\pi_j^*)^\beta & \text{if} \quad \beta > -1,\ \beta \neq 0 \\ \Sigma f_i \log(\pi_i^*)/\Sigma b_j \log(\pi_j^*) & \text{if} \quad \beta = 0, \end{cases}$$

and

$$\sigma(\Delta_\omega^{H-S};\ C;\ \underset{\sim}{b},\ \underset{\sim}{f}) = \Sigma f_i (1-\pi_i^*)^\omega/\Sigma b_j (1-\pi_j^*)^\omega \ \text{if}\ \omega > 0.$$

The next example illustrates the behavior of the sensitivity of $\Delta_\rho^{(\text{rank})}$ and of Δ_β when C is a three-species community. Following the example, we show that the features illustrated hold, in fact, for multi-species communities.

Example 1.6: (Three-species community). Take $\pi^* = (\pi_1^*, \pi_2^*, \pi_3^*)$, $\underset{\sim}{b} = (-1, 1, 0)$ and $\underset{\sim}{f} = (0, -1, 1)$. By Theorem 1.10, $\sigma(\Delta_\rho^{(\text{rank})}; C; \underset{\sim}{b}, \underset{\sim}{f}) = \rho$ which is a strictly increasing function of ρ and approaches infinity as $\rho \to \infty$. The sensitivity of Δ_β depends on the community C which may vary over the region XYZ of the abundance simplex (see Figure 1.6). Letting $\phi = (\pi_2^* - \pi_3^*)/(\pi_1^* - \pi_2^*)$, reparametrize C with the pair (ϕ, π_3^*), $0 \leq \phi \leq \infty$, $0 \leq \pi_3^* < 1/3$. When ϕ is fixed, C varies along a straight line passing through the centroid of the simplex as shown in Figure 1.6. In Figure 1.14, the sensitivity of Δ_β is plotted against β for several values of π_3^* when $\phi = 1$. The following features should be noted:

(i) The sensitivity of the Simpson index is identically 1 (= ϕ) independent of π_3^*.

(ii) When $\pi_3^* \neq 0$, the sensitivity of Δ_β is a strictly decreasing function of β which approaches zero as $\beta \to \infty$ and which has a finite positive limit as β approaches -1 from above.

(iii) The sensitivity to the introduction of a species is covered by the case $\pi_3^* = 0$. The sensitivity is infinite for $-1 \leq \beta \leq 0$; over the interval $0 < \beta < \infty$, the sensitivity is is a strictly decreasing function of β which approaches infinity as $\beta \to 0$ and zero as $\beta \to \infty$.

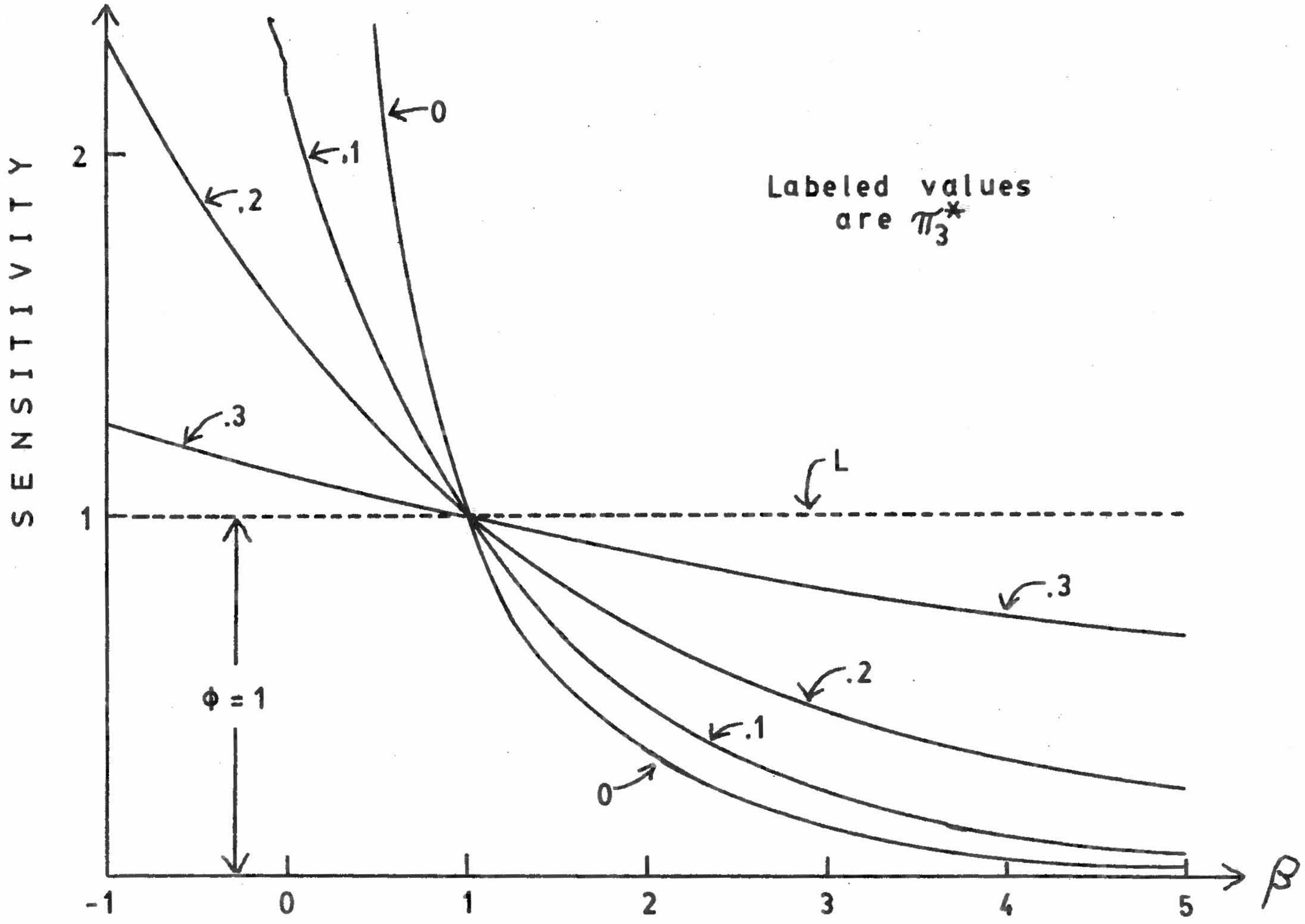


Figure 1.14 Sensitivity of $\Delta_\beta(\phi = 1$ and $\pi_3^* = 0, .1, .2, .3)$

For other finite, nonzero values of ϕ, these qualitative features of Figure 1.14 are unchanged except that the horizontal line L is at a height ϕ above the β-axis. The value $\phi = 0$ is excluded by the requirement that $f_2 \geq f_3$ when $\pi_2^* = \pi_3^*$. When $\phi = \infty$, the sensitivity is infinite for all β. Note that in this case there is a tie between the first and second ranked species ($\pi_1^* = \pi_2^*$) and that the sum of the corresponding components of $\underset{\sim}{b}$ vanishes.

For multispecies communities, a preliminary lemma is needed. The proof is given in Appendix A.1.

Definition 1.5: Let $\{a_i\}_{i=1}^{\infty}$ be a sequence of extended real numbers. A tied set of $\{a_i\}$ is a maximal set I of subscripts such that $a_i = a_j$ whenever $i, j \in I$. A transfer vector $\underset{\sim}{b}$ vanishes over tied sets of $\{a_i\}$ if $\sum\limits_{i \in I} b_i = 0$ for each tied set I of $\{a_i\}$.

Lemma 1.2: Let $(\underset{\sim}{b}, \underset{\sim}{f})$ be a pair of nonzero transfer vectors which are separated at m and $\infty \neq \lambda_i \geq \lambda_2 \geq \ldots \geq -\infty$ a monotonic sequence of extended real numbers. Assume that λ_m is finite and that $f_j \geq f_{j+1}$ whenever $\lambda_j = \lambda_{j+1}$. For $-\infty < x < \infty$, define

$$\sigma(x) = \begin{cases} \Sigma f_i \exp(\lambda_i x)/\Sigma b_j \exp(\lambda_j x) & \text{if } x \neq 0 \\ \Sigma f_i \lambda_i / \Sigma b_j \lambda_j & \text{if } x = 0. \end{cases}$$

Then

a) $\sigma(x) \equiv \infty$ when $\underset{\sim}{b}$ vanishes over tied sets of $\{\lambda_i\}$.

b) Suppose $\underset{\sim}{b}$ does not vanish over tied sets of $\{\lambda_i\}$ and that $f_i = 0$ whenever $\lambda_i = -\infty$. Then $\sigma(x)$ is a strictly decreasing function of x with $\lim\limits_{x \to -\infty} \sigma(x) = \infty$ and $\lim\limits_{x \to \infty} \sigma(x) = 0$.

c) Suppose $\underset{\sim}{b}$ does not vanish over tied sets of $\{\lambda_i\}$ and that there is an i for which $f_i \neq 0$ and $\lambda_i = -\infty$. Then $\sigma(x) = \infty$ for $x \leq 0$; over the interval $0 < x < \infty$, $\sigma(x)$ is a strictly decreasing function of x with $\lim_{x \to 0} \sigma(x) = \infty$ and $\lim_{x \to \infty} \sigma(x) = 0$.

Now let $C = (s, \underset{\sim}{\pi}^*)$ be an s-species community and $(\underset{\sim}{b}, \underset{\sim}{f})$ a pair of nonzero transfer vectors which are separated at m $(1 < m \leq s)$. Assume that $f_i \geq f_{i+1}$ whenever $\pi_i^* = \pi_{i+1}^*$ and that $b_i \leq b_{i+1}$ whenever both $\pi_i^* = \pi_{i+1}^*$ and $f_i = f_{i+1}$.

Theorem 1.12: The sensitivity of $\Delta_\rho^{(\text{rank})}$ is a strictly increasing function of ρ which approaches zero as $\rho \to 0$ and which approaches infinity as $\rho \to \infty$.

Proof: Write $\rho = \exp(-x)$ and apply Lemma 1.2(b) with $\lambda_i = -(i - 1)$.

Theorem 1.13: Let $\sigma(\beta) = \sigma(\Delta_\beta; C; \underset{\sim}{b}, \underset{\sim}{f})$, $\beta > -1$, be the sensitivity of Δ_β. Then $\sigma(\beta) \equiv \infty$ when $\underset{\sim}{b}$ vanishes over tied sets of $\underset{\sim}{\pi}^*$. If $\underset{\sim}{b}$ does not vanish over tied sets of $\underset{\sim}{\pi}^*$, there are two subcases depending on whether a species is introduced: the behavior of $\sigma(\beta)$ is described under item (ii) of Example 1.6 when $f_i = 0$ for all $i > s$ and under item (iii) of Example 1.6 otherwise.

Proof: Write $\pi_i^* = \exp(\lambda_i)$ and apply Lemma 1.2 with $x = \alpha$.

Theorem 1.14: Let $\sigma(\omega) = \sigma(\Delta_\omega^{H-S}; C; \underset{\sim}{b}, \underset{\sim}{f})$, $\omega > 0$, be the sensitivity of the Hurlbert-Smith index. Then $\sigma(\omega) = \infty$ when $\underset{\sim}{b}$ vanishes over tied sets of $\underset{\sim}{\pi}^*$; otherwise $\sigma(\omega)$ is a strictly increasing function of ω which has a finite positive limit as $\omega \to 0$ and which approaches infinity as $\omega \to \infty$.

Proof: Write $1 - \pi_i^* = \exp(-\lambda_i)$ and apply Lemma 1.2 with $x = -\omega$.

Next we state the crossing point theorem which was illustrated in Example 1.5.

<u>Theorem 1.15</u>: Let $C = (s, \underset{\sim}{\pi}^*)$ and $C'' = (s'', \underset{\sim}{\pi}''^*)$ be two communities with C'' more diverse in the tail than C.

a) Let $g(\rho) = \Delta_\rho^{(\mathrm{rank})}(C'') - \Delta_\rho^{(\mathrm{rank})}(C)$, $\rho \geq 0$. Then $g(0) = \pi_1^* - \pi_1''^* \leq 0$. On the interval $0 < \rho < \infty$, $g(\rho)$ is negative for small ρ, positive for large ρ and has exactly one zero.

b) Let $g(\beta) = \Delta_\beta(C'') - \Delta_\beta(C)$, $\beta \geq -1$. Then $g(-1) = s'' - s \geq 0$. On the interval $-1 < \beta < \infty$, either (i) $g(\beta)$ is negative for all β or (ii) $g(\beta)$ is positive for small β, negative for large β and has exactly one zero. Case (ii) must apply if $s'' > s$.

c) Let $g(\omega) = \Delta_\omega^{H-S}(C'') - \Delta_\omega^{H-S}(C)$, $\omega \geq 0$. Then $g(0) = 0$ and $g(\infty) = s'' - s \geq 0$. On the interval $0 < \omega < \infty$, either (i) $g(\omega)$ is positive for all ω or (ii) $g(\omega)$ is negative for small ω, positive for large ω and has exactly one zero.

<u>Proof</u>: (a) It is easy to see that $g(\rho)$ is positive for large ρ and negative for small ρ (see Remark 1.8, below). Thus $g(\rho)$ has at least one positive zero and we need only show that this zero is unique. In fact, if ρ_0 is any positive zero, we show that $g(\rho) > 0$ for $\rho > \rho_0$. Referring to Figure 1.13, let Γ and Γ_0 be the intersections with the hk-plane of the contours through C of $\Delta_\rho^{(\mathrm{rank})}$ and of $\Delta_{\rho_0}^{(\mathrm{rank})}$. Since $\rho_0 > 0$, Γ and Γ_0 can be shown to be one dimensional. At C, Γ has a strictly greater slope than Γ_0 so that, if $g(\rho)$ were nonpositive, Γ would have to cross Γ_0 at some point between C and C''. But at the crossing point the slope of Γ will be less than or equal to that of Γ_0 and this contradicts the strict monotonicity of the sensitivity.

(b) Let the first nonzero component of the backward transfer vector $\underset{\sim}{b}$ occur in the nth position so that $\pi_i''^* = \pi_i^*$ for $i < n$ and $\pi_n''^* > \pi_n^* \geq \pi_{n+1}^* \geq \pi_{n+2}^* \geq \cdots$. Then $\beta : g(\beta)$ equals

$$-(\pi''^{*}_{n})^{\beta+1}[1 + \sum_{i>n} (\pi''^{*}_{i}/\pi''^{*}_{n})^{\beta + 1} - \sum_{i\geq n} (\pi^{*}_{i}/\pi''^{*}_{n})^{\beta + 1}].$$

Since the term in brackets is asymptotically positive, $g(\beta)$ must be negative for large β. The proof that the crossing point is unique if it exists is similar to (a). The communities C and C" of Figure 1.8 show that $g(\beta)$ need not have a positive zero. Part (c) of the theorem is proved in the same manner as part (b); here again the crossing point need not exist.

Remark 1.8: There is an alternative proof of Theorem 1.15(a) which yields more information. Using 1.14, $g(\rho)$ may be expressed in the form $T_2 + T_3\rho + T_4\rho^2 + \ldots$ where $\{T_k\}$ is the sequence of tail sum differences. By Descartes' rule of signs, the number of positive zeros of $g(\rho)$, counting multiplicity, does not exceed the number of sign changes of $\{T_k\}$ and the difference between the two is an even integer. Thus $g(\rho)$ has exactly one positive zero when $\{T_k\}$ has one sign change.

The method of this section can also be used to define the sensitivity to abundant species and the sensitivity to intermediate species. The sensitivity to abundant species turns out to be the reciprocal of the sensitivity to rare species and gives nothing new. The sensitivity to intermediate species would be defined by making forward transfers among the intermediate species and offsetting backward transfers among both the rare and the abundant species. So defined, the sensitivity of $\Delta_{\rho}^{(\text{rank})}$ to intermediate species is a unimodal function of ρ which vanishes at both $\rho = 0$ and $\rho = \infty$. (We conjecture, but have not proved, that the sensitivity of Δ_{β} to intermediate species is a unimodal function of β.) This may provide some insight into the behavior of the function $g(\rho)$ of Example 1.4. In that example, C" is less diverse than C among the abundant and also among the rare species but is more diverse among the intermediate species. For small ρ, $\Delta_\rho^{(\text{rank})}$ is sensitive to the abundant species and $g(\rho)$ is negative. As ρ increases, $\Delta_\rho^{(\text{rank})}$ becomes more sensitive to the intermediate species and $g(\rho)$ increases. Finally, for large values of ρ, $\Delta_\rho^{(\text{rank})}$ is sensitive to the rare species and $g(\rho)$ begins to decrease.

1.7.2 Response to Perturbations. Consider a community C and a subcollection of species from this community. Label the members of the subcollection as $i = 1,2,\ldots,n$, denote their abundances by $\pi_1, \pi_2,\ldots,\pi_n$ and let $\pi = \pi_1 + \pi_2 +\ldots+ \pi_n$ be their combined abundance. The subcollection forms a community C' with abundance vector $(\pi_1/\pi, \pi_2/\pi,\ldots,\pi_n/\pi)$. Keeping π fixed, let the subcollection be subjected to an infinitesmal perturbation of form

$$\tilde{\pi}_i = \pi_i + hb_i, \; i = 1,\ldots,n, h > 0, \; \Sigma b_i = 0.$$

The perturbation induces a change $\delta\Delta(C)$ in the diversity measure of C and also a change $\delta\Delta(C')$ in that of C'. We wish to compare the magnitudes of these two changes and we define the response function of Δ to be the ratio $\delta\Delta(C)/\delta\Delta(C')$. The dependence of the response function on the combined abundance π will be of particular interest. The most natural dependence occurs when the response function equals π (i.e., $\delta\Delta(C) = \pi\cdot\delta\Delta(C')$). More generally write the response function as $\pi\cdot W$ and call W the deflation factor. (Values of W greater than one act as an inflation factor.) In general, W depends on the direction vector $\underset{\sim}{b}$ as well as the parameters of C and of C'. We also remark that W is invariant to affine transformations of the index but not to monotone transformations in general.

An easy calculation shows that the dichotomous index with auxiliary function V has deflation factor $W = \Sigma b_i V'(\pi_i)/\Sigma b_i V'(\pi_i/\pi)$. For the index Δ_β, this expression reduces to $W = \pi^\beta$ which depends only on π. The deflation factor of the rank type index is found to be $W = \Sigma b_j R(i_j)/\Sigma b_j R(j)$ where $i_1 < i_2 <\ldots< i_n$ are the ranks (within C) of the various species in the subcollection. In particular, $\Delta_\rho^{(rank)}$ has deflation factor $W = \Sigma b_j \rho^{(i_j-1)}/\Sigma b_j \rho^{j-1}$; in general, this expression depends on $\underset{\sim}{b}$ but when the species are consecutively ranked within the community C ($i_1 = i$, $i_2 = i + 1, \ldots, i_n = i + n - 1$), it reduces to ρ^{i-1} which depends only on the rank of the most abundant species. In fact, the indices $\Delta_\rho^{(rank)}$ are characterized by this last property.

<u>Theorem 1.16</u>: Let Δ be a rank type index whose deflation factor depends only on the rank of the most abundant species when the species are consecutively ranked. Then there is a real number ρ such that Δ is an affine transformation of $\Delta_\rho^{(rank)}$.

<u>Proof</u>: Let i be the rank of the most abundant species and let k be a positive integer. Taking $b_1 = -1$, $b_2 = b_3 =\ldots= b_k = 0$, $b_{k+1} = 1$, we find that

$$W = W(i) = (R(i + k) - R(i))/(R(k + 1) - R(1)).$$

Thus R satisfies the proportionality equation of Section 1.6 and the theorem follows from (1.13).

With suitable differentiability assumptions on V, one may similarly characterize Δ_β as the only index whose deflation factor depends only on the combined abundance π. But in Section 1.8, we give a characterization of Δ_β which requires only very mild regularity assumptions.

1.8 Diversity Decomposition

The calculated value of a diversity index provides only limited information concerning the overall structure of a community. Often it is possible to decompose the community in some natural way, and as in the analysis of variance, apportion the total diversity among and between the various components. We consider two types of decompositions: two-way classifications and mixtures. Allan [26] has discussed both types in the case of the Shannon index.

As measured by Shannon index, the total diversity of a two-way classification decomposes into the sum of two terms: the diversity of the row marginals and the average diversity of the normalized rows. In Section 1.8.2 we replace this last average by a "deflated" average and extend the decomposition to a wider class of indices. In the important special case of hierarchial classification, the total diversity equals the diversity of the column marginals and the decomposition becomes a decomposition of marginal diversity. Pielou [27], Lloyd et al. [28] and Wilson [29] have used this hierarchial decomposition to apportion taxonomic diversity. Theil [30, 31] has discussed numerous applications in the social and administrative sciences.

Lewontin [21] has pointed out a general method of decomposing marginal diversity even for nonhierarchial classifications. In a two-way classification, the vector of column marginals is a mixture of the normalized rows and, when the index is concave, the diversity T of the column marginals will be greater than or equal to the average diversity W of the normalized rows. Lewontin proposes that the "within" and the "between" components of diversity be defined as W and T-W, respectively.

Lewontin explicitly gives the decomposition only for the Shannon index while Nei [32] has given it for the Simpson index. In Section 1.8.3 we discuss this mixture decomposition for the general dichotomous index. However, in the case of a hierarchial classification, the mixture decomposition coincides with the hierarchial decomposition only for the Shannon index. At present, we are aware of no reason for preferring one decomposition over the other.

Since the decompositions of this section bear a formal resemblance to the analysis of variance, we begin by striking an analogy between diversity and variance; the analogy is strengthened in Section 1.12.5.

1.8.1 Diversity and Variance.

Let X_1 and X_2 be independent realizations of a random variable X. The variance of X is given by $\sigma^2 = (1/2)E[(X_1-X_2)^2]$, and may be interpreted as the average difference (or differentness) between the two independent observations. By analogy, define the differentness between species i and species j to be $1 - \delta_{ij}$ where (δ_{ij}) is the identity matrix. Simpson's index can be written as $\Delta_1 = \Sigma\pi_i(1 - \delta_{ij})\pi_j$ so that diversity (at least as measured by Simpson's index) is the average differentness between two randomly selected members of the community.

Now Δ_1 is a measure of diversity while σ^2 is a measure of spread. The analogy may be extended to one between diversity and spread in general. Given a ranked abundance vector $\underset{\sim}{\pi}^*$, define a symmetric unimodal probability distribution on the integers by symmetrizing $\underset{\sim}{\pi}^*$, i.e., $P(0) = \pi_1^*$ and $P(i) = P(-i) = \pi_{i+1}^*/2$ for $i = 1,2,3,\ldots$. It is intuitively clear that introducing a species or transferring abundance will increase the spread of this distribution with any reasonable interpretation of spread. In fact, some of the standard measures of spread, when applied to this distribution, become identified with previously considered diversity indices. The half range, for example, is the Species Count while the mean absolute deviation is the reduced Average Rank.

1.8.2. Diversity of a Two-Way Classification. Pielou [33, p. 7] has listed three desirable properties of a diversity index:

P1. For a given number of species, the index should assign maximum diversity to the completely even community.

P2. Given two completely even communities, the one with more species should be assigned the greater diversity.

P3. An ANOVA formula (1.16) with $W(\pi) = 1$ should hold for two-way classifications.

Theorem 1.17: Any diversity index Δ which satisfies Criterion C2 also satisfies properties P1 and P2.

Proof: Let $C' = (s', \underset{\sim}{\nu})$ and $C = (s, \underset{\sim}{\pi})$ be two communities with C' completely even. Clearly $\underset{\sim}{\nu}$ is a convex linear combination of permutations of $\underset{\sim}{\pi}$ when $s \leq s'$. Applying Theorem 1.7, it follows that $\Delta(C') \geq \Delta(C)$.

As has been shown by Khinchin (see [34, p. 67]), the three properties P1, P2 and P3 together with some regularity assumptions characterize the Shannon index up to a constant multiple. However, deflated ANOVA formulas can be associated with certain other indices. Consider a two-way classification AxB with cell proportions π_{ij}, $i = 1,2,\ldots,a$, $j = 1,2,\ldots,b$. Without loss of generality, we may assume that the marginals $\pi_{i.}$, $i = 1,2,\ldots,a$, and $\pi_{.j}$, $j = 1,2,\ldots,b$, are all nonzero. For now we restrict attention to dichotomous indices with rarity measure R. Putting

$$\Delta(A) = \sum_{i=1}^{a} \pi_{i.} R(\pi_{i.}),$$

$$\Delta(B|A_i) = \sum_{j=1}^{b} (\pi_{ij}/\pi_{i.}) R(\pi_{ij}/\pi_{i.}),$$

and

$$\Delta(AxB) = \sum_{i=1}^{a} \sum_{j=1}^{b} \pi_{ij} R(\pi_{ij}),$$

the total diversity of the table may be written as

$$\Delta(AxB) = \Delta(A) + \sum_{i=1}^{a} \pi_{i.} \sum_{j=1}^{b} (\pi_{ij}/\pi_{i.})[R(\pi_{ij}) - R(\pi_{i.})]. \quad (1.15)$$

For the index Δ_β with rarity measure $R(\pi) = (1 - \pi^\beta)/\beta$ and with deflation factor $W(\pi) = \pi^\beta$, (1.15) reduces to the <u>deflated ANOVA formula</u>

$$\Delta(AxB) = \Delta(A) + \sum_{i=1}^{a} \pi_{i.} W(\pi_{i.}) \Delta(B|A_i),$$

or, more schematically,

$$\Delta(AxB) = \Delta(A) + E_A[W(\pi_{i.})\Delta(B|A_i)]. \quad (1.16)$$

Since $\Delta(AxB) = \Delta(BxA)$, interchanging the roles of A and B in (1.16) leads to the equation

$$\Delta(B) = \Delta(A) + E_A[W(\pi_{i.})\Delta(B|A_i)] - E_B[W(\pi_{.j})\Delta(A|B_j)]. \quad (1.17)$$

With AxB classification as habitat x species, ecologists [26, 35, 36, 37], using the Shannon (or Brillouin) index, have attempted to interpret the last two terms of (1.17) as measures of "niche overlap" and "niche width," respectively. These interpretations are also reasonable for the Species Count and for the Simpson index; for example, with the Simpson index, $E_A[W(\pi_{i.})\Delta(B|A_i)]$ is the probability that two randomly selected organisms belong to the same habitat but different species. Accepting these interpretations, (1.17) states that

$$\text{species diversity} = \text{habitat diversity} + \text{niche overlap} - \text{niche width}. \qquad (1.18)$$

Arguing from the Lotka-Volterra competition equations, MacArthur [38, p. 195] has obtained a multiplicative version of (1.18).

A two-way classification is hierarchial if the rows are mutually orthogonal or, more specifically, if, for each j, all but one of the proportions π_{ij}, $i = 1,2,\ldots,a$, are zero. In this case $\Delta(A\text{x}B) = \Delta(B)$ and (1.16) becomes a decomposition of marginal diversity,

$$\Delta(B) = \Delta(A) + E_A[W(\pi_{i.})\Delta(B|A_i)]. \qquad (1.19)$$

For example, taking the AxB classification as genus x species, (1.19) expresses the overall species diversity as the sum of the genus diversity and a deflated within-genus species diversity.

The indices Δ_β are essentially the only indices satisfying the deflated ANOVA formula (1.16). This may be proved under a variety of regularity assumptions (see Aczel and Daroczy [34, chap. 6] and the references therein); here we present two such characterizations both of whose proofs reduce to the proportionality equation (1.3). In the first version the index is assumed to be of the dichotomous type.

Theorem 1.18: Suppose that R is measurable and nonconstant on the interval (0,1]. If the dichotomous index $\Delta = \Sigma\pi_i R(\pi_i)$ satisfies the deflated ANOVA formula (1.16) for some deflation factor W then there is a real number β (possibly $\beta = +\infty$) for which $W(\pi) = \pi^\beta$ and

$$R(\pi) = \begin{cases} \text{constant}\cdot(1 - \pi^\beta) & \text{if } \beta \neq 0 \\ \text{constant}\cdot\log(\pi) & \text{if } \beta = 0. \end{cases}$$

If, in addition, W is assumed to be measurable then (1.16) need be assumed only for independent classifications.

Proof: First suppose that $R(1) = 0$. Let $\underset{\sim}{\pi}$ be an arbitrary n-dimensional probability vector and $0 < q \leq 1$. Applying (1.16) to the 2xn classification

$$\begin{array}{l} q\pi_1 \; q\pi_2 \; q\pi_3 \; \cdots \; q\pi_n \\ 1 - q \quad 0 \quad 0 \; \ldots \; 0 \end{array}$$

gives $\Sigma\pi_i[R(q\pi_i) - R(q) - W(q)R(\pi_i)] = 0$. For fixed q, the expression in brackets is measurable; using Theorem 1.1, the pair (R,W) is then a solution of the proportionality equation (1.3). Theorem 1.2 now gives the result. Next suppose that $R(1) \neq 0$. Applying (1.16) to nx1 classifications shows that $R(1)\Sigma\pi_i W(\pi_i) = 0$ for all probability vectors $\underset{\sim}{\pi}$. It follows from this that the rarity measure $R(\pi) - R(1)$ meets all the assumptions of the theorem and with the same deflation factor W. By the first part of the proof, $W(\pi) = \pi^{\beta}$ so that $W(1) = 1$. Applying (1.16) to the 1x1 classification gives $0 = R(1)W(1) = R(1)$ which is a contradiction. Finally, if W is assumed to be measurable, apply (1.16) to the general independent two-way classification and proceed as above but with a double application of Theorem 1.1.

For the second characterization we make no specific assumptions concerning the functional form of the index.

Theorem 1.19: Suppose there is a deflation factor W such that the index $\Delta = \Delta(\underset{\sim}{\pi})$ satisfies (1.16) for all independent classifications. Also assume that (i) W is measurable and (ii) $\Delta(\underset{\sim}{\pi})$ is nonconstant and permutation invariant. Then there is a real number β (possible $\beta = +\infty$) such that $W(\pi) = \pi^{\beta}$ and, provided $\beta \neq 0$, $\Delta = \text{constant}\cdot\Delta_{\beta}$.

(If $\beta = 0$, additional substantive assumptions are needed to characterize the Shannon index. A simple counterexample is provided by $\Delta = \log(s)$. See [34, chap. 3; 39] for some possible additional assumptions.)

Proof: Write $H(\underset{\sim}{\pi}) = \Sigma\pi_i W(\pi_i)$. For independent classifications AxB, (1.16) takes the form of a multivariate proportionality equation $\Delta(AxB) = \Delta(A) + H(A)\Delta(B)$. Consider an arbitrary independent classification AxBxC where $\Delta(C) \neq 0$; expanding both sides of $(Ax(BxC)) = ((AxB)xC)$ gives $H(AxB) = H(A)H(B)$. A double application of Theorem 1.1 to this last equation shows that $W(xy) = W(x)W(y)$ for $0 < x, y < 1$. It is easy to see that W is not identically zero so that $W(x) = x^{\beta}$ for some β. Take AxB as the independent two-way classification with marginals $\underset{\sim}{\pi}$ and (1/2, 1/2). If $\beta \neq 0$, then also $1 - H(B) = 1 - 1/2^{\beta} \neq 0$. Using (1.16) to expand both sides of $\Delta(AxB) = \Delta(BxA)$ and solving for $\Delta(A)$ gives $\Delta(A) = [\Delta(B)/(1 - H(B))](1 - H(A)) = \text{constant}\cdot\Delta_{\beta}(A)$.

1.8.3. Diversity of a Mixture. For random variables, the variance of a mixture is greater than or equal to the average variance of the components; the difference is the between-component variance and is nonnegative. Here, we exploit the analogy between diversity and variance to effect an ANOVA-like decomposition for the diversity of a community $\overline{C} = (\overline{s}, \overline{\underset{\sim}{\pi}})$ which is a mixture of subcommunities $C_1 = (s_1, \underset{\sim}{\pi}^{(1)})$, $C_2 = (s_2, \underset{\sim}{\pi}^{(2)}), \ldots, C_n = (s_n, \underset{\sim}{\pi}^{(n)})$ with $w_1, w_2, \ldots, w_n$ as the mixing proportions. Note that $\overline{\underset{\sim}{\pi}} = \Sigma w_k \underset{\sim}{\pi}^{(k)}$ and $\Sigma w_k = 1$.

Example 1.7: In the two-way classification habitat x species take $\underset{\sim}{\pi}^{(1)}, \underset{\sim}{\pi}^{(2)}, \ldots, \underset{\sim}{\pi}^{(n)}$ as the normalized rows, the mixing proportions as the row marginals and $\overline{\underset{\sim}{\pi}}$ as the vector of column marginals.

Throughout this section the diversity index is assumed to satisfy Criterion C3 so that $\Delta(\overline{C}) \geq \Sigma w_k \Delta(C_k)$. The within-community diversity may then be defined as the average $\Sigma w_k \Delta(C_k)$ and the between-community diversity as the difference $\Delta(\overline{C}) - \Sigma w_k \Delta(C_k)$. The structure of the between-community component of diversity will be discussed only for the dichotomous type of index $\Delta = \Sigma \pi_i R(\pi_i)$. Randomly select an individual from the subcommunity C_k. This individual may be regarded as a member of either C_k or $\overline{C}$. Comparing these two viewpoints, the difference in average rarity is

$$\Gamma(C_k; \overline{C}) = \Sigma_i \pi_i^{(k)} [R(\overline{\pi}_i) - R(\pi_i^{(k)})],$$

which we call the rarity gain. With Δ as the Shannon index, $\Gamma(C_k; \overline{C})$ is known as the "information gain (of order 1)" [11, 40] and measures the increase in information attributable to the knowledge that the individual is a member of the particular subcommunity C_k. While the information gain is known to be nonnegative, the rarity gain may be negative for indices other than the Shannon index (see [34, p. 114]). However, as stated by the next theorem, the between-community diversity is the rarity gain averaged over all the subcommunities so that this average is nonnegative.

Theorem 1.20: Assume that the dichotomous index Δ satisfies Criterion C3. Then $\Delta(\overline{C}) = \Sigma w_k \Delta(C_k) + \Sigma w_k \Gamma(C_k; \overline{C})$.

Proof: Straightforward.

Nei [32] has studied the decomposition of Theorem 1.20 in the case of Simpson index and has shown that the between-community diversity may be written as

$$\Sigma w_k \Gamma(C_k; \overline{C}) = (1/2) \Sigma w_i w_j |\underset{\sim}{\pi}^{(i)} - \underset{\sim}{\pi}^{(j)}|^2 \qquad (1.20)$$

where $|\cdot|$ is the ordinary Euclidian distance. It is clear from (1.20) that (for the Simpson index) the between-community diversity is zero if and only if all the subommunities with nonzero mixing proportions have identical abundance vectors. In fact this will be true for any dichotomous index whose auxiliary function $V(\pi) = \pi R(\pi)$ is strictly concave. In particular, it is true for the index Δ_β when $\beta > -1$.

The decomposition of Theorem 1.20 is easily extended to higher order mixtures. For example, with a doubly indexed family of subcommunities C_{ij} and with mixing proportions w_{ij}, one obtains, with the obvious notation,

$$\Delta(\overline{C}_{..}) = \Sigma w_{ij}\Delta(C_{ij}) + \Sigma w_{ij}\Gamma(C_{ij};\ \overline{C}_{i.}) + \Sigma w_{i.}\Gamma(\overline{C}_{i.};\ \overline{C}_{..}). \qquad (1.21)$$

1.8.4. Disorder, Hardy-Weinberg Equilibrium, and a Law of Increasing Entropy. This section attempts to show that entropy and diversity are not equivalent concepts, that entropy measures may be used as diversity measures but not conversely.

There appears to be no precise definition, in the scientific literature, of entropy *as a concept* and none is offered here. Webster's Third International Dictionary [41, p. 759] states that entropy is "... the absence of form, pattern, hierarchy, or differentiation... the general trend of the universe toward final death and disorder." The key phrases here are "absence of pattern" and "disorder." A standard example from physics clarifies the point. Consider a chamber occupied by gas molecules. If most of the molecules are initially located in the left half of the chamber there will be a net motion to the right and work may be extracted from the system. Gradually the molecules will become more uniformly distributed throughout the chamber and once the state of complete uniformity (disorder) is achieved, the energy of the system is no longer available for doing work.

A similar trend toward disorder occurs when a biological population approaches Hardy-Weinberg equilibrium. The Hardy-Weinberg law states that, under certain restrictive assumptions, the marginal gene frequencies remain constant from generation to generation while the gamete frequencies converge to a product probability measure. The restrictive assumptions [42, p. 60] include random mating, infinite population and no mutation, migration or selection. These assumptions are not likely to be met in practice; evolutionary dynamics resists the trend toward disorder by organizing the gene combinations into clusters that are best adapted to particular environmental conditions (Lewontin [21], Dobzhansky [43, p. 21]). It is shown below that the Shannon measure (but not the Simpson measure) of the gamete frequencies increases during the approach to Hardy-Weinberg equilibrium. Presumably, this law of increasing entropy is well-known to geneticists, but there seems to be no ready reference.

The simple case of two loci each having two alleles is considered first. Suppose the alleles at these two loci are represented by a_1, a_2 and b_1, b_2, respectively. The possible gametes are then a_1b_1, a_1b_2, a_2b_1, and a_2b_2 and the gamete frequencies in the initial generation may be displayed in a 2x2 contingency table:

$$\begin{array}{c|cc|} & b_1 & b_2 \\ \hline a_1 & \pi_{11} & \pi_{12} \\ a_2 & \pi_{21} & \pi_{22} \\ \hline \end{array} \tag{1.22}$$

Let $\gamma_0 = \pi_{11}\pi_{22} - \pi_{12}\pi_{21} = \pi_{11} - \pi_{1\cdot}\pi_{\cdot 1}$ be the covariance of this table when the value zero is assigned to a_1 and b_1 and the value one to a_2 and b_2. Also suppose that $0 < \lambda < 1$ is the recombination fraction. In the next generation, the covariance is

$$\gamma_1 = (1 - \lambda)\gamma_0, \tag{1.23}$$

and the gamete frequencies are:

$$\begin{array}{c|cc|} & b_1 & b_2 \\ \hline a_1 & \pi_{11} - \lambda\gamma_0 & \pi_{12} + \lambda\gamma_0 \\ a_2 & \pi_{21} + \lambda\gamma_0 & \pi_{22} - \lambda\gamma_0 \\ \hline \end{array} . \tag{1.24}$$

(For details of the derivation, see Elandt-Johnson [42, Section 4.51] or Kempthorne [44, Section 2.8].) Notice that (1.22) and (1.24) have the same marginals. Iterating (1.23), the covariance in the nth generation is $(1 - \lambda)^n\gamma_0$ which converges to zero. Thus, in the limit, the gamete frequencies become independent, which is the state of maximum disorder given the requirement of fixed marginals.

The transition from (1.22) to (1.24) will be called a <u>Hardy-Weinberg transfer</u>. Hoeffding [45] and Tchen [46] have considered a similar transfer of probability in attempting to formalize the concept of correlation.

<u>Theorem 1.21</u>: The Shannon index increases under a Hardy-Weinberg transfer. The Simpson index does not, in general.

<u>Proof</u>: Let $\pi_{ij}(i, j = 1,2)$ be fixed and write $\pi_{ij}(\lambda) = \pi_{ij} - (-1)^{i+j}\lambda\gamma_0$ where $0 < \lambda < 1$. Now,

$$d[-\Sigma\pi_{ij}(\lambda)\log \pi_{ij}(\lambda)]/d\lambda = \gamma_0 \log \omega, \tag{1.25}$$

where $\omega = \pi_{11}(\lambda)\pi_{22}(\lambda)/(\pi_{21}(\lambda)\pi_{12}(\lambda))$ is the odds ratio. But the odds ratio is greater than, less than, or equal to one exactly when the covariance γ_0 is positive, negative, or zero. Thus, the right hand side of (1.25) is always nonnegative. A counterexample for the Simpson index is given in the next example.

Example 1.8: Let the initial gamete frequencies by $\pi_{11} = .26$, $\pi_{12} = .22$, $\pi_{21} = .27$, and $\pi_{22} = .25$ and suppose the recombination fraction is 1/2. The equilibrium frequencies are given in Table 1.2(a). The values of the Shannon index and the Simpson index for successive generations are given in Table 1.2(b). Note that the Simpson index first increases but then decreases. Examples can also be given in which the Simpson index steadily increases, steadily decreases, or first decreases and then increases. In all cases, however, the Simpson index is unimodal.

The next theorem states that any dichotomous index which increases under Hardy-Weinberg transfers may also be used as a diversity index. No attempt has been made to identify the best possible regularity assumptions.

Theorem 1.22: Let Δ be a dichotomous index whose auxiliary function $V(\pi) = \pi R(\pi)$ is continuously differentiable for $0 \leq \pi \leq 1$. The possibility that $V'(\pi)$ is infinite at $\pi = 0$ is not excluded. If Δ increases under all Hardy-Weinberg transfers, then it also satisfies Criterion C2.

Proof: Let $0 \leq y < x \leq 1$ and $x + y \leq 1$. In Theorem 1.5, π_i and π_j correspond to x and y respectively. First assume that $x + y < 1$ and put $z = (1 - x - y)/2 > 0$. The first table immediately below has positive covariance so that the following is a Hardy-Weinberg transfer for all sufficiently small $h > 0$:

$$\begin{array}{|cc|} \hline x & y \\ z & z \\ \hline \end{array} \rightarrow \begin{array}{|cc|} \hline x - h & y + h \\ z + h & z - h \\ \hline \end{array} .$$

Table 1.2 Numerical Example of the Approach to Hardy-Weinberg Equilibrium ($\lambda = 1/2$)

(a) Gamete Frequencies

Initial

	b_1	b_2	
a_1	.26	.22	.48
a_1	.27	.25	.52
	.53	.47	

Equilibrium

	b_1	b_2	
a_1	.2544	.2256	.48
a_2	.2756	.2444	.52
	.53	.47	

(b) Parameter Changes

Generation	Covariance	Shannon Index	Simpson Index
0	0.0056 0000	1.3834 4084	.7486 0000
1	0.0028 0000	1.3836 3002	.7486 8064
2	0.0014 0000	1.3836 7730	.7486 9744
3	0.0007 0000	1.3836 8913	.7486 9996
4	0.0003 5000	1.3836 9208	.7486 9975
5	0.0001 7500	1.3836 9282	.7486 9928
6	0.0000 8750	1.3836 9301	.7486 9895
7	0.0000 4375	1.3836 9305	.7486 9876
∞	zero	1.3836 9307	.7486 9856

We must have $d\Delta/dh|_{h=0} \geq 0$ where $\Delta = \Delta(h)$ is the index evaluated on the second table. But this derivative is $V'(y) - V'(x) + V'(z) - V'(z) = V'(y) - V'(x)$. Thus $V'(y) \geq V'(x)$. Taking limits, this inequality also holds when $x + y = 1$. Theorem 1.5 now gives the result.

The notion of a Hardy-Weiberg transfer can be extended to the case of multiple alleles and/or multiple loci. However, the notation becomes very complicated and a different approach will be taken. The idea is to write each generation's gamete frequencies as a convex linear combination of marginal products for the preceding generation and then apply Criterion C3. We first illustrate the method in the previous case of two loci each having two alleles. Let $\underset{\sim}{\pi} = (\pi_{ij})$ be the initial generation's frequencies, $\underset{\sim}{\pi}^{(1,2)} = (\pi_{i\cdot}\pi_{\cdot j})$ the equilibrium frequencies, and $\underset{\sim}{\hat{\pi}}$ the next generation's frequencies. Now (1.24) may be rewritten in the form $\underset{\sim}{\hat{\pi}} = (1 - \lambda)\underset{\sim}{\pi} + \lambda\underset{\sim}{\pi}^{(1,2)}$. Since the Shannon index Δ_0 satisfies Criterion C3, this implies that $\Delta_0(\underset{\sim}{\hat{\pi}}) \geq (1 - \lambda)\ \Delta_0(\underset{\sim}{\pi}) + \lambda\Delta_0(\underset{\sim}{\pi}^{(1,2)})$. But it is shown in the next lemma that $\Delta_0(\underset{\sim}{\pi}^{(1,2)}) \geq \Delta_0(\underset{\sim}{\pi})$ so that $\Delta_0(\underset{\sim}{\hat{\pi}}) \geq \Delta_0(\underset{\sim}{\pi})$, i.e., the Shannon index increases from generation to generation.

Lemma 1.3: Let AxB be a two-way classification and A*B be the independent classification with the same marginals. Let Δ_0 be the Shannon index. Then

a) $\Delta_0(A*B) = \Delta_0(A) + \Delta_0(B)$,

b) $\Delta_0(A*B) \geq \Delta_0(AxB)$.

Proof: The first assertion follows from the ANOVA decomposition (1.16) since $W(\pi) \equiv 1$ for the Shannon index. But by Criterion C3, $\Delta_0(B) \geq E_A[\Delta_0(B|A_i)]$. Comparing part (a) with (1.16) gives the second assertion.

The general proof that the Shannon index increases from generation to generation is now easy. Label the different loci as 1, 2, 3,...,m and let $\underset{\sim}{\pi} = (\pi_{i_1, i_2 --- i_m})$ be the initial gamete frequencies. For each partition of {1, 2, 3,...,m} into two nonempty disjoint subsets U and V define a marginal product $\underset{\sim}{\pi}^{(U,V)}$ as follows:

$$\pi^{(U,V)}_{i_1 i_2 --- i_m} = \pi_{u_1 u_2 --- u_m} \pi_{v_1 v_2 --- v_m}$$

where the symbol u_j stands for i_j if j is in U and for • (dot) if j is in V and where v_j stands for i_j if j is in V and for • (dot) if j is in U. For example, if m = 4 and if the partition consists of the two sets {1, 3} and {2, 4}, then

$$\pi^{(U,V)}_{hijk} = \pi_{h \cdot j \cdot} \pi_{\cdot i \cdot k} \cdot$$

Finally, let $\hat{\underset{\sim}{\pi}}$ be the gamete frequencies in the next generation. It is not hard to see that $\hat{\underset{\sim}{\pi}}$ is a convex linear combination of $\underset{\sim}{\pi}$ and the marginal products $\underset{\sim}{\pi}^{(U,V)}$. Since $\Delta_0(\underset{\sim}{\pi}^{(U,V)}) \geq \Delta_0(\underset{\sim}{\pi})$, it follows as above that $\Delta_0(\hat{\underset{\sim}{\pi}}) \geq \Delta_0(\underset{\sim}{\pi})$.

Remark 1.9: The above ideas can be modified to produce a simple general proof of the Hardy-Weinberg law (provided, of course, that no two loci are rigidly linked). Just as before, a marginal product can be defined for a partition consisting of any number of sets. Thus each partition may be thought of as an operator acting on the class of possible gamete frequencies. The recombination laws induce a Markov chain whose states are the partitions. Transitions from one partition to another are possible only if the second partition is the same as the first or is finer than the first. Moreover, the probability of a self-transition is strictly less than one if no two loci are rigidly linked. Thus the finest partition is the unique absorbing state; but this partition induces the product probability measure.

1.9 Estimation of Simpson's Index

A problem associated with the use of diversity indices as indicators of environmental quality is the time and level of professional expertise required for a taxonomic classification of the sample. Cairns et al. [47, 48] have developed an ingenious technique to overcome this difficulty. Their approach is a nice illustration of the concept of inter- and intraspecific encounters discussed in Section 1.3.1. Given a random sample $A_1, A_2, \ldots, A_N, A_{N+1}$ of specimens, define a <u>run</u> to be a maximal sequence of consecutive specimens of the same species. Cairns suggests the ratio, (# runs)/(N + 1), as a measure of the diversity of the sample. In implementing the technique, the investigator need only make the successive comparisons A_1 vs. A_2, A_2 vs. A_3, A_3 vs. A_4, etc., so that the method is rapid and does not call for sophisticated taxonomic skill.

In what follows, it is shown that, with a minor bias correction, Cairns diversity measure becomes an unbaised estimator of Simpson's index Δ_1. Asymptotic normality is also established. The unbiased version is obtained as CL = (# runs - 1)/N and will be called <u>Cairns linked estimator</u>. The statistical analysis is facilitated by introducing indicator random variables T_i with the property that each occurrence of

$T_i = 1$ signals the start of a new run. The T_i are defined by

$$T_i = \begin{cases} 1 & \text{if } A_i \text{ and } A_{i+1} \text{ belong to different species} \\ 0 & \text{otherwise.} \end{cases}$$

Then $T_1, T_2, \ldots, T_N$ are identical 0-1 random variables, but adjacent T_i need not be independent (since the comparisons are linked). Let ρ be the correlation between T_1 and T_2.

Theorem 1.23: a) $E[T_1] = \Delta_1$, b) $\mathrm{Var}[T_1] = \Delta_1(1 - \Delta_1)$, c) $\mathrm{Cov}(T_1, T_2) = \Sigma\pi_i^3 - (\Sigma\pi_i^2)^2$, d) $0 \leq \rho \leq 1/2$. The lower bound $\rho = 0$ is achieved only for a completely even community while the upper bound $\rho = 1/2$ is approached as $\pi_1^* \to 1$. (ρ is undefined when $\pi_1^* = 1$.)

Proof: a) and b) are obvious since T_1 is a 0-1 random variable with $P(T_1 = 1) = \Delta_1$. c) $\mathrm{Cov}(T_1, T_2) = \mathrm{Cov}(1 - T_1, 1 - T_2) = P(T_1 = 0, T_2 = 0) - (P(T_1 = 0))^2 = \Sigma\pi_i^3 - (\Sigma\pi_i^2)^2$. d) The covariance has the form of a variance, $\Sigma\pi_i \cdot \pi_i^2 - (\Sigma\pi_i \cdot \pi_i)^2$. Hence $\rho > 0$ with equality <=> the nonzero π_i are all equal. We employ a standard inequality to show that $\rho \leq 1/2$:

$$\begin{aligned} \Sigma\pi_i^4 &\leq (\Sigma\pi_i^2)^2 \quad \text{([6, p. 28])} \\ &\leq \Sigma\pi_i^2(1 - \pi_i)^2 + (\Sigma\pi_i^2)^2 \\ &\leq \Sigma\pi_i^2 - 2\Sigma\pi_i^3 + \Sigma\pi_i^4 + (\Sigma\pi_i^2)^2. \end{aligned}$$

Therefore,

$$2\Sigma\pi_i^3 - 2(\Sigma\pi_i^2)^2 \leq \Sigma\pi_i^2 - (\Sigma\pi_i^2)^2 = \Delta_1(1 - \Delta_1).$$

Remark 1.10: The bounds $0 \leq \rho \leq 1/2$ can be improved if the value of Δ_1 is known. For given Δ_1, we have obtained sharp upper and lower bounds on ρ. The upper bound confirms the intuition that ρ tends to be small for highly diverse communities.

Theorem 1.24: a) $CL = \sum_1^N T_i/N$, b) $E[CL] = \Delta_1$,

c)
$$\begin{aligned} \mathrm{Var}[CL] &= [1 + 2\rho - 2\rho/N]\Delta_1(1 - \Delta_1)/N \\ &\simeq [1 + 2\rho]\Delta_1(1 - \Delta_1)/N \text{ for large } N \\ &\leq 2\Delta_1(1 - \Delta_1)/N, \end{aligned}$$

d) CL is asymptotically normal as $N \to \infty$.

Proof: a) and b) are obvious and c) is a routine calculation once it is noted that nonadjacent T_i are independent. The asymptotic normality follows from Noether's central limit theorem which is stated next.

Theorem 1.25: (Noether [49]). Let $Z_1, Z_2, Z_3, \ldots$ be independent random variables. a) Let $T_1, T_2, T_3, \ldots$ be uniformly bounded random variables with T_i a function of Z_i and Z_{i+1} only. Then $S_N = \sum_{i=1}^N T_i$ is asymptotically normal provided $\mathrm{Var}[S_N]$ is of exact order N. b) Let T_{ij}, i, j = 1,2,... be uniformly bounded random variables with T_{ij} a function of Z_i and Z_j only. Then $S_N = \sum_{i,j=1}^N T_{ij}$ is asymptotically normal provided $\mathrm{Var}[S_N]$ is of exact order N^3.

Remark 1.11: Noether presents a proof of b). The proof of a) requires only slight modifications in his argument.

Remark 1.12: Theorem 1.24 is related to a result of Mood [50] who examined the distribution of the number of runs when a fixed sample is subjected to random permutations.

Use of the exact formula for Var[CL] requires estimation of the

the correlation ρ. The pairs (T_1,T_2), (T_4,T_5), (T_7,T_8)... constitute approximately N/3 independent observations on the bivariate distribution of (T_1,T_2) from which an estimate may be obtained.

The nonnegativity of ρ indicates that linking the consecutive comparisons reduces the efficiency of the estimate and suggests an estimate based upon independent pairs of specimens. For M such pairs (2M specimens), define Cairns unlinked estimator as

$$CU = (\# \text{ of unlike pairs})/M.$$

Then, trivially,

Theorem 1.26: a) $E[CU] = \Delta_1$, b) $Var[CU] = \Delta_1(1 - \Delta_1)/M$, c) CU is asymptotically normal as $M \to \infty$.

Efficiency comparisons of the linked and unlinked estimators require a common yardstick. The number of specimens in the sample is a natural yardstick if specimens are difficult to obtain. For $n = 2M = N + 1$ specimens,

$$\frac{Var[CL]}{Var[CU]} = \frac{n}{n-1}\left[\frac{1}{2} + \rho\,\frac{n-2}{n-1}\right] \simeq \frac{1}{2} + \rho$$

and (approximately),

$$\frac{1}{2} \leq \frac{Var[CL]}{Var[CU]} \leq 1.$$

Thus CL is at least as efficient as CU in these circumstances and may be twice as efficient.

In the event that specimens are easily obtained, the number of comparisons seems to be a reasonable statistical yardstick. Here the situation is reversed. For N = M comparisons,

$$\frac{Var\ [CL]}{Var\ [CU]} = 1 + 2\rho(1 - 1/N) \simeq 1 + 2\rho$$

and the relative efficiency is bounded by

$$1 \leq \frac{\text{Var [CL]}}{\text{Var [CU]}} \leq 2.$$

Operational considerations may give preference to the linked estimator, however; this is especially true for a highly diverse community where ρ may be expected to be small.

We next turn our attention to the case of a complete taxonomic classification of the sample, which is taken to consist of t species and n specimens with $x_1, x_2, \ldots, x_t$ as the species counts. Simpson [3] has shown that $D = \Sigma x_i(n - x_i)/(n(n - 1))$ is an unbiased estimator of Δ_1. On the basis of the asymptotic behavior of the third and fourth moments, he also concluded that D was likely to be asymptotically normal (provided $\rho \neq 0$). Bowman et al. [51] have also examined the moments of D. We show how Noether's central limit theorem can be used to establish the asymptotic normality without the need for laborious moment calculations.

<u>Theorem 1.27</u>: a) $E[D] = \Delta_1$, b) $\text{Var [D]} = \frac{2\Delta_1(1 - \Delta_1)}{n(n - 1)}[1 + 2(n - 2)\rho]$, c) If $\rho \neq 0$, D is asymptotically normal.

<u>Proof</u>: Let $S_n = \sum_{i,j=1}^{n} T_{ij}$, where

$$T_{ij} = \begin{cases} 1 \text{ if } i < j \text{ and specimens } i \text{ and } j \text{ are of different species} \\ 0 \text{ otherwise.} \end{cases}$$

It may be seen that $D = 2S_n/(n(n - 1))$. Using Theorem 1.23, the calculation of E[D] and Var[D] is routine after it is noted that T_{ij} and T_{km} are independent whenever $\{i,j\}$ and $\{k,m\}$ are disjoint. Part b) implies that $\text{Var}[S_n]$ is of exact order n^3 when $\rho \neq 0$. The asymptotic normality of D now follows from Noether's theorem.

Remark 1.13: For small sample size, Nei and Roychoudhury [52] have found that the biased estimator $\Sigma x_i(n - x_i)/n^2$ has a smaller mean square error than D.

1.10 Estimation of Species Richness

Estimation of the number of species in the community is one of the more interesting and intriguing problems facing the ecologist. Usually, he is not in a position to establish sharp boundaries for the community and finds it impossible to describe his sampling scheme with any degree of statistical precision. Likewise we will not be precise about the sampling method, but shall suppose that the species are represented in the sample independently of one another and in accord with their own individual probability distributions. As usual, s is the number of species in the community and $\underset{\sim}{\pi}$ is the species abundance vector.

Let $p_i(x)$, $x = 0,1,2,\ldots, i = 1,2,\ldots,s$, be the probability that species i is represented in the sample by x individuals. The probability distributions $p_1(x), p_2(x),\ldots,p_s(x)$ depend in an unknown way upon the vector $\underset{\sim}{\pi}$, the sampling intensity, and the response of the various species to the sampling effort (catchability). The number of species actually present in the sample (i.e. represented by a positive number of individuals) is itself a random variable. This random variable (and sometimes its observed value) is denoted by t. We have

$$E[t] = \sum_{i=1}^{s} [1 - p_i(0)] = s[1 - \frac{1}{s} \sum_{i=1}^{s} p_i(0)] = s[1 - \overline{p}(0)],$$

where $\overline{p}(0)$ is the average species absence probability. If an estimate $\tilde{p}(0)$ of $\overline{p}(0)$ is available, then s may be estimated as $\tilde{s} = t/(1 - \tilde{p}(0))$.

Estimation Method 1: Assume that the probability distributions $p_1(x),\dots,p_s(x)$ are members of a known parametric family $p(x;\ \theta)$ with possibly different values of θ: $p_i(x) = p(x;\ \theta_i)$, $i = 1,\dots,s$. Further assume that θ is a scalar parameter and can be estimated from a single observation. Let $i_1, i_2,\dots,i_t$ be the (unknown) labels of the species actually present in the sample. For $j = i_1, i_2,\dots,i_t$, estimate θ_j by, say, $\tilde{\theta}_j$ and $p(0;\ \theta_j)$ by $p(0;\ \tilde{\theta}_j)$. Take $\sum_j p(0;\ \tilde{\theta}_j)/t$ as the estimate of $\overline{p}(0)$.

Estimation Method 2: Assume that the probability distributions $p_1(x),\dots,p_s(x)$ are all identical and equal to $p(x;\ \underset{\sim}{\theta})$ where $p(x;\ \underset{\sim}{\theta})$ is a known parametric family and $\underset{\sim}{\theta}$ is a vector of parameters. Let $X_1, X_2,\dots,X_t$ be the observed (nonzero) species counts in the sample, listed in some random order. It may be shown that, conditional on t, $X_1, X_2,\dots,X_t$ are independent and identically distributed with common distribution $p(x;\ \underset{\sim}{\theta})/(1 - p(0;\ \underset{\sim}{\theta}))$, which gives an estimate $\tilde{\underset{\sim}{\theta}}$ of $\underset{\sim}{\theta}$. Estimate $\overline{p}(0) = p(0;\ \underset{\sim}{\theta})$ by $p(0;\ \tilde{\underset{\sim}{\theta}})$.

The above methods are standard in the literature. We suggest a third method that seems to be new and which leads to modified versions of Method 1 and Method 2.

Estimation Method 3: Let the sample consist of n individuals of which n_1 are singletons. Assume the sample is representative of the community in the sense that the 'rare' species (i.e. singletons) in the sample correspond to the 'rare' species (i.e. unobserved) in the community. The singletons taken together comprise a fraction n_1/n of the sample and they divide this fraction among themselves into n_1 equal parts. Extrapolating (not interpolating!) to the community, conclude that the unobserved species comprise a fraction n_1/n of the community and that they divide this fraction among themselves into n_1 equal parts. In particular, estimate the number of unobserved species as n_1 and the total number of species as $t + n_1$. To correct for any bias of this estimator, note that

$$E[t] = s[1 - \bar{p}(0)],\ E[n_1] = \sum_{i=1}^{s} p_i(1) = s\bar{p}(1),$$

$$E[t + n_1] = s[1 - \bar{p}(0) + \bar{p}(1)].$$

Under the assumptions of either Method 1 or Method 2, both $\bar{p}(0)$ and $\bar{p}(1)$ may be estimated by, say, $\tilde{p}(0)$ and $\tilde{p}(1)$. Take $\tilde{s} = (t + n_1)/(1 - \tilde{p}(0) + \tilde{p}(1))$ as a modified estimate of s.

Example 1.9: Both Method 1 and Method 2 and their modified versions were applied to the Rothamsted light trap data reported by Bliss [53]. The underlying distribution for these methods was assumed to be Poisson. The results are presented in Table 1.3; for comparison the last column gives the estimate of s which Bliss obtained by fitting the lognormal distribution (Method 2). For the year 1934 Bulmer [54] has also obtained the estimate $\tilde{s} = 226$ by fitting the Poisson-lognormal distribution.

Table 1.3 Estimates of Species Richness

			Method				
Year	t	n_1	(1)	Modified(1)	(2)	Modified(2)	Lognormal
1933	183	32	232	250	183	215	208
1934	176	34	226	252	176	210	199
1935	202	39	260	289	202	241	239
1936	157	51	243	296	157	208	222

Remark 1.14: In a different context and with a different viewpoint, Robbins [55] has suggested an estimator similar to n_1/n for estimating the proportion of unobserved outcomes. Also see Good [14].

1.11 The Diversity Concept in Other Fields

The previous sections have emphasized ecological diversity. Here we survey several additional areas in which the diversity concept has been found to be of some value.

1.11.1 Genetic Diversity. What proportion of human genetical variation is accounted for by a system of racial classification which is largely based on morphological characteristics? The question has generated a good deal of controversy. To obtain some quantitative answers, Lewontin [21] subdivides the human species into races and the races into populations and develops the mixture decomposition (1.21) for the Shannon index. Examining the gene frequencies at 17 loci, he finds that, on the average, 6.3% of the genetic diversity is accounted for by racial classification, 8.3% by population differences within a race, and 85.4% by variation among individuals. His final conclusion:

> It is clear that our perception of relatively large differences between human races and subgroups, as compared to the variation within these groups, is indeed a biased perception and that, based on randomly chosen genetic differences, human races are remarkably similar to each other, with the largest part by far of human variation being accounted for by the differences between individuals [21, p. 397].

1.11.2. Linguistic Diversity. In a quite interesting paper, Greenberg [56] describes eight diversity indices which might be applied to the measurement of the communication potential among the inhabitants of a geographical region. The languages spoken within the region correspond to the species; π_i is the proportion of inhabitants who speak language i.

Greenberg's approach is to identify the special features of his problem and adapt Simpson's index to those features. He first observes that some languages are quite similar to one another and introduces a resemblance factor r_{ij} which measures the similarity between language i and language j. The monolingual weighted index is then given by $B = 1 - \Sigma\pi_i r_{ij} \pi_j$. As in Section 1.8.1, B is the average linguistic differentness between two randomly selected inhabitants.

The mathematical form of the monolingual weighted index also arises in assessing the effect on Simpson's index of taxonomic misclassification. Suppose the taxonomist classifies species i as species j with probability P_{ij} so that $\underset{\sim}{P} = (P_{ij})$ is a (row wise) stochastic matrix. The true species abundance vector $\underset{\sim}{\pi}^t = (\pi_1, \pi_2, \ldots, \pi_s)$ is then observed as $\underset{\sim}{\pi}^t \underset{\sim}{p}$ and Simpson's index as $1 - \underset{\sim}{\pi}^t \underset{\sim}{P}\underset{\sim}{P}^t \underset{\sim}{\pi}$. (Notice that the (i,j) entry of $\underset{\sim}{P}\underset{\sim}{P}^t$ is likely to be large when species i and j closely resemble one another.) There is no distortion in Simpson's index precisely when $\underset{\sim}{P}$ is an orthogonal matrix. However, an orthogonal stochastic matrix is necessarily a permutation matrix and for all practical purposes misclassification will distort Simpson's index. In some cases, it should be possible to obtain prior estimates of $\underset{\sim}{P}$ and to correct for the distortion.

A second special feature identified by Greenberg is the possibility that n inhabitant might speak more than one language. He proposes several solutions; for example, in the random-speaker-hearer method, a randomly selected individual speaks one of his languages at random. The index is the probability that a second randomly chosen individual cannot understand this language. As Greenberg points out, this does not take into account the relative frequency which which the speaker actually employs the various languages in his repertoire.

Observing that a geographical region may be an aggregate of several subregions, Greenberg also raises the question, as in the mixture decomposition of Section 1.8.3, of how the diversity of the whole can be computed from knowledge of the parts. He cites the example of Mexico and its 32 provinces and points out that the diversity of the whole may exceed the diversity of any or of all of its parts but he carries the analysis no further. Lieberson [57, 58] has extended Greenberg's indices to measures of the "diversity between two subpopulations."

Greenberg suggests that high linguistic diversity will often be associated with inadequate communication, lowered economic productivity, and political instability. By contrast, diversity is generally considered desirable in ecology.

Diversity also occurs in word frequency studies as an index of literary style (cf. Guiraud [59], Herdan [60, 61] and Yule [62]). The word-types in an author's lexicon correspond to the species and the relative frequency of occurrence of these word-types to the species proportions. Literary text has a distinguishing feature though: the words are arranged in a linear order for which a stochastic model might be constructed. Brainerd [63], for example, has attempted to fit an inhomogeneous pure birth process.

<u>1.11.3 Industrial Concentration</u>. Consider s business firms whose proportionate shares of the market are $\pi_1, \pi_2, \ldots, \pi_s$. According to Horowitz [64], the diversity of the market shares will serve as a meaningful index of industrial competition. Usually economists prefer the dual notion of concentration but, to avoid confusion, we will stick to diversity. Both the Shannon index and the Simpson index are widely used

[8, 10, 64, 65, 66, 67], the latter under the name of Herfindahl index. Theil [30, 31] is a strong advocate of the Shannon index because of its ANOVA decomposition. Hall and Tideman [68] have suggested a variant of the average rank; they also advocate the principle of transfers.

Arguing that the competitive patterns within an industry are largely determined by the dominant firms, Hart [9, 69] claims that the Simpson index is sometimes overly sensitive to the entry of a few small firms. (As discussed in Section 1.7.1, the Shannon measure is even more sensitive.) This is interesting since the highly sensitive index s - 1 is fashionable among ecologists.

1.12 Income Inequality

1.12.1 Inequality as Relative Unevenness. Suppose there are s individuals in a population and that the ith individual receives a fraction π_i of the total income. At first glance it might seem that the diversity of $\underset{\sim}{\pi} = (\pi_1, \pi_2, \ldots, \pi_s)$ would serve as an inverse measure of income inequality. Indeed, Theil [30, p. 128] asserts that "concentration and inequality are essentially the same concepts" and that $\Sigma\pi_i^2$ (which is a decreasing transformation of the Simpson index) can be used as a measure of inequality. This is not the case, however. Perfect equality occurs when the income shares π_i are all the same; in these circumstances, a measure of inequality should take on the same value irrespective of the number of recipients. On the other hand, industrial concentration decreases with an increase in the number of firms in an equally distributed market. Thus $\Sigma\pi_i^2$ is a suitable measure of concentration but not of inequality.

We have pointed out that diversity is influenced by two factors: evenness and richness. In light of the preceding paragraph, it appears that a measure of income inequality should take into account only the unevenness of the income distribution and not the number of recipients. In fact, it is desirable to relativize this last statement: if the population is partitioned into subpopulations, we identify income inequality with the unevenness of the income shares of these subpopulations relative to their population shares. Consider, for instance, two subpopulations A and B with respective total incomes of \$40,000 and \$20,000. While total income is different for the two subpopulations, per capita income is the same if A has four members and B only two. In general, we suppose that there are s subpopulations and let π_i be the fraction of total income and ν_i the proportion of individuals in the ith subpopulation. Some of the π_i may be zero; but, in any case, $\underset{\sim}{\pi}$ is absolutely continuous with respect to $\underset{\sim}{\nu}$. As Theil [30, p. 102] points out, per capita income in the ith subpopulation is proportional to the likelihood ratio π_i/ν_i; the constant of proportionality is the per capita income in the overall population. We refer to π_i/ν_i as the <u>standardized per capita income</u>.

<u>1.12.2 The Lorenz Ordering</u>. An intrinsic inequality ordering, similar to the diversity ordering of Section 1.5, may be defined on the set of pairs $(\underset{\sim}{\pi}, \underset{\sim}{\nu})$ by considering finite sequences of the following operations:

I1. Simultaneously permuting the components of $\underset{\sim}{\pi}$ and $\underset{\sim}{\nu}$ does not change the inequality of $\underset{\sim}{\pi}$ relative to $\underset{\sim}{\nu}$.

I2. Combining two subpopulations which have the same per capita income does not change inequality. Specifically, if $\pi_i/\nu_i = \pi_j/\nu_j$, this operation replaces the two components π_i and π_j by the single component $\pi_i + \pi_j$ and also replaces ν_i and ν_j by $\nu_i + \nu_j$. Conversely, splitting a subpopulation into two subpopulations with the same per capita income does not change inequality.

I3. (Principle of transfers). Inequality is increased by a transfer of income from one subpopulation to another provided the second has a larger per capita income. This operation changes $\underset{\sim}{\pi}$ but not $\underset{\sim}{\nu}$.

Operations I1 and I2 are symmetrical in $\underset{\sim}{\pi}$ and $\underset{\sim}{\nu}$ but I3 is not; consequently, measures of inequality need not be symmetrical in $\underset{\sim}{\pi}$ and $\underset{\sim}{\nu}$. Operation I2 permits the comparison of populations with different numbers of individuals and was suggested to us by axiom A5 of Dasgupta, Sen and Starrett [17].

The intrinsic diversity ordering admitted a simple geometric representation on a simplex. The inequality ordering can also be represented on the cartesian product of two simplices but this requires too many dimensions to be useful. Easier to visualize is the Lorenz curve which is discussed by Theil [30, p. 121] for discrete distributions, by Kendall and Stuart [70, p. 48] for continuous distributions, and by Thompson [71] in general. Thompson [71] also suggests some interesting biological applications. Arranging the subpopulations in order of decreasing per capita income

$$\pi_1/\nu_1 \geq \pi_2/\nu_2 \geq \dots \geq \pi_s/\nu_s,$$

the Lorenz curve is the polygonal path joining the successive points $P_1, P_2, \dots, P_{s+1}$ where $P_{s+1} = (0, 0)$ and

$$P_k = (\sum_{i \geq k} \nu_i, \sum_{i \geq k} \pi_i), \; k = 1,2,\ldots,s.$$

Refer to Figure 1.15. The slope of the segment $P_k P_{k+1}$ is the standardized per capita income π_k/ν_k of the kth subpopulation. Operation I2 has the effect of inserting or removing "extra" vertices (e.g., P_3 of Figure 1.15) on any straight line segment. Transferring income lowers the Lorenz curve as shown in Figure 1.15. A simple induction argument shows that $(\underset{\sim}{\pi}', \underset{\sim}{\nu}')$ is intrinsically more unequal than $(\underset{\sim}{\pi}, \underset{\sim}{\nu})$ <=> the Lorenz curve of $(\underset{\sim}{\pi}', \underset{\sim}{\nu}')$ is uniformly below that of $(\underset{\sim}{\pi}, \underset{\sim}{\nu})$. This generalizes a previous result of Kolm [72].

1.12.3. Measures of Inequality. An obvious measure of income inequality is twice the area between the Lorenz curve and the 45° line. This number is known as the Gini coefficient $G(\underset{\sim}{\pi}; \underset{\sim}{\nu})$ and varies between zero and one. Still supposing that the subpopulations are arranged in order of decreasing per capita income, Theil [30, p. 121] and others have shown that $G(\underset{\sim}{\pi}; \underset{\sim}{\nu})$ is symmetrical in $\underset{\sim}{\pi}$ and $\underset{\sim}{\nu}$ and that

$$G(\underset{\sim}{\pi}; \underset{\sim}{\nu}) = (1/2)\Sigma \nu_i \nu_j |\pi_i/\nu_i - \pi_j/\nu_j| \tag{1.26}$$

and

$$G(\underset{\sim}{\pi}; \underset{\sim}{\nu}) = 1 - 2\Sigma_i \pi_i (\sum_{j \leq i} \nu_j) + \Sigma_i \pi_i \nu_i. \tag{1.27}$$

From (1.26), the Gini coefficient is one half the mean absolute difference between the standardized per capita incomes of two randomly selected subpopulations. When $\underset{\sim}{\nu}$ is completely even (i.e., when each $\nu_i = 1/s$), (1.27) simplifies to

$$G(\underset{\sim}{\pi}; \underset{\sim}{\nu}) = 1 - [2\Sigma(i\pi_i) - 1]/s, \tag{1.28}$$

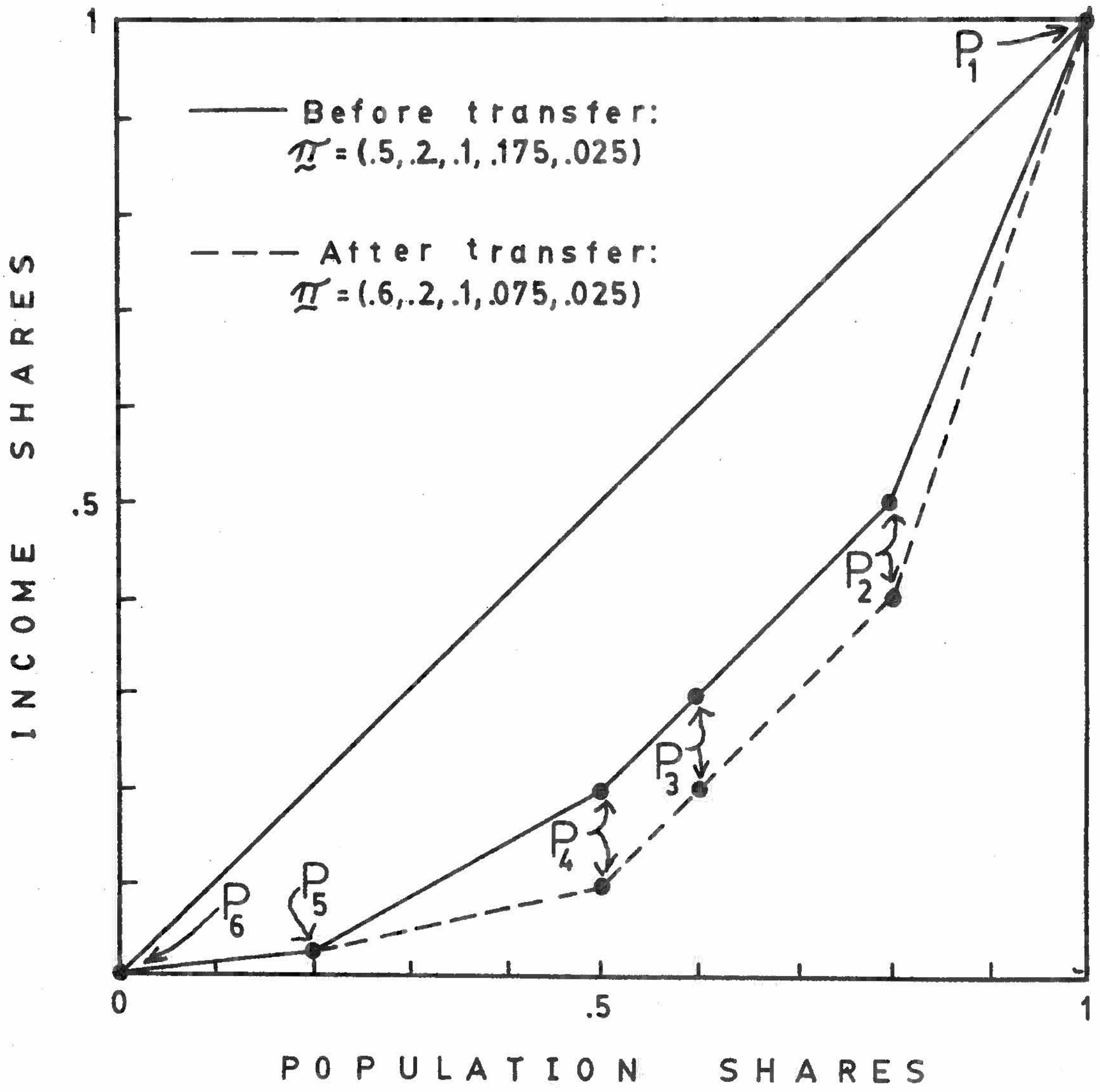


Figure 1.15 Effect on the Lorenz Curve of an Income Transfer

($\underset{\sim}{\nu}$ = (.2, .2, .1, .3, .2)) where $\pi_1 \geq \pi_2 \geq \ldots \geq \pi_s$. Sen [73] gives an axiomatic characterization of $G(\underset{\sim}{\pi}; \underset{\sim}{\nu})$ when $\underset{\sim}{\nu}$ is completely even.

The Gini coefficient and several other standard measures of inequality can be constructed by adapting a method which has been employed by ecologists to measure the evenness of a distribution. They use the ratio of the actual value of a diversity index to the maximum value that the index could assume for a community with the same number of species (Pielou [33, p. 15]). This ratio, however, is not invariant to monotone transformations of the index and it is preferable to first convert to the numbers equivalent and use the ratio S_Δ/s as a measure of (absolute) evenness. To convert this ratio to a measure of inequality several decreasing transformations are possible: $1 - (S_\Delta/s)$, $\log(s/S_\Delta)$, or $(s/S_\Delta) - 1$. All three are instances of the transformation

$$S_\Delta/s \rightarrow [(s/S_\Delta)^\beta - 1]/\beta, \beta \text{ real}. \qquad (1.29)$$

Applying (1.29) with $\beta = -1$ to the Average Rank index gives Gini's coefficient in the form (1.28). Applying (1.29) to the index Δ_β gives

$$[\Sigma\pi_i(\pi_i/s^{-1})^\beta - 1]/\beta \qquad (1.30)$$

as a measure of absolute inequality. To obtain a measure of relative inequality, we replace s^{-1} by ν_i in (1.30) yielding

$$\delta_\beta(\underset{\sim}{\pi}; \underset{\sim}{\nu}) = [\Sigma\pi_i(\pi_i/\nu_i)^\beta - 1]/\beta. \qquad (1.31)$$

The measure δ_β is proportional to the directed divergence of type $\beta + 1$ [34, p. 208; 74] and $\log(1 + \beta\delta_\beta)/\beta$ is Renyi's information gain of order $\beta + 1$ [11, 40].

Remark 1.15: Hill [12] has proposed the ratio S_β/S_α, $\alpha, \beta \geq -1$, as a measure of evenness. (S_β is the numbers equivalent of Δ_β.)

Peet [25, p. 301] remarks that "These ratios do not measure equitability (evenness) as it is normally defined... ." While it is unclear what Peet means by this statement, it may be verified that Hill's ratios violate the principle of transfers unless $\alpha = -1$ or, trivially, unless $\beta = \alpha$. Thus only the ratio S_β/s is a suitable measure of evenness.

When $\beta = -1$, $\delta_\beta(\underset{\sim}{\pi}; \underset{\sim}{\nu})$ reduces to the proportion of the population which has no income. Here we have used the convention that $0 \cdot \infty = 0$. Some other special cases should be mentioned:

Pearson's X^2: Putting $\beta = 1$ in (1.31) gives

$$\begin{aligned}\delta_1(\underset{\sim}{\pi}; \underset{\sim}{\nu}) &= \Sigma(\pi_i - \nu_i)^2/\nu_i \\ &= \Sigma\nu_i(\pi_i/\nu_i - 1)^2,\end{aligned}$$

which is the variance of the standardized per capita incomes or, equivalently, the squared coefficient of variation of the unstandardized per capita incomes.

Theil index: With $\beta = 0$, (1.31) becomes

$$\delta_0(\underset{\sim}{\pi}; \underset{\sim}{\nu}) = \Sigma\pi_i \log(\pi_i/\nu_i),$$

which has been used by Theil as a measure of income inequality.

Bhattacharyya's divergence: Putting $\beta = -.5$ in (1.31) gives

$$\begin{aligned}\delta_{-.5}(\underset{\sim}{\pi}; \underset{\sim}{\nu}) &= 2[1 - \Sigma(\pi_i\nu_i)^{1/2}] \\ &= 4\sin^2(\theta/2) \simeq \theta^2,\end{aligned}$$

where θ is the angle between the two unit vectors $(\sqrt{\pi_1}, \sqrt{\pi_2}, \ldots, \sqrt{\pi_s})$ and $(\sqrt{\nu_1}, \sqrt{\nu_2}, \ldots, \sqrt{\nu_s})$. Bhattacharyya [75] uses θ^2 as a measure of the distance between two multinomial populations. Note that $\beta = -.5$ is the only instance in which $\delta_\beta(\underset{\sim}{\pi}; \underset{\sim}{\nu})$ is symmetrical in $\underset{\sim}{\pi}$ and $\underset{\sim}{\nu}$.

Assuming $\beta \geq -1$, we have the following properties: (i) $\delta_\beta(\underset{\sim}{\pi};\ \underset{\sim}{\nu});\ \geq 0$; (ii) $\delta_\beta(\underset{\sim}{\pi};\ \underset{\sim}{\nu}) = 0 \iff \underset{\sim}{\pi} = \underset{\sim}{\nu}$, provided $\beta \neq -1$; (iii) $\delta_\beta(\underset{\sim}{\pi}';\ \underset{\sim}{\nu}') \geq \delta_\beta(\underset{\sim}{\pi};\ \underset{\sim}{\nu})$ whenever $(\underset{\sim}{\pi}',\ \underset{\sim}{\nu}')$ is intrinsically more unequal than $(\underset{\sim}{\pi},\ \underset{\sim}{\nu})$. There is also a decomposition for two-way classifications AxB (occupation x age group, for instance); it takes the form

$$\delta_\beta(\underset{\sim}{\pi};\ \underset{\sim}{\nu}) = \delta_\beta(\underset{\sim}{\pi}(A);\ \underset{\sim}{\nu}(A)) + \Sigma\pi_i(A)[\pi_i(A)/\nu_i(A)]^\beta\ \delta_\beta(\underset{\sim}{\pi}|A_i;\ \underset{\sim}{\nu}|A_i) \tag{1.32}$$

where: $\underset{\sim}{\pi}$ and $\underset{\sim}{\nu}$ are the arrays of income shares and population shares, respectively; $\underset{\sim}{\pi}(A)$ and $\underset{\sim}{\nu}(A)$ are the vectors of row marginals; $\underset{\sim}{\pi}|A_i$ and $\underset{\sim}{\nu}|A_i$ are the normalized ith row vectors. The last term of (1.32) is an average with respect to the income shares but, when β <u>strictly</u> greater than -1, it may be written more naturally as an average with respect to the population shares:

$$\Sigma\nu_i(A)[\pi_i(A)/\nu_i(A)]^{\beta+1}\ \delta_\beta(\ldots).$$

The factor $[\pi_i(A)/\nu_i(A)]^{\beta+1}$ inflates the contribution of inequality within high per capita income groups while deflating that of low per capita income groups. Such an effect is desirable in a measure of industrial concentration (Section 1.11.3) but may be undesirable in a measure of income inequality (Chiswick [76]). If desired, the contribution of low per capita income groups can be inflated by interchanging $\underset{\sim}{\pi}$ and $\underset{\sim}{\nu}$ and adopting $\delta^*_\beta(\underset{\sim}{\pi};\ \underset{\sim}{\nu}) = \delta_\beta(\underset{\sim}{\nu};\ \underset{\sim}{\pi})$, $\beta \geq 0$, as the inequality measure. This measure preserves the Lorenz ordering on the pairs $(\underset{\sim}{\pi},\ \underset{\sim}{\nu})$ but becomes infinite when any of the subpopulations have zero income.

1.12.4 The Incomes Equivalent Approach. We have motivated the inequality measure δ_β by transforming a numbers equivalent. Atkinson [77] has given a similar justification for δ_β^* by transforming an incomes equivalent. He considers a utility function $U(\cdot)$ so that per capita social welfare is $\Sigma\nu_i U(\lambda\pi_i/\nu_i)$, where λ is the per capita income for the entire population. The incomes equivalent λ_U is defined as that level of per capita income which would achieve the same per capita social welfare if income were evenly distributed, i.e.,

$$U(\lambda_U) = \Sigma\nu_i U(\lambda\pi_i/\nu_i).$$

With the usual assumption that the utility function is concave and increasing, it is not hard to show that the standardized incomes equivalent λ_U/λ is an inverse measure of inequality provided λ is held constant; i.e., λ_U/λ is unchanged by operations I1 and I2 and decreases under operation I3. Requiring λ_U/λ to be independent of λ further restricts the utility function to be a positive affine transformation of

$$U_\beta(t) = (1 - t^{-\beta})/\beta, \ \beta \geq -1.$$

The inequality measure δ_β^* is obtained by taking $U = U_\beta$ and applying the transformation (1.29) to the ratio λ_U/λ. Atkinson [77] suggests using $1 - (\lambda_U/\lambda)$ as the measure of inequality, pointing out that it is the proportion of national income which could be "saved" if income were evenly distributed.

1.12.5 Completing the Circle. Our starting point for the inequality measure δ_β was the diversity index Δ_β. Conversely, Δ_β may be obtained from δ_β by regarding the latter as a measure of distance and by further exploiting the analogy between diversity and variance. Recall that the variance of a random variable X is the minimum value of the mean square error $E[(X-a)^2]$, and that the minimum is achieved when $a = \mu = E[X]$. This is usually proved by noting the decomposition

$$E[(X-a)^2] = (\mu - a)^2 + \sigma_X^2.$$

Consider an abundance vector $\underset{\sim}{\pi}$ and write it in the form $\underset{\sim}{\pi} = \Sigma\pi_k\underset{\sim}{\varepsilon}^{(k)}$ where $\underset{\sim}{\varepsilon}^{(k)}$ is the kth coordinate vector. The mean deflated distance between an arbitrary abundance vector $\underset{\sim}{\nu}$ and a random observation on $\underset{\sim}{\pi}$ decomposes as

$$\Sigma\pi_k\cdot\pi_k^{\beta}\cdot\delta_\beta(\underset{\sim}{\varepsilon}^{(k)};\ \underset{\sim}{\nu}) = \delta_\beta(\underset{\sim}{\pi};\ \underset{\sim}{\nu}) + \Delta_\beta(\underset{\sim}{\pi}).$$

The minimum value of this expression is $\Delta_\beta(\underset{\sim}{\pi})$ and the minimum is achieved when $\underset{\sim}{\nu} = \underset{\sim}{\pi} = \Sigma\pi_k\underset{\sim}{\varepsilon}^{(k)}$.

Using the mixture notation of Section 1.8.3, one might similarly attempt to minimize, with respect to $\underset{\sim}{\nu}$, the expression

$$\Sigma w_k\cdot w_k^{\beta}\cdot\delta_\beta(\underset{\sim}{\pi}^{(k)};\ \underset{\sim}{\nu}). \tag{1.33}$$

Here the situation becomes somewhat unpleasant; in particular, (1.33) is minimized by $\underset{\sim}{\nu} = \underset{\sim}{\pi} = \Sigma w_k\underset{\sim}{\pi}^{(k)}$ in only two cases: (i) when $\beta = 0$ (the Shannon index) or (ii) when the abundance vectors $\underset{\sim}{\pi}^{(1)}$, $\underset{\sim}{\pi}^{(2)},\ldots,\underset{\sim}{\pi}^{(n)}$ are mutually orthogonal. This anomoly appears to be related to the failure of the mixture decomposition (Theorem 1.20) to coincide with the hierarchial decomposition (1.19). Sibson [40] or Jardine and Sibson [78, chap. 2] may be consulted for more details concerning the problem of minimizing (1.33).

The principal results and conclusions of this section on income inequality may be summarized as follows:

1. Inequality is the opposite of evenness not of diversity. Inequality is best regarded as a relative concept: the inequality of the income shares relative to the population shares.

2. An intrinsic inequality ordering similar to the intrinsic diversity ordering, has been defined and shown to be equivalent to the Lorenz ordering.

3. In general, the measures of equitability proposed by Hill [12] do not preserve the intrinsic inequality ordering (in the opposite sense).

4. A technique has been given for associating inequality measures with diversity indices. The association pairs the Gini coefficient with the Average Rank index, the Theil index with the Shannon index, and the coefficient of variation with the Simpson index. More generally, the directed divergence of type $\beta + 1$, denoted δ_β, is paired with the diversity index Δ_β.

5. A decomposition, applicable to two-way classifications, has been given for δ_β. The decomposition reveals that δ_β inflates the contribution to inequality due to high per capita income groups and deflates that due to low per capita income groups. A modified form of δ_β inflates the contribution of low per capita income groups. The modified index is a transformation of the incomes equivalent corresponding to a certain family of social welfare functions.

6. There has been established a diversity analogue of the familiar decomposition

$$E[(X - a)^2] = (E[X] - a)^2 + \mathrm{Var}[X]$$

for random variables. The directed divergence δ_β corresponds to the first term $(X - a)^2$ while the diversity Δ_β corresponds to the variance.

CHAPTER 2

DIVERSITY OF RANDOM COMMUNITIES

2.1 Introduction

Throughout Chapter 1, the community abundance vector was assumed to be fixed. Often it will be realistic to regard the abundance vector as a realization of some stochastic generating mechanism. In this case, our interest will focus on the "diversity producing capacity" of the mechanism and not on the diversity of particular realizations.

As in Chapter 1, two separate problems can be distinguished. First of all, we may wish only to rank random communities: to be able to say that one random community is more diverse than another. This is the ordinal problem and is dealt with in Section 2.4. The problem of quantification, on the other hand, calls for a numerical measure of the diversity of random communities. Two possibilities immediately suggest themselves: $E[\Delta(\underset{\sim}{\pi})]$ and $\Delta(E[\underset{\sim}{\pi}^*])$. These will be called the $E\Delta$ and ΔE measures corresponding to the diversity index Δ. However, if the abundance vector is a member of some parametric family, it may be more natural and, from the inferential standpoint, more appropriate to adopt some function of the parameters as the diversity measure. Usually this function will pertain specifically to the underlying model and will not fall into the framework considered in Chapter 1. Of course, some justification must be given for the function that is chosen. A central theme of Section 2.2 and Section 2.4 is the justification of Fisher's "alpha" as a diversity measure associated with the (symmetric) Dirichlet model.

In addition to its attractive theoretical properties, the Dirichlet model often provides an acceptable fit to empirical data.[1] One usually finds the index k to be small and the number of species s to be large. This leads one to idealize the model by letting $k \to 0$ and $s \to \infty$ with the product $\alpha = sk$ held fixed. While the negative binomial sampling distribution converges under this limiting scheme, the symmetric Dirichlet does not. This would seem to lead to the anomoly of a sampling distribution without a valid underlying population model. In Section 2.2, we make this limit rigorous at the population level by introducing a new concept called size-biased permutation. It is shown that the size-biased permutation of the Dirichlet model does have a valid limit. Moreover, this limiting Dirichlet is itself invariant under size-biased permutation and can be simply described in terms of iid random variables by means of what we call a residual allocation model. Further, the limiting Dirichlet is the only residual allocation model which is invariant under size-biased permutation.

The limiting Dirichlet has but one parameter--Fisher's α. We show that the result of randomly deleting a fraction $1 - p$ of the species is still a limiting Dirichlet but with parameter $p\alpha$. This provides some justification for interpreting α as a species richness parameter.

Treating a community as random introduces an additional component of variability into the sampling distribution. In an effort to eliminate this component, there has been some interest in identifying fixed models that closely resemble realizations of the limiting Dirichlect. Two

---

[1] The Dirichlet model predicts a negative binomial sampling distribution under the additional assumption that the total (unnormalized) abundance has an independent Gamma distribution. The log series sampling distribution results after taking Fisher's limit. See Section 2.3.

criteria for such fixations are considered in Section 2.3 and both the geometric series model[2] and the exponential integral model[2] are assessed against these criteria. First, it is shown that, every realization of the limiting Dirichlet is asymptotically geometric in a certain weak sense. Both models meet this requirement--but in a much stronger sense. The second criterion requires that the sampling distribution[3] match the log series distribution, at least approximately. The exponential integral model meets this requirement but the geometric series model does not. It is further shown that the sampling distribution corresponding to the geometric series model is a smoothly truncated rectangular hyperbola and has a tail much shorter than the log series. Empirical data usually requires a tail at least as long as the log series and, for this reason, the geometric series model will seldom provide an adequate fit. We also show that the exact sampling distributions of these models exhibit a periodicity that makes likelihood estimation impractical.

Section 2.4 extends the intrinsic diversity ordering to random communities. Four possible definitions are considered and these are shown to be equivalent for fixed, but not for random, communities. One of these formulations permits us to show that the obvious estimator[4] of a diversity index Δ is always negatively biased when Δ satisfies Criterion C3.

As was the case for fixed communities, the stochastic diversity ordering is only partial and two communities need not be comparable.

---

[2] This model is described in Section 2.3.

[3] The term "distribution" refers to the expected frequencies. This rather peculiar terminology has become standard in the present context.

[4] Replace population proportions with sample proportions.

In fact, comparability is the exception not the rule. However, it is shown for the Dirichlet model that the pair of parameters (s, k) is approximately[5] a complete diversity measure. Further, any two limiting Dirichlet models are comparable and here α is a complete diversity parameter.

The conventions for this chapter are as given in Section 1.1.1 except that infinite dimensional vectors are sometimes considered.

2.2 Random Communities

By a <u>random community</u> is meant any community whose abundance vector $\underset{\sim}{\pi}$ is random. In general, the number of species may vary from realization to realization. This section gives a number of examples of random communities which are referred to in the sequel. Also, it is shown how Fisher's limit can be made rigorous and how Fisher's α can be interpreted as a species richness measure.

<u>Example 2.1</u>: (Fixed community.) Any fixed community is also a random community whose abundance vector has a degenerate distribution.

<u>Example 2.2</u>: Let $\underset{\sim}{\pi}$ take the values (1, 0, 0), (.5, .5, 0), and (.4, .3, .3), each with probability 1/3. In this example, the number of species varies with the realization.

<u>Example 2.3</u>: Let $\underset{\sim}{\pi}$ assume the two values (.5, .3, .2) and (.6, .31, .09), each with probability 1/2. As examples of EΔ and ΔE measures, take Δ to be Simpson's index. One finds that $E[\Delta_1(\underset{\sim}{\pi})] = .5779$ and $\Delta_1(E[\underset{\sim}{\pi}^*]) = .58345$.

---

[5] See Theorem 2.18 for a precise statement.

Example 2.4: Let $\underset{\sim}{\nu}$ assume the two values (.5, .4, .1) and (.2, .2, .6), each with probability 1/2. Then $E[\underset{\sim}{\nu}] = (.35, .3, .35)$ and $E[\underset{\sim}{\nu}^*] = (.55, .3, .15)$. Notice that $E[\underset{\sim}{\nu}]$ is intrinsically more diverse than any realization of $\underset{\sim}{\nu}$ and is not at all representative of $\underset{\sim}{\nu}$. On the other hand, $E[\underset{\sim}{\nu}^*]$ gives a better overall description of the community structure without regard to species identity. Here, $E[\Delta_1(\underset{\sim}{\nu})] = .570$ and $\Delta_1(E[\underset{\sim}{\nu}^*]) = .585$. Note that the $E\Delta_1$ and $\Delta_1 E$ measures inconsistently rank $\underset{\sim}{\nu}$ in comparison with the community $\underset{\sim}{\pi}$ of the previous example. In each case, though, the $E\Delta_1$ measure is less than the $\Delta_1 E$ measure. It is a simple consequence of Jensen's inequality that this will always be so when the index satisfies Criterion C3.

Example 2.5: (Random sample.) Let $\underset{\sim}{Y} = (Y_1, Y_2, Y_3, \ldots)$ be a random sample from a fixed community $\underset{\sim}{\pi}$ where Y_i is the number of times the ith ranked species is represented in the sample. Then $\underset{\sim}{Y}/N$ is a random community whose abundance vector has a rescaled multinomial distribution. Note that $E[\underset{\sim}{Y}/N) = \underset{\sim}{\pi}^*$.

Example 2.6: (Random permutation.) Let $\underset{\sim}{\nu}$ be obtained from the random community $\underset{\sim}{\pi}$ by subjecting the components of $\underset{\sim}{\pi}$ to a permutation. In general, the permutation may be random and its distribution may depend on the realized value of $\underset{\sim}{\pi}$. Species identity may be lost after a random permutation since, for example, ν_1 may reference different species in different realizations. But species identity is unimportant for questions of diversity so that $\underset{\sim}{\pi}$ and $\underset{\sim}{\nu}$ may be regarded as equivalent.

Two types of random permutations are quite useful. The ranked permutation $\underset{\sim}{\nu} = \underset{\sim}{\pi}^*$ arranged the components of $\underset{\sim}{\pi}$ in descending order and is canonical in the sense that $\underset{\sim}{\pi}^*$ and $\underset{\sim}{\pi}'^*$ are equal in distribution whenever $\underset{\sim}{\pi}'$ is a random permutation of $\underset{\sim}{\pi}$.

The size-biased permutation $\underset{\sim}{\nu} = \underset{\sim}{\pi}^{\#}$ is obtained as follows: Randomly select an individual from the community and put $\pi_1^{\#} = \pi_{i_1}$ where i_1 is the label of the species to which the selected individual belongs. Now remove species i_1 from the community and randomly select a second individual. Put $\pi_2^{\#} = \pi_{i_2}$ where i_2 is the label of the species to which the second individual belongs. Remove species i_1 and i_2 and randomly select a third individual, etc. Alternatively, the partial sums π_1, $\pi_1 + \pi_2$, $\pi_1 + \pi_2 + \pi_3, \ldots$ partition the unit interval into the subintervals $I_i = (\pi_1+\ldots+\pi_{i-1}, \pi_1+\ldots+\pi_i]$, $i = 1,2,3,\ldots$. Now let $\underset{\sim}{U} = (U_1, U_2, U_3, \ldots)$ be a random sample from the Uniform (0,1) distribution where $\underset{\sim}{U}$ is independent of $\underset{\sim}{\pi}$. Let $I_1^{\#}$ be the subinterval containing U_1 and, by induction, let $I_{m+1}^{\#}$ be the subinterval which contains the first component of $\underset{\sim}{U}$ that is not contained in the union of $I_1^{\#}, I_2^{\#}, \ldots, I_m^{\#}$. Put $\pi_i^{\#}$ equal to the length of $I_i^{\#}$.

It should be clear that the size-biased permutation is cononical in the sense defined above. In particular, $\underset{\sim}{\pi}^{\#}$, $\underset{\sim}{\pi}^{\#\#}$ and $\underset{\sim}{\pi}^{*\#}$ are equal in distribution.

The size-biased permutation arises naturally in the problem of "heaps" which has been considered by Kingman [79]. A number of items labeled $i = 1,2,\ldots,N$ are stored in a heap and are demanded from time to time. After being used, the item is replaced on the top of the heap. The successive arrangements of items in the heap form an irreducible aperiodic Markov chain with N! states. The equilibrium distribution is easily seen to be the distribution of $\underset{\sim}{\pi}^{\#}$ where π_i is the proportional demand for item i.

As a numerical example of size-biased permutation, take $\underset{\sim}{\pi}$ to be the fixed community (.6, .3, .1). The probability distribution of the random community $\underset{\sim}{\pi}^{\#}$ is given in Table 2.1. Notice that $\underset{\sim}{\pi}^{\#}$ arranges the components of $\underset{\sim}{\pi}$ so that large abundances are likely, but not certain, to appear early in the list. The next theorem makes this precise.

Table 2.1: Probability Distribution of $\underset{\sim}{\pi}^{\#}$

$\underset{\sim}{x}$	$P(\underset{\sim}{\pi}^{\#} = \underset{\sim}{x})$
(.6, .3, .1)	.450
(.6, .1, .3)	.150
(.3, .6, .1)	.257
(.3, .1, .6)	.043
(.1, .6, .3)	.067
(.1, .3, .6)	.033

<u>Theorem 2.1</u>: Let $\underset{\sim}{\pi}^{\#}$ be the size-biased permutation of the random community $\underset{\sim}{\pi}$. Then $\pi_i^{\#}$ is stochastically greater than $\pi_{i+1}^{\#}$, $i = 1,2,3,\ldots$. Consequently, $E[\pi_1^{\#}] \geq E[\pi_2^{\#}] \geq \ldots$ and the components of the fixed community $E[\underset{\sim}{\pi}^{\#}]$ are already arranged in descending order.

<u>Proof</u>: By conditioning on the value of $\underset{\sim}{\pi}$, it will suffice to prove the theorem when $\underset{\sim}{\pi}$ is a fixed community. Also, it is only necessary to prove that $\pi_1^{\#}$ is stochastically greater than $\pi_2^{\#}$ since the general case than follows by conditioning on $\pi_1^{\#}, \pi_2^{\#}, \ldots, \pi_{i-1}^{\#}$. Let $x_1 \geq x_2 \geq x_3 \geq \ldots$ be the components of $\underset{\sim}{\pi}$ listed in descending order and, for the moment, assume that there are no ties among the nonzero x_i. Now, $P(\pi_1^{\#} \geq x_m) = x_1 + x_2 + \ldots + x_m \equiv y$, while

$$P(\pi_2^{\#} \geq x_m) = \sum_{i \leq m} x_i(y - x_i)/(1 - x_i) + \sum_{i>m} x_i y/(1 - x_i). \qquad (2.1)$$

Let $0 < y \leq 1$ be fixed. The function $x(y - x)/(1 - x)$ is concave on the unit interval. Applying the principle of forward transfers, the first term of (2.1) is maximized when $x_1 = x_2 = \ldots = x_m = y/m$ and has maximum possible value $y(y - y/m)/(1 - y/m)$. On the other hand, the function $x/(1 - x)$ is convex on the unit interval so that backward transfers increase the second sum in (2.1). Holding x_m fixed and bearing in mind the constraint $x_m \geq x_{m+1} \geq x_{m+2} \geq \ldots$, the second sum is maximized when $x_m = x_{m+1} = x_{m+2} = \ldots = x_{m+n}$ and $x_{m+n+1} = \delta$ where $1 - y = nx_m + \delta$ with $0 \leq \delta < x_m$. Therefore,

$$\begin{aligned}\sum_{i>m} x_i y/(1 - x_i) &\leq y[nx_m/(1 - x_m) + \delta/(1 - \delta)] \\ &\leq y[nx_m/(1 - x_m) + \delta/(1 - x_m)] \\ &\leq y(1 - y)/(1 - x_m) \\ &\leq y(1 - y)/(1 - y/m).\end{aligned}$$

Putting these two upper bounds together gives $P(\pi_2^{\#} \geq x_m) \leq y = P(\pi_1^{\#} \geq x_m)$ which proves that $\pi_1^{\#}$ is stochastically greater than $\pi_2^{\#}$. In the event of ties among the nonzero x_i, first break these ties by a slight deformation, then apply the above result and take the limit as the deformation goes to zero. For example, if $x_1 = x_2 > x_3 \geq \ldots$, we have

$$\begin{aligned}P(\pi_2^{\#} \geq x_1 | \underset{\sim}{\pi}^* = \underset{\sim}{x}) &= \lim_{\substack{h \to 0 \\ h>0}} P(\pi_2^{\#} \geq x_2 - h | \underset{\sim}{\pi}^* = (x_1 + h, x_2 - h, x_3, \ldots)) \\ &\geq \lim_{\substack{h \to 0 \\ h>0}} P(\pi_1^{\#} \geq x_1 - h | \underset{\sim}{\pi}^* = (x_1 + h, x_2 - h, x_3, \ldots)) \\ &= P(\pi_1^* \geq x_1 | \underset{\sim}{\pi}^* = \underset{\sim}{x}).\end{aligned}$$

<u>Example 2.7</u>: (Symmetric Dirichlet community.) Let $\underset{\sim}{\pi}$ have a symmetric Dirichlet distribution with s components and with index $k > 0$. Write $\underset{\sim}{\pi} \sim D(s, k)$. With probability one, every realization has exactly s species. Recall that $\underset{\sim}{\pi}$ can be obtained as $(\lambda_1, \lambda_2, \ldots, \lambda_s)/\Sigma\lambda_i$ where $\lambda_1, \lambda_2, \ldots, \lambda_s$ are iid Gamma random variables with common index k and with common, but arbitrary, scale parameter. A useful fact is that the fractions are jointly independent of the total, i.e., $\underset{\sim}{\pi}$ is independent of $\lambda_1 + \lambda_2 + \ldots + \lambda_s$. The next theorem gives the $E\Delta$ measures of the Dirichlet community for the indices $\Delta_\beta = \Sigma\pi_i(1 - \pi_i^\beta)/\beta$, $\Delta_\omega^{H-S} = \Sigma(1 - \pi_i)[1 - (1 - \pi_i)^\omega]$, and $\Delta_1^{(\text{rank})} = \Sigma(i - 1)\pi_i^*$. The function ψ is the digamma function:

$$\psi(z + 1) = d \log \Gamma(z + 1)/dz = -\gamma + \sum_{n=1}^{\infty} z/[n(n + z)].$$

<u>Theorem 2.1</u>: Let $\underset{\sim}{\pi} = (\pi_1, \pi_2, \ldots, \pi_s) \sim D(s, k)$. The components $\pi_1, \pi_2, \ldots, \pi_s$ are identically distributed and the common distribution is a Beta of the first kind with parameters k and $(s - 1)k$. Furthermore,

$$\text{a)} \quad E[\Delta_\beta(\underset{\sim}{\pi})] = \begin{cases} \beta^{-1}[1 - \Gamma(sk + 1)\Gamma(k + 1 + \beta)/(\Gamma(sk + 1 + \beta)\Gamma(k + 1))] & \text{if } \beta \geq -1,\ \beta \neq 0 \\ \psi(sk + 1) - \psi(k + 1) & \text{if } \beta = 0, \end{cases}$$

$$\text{b)} \quad E[\Delta_\omega^{H-S}(\underset{\sim}{\pi})] = \frac{\Gamma(sk + 1 + \omega)\Gamma(sk + 1 - k) - \Gamma(sk + 1)\Gamma(sk + 1 + \omega - k)}{k\Gamma(sk + 1 + \omega)\Gamma(sk - k)}$$

$$\text{if } \omega \geq 0,$$

$$\text{c)} \quad E[\Delta_1^{(\text{rank})}(\underset{\sim}{\pi})] = \frac{s - 1}{2}[1 - \frac{\Gamma(2k + 1)}{4k\Gamma(k + 1)^2}].$$

<u>Proof</u>: The first assertion is clear and parts a) and b) are easy calculations since the nonzero components of $\underset{\sim}{\pi}$ are identically distributed.

The expression for $E[\Delta_1^{(rank)}(\underset{\sim}{\pi})]$ has been obtained by Kingman [79, equation (14)] in a different context.

Remark 2.1: In general, there are no simple closed form expressions for $E[\Delta_\rho^{(rank)}(\underset{\sim}{\pi})]$. If there were, using (1.14), one could obtain closed form expressions for the expected order statistics from a Gamma distribution.

Sometimes one wishes to replace a random community with a "typical" fixed community. As has been pointed out, $E[\underset{\sim}{\pi}]$ is not an appropriate interpretation of "typical." In fact, $E[\underset{\sim}{\pi}]$ is a completely even community when the nonzero components of $\underset{\sim}{\pi}$ are identically distributed as in the Dirichlet model. Better choices are $E[\underset{\sim}{\pi}^{\#}]$ or $E[\underset{\sim}{\pi}^{*}]$. (See Section 1.3 and Section 1.4, however.) The size-biased permutation of the Dirichlet community is quite manageable and a number of its properties are established below. On the other hand, $\underset{\sim}{\pi}^{*}$ is usually intractable; the Dirichlet community with $k = 1$ is an exception and is considered next.

Example 2.8: (MacArthur's model.) MacArthur [80] supposed that s species were competing among themselves for a fixed resource. The total available resource was represented by the unit interval and was apportioned among the species by selecting $s - 1$ points at random from the unit interval and arranging these points in increasing order: $0 \equiv V_0 \leq V_1 \leq V_2 \leq \ldots \leq V_{s-1} \leq V_s \equiv 1$. Species i then received an amount of resource equal to $\pi_i = V_i - V_{i-1}$. Defined in this way, the community $\underset{\sim}{\pi} = (\pi_1, \pi_2, \pi_3, \ldots, \pi_s)$ is called MacArthur's random model with s species while $E[\underset{\sim}{\pi}^{*}]$ is MacArthur's fixed model. The next theorem was proved by Cohen [81].

Theorem 2.3: MacArthur's random model with s species is equal in distribution to the symmetric Dirichlet model with s species and with index k = 1.

Proof: Let $\underset{\sim}{\nu} \sim D(s, k)$ and obtain $\underset{\sim}{\nu}$ as $(\lambda_1, \lambda_2,\ldots,\lambda_s)/\lambda$ where the λ_i are iid exponential random variables and where $\lambda = \lambda_1 + \lambda_2 + \ldots + \lambda_s$. Conditional on λ, the variables $W_1 = \lambda_1$, $W_2 = \lambda_1 + \lambda_2,\ldots,W_{s-1} = \lambda_1 + \lambda_2 + \ldots + \lambda_{s-1}$ are jointly distributed as the order statistics of a random sample of size s - 1 from a Uniform $(0,\lambda)$ distribution. It follows that, conditional on λ, the variables $V_i = W_i/\lambda$ are jointly distributed as the order statistics from a Uniform (0,1) distribution. Moreover the qualifying phrase "conditional on λ" can be dropped from the previous sentence since $(V_1, V_2,\ldots,V_{s-1})$ is independent of λ. With $V_0 = 0$ and $V_s = 1$, ν_i is obtained as $V_i - V_{i-1}$, just as in MacArthur's model.

Corollary 2.1: For MacArthur's random model with s species, $E[\Delta_\beta(\underset{\sim}{\pi})]$, $E[\Delta_\omega^{H-S}(\underset{\sim}{\pi})]$, and $E[\Delta_1^{(rank)}(\underset{\sim}{\pi})]$ are obtained by putting k = 1 in Theorem 2.1.

Corollary 2.2: Let $\underset{\sim}{\nu} = E[\underset{\sim}{\pi}^*]$ be MacArthur's fixed model with s species and define $H_0 = 0$ and $H_m = 1 + 1/2 + 1/3 + \ldots + 1/m$. Then $\nu_i = (H_s - H_{i-1})/s$, $i = 1,2,\ldots,s$.

Proof: Let $\underset{\sim}{\pi}$ be obtained as $(\lambda_1, \lambda_2,\ldots,\lambda_s)/\lambda$ as in the proof of Theorem 2.3. Without loss of generality suppose $E[\lambda_i] = 1$ and $E[\lambda] = s$. Then $\pi_i^* = \lambda_i^*/\lambda$ where $\lambda_1^* \geq \lambda_2^* \geq \lambda_3^* \geq \ldots \geq \lambda_s^*$ are the descending order statistics from the standard exponential distribution. Since the fractions are independent of the total, it follows that $E[\pi_i^*] = E[\lambda_i^*/\lambda] = E[\lambda_i^*]/E[\lambda] = E[\lambda_i^*]/s$. But $E[\lambda_i^*] = H_s - H_{i-1}$.

Feller [82, p. 20] gives an intuitive as well as a rigorous proof of the last assertion.

MacArthur's fixed model is discussed further in Appendix A.2 where there is given a mnemonic device for recalling the mathematical form of $E[\underset{\sim}{\pi}^*]$.

Example 2.9: (Residual allocation model.) Because of the constraint $\Sigma\pi_i = 1$, the components of $\underset{\sim}{\pi}$ are necessarily dependent. Calculations are simplified if the distribution of $\underset{\sim}{\pi}$ can be described in terms of independent random variables. The Dirichlet distribution, for example, can be described in terms of independent Gammas as in Example 2.7. Here we consider another method that sometimes works.

Given the random community $\underset{\sim}{\pi} = (\pi_1, \pi_2, \ldots, \pi_s)$, define the residual fractions $Q_1, Q_2, \ldots, Q_s$ as

$$\begin{aligned} Q_1 &= \pi_1 \\ Q_2 &= \pi_2/(1 - \pi_1) \\ Q_3 &= \pi_3/(1 - \pi_1 - \pi_2) \\ &\vdots \\ Q_s &= \pi_s/(1 - \pi_1 - \pi_2 - \ldots - \pi_{s-1}) \equiv 1. \end{aligned} \tag{2.2}$$

The residual fractions specify a sequential resource allocation scheme: Let the total resource be represented by the unit interval as in MacArthur's model. Species 1 receives a fraction Q_1 of the total, then species 2 receives a fraction Q_2 of the residual $1 - \pi_1$, then species 3 receives a fraction Q_3 of the new residual $1 - \pi_1 - \pi_2$, etc.

Theorem 2.4: Let $0 < \pi_i < 1$, $i = 1,2,\ldots,s - 1$, subject to $\pi_1 + \pi_2 + \ldots + \pi_{s-1} < 1$ and $0 < Q_i < 1$, $i = 1,2,\ldots,s - 1$. Then the transformation $(\pi_1, \pi_2, \ldots, \pi_{s-1}) \rightarrow (Q_1, Q_2, \ldots, Q_{s-1})$ given by (2.2) in one-one, onto, and has Jacobian

$$\partial(\pi_1,\pi_2,\ldots,\pi_{s-1})/\partial(Q_1,Q_2,\ldots,Q_{s-1}) = (1-\pi_1)(1-\pi_1-\pi_2)\ldots(1-\pi_1-\pi_2-\ldots-\pi_{s-2}). \tag{2.3}$$

The residuals are given by

$$\begin{aligned}
1 - \pi_1 &= 1 - Q_1 \\
1 - \pi_1 - \pi_2 &= (1 - Q_1)(1 - Q_2) \\
1 - \pi_1 - \pi_2 - \pi_3 &= (1 - Q_1)(1 - Q_2)(1 - Q_3) \\
&\vdots \\
1 - \pi_1 - \pi_2 - \ldots - \pi_{s-1} &= (1 - Q_1)(1 - Q_2)\ldots(1 - Q_{s-1}),
\end{aligned} \tag{2.4}$$

and the components of $\underset{\sim}{\pi}$ by

$$\begin{aligned}
\pi_1 &= Q_1 \\
\pi_2 &= (1 - Q_1)Q_2 \\
\pi_3 &= (1 - Q_1)(1 - Q_2)Q_3 \\
&\vdots \\
\pi_s &= (1 - Q_1)(1 - Q_2)\ldots(1 - Q_{s-1})Q_s.
\end{aligned} \tag{2.5}$$

Proof: All parts of the theorem are a straightforward calculation using equation (2.2).

A random community $\underset{\sim}{\pi} = (\pi_1, \pi_2,\ldots,\pi_s)$ is called a residual allocation model when the residual fractions are independently distributed with $P(0 < Q_i < 1) = 1$, $i = 1,2,\ldots,s - 1$, and $P(Q_s = 1) = 1$.

Theorem 2.5: Let the random community be a residual allocation model with residual fractions $Q_1, Q_2,\ldots,Q_s$. The fixed community $E[\underset{\sim}{\pi}]$ is then a residual allocation model with degenerate residual fractions $E[Q_1], E[Q_2],\ldots,E[Q_s]$.

Proof: Take expectations of (2.5) and use the independence of the residual fractions.

Theorem 2.6: Let $\underset{\sim}{\pi} \sim D(s, k)$. Then $\underset{\sim}{\pi}$ is a residual allocation model with residual fractions given by

$$\begin{aligned} Q_1 &\sim \mathrm{Beta}_1(k, (s-1)k) \\ Q_2 &\sim \mathrm{Beta}_1(k, (s-2)k) \\ Q_3 &\sim \mathrm{Beta}_1(k, (s-3)k) \\ &\vdots \\ Q_s &\sim \mathrm{Beta}_1(k, (s-s)k) \equiv 1. \end{aligned}$$

Proof: This follows by a change of variables using Theorem 2.4.

The next theorem, which is a principal result of this section, asserts that the size-biased permutation of the Dirichlet model is also a residual allocation model.

Theorem 2.7: Let $\underset{\sim}{\pi}^{\#}$ be the size-biased permutation of $\underset{\sim}{\pi}$ where $\underset{\sim}{\pi} \sim D(s, k)$. Then $\underset{\sim}{\pi}^{\#}$ is a residual allocation model with

$$\begin{aligned} Q_1 &\sim \mathrm{Beta}_1(k+1, (s-1)k) \\ Q_2 &\sim \mathrm{Beta}_1(k+1, (s-2)k) \\ Q_3 &\sim \mathrm{Beta}_1(k+1, (s-3)k) \\ &\vdots \\ Q_s &\sim \mathrm{Beta}_1(k+1, (s-s)k) \equiv 1. \end{aligned}$$

Proof: By Theorem 2.4, the sample space of $Q_1, Q_2, \ldots, Q_{s-1}$ is a Cartesian product. For notational simplicity take $s = 4$ and define the function $h_x(t)$ to be one when $t \leq x$ and zero otherwise. Suppose that $0 \leq x_1, x_2, x_3 \leq 1$. Then the joint probability distribution function of the residual fractions Q_1, Q_2, and Q_3 is given by

$$P(Q_1 \leq x_1,\ Q_2 \leq x_2,\ Q_3 \leq x_3) = E[\Sigma\pi_i \cdot \pi_j/(1 - \pi_i) \times$$
$$\pi_m/(1 - \pi_i - \pi_j)\cdot h_{x_1}(\pi_i)h_{x_2}(\pi_j/(1 - \pi_i))h_{x_3}(\pi_m/(1 - \pi_i - \pi_j))], \quad (2.6)$$

where the sum ranges over all triplets (i, j, m) of distinct integers with $1 \leq i, j, m \leq 4$. Now each term in the sum is identically distributed so that the right hand side of (2.6) is proportional to

$$E[\pi_1\cdot\pi_2/(1 - \pi_1)\cdot\pi_3/(1 - \pi_1 - \pi_2)\cdot h_{x_1}(\pi_1)h_{x_2}(\pi_2/(1 - \pi_1))h_{x_3}(\pi_3/(1-\pi_1-\pi_2))],$$

which, in turn, is proportional to

$$\int_R (\pi_1\pi_2\pi_3)\pi_1^{k-1}\pi_2^{k-2}\pi_3^{k-1}(1 - \pi_1 - \pi_2 - \pi_3)^{k-1}/[(1 - \pi_1)(1 - \pi_1 - \pi_2)]d\pi_1 d\pi_2 d\pi_3,$$

where R is the region defined by

$$0 \leq \pi_1 \leq x_1,$$
$$0 \leq \pi_2/(1 - \pi_1) \leq x_2,$$
$$0 \leq \pi_3/(1 - \pi_1 - \pi_2) \leq x_3.$$

Make the change of variables $z_1 = \pi_1$, $z_2 = \pi_2/(1 - \pi_1)$, and $z_3 = \pi_3/(1 - \pi_1 - \pi_2)$ and apply Theorem 2.4 to find that $P(Q_1 \leq x_1, Q_2 \leq x_2, Q_3 \leq x_3)$ is proportional to a product of integrals,

$$\int_0^{x_1} z_1^k(1 - z_1)^{3k-1}dz_1 \int_0^{x_2} z_2^k(1 - z_2)^{2k-1} \int_0^{x_3} z_3^k(1 - z_3)^{k-1}dz_3,$$

which completes the proof.

From Theorem 2.7 and from (2.5), it follows that $\pi_i^{\#}$, $i = 1,2,\ldots,s$, is a product of independent Beta_1 random variables. However, the factors do not have the structure necessary for their product to possess a Beta_1 distribution. See Kotlarsky [83] for results concerning the product of independent Beta_1 random variables.

Open Problem: Characterize the class of residual allocation models whose size-biased permutation is also a residual allocation model.

Example 2.10: (Infinite residual allocation model.) Let Q_1, Q_2, $Q_3, \ldots$ be an infinite sequence of independent random variables with $P(0 < Q_i < 1) = 1$. Let $\pi_1, \pi_2, \pi_3, \ldots$ be defined as in (2.5). Then $\underset{\sim}{\pi} = (\pi_1, \pi_2, \pi_3, \ldots)$ is called an infinite residual allocation model provided

$$P(\lim_{n \to \infty} (1 - \pi_1 - \pi_2 - \ldots - \pi_n) = 0) = 1. \tag{2.7}$$

Theorem 2.8: Equation (2.7) holds when Q_1, Q_2, $Q_3, \ldots$ are independent and identically distributed.

Proof: Take the negative logarithm of both sides of (2.4) to find that $-\log(1 - \pi_1 - \pi_2 - \ldots - \pi_n) = \sum_{i=1}^{n} -\log(1 - Q_i)$. Since the $-\log(1 - Q_i)$ are iid positive random variables, this last sum diverges to $+\infty$ with probability one. Now (2.7) is immediate.

Remark 2.1: This is the first time we have considered communities with infinitely many species. A measure theoretic framework adequate for our purposes is sketched in Appendix A.3. Measurability questions will be ingnored in the body of the text.

Example 2.11: (Engen's model.) This is the infinite residual allocation model $\underset{\sim}{\pi}$ whose residual fractions are iid with common distribution $\text{Beta}_1(1, \alpha)$ where $\alpha > 0$. From (2.5), $E[\pi_1]$, $E[\pi_2]$, $E[\pi_3], \ldots$ forms a geometric progression with ratio $\theta = \alpha/(\alpha + 1)$ and odds ratio $\theta/(1 - \theta) = \alpha$. By matching moments of first moment distributions, Engen [84] has associated his model with Fisher's limiting Dirichlet. The next example makes this association a bit more precise.

Example 2.12: (Fisher's limiting Dirichlet.) For the Dirichlet model of Example 2.7, Fisher [85] introduced the product $\alpha = sk$ as a diversity measure. For fixed k, α is proportional to the number of species and may be regarded as a species richness parameter. As justification for the use of α, Kempton & Taylor [86] point out that, for most data sets, it is not possible to efficiently estimate s and k separately although the product $\alpha = sk$ can be efficiently estimated. Fisher found that s was usually very large and k very small which led him to idealize the model by taking the formal limit as $s \to \infty$ and $k \to 0$ with the product $\alpha = sk$ held fixed. The sampling distribution considered by Fisher did converge (to the log series distribution) permitting the development of inferential techniques. But the underlying population model does not converge. More precisely, by Theorem 2.2, each component of the Dirichlet model converges to zero. The existence of a sampling distribution with, apparently, no valid underlying population model has caused some difficulty of interpretation especially concerning the parameter α.

The reason the Dirichlet fails to converge becomes clear once we note that there is no community $\underset{\sim}{\pi}$ with infinitely many species and whose components π_i are identically distributed. As Kingman [79] points out, such a community would have to satisfy the imcompatible requirements

$$E[\pi_1] = E[\pi_2] = E[\pi_3] = \ldots,$$
$$\Sigma\, E[\pi_i] = 1.$$

This difficulty can be avoided by subjecting the Dirichlet community to an appropriate canonical random permutation before taking the limit. For example, results of Kingman [79] imply that the ranked permutation

$\underset{\sim}{\pi}^{*}$ does converge in distribution under Fisher's limiting scheme.[1] Unfortunately, the limiting distribution of $\underset{\sim}{\pi}^{*}$ is quite intractable (Kingman [79, Appendix I]). On the other hand, our next theorem shows the the size-biased permutation converges in distribution to Engen's model. Furthermore, the ranked permutation of Engen's model is equal in distribution to Kingman's limit so that Engen's model and Kingman's limit can be regarded as alternative descriptions of the same underlying community structure; this will be provided in Chapter 4 using the Gamma process.

<u>Theorem 2.9</u>: Let $\underset{\sim}{\pi} \sim D(s, k)$. Then $\underset{\sim}{\pi}^{\#}$ converges in distribution to Engen's model with parameter α when $s \to \infty$ and $k \to 0$ such that $sk = \alpha > 0$. Here convergence in distribution means weak convergence of the finite dimensional marginals.

<u>Proof</u>: This follows from Theorem 2.7 and (2.5) using the fact that convergence in distribution is preserved by a continuous transformation.

<u>Open Problem</u>: Let the sequence $\underset{\sim}{\pi}_1, \underset{\sim}{\pi}_2, \underset{\sim}{\pi}_3, \ldots$ of random communities converge in distribution to the random community $\underset{\sim}{\pi}$. Prove or disprove: $\underset{\sim}{\pi}_1^{*}, \underset{\sim}{\pi}_2^{\#}, \underset{\sim}{\pi}_3^{\#}, \ldots$ converges in distribution to $\underset{\sim}{\pi}^{\#}$. Using (2.6), this <u>can</u> be proved when the number of species is uniformly bounded. If it is true in general, it has the important implication that $\underset{\sim}{\pi}$ and $\underset{\sim}{\pi}^{\#}$ are equal in distribution when $\underset{\sim}{\pi}$ is Engen's model (use Theorem 2.9 and the equality in distribution of $\underset{\sim}{\pi}^{\#}$ and $\underset{\sim}{\pi}^{\#\#}$). Actually, it will be proved

---

[1] The possibility of a connection between Kingman's work and Fisher's limiting Dirichlet was raised by Cox [87] and, curiously, Kingman [79, p. 22] replied that he saw no connection.

later that Engen's model is invariant under size-biased permutation. Anticipating this result, we next give a characterization of Engen's model.

Theorem 2.10: Engen's model is the only infinite residual allocation model whose residual fractions are iid and which is invariant under size-biased permutation.

Proof: Let $\underset{\sim}{\nu}$ be a random community meeting the assumptions of the theorem. Put $A(n) = E[Q_1^n]$ and $B(n) = E[(1 - Q_1)^n]$, $n = 1,2,3,\ldots$. We use the fact that the distribution of the bounded random variable Q_1 is uniquely determined by its moment sequence $A(n)$ (Feller [82, p. 225]). Notice that $B(1) = 1 - A(1)$, $B(2) = 1 - 2A(1) + A(2)$ and in general, $B(n + 1) = L(n) + (-1)^{n+1}A(n + 1)$ where $L(n)$ is a linear combination of $1, A(1), A(2),\ldots,A(n)$. Now

$$\begin{aligned} A(n) &= E[Q_1^n] = E[\nu_1^n] = E[(\nu_1^{\#})^n] \\ &= E[\Sigma\, \nu_i^{n+1}] = \Sigma E[\nu_i^{n+1}] \\ &= E[Q_1^{n+1}] + E[(1 - Q_1)^{n+1}Q_2^{n+1}] + E[(1 - Q_1)^{n+1}(1 - Q_2)^{n+1}Q_3^{n+1}]+\ldots \\ &= A(n + 1)[1 + B(n + 1) + B^2(n + 1)+\ldots] \\ &= A(n + 1)/[1 - B(n + 1)] = A(n + 1)/[1 - L(n) + (-1)^n A(n + 1)]. \end{aligned}$$

Solving for $A(n + 1)$, one finds that $[1 - (-1)^n A(n)]A(n + 1)$ is a polynomial function of $A(1), A(2),\ldots,A(n)$. By assumption $P(0 < Q_1 < 1) = 1$ so that $0 < A(n) < 1$ and the coefficient $1 - (-1)^n A(n)$ is nonzero. Thus the moment sequence is uniquely determined by the first moment $A(1)$. But for Engen's model, the first moment is $A(1) = 1/(\alpha + 1)$ which ranges over the entire open unit interval as α varies from zero to infinity. The proof is now complete since Engen's model meets the assumptions of the theorem. Note that it was only necessary to use the equality in distribution of the first components ν_1 and $\nu_1^{\#}$. We conjecture that the assumption that the residual fractions are identically distributed can be dropped when the full force of invariance is used.

Example 2.13: (Random deletion of species.) Let $0 < p < 1$ and suppose $\underset{\sim}{\nu}$ is a random community with $P(0 < \nu_i < 1) = 1$ for $i = 1,2,3,\ldots$. Further suppose $\varepsilon_1, \varepsilon_2, \varepsilon_3,\ldots$ are independent Bernoulli random variables with $P(\varepsilon_i = 0) = 1 - p$ and $P(\varepsilon_i = 1) = p$. Put $\pi_i = \nu_i\varepsilon_i/\Sigma\nu_j\varepsilon_j$. Then $\underset{\sim}{\pi} = (\pi_1, \pi_2, \pi_3,\ldots)$ is a random community obtained by screening $\underset{\sim}{\nu}$. Call $\varepsilon_1,\varepsilon_2,\varepsilon_3,\ldots$ the selection variables. Defined in this way, $\underset{\sim}{\pi}$ has infinitely many components equal to zero. If desired, the zero components can be omitted by the following device: Let $X(1)$, $X(2)$, $X(3),\ldots$ be independent geometric random variables with $P(X(i) = x) = p(1 - p)^{x-1}$, $x = 1,2,3,\ldots$. Put $\tau_i = \nu_{X(1)+X(2)+\ldots+X(i)}$ and $\pi_i = \tau_i/\tau$ where $\tau = \tau_1 + \tau_2 + \tau_3 + \ldots$. Part (a) of the next theorem is proved in Chapter 4; a direct verification appears to be difficult.

Theorem 2.11: Let $\underset{\sim}{\nu}$ be Engen's model with parameter α. Then

a) $\tau = \Sigma\tau_i$ has a $\text{Beta}_1(p\alpha, (1 - p)\alpha)$ distribution and is independent of $\underset{\sim}{\pi} = (\tau_1, \tau_2, \tau_3,\ldots)/\tau$,

b) $\underset{\sim}{\pi}$ is equal in distribution to Engen's model with parameter $p\alpha$.

Proof: b) For reference, note that $E[Q^t(1 - Q)^T] = \alpha\,\Gamma(t + 1)\Gamma(\alpha + T)/\Gamma(\alpha + t + T + 1)$ when $Q \sim \text{Beta}_1(1, \alpha)$. The joint distribution of a uniformly bounded collection of positive random variables is known to be uniquely determined by the joint Mellin transform of the finite dimensional marginals. Now, for Engen's model with parameter α, the joint Mellin transform of $\nu_1, \nu_2,\ldots,\nu_n$ is

$$E[\prod_{j=1}^{n} \nu_j^{t(j)}] = E[\prod_{j=1}^{n} Q_j^{t(j)}(1 - Q_j)^{t(j+1)+t(j+2)+\dots+t(n)}], \quad (t(j) \geq 0).$$

Write $T(i) = t(i) + t(i+1)+\dots+t(n)$, $i = 1,2,\dots,n$ and $T(n + 1) = 0$. Using the independence of $Q_1, Q_2, Q_3,\dots$, the joint Mellin transform simplifies to

$$E[\prod_{j=1}^{n} \nu_j^{t(j)}] = \prod_{j=1}^{n} E[Q_j^{t(j)}(1 - Q_j)^{T(j+1)}] \tag{2.8}$$

$$= \prod_{j=1}^{n} [\alpha\Gamma(t(j) + 1)\Gamma(\alpha + T(j + 1))/\Gamma(\alpha + T(j) + 1)]$$

$$= \prod_{j=1}^{n} [\alpha\Gamma(t(j) + 1)/(\alpha + T(j))]\cdot \prod_{j=1}^{n} [\Gamma(\alpha + T(j+1))/\Gamma(\alpha+T(j))]$$

$$= \prod_{j=1}^{n} [\alpha\Gamma(t(j) + 1)/(\alpha + T(j))]\cdot\Gamma(\alpha)/\Gamma(\alpha + T(1)). \tag{2.9}$$

On the other hand, the joint Mellin transform of $\pi_1, \pi_2,\dots,\pi_n$ is

$$E[\prod_{j=1}^{n} \pi_j^{t(j)}] = E[(\prod_{j=1}^{n} \tau_j^{t(j)})/\tau^{T(1)}].$$

But $\underset{\sim}{\pi}$ is idependent of τ so that the expectation of the quotient is the quotient of the expectations, giving

$$E[\prod_{j=1}^{n} \pi_j^{t(j)}] = E[\prod_{j=1}^{n} \tau_j^{t(j)}]/E[\tau^{T(1)}].$$

By part a), $\tau \sim \mathrm{Beta}_1(p\alpha, (1 - p)\alpha)$ so that

$$E[\tau^{T(1)}] = [\Gamma(\alpha)/\Gamma(\alpha + T(1))][\Gamma(p\alpha + T(1))/\Gamma(p\alpha)]. \tag{2.10}$$

Next let $X(1), X(2), X(3),\dots$ be geometric random variables as in the discussion preceding the statement of the theorem. With the aid of (2.5), it is not hard to see that $\prod_{j=1}^{n} \tau_j^{t(j)}$ can be written as

$$[\prod_{j=1}^{n} \prod_{i=1}^{X(j)-1} (1 - Q_{X(1)+\ldots+X(j-1)+i})^{T(j)}] \prod_{j=1}^{n} [Q_{X(1)+\ldots+X(j)}^{t(j)} \times$$

$$(1 - Q_{X(1)+\ldots+X(j)})^{T(j+1)}] \tag{2.11}$$

Given X(1), X(2),...,X(n), the conditional expectation of (2.11) is (use (2.8) and (2.9) for the last factor)

$$\prod_{j=1}^{n} [\alpha/(\alpha + T(j))^{X(j)-1} \cdot \prod_{j=1}^{n} [\alpha\Gamma(t(j) + 1)/(\alpha + T(j))]\cdot\Gamma(\alpha)/\Gamma(\alpha + T(1)). \tag{2.12}$$

Now take the expectation of (2.12) with respect to X(1), X(2),...,X(n) to find that

$$E[\prod_{j=1}^{n} \tau_j^{t(j)}] =$$

$$\prod_{j=1}^{n} [p/(1 - (1 - p)\alpha/(\alpha + T(j))]\cdot \prod_{j=1}^{n} [\alpha\Gamma(t(j) + 1)/(\alpha + T(j))] \times$$

$$\Gamma(\alpha)/\Gamma(\alpha + T(1))$$

$$= \prod_{j=1}^{n} [p\alpha\Gamma(t(j) + 1)/(\alpha + T(j) - (1 - p)\alpha)]\cdot\Gamma(\alpha)/\Gamma(\alpha + T(1))$$

$$= \prod_{j=1}^{n} [p\alpha\Gamma(t(j) + 1)/(p\alpha + T(j))]\cdot\Gamma(\alpha)/\Gamma(\alpha + T(1)). \tag{2.13}$$

Dividing (2.13) by (2.10) gives the Mellin transform (2.9) of Engen's model with parameter $p\alpha$. This completes the proof.

Remark 2.3: If the zero components of $\underset{\sim}{\pi}$ are not omitted, then it is false that $\underset{\sim}{\tau}$ is independent of $\underset{\sim}{\pi}$. In fact the sequence of selection variables ε_1, ε_2,... is uniquely determined when one knows which components of $\underset{\sim}{\pi}$ are zero. But knowledge of the ε_i provides information about $\tau = \Sigma\nu_i\varepsilon_i$ since the distribution of $(\nu_1, \nu_2, \nu_3, \ldots)$ is not exchangeable.

Remark 2.4: It has been pointed out previously that the parameter $\alpha = sk$ is proportional to the number of species for the Dirichlet model. Theorem 2.11(b) implies that this interpretation remains viable even after taking Fisher's limit.

Remark 2.5: Theorem 2.11 is also true when Engen's model is replaced with Kingman's limit.

The Mellin transform (2.9) can also be used to prove the invariance of Engen's model under sized-biased permutation and, in the process, to establish the well-known negative Polya identity. Write $\underset{\sim}{\pi} = \underset{\sim}{\nu}^{\#}$ where $\underset{\sim}{\nu}$ is Engen's model with parameter α. Given $\underset{\sim}{\nu}$, the first component π_1 takes the value ν_i with probability ν_i and the Mellin transform of π_1 becomes

$$E[\pi_1^t] = E[\Sigma\nu_i\nu_i^t] = E[\Sigma\ \nu_i^{t+1}] \quad (t \geq 0).$$

By the monotone convergence theorem, the expectation can be computed termwise. Using (2.9) this gives

$$\begin{aligned} E[\pi_1^t] &= \sum_{i=1}^{\infty} [\alpha/(\alpha + t + 1)]^i\ \Gamma(t + 2)\Gamma(\alpha)/\Gamma(\alpha + t + 1) \\ &= [\alpha/(t + 1)]\Gamma(t + 2)\Gamma(\alpha)/\Gamma(\alpha + t + 1) \\ &= \Gamma(t + 1)\Gamma(\alpha + 1)/\Gamma(\alpha + t + 1). \qquad (2.14) \end{aligned}$$

Since (2.14) is the Mellin transform of ν_1, the equality in distribution of $\pi_1 = \nu_1^{\#}$ and ν_1 has been established.

Next consider the first two components (π_1, π_2) of $\underset{\sim}{\pi}$. Conditional on $\underset{\sim}{\nu}$, these take the value (ν_i, ν_j) with probability $\nu_i\nu_j/(1 - \nu_i)$ when $i \neq j$ and with probability zero when $i = j$. The joint Mellin transform is then

$$E[\pi_1^t \, \pi_2^u] = E[\sum_{i\neq j} \nu_i^{t+1} \, \nu_j^{u+1}/(1 - \nu_i)]$$

$$= E[\sum_{i\neq j} \sum_{x=1}^{\infty} \nu_i^{t+x} \, \nu_j^{u+1}]$$

$$= \sum_{x=1}^{\infty} (\sum_{i<j} + \sum_{j<i})E[\nu_i^{t+x} \, \nu_j^{u+1}] \quad (t, u \geq 0).$$

But by (2.9),

$$\sum_{i<j} E[\nu_i^{t+x}\nu_j^{u+1}] = \sum_{i<j} [\alpha/(\alpha + t + x + u + 1)]^i[\alpha/(\alpha + u + 1)]^{j-i}\Gamma(t + x + 1) \times$$

$$\Gamma(u + 2) \, \Gamma(\alpha)/\Gamma(\alpha + u + t + x + 1)$$

$$= [\alpha/(t+x+u+1)][\alpha/(u+1)]\Gamma(t+x+1)\Gamma(u+2)\Gamma(\alpha)/\Gamma(\alpha+u+t+x+1). \tag{2.15}$$

Interchanging $t + x$ and $u + 1$ in (2.15) gives

$$\sum_{j<i} E[\nu_i^{t+x}\nu_j^{u+1}] = [\alpha/(t + x + u + 1)][\alpha/(t + x)]\times$$
$$\Gamma(t + x + 1)\Gamma(u + 2)\Gamma(\alpha)/\Gamma(\alpha + u + t + x + 1). \tag{2.16}$$

Adding (2.15) and (2.16), the Mellin transform of (π_1, π_2) is

$$E[\pi_1^t \, \pi_2^u] = \sum_{x=1}^{\infty} \alpha^2\Gamma(t + x)\Gamma(u + 1)\Gamma(\alpha)/\Gamma(\alpha + u + t + x + 1). \tag{2.17}$$

On the other hand, the Mellin transform of (ν_1, ν_2) is, by (2.9),

$$E[\nu_1^t\nu_2^u] = \alpha^2\Gamma(t + 1)\Gamma(u + 1)\Gamma(\alpha)/[(\alpha + u)\Gamma(\alpha + u + t + 1)]. \tag{2.18}$$

After some cancellations, the equality

$$E[\pi_1^t\ \pi_2^u] = E[\nu_1^t\ \nu_2^u], \tag{2.19}$$

is equivalent to

$$\sum_{x=1}^{\infty} \Gamma(t + x)/\Gamma(\alpha + u + t + x + 1) = \Gamma(t + 1)/[(\alpha + u)\Gamma(\alpha + u + t + 1)]. \tag{2.20}$$

Now here is the trick. The equality in distribution of π_1 and ν_1 has already been established so that (2.19) is true when $u = 0$. Thus (2.20) is true when $u = 0$ and <u>for all</u> $\alpha > 0$. But since (2.20) involves only the sum $\alpha + u$, it must be true for all $\alpha > 0$ and all $u \geq 0$.

It may be remarked that (2.20) is equivalent to

$$\sum_{x=1}^{\infty} x^{-1} \binom{t + 1 + x - 1}{x - 1}\binom{\alpha + u}{1}\bigg/\binom{\alpha + u + t + x + 1}{x} = 1,$$

which is the negative Polya identity.

The above procedure can be iterated (the next step, for example, leads to the bivariate negative Polya identity) but the calculations become very complicated. A simple and general proof of the equality in distribution of $\underset{\sim}{\pi}$ and $\underset{\sim}{\nu}$ will be given in Chapter 4.

As an illustration of the usefulness of the preceding result, consider the problem of calculating $E[\Delta(\underset{\sim}{\nu})]$ where $\underset{\sim}{\nu}$ is now an arbitrary random community and where

$$\Delta(\underset{\sim}{\nu}) = \Sigma\ \nu_i\ R(\nu_i) \tag{2.21}$$

is a dichotomous index with rarity measure R. Given $\underset{\sim}{\nu}$, the right hand side of (2.21) is the conditional expectation of $R(\nu_1^{\#})$ so that

$$E[\Delta(\underset{\sim}{\nu})] = E[R(\nu_1^{\#})], \tag{2.22}$$

provided the expectation on the right exists, possibly with value $+\infty$. The existence of this expectation is guaranteed in the usual case that R is nonnegative.

When $\underset{\sim}{\nu}$ is Engen's model with parameter α, we have shown that $\nu_1^{\#} \equiv \nu_1 \sim \mathrm{Beta}_1(1, \alpha)$, so that (2.22) becomes

$$E[\Delta(\underset{\sim}{\nu})] = \alpha \int_0^1 R(x)(1 - x)^{\alpha-1}dx. \tag{2.23}$$

Equation (2.23) may be compared with equation (20) of Kingman [79]. Given the index, the right hand side of (2.23) depends only on α. The next theorem provides further justification for the use of α as a diversity measure.

Theorem 2.12: Assume that R is decreasing on the interval (0,1] (Criterion C1). Then $E[\Delta(\underset{\sim}{\nu})]$ is an increasing function of α where $\underset{\sim}{\nu}$ is Engen's model with parameter α.

Proof: This is immediate from (2.23) since the distribution $\mathrm{Beta}_1(1, \alpha)$ is stochastically decreasing in α.

Remark 2.6: Theorem 2.12 is true for any index, dichotomous or not, which satisfies Criterion C3. This is proved in Section 2.4.

Computation of some of the standard EΔ diversity measures for Engen's model is reduced to the evaluation of the integral in (2.23). Notice the power of this result; Shannon's index $-\Sigma\nu_i \log(\nu_i)$, for example, is an infinite series and it is not at all apparent that the series converges with probability one, let alone has a finite expectation.

Theorem 2.13: Let $\underset{\sim}{\nu}$ be Engen's model with parameter α. Then

a) $E[\nu_1]$, $E[\nu_2]$, $E[\nu_3]$ forms a geometric series with ratio $\theta = \alpha/(1 + \alpha)$ and odds ratio $\theta/(1 - \theta) = \alpha$,

b) $E[\Delta_\beta(\underset{\sim}{\nu})] = \beta^{-1}[1 - \Gamma(\beta + 1)\Gamma(\alpha + 1)/\Gamma(\alpha + \beta + 1)]$ if $\beta > -1$, $\beta \neq 0$,

c) $E[\Delta_0(\underset{\sim}{\nu})] = \psi(\alpha + 1) - \psi(1)$

$$= \sum_{n=1}^{\infty} \alpha/(n(n + \alpha)], \tag{2.24}$$

d) when $\omega \geq 0$,

$$E[\Delta_\omega^{H-S}(\underset{\sim}{\nu})] = \alpha[\psi(\alpha + \omega + 1) - \psi(\alpha + 1)]$$

$$= \sum_{n=1}^{\infty} (\alpha\omega)/[n + \alpha)(n + \alpha + \omega)], \tag{2.25}$$

e) $E[\Sigma(i - 1)\nu_i] = \alpha$,

f) $E[\Sigma(i - 1)\nu_i^*] = \alpha \log 2 \simeq .69\, \alpha$.

Proof: Part a) is obvious and has been previously pointed out. Parts b), c) and d) follow from (2.23). Part e) follows from the monotone convergence theorem and a). Notice that $\Sigma(i - 1)\nu_i$ can be rewritten as

$$(1 - \nu_1) + (1 - \nu_1 - \nu_2) + (1 - \nu_1 - \nu_2 - \nu_3) + \ldots,$$

or, by (2.5), as

$$(1 - Q_1) + (1 - Q_1)(1 - Q_2) + (1 - Q_1(1 - Q_2)(1 - Q_3) + \ldots, \tag{2.26}$$

where the $1 - Q_i$ are iid with common distribution $\mathrm{Beta}_1(\alpha, 1)$. It follows that the series (2.26) converges will probability one and has finite expectation equal to α. This series with reappear in Chapter 4. The proof of f) is deferred until Chapter 4.

Remark 2.7: When α is a positive integer, (2.24) simplifies to $1 + 1/2 + 1/3 + \ldots + 1/\alpha$. When ω is a positive integer, (2.25) simplifies to $\alpha[(1 + \alpha)^{-1} + (2 + \alpha)^{-1} + \ldots + (\omega + \alpha)^{-1}]$. In particular, we have $E[\Delta_1(\underset{\sim}{\nu})] = E[\Delta_1^{H-S}(\underset{\sim}{\nu})] = \alpha/(\alpha + 1)$.

Remark 2.8: The expressions in b), c) and d) result when Fisher's limit is formally applied to the formulae given in Theorem 2.2. In fact, for any dichotomous index, $E[\Delta(\underset{\sim}{\nu})]$ may be obtained by first calculating this expectation for the Dirichlet model and then applying Fisher's limit. This follows from a comparison of Equation (2.23) with (16) and (20) of Kingman [79].

Remark 2.9: For the rank type index $\Delta_\rho^{(rank)}$, one obtains

$$E[\Sigma\, \nu_i(1 - \rho^{i-1})/(1 - \rho)] = \begin{cases} \alpha/(1 + \alpha - \rho\alpha) \text{ if } 0 \leq \rho < (\alpha + 1)/\alpha \\ \\ \infty \quad \text{if } (\alpha + 1)/\alpha \leq \rho. \end{cases}$$

On the other hand, closed form expressions for the expected value $E[\Delta_\rho^{(rank)}(\underset{\sim}{\nu})] = E[\Sigma \nu_i^*(1 - \rho^{i-1})/(1 - \rho)]$ are not known.

2.3 Fixing the Limiting Dirichlet Model

A number of authors, most notably Engen [84, 88], have suggested that it may be preferable to draw inferences about the physically realized community instead of underlying generating mechanisms. To this end, several attempts have been made to identify fixed communities that closely resemble realizations of the limiting Dirichlet. Most of these attempts fall into two general classes which may be termd the *method of expectations* and the *method of quantiles*.

The first approach (method of expectations) replaces the random community $\underset{\sim}{\pi}$ with the expected value of some canonical, stochastically decreasing permutation of $\underset{\sim}{\pi}$. For instance, either $\underset{\sim}{\nu} = E[\underset{\sim}{\pi}^*]$ or $\underset{\sim}{\nu} = E[\underset{\sim}{\pi}^\#]$ might be adopted as the fixed model. The ranked permutation of the limiting Dirichlet is intractable and will not be discussed here. The size-biased permutation of the limiting Dirichlet leads to the geometric series model with ratio θ,

$$\nu_i = \theta^{i-1}(1 - \theta), \ i = 1,2,3,\ldots, \ 0 < \theta < 1, \tag{2.27}$$

and has been recommended by Engen [84]. The ratio is a complete diversity parameter for this model, but in practice, it may be preferable to use the odds ratio $\alpha' = \theta/(1 - \theta) = \Sigma(i - 1)\nu_i$ since it (apparently) corresponds to Fisher's α by Theorem 2.13 (a,e). Estimation of α' for multinomial sampling has been discussed by Engen [84], but his justification of jackknifing appears to be invalid. See Appendix A.6.

The second appraoch (method of quantiles) begins with the assumption that the species abundances are independent realizations of the standard Gamma distribution with index k. These random abundances are then approximated by the quantiles of the Gamma distribution. Assuming there are s species, the quantiles ν_i are defined implicitly by

$$\int_{\nu_i}^{\infty} x^{k-1}e^{-x}/\Gamma(k)\,dx = i/(s+1), \ i = 1,2,\ldots,s. \tag{2.28}$$

Dividing both sides of (2.28) by k and applying Fisher's limit gives

$$\int_{\nu_i}^{\infty} x^{-1}e^{-x}dx = i/\alpha, \ i = 1,2,3,\ldots \ . \tag{2.29}$$

The solution of (2.29) is obtained as

$$\nu_i = E_1^{-1}(i/\alpha),\ i = 1,2,3,\ldots, \tag{2.30}$$

where E_1^{-1} is the inverse function of the exponential integral function. Tables and properties of $E_1(x)$ can be found in Abramowitz and Stegun [89]. It may be shown that $\Sigma\nu_i < \infty$; in what follows there is no need to normalize the ν_i to unity. The preceding derivation of (2.30) is due to Watterson [90]; variants have been considered by Engen [88] and Holgate [91].

Both (2.27) and (2.30) are candidates for "fixations" of the limiting Dirichlet and we wish to judge how closely they resemble realizations of this random model. A first standard of comparison is provided by the next theorem.

Theorem 2.14: Let $\underset{\sim}{\pi}$ be the limiting Dirichlet as represented by Engen's random model with parameter α. Then, with probability one,

$$\lim_{i\to\infty} \log(\pi_i)/i = -1/\alpha. \tag{2.31}$$

Thus the tail of $\underset{\sim}{\pi}$ is approximately geometric with ratio $\theta = \exp(-1/\alpha)$ and odds ratio $\alpha' = 1/[\exp(1/\alpha) - 1]$.

Proof: Use (2.5) to write the components of $\underset{\sim}{\pi}$ as

$$\pi_i = (1 - Q_1)(1 - Q_2)\ldots(1 - Q_{i-1})Q_i, \tag{2.32}$$

where the Q_i are independent realizations of the $\text{Beta}_1(1, \alpha)$ distribution. Take logarithms, divide by i, and apply the strong law of large numbers and the Borel-Cantelli lemma to conclude that, with probability one, $\log(\pi_i)/i$ converges to $E[\log(1 - Q_1)]$. After the change of variable $V = -\alpha \log(1 - Q_1)$, the expectation is easily seen to be $-1/\alpha$.

Remark 2.10: Kingman [79] has shown that (2.31) also holds when $\underset{\sim}{\pi}$ is the ranked permutation of Engen's model.

The result (2.31) is satisfying in that it holds with probability one, but a word of caution is in order since the asymptotic behavior of $\underset{\sim}{\pi}$ can be very nongeometric. From (2.32), ratio π_{i+1}/π_i equals $(1 - Q_i)Q_{i+1}/Q_i$ so that π_2/π_1, π_4/π_3, $\pi_6/\pi_5, \ldots$ are independent and identically distributed unbounded random variables. It follows that the ratios form a dense subset of the positive real axis and have no limit as $i \to \infty$.

At least three levels of asymptotic geometricity can be identified. Let $0 < \theta < 1$. A sequence $a_1, a_2, a_3, \ldots$ of positive real numbers is said to be <u>asymptotically geometric</u> (AG) in the strong, intermediate, or weak sense when:

i) (strong AG) $0 < \lim\limits_{i \to \infty} a_i/\theta^i < \infty$,

ii) (intermediate AG) $\lim\limits_{i \to \infty} a_{i+1}/a_i = \theta$,

iii) (weak AG) $\lim\limits_{i \to \infty} \log(a_i)/i = \log(\theta)$.

In all three cases, call θ the <u>asymptotic ratio</u>. As the terminology suggests, strong AG implies intermediate AG implies weak AG and each implies that the series Σa_i is convergent. Notice that intermediate AG is equivalent to the requirement that the differences $\log(a_{i+1}) - \log(a_i)$ converge to $\log \theta$. On the other hand, by telescoping, weak AG requires only that the Caesaro means of these differences converge to $\log \theta$. Weak AG is thus seen to be weak indeed and is about all that can be expected if the a_i are random or, for that matter, are any physically realized quantities. See Appendix A.4 for further discussion of asymptotic geometricity.

How compatible with Theorem 2.14 are the two fixed models (2.27) and (2.30)? The geometric series model is evidently strong AG except that the odds ratio should be taken as $\alpha' = 1/[\exp(1/\alpha) - 1]$ and not α as might be suggested by Theorem 2.13. However the correction is small since $\alpha - 1/2 \leq \alpha' \leq \alpha$, with the lower bound being the better approximation for reasonable values of α. A numerical comparison is given in Table 2.2. The correction has the effect of reducing the ratio and hence the diversity. Even so, the geometric series model is intrinsically more diverse than Watterson's exponential integral model for a given value of α (see Appendix A.7). The relative frequencies of the two models are plotted in Figure 2.1 when $\alpha = 5$. As the plot shows, the exponential integral model assigns greater abundance to the common species and less abundance to the rare species than does the geometric series model.

Table 2.2: Lower and Upper Bounds for the Corrected Odds Ratio

$\alpha - 1/2$	$1/[\exp(1/\alpha) - 1]$	α
0.500	0.582	1
1.500	1.541	2
2.500	2.528	3
4.500	4.517	5
7.500	7.510	8
9.500	9.508	10
19.500	19.504	20
49.500	49.502	50
99.500	99.501	100

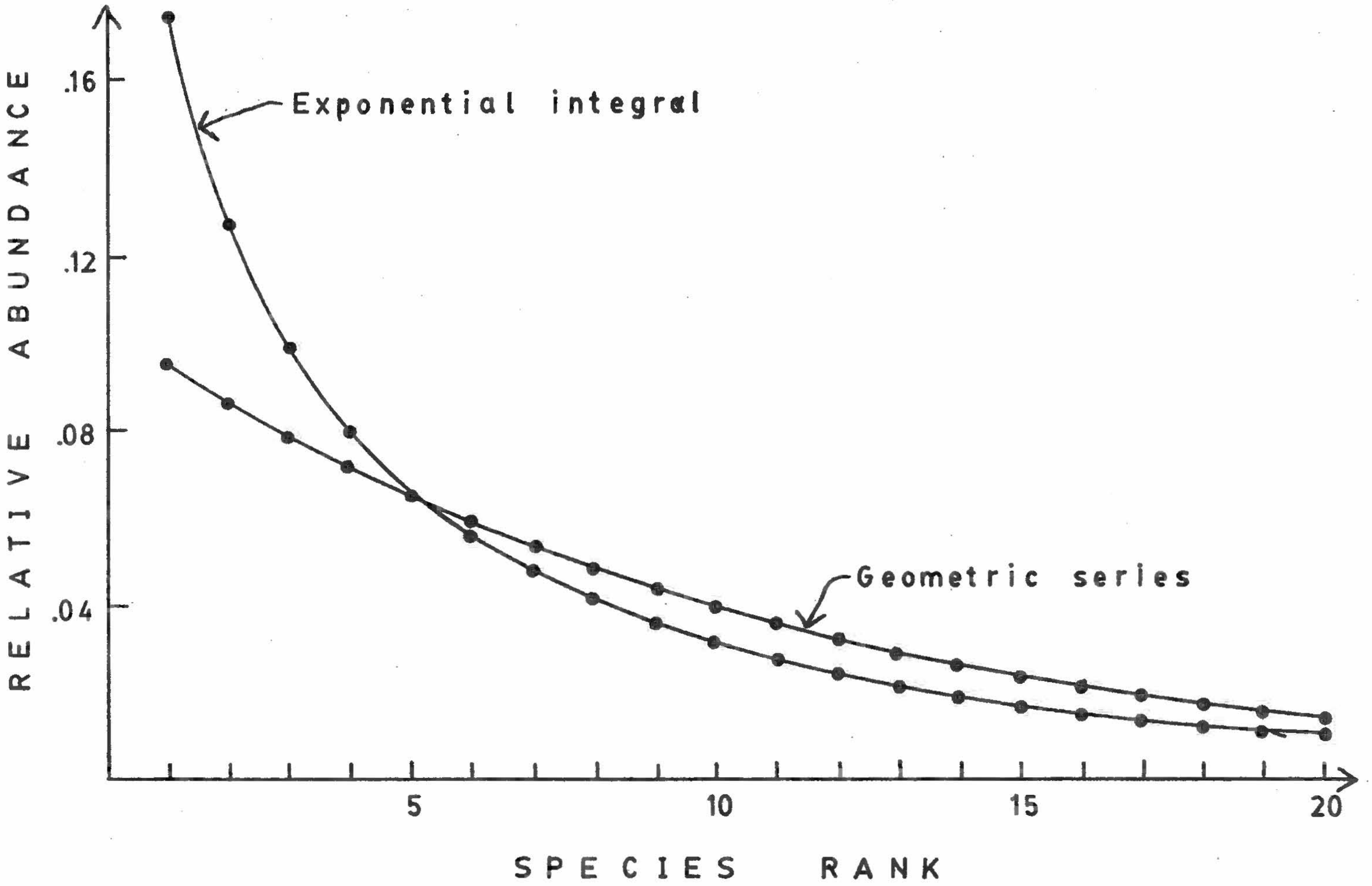


Figure 2.1 Relative Frequencies of the Exponential Integral

Model and the Geometric Series Model ($\alpha = 5$)

It is possible to show that the sequence defined by (2.30) is also strong AG with ratio θ given by $-\log \theta = 1/\alpha$. This establishes that $\Sigma\nu_i < \infty$ as pointed out earlier. Both models thus satisfy (2.31) albeit in a much smoother sense than would be observed in any realization of the limiting Dirichlet. But asymptotic geometricity is only part of the story. Chapter 4 develops a wide class of random communities which satisfy this condition with probability one, so that, while asymptotic geometricity is a helpful guide in identifying possible fixations of the limiting Dirichlet, it alone does not suffice.

A second, and more powerful, standard of comparison was initiated by Watterson [90] in connection with the exponential integral model. It consists of comparing the sampling distribution of the fixed model with Fisher's log series distribution.

Several version's and/or interpretations of the log series exist and have been surveyed by Watterson [90]. The version of interest here (Watterson's Version 2) is due to Anscombe [92], who obtained it by supposing that the species abundances $\lambda_1, \lambda_2, \ldots, \lambda_s$ were independent realizations of the standard Gamma distribution with index k. The observed number of representatives of species i was further assumed to be a Poisson variate with mean $A\lambda_i$ where A is the sampling intensity. Letting n_x, $x = 1,2,3,\ldots$, be the number of species having x representatives in the sample, the vector of observations $(n_1, n_2, n_3,\ldots)$ is a sufficient statistic whose likelihood may be written down. The number of species in the community enters this likelihood as an unknown parameter. After applying Fisher's limit to the likelihood, the n_x become independent Poisson variates whose means are given by

$$E[n_x] = \alpha[A/(1 + A)]^x/x. \qquad (2.33)$$

The observed number of species Σn_x is, *as a consequence*, also Poisson with mean $\alpha \log (1 + A)$.

Notice that the preceding derivation does not involve truncation. Some authors, Pielou [33, p. 44], for instance, prefer to truncate the negative binomial before letting $k \to 0$. The *population parameter* α is then introduced in an obscure way that makes it appear to be at once a random variable and a function of the sampling parameter A.

Now consider Poisson sampling from a fixed model $\underset{\sim}{\nu}$: assume the observed number of representatives from the different species are independent Poisson variates whose means $A\nu_i$ are proportional to the respective species abundances. Because of the confounding between A and $\underset{\sim}{\nu}$, there is no need to normalize $\underset{\sim}{\nu}$ to unity. The subsequent algebra is somewhat simplified if the geometric series model is specified by

$$\nu_i = \theta^i \tag{2.34}$$

in place of (2.27).

The joint probability generating function of $n_1, n_2, n_3, \ldots$ is

$$E[t_1^{n_1} t_2^{n_2} t_3^{n_3} \ldots] = \prod_{i=1}^{\infty} [1 - \sum_{x=1}^{\infty} (1 - t_x)(A\nu_i)^x \exp(-A\nu_i)/x!] \tag{2.35}$$

and, in particular,

$$E[t_1^{n_1}] = \prod_{i=1}^{\infty} [1 - (1 - t_1)A\nu_i \exp(-A\nu_i)]. \tag{2.36}$$

From (2.35), it follows that the n_x are not independent and their marginal distributions are not Poisson. At least to this extent, the sampling distribution of a fixed model can never be Fisher's log series. This is not surprising: a central purpose of fixation is the elimination of the variance component due to community variability and this elimination, to be effective, must change the sampling distribution.

One is still free to demand that the n_x have expectations given by (2.33), at least approximately, and this is the second requirement to be satisfied by any proposed fixation of the limiting Dirichlet model.

Either directly or from (2.35), one finds that

$$E[n_x] = \sum_{i=1}^{\infty} (A\nu_i)^x \exp(-A\nu_i)/x!. \tag{2.37}$$

Often ν_i is formally defined for all positive real i and is a continuous decreasing function of i. This is the case for both (2.30) and (2.34). The sum in (2.37) can then be approximated by an integral to give

$$E[n_x] \simeq I_x \equiv \int_0^{\infty} (A\nu_i)^x \exp(-A\nu_i)/x!\ di. \tag{2.38}$$

Watterson [90] has established the bounds

$$I_x - 1 < E[n_x] < I_x + 2, \tag{2.39}$$

although the approximation is usually much better than might be suggested by (2.39), especially if A is large. The integral I_x can be evaluated by a change of variable to produce the following two results, the first of which was found by Watterson [90]:

i) When $\nu_i = E_1^{-1}(i/\alpha)$,

$$E[n_x] \simeq I_x = \alpha[A/(1 + A)]^x/x, \tag{2.40}$$

ii) When $\nu_i = \theta^i$,

$$E[n_x] \simeq I_x = (-1/\log \theta)\Gamma(x;A)/x, \tag{2.41}$$

where $\Gamma(x;A)$ is the incomplete Gamma function defined by

$$\Gamma(x;A) = \int_0^A u^{x-1}e^{-u}du/\Gamma(x)$$

$$= 1 - e^{-A}[1 + A + A^2/2!+\ldots+A^{x-1}/(x - 1)!].$$

The coefficient $-1/\log\theta$ in (2.41) may be identified with Fisher's α by Theorem 2.14.

These results show that Watterson's exponential integral model is an acceptable fixation of the limiting Dirichlet but that Engen's geometric series model is not. The random community corresponding to the geometric series model is identified in Chapter 4; the limiting Dirichlet is a randomization of this community.

The earliest plots of n_x versus x are apparently due to J. C. Willis [93] who observed a consistent pattern which he called the "hollow curve." Subsequently, Chamberlin [94] described the ideal hollow curve as the rectangular hyperbola α/x. Both (2.40) and (2.41) are weighted rectangular hyperbolas of form $\alpha w(x;A)/x$. The unweighted rectangular hyperbola is impossible in the present context since the number of species in the sample is a Poisson random variable with mean $\Sigma E[n_x] < \infty$. Fisher [85] introduced the exponential convergence factor of (2.40). The Gamma convergence factor of (2.41) seems to be new and has the effect of smoothly truncating the rectangular hyperbola at about $x = A$. Figure 2.2 provides a comparison of the two convergence factors when $A = 50$ and $A = 100$.

The geometric series model will rarely be useful for graduating species frequency data since the latter is often found to have an even longer tail than the log series (Kempton & Taylor [86], Anscombe [92], and Kempton [95]). It may possibly find application in word frequency studies where the truncated rectangular hyperbola has been used.

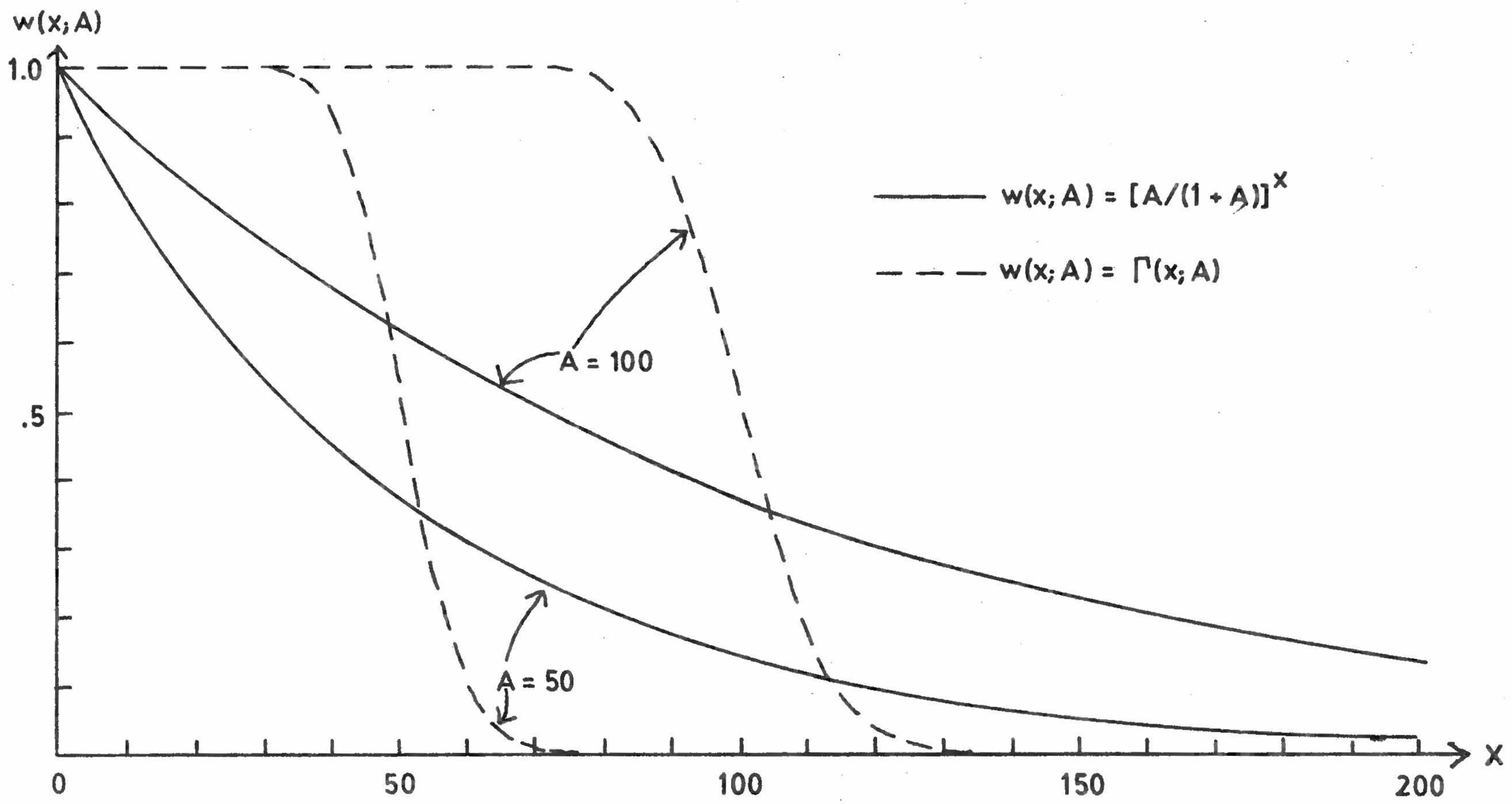


Figure 2.2 Comparison of the Exponential Convergence Factor of (2.40) with the Gamma Convergence Factor of (2.41) (A = 50, 100)

In principle, the sampling distribution corresponding to a fixed model $\underset{\sim}{\nu}$ is determined by the generating function (2.35). However, the distribution is extremely complicated and, for realistic sized samples, it will be impossible to explicitly write down the likelihood. Some information can still be extracted from the generating function. For simplicity, consider (2.36). Since $\Sigma A\nu_i \exp(-A\nu_i)|1 - t_1| < \infty$, the infinite product in (2.36) is absolutely convergent for all t_1 and defines an analytic function of the complex variable t_1. Moreover, this function has zeros at the values $t_1 = 1 - 1/[A\nu_i \exp(-A(\nu_i)]$, $i = 1,2,3,\ldots$. When the sequence ν_i is strong AG with ratio θ, these zeros do not stabilize as $A \to \infty$ but vary approximately as a periodic function of $-\log_\theta A$. It follows that the likelihood is multimodal for large A, making maximum likelihood estimation difficult even if the likelihood could be written down.

The limiting periodicity can be seen quite clearly for the geometric series model. Putting $\nu_i = \theta^i$ in (2.37) gives

$$E[n_1] = \sum_{i=1}^{\infty} A\theta^i \exp(-A\theta^i). \tag{2.42}$$

When A goes to ∞ through the successive values $a\theta^{-n}$, $n = 1,2,3,\ldots$, this expression converges to

$$\sum_{i=-\infty}^{\infty} a\theta^i \exp(a\theta^i),$$

which is clearly a periodic function of $-\log_\theta a$.

When $\theta = 1/2$, $E[n_1]$ has been plotted against $\log_2 A$ in Figure 2.3 and Figure 2.4. These figures display the graphs of both (2.42) and the approximation (2.41). Notice that the oscillation does not damp out as $A \to \infty$ although the amplitude (about .000015) of the oscillation is very small.

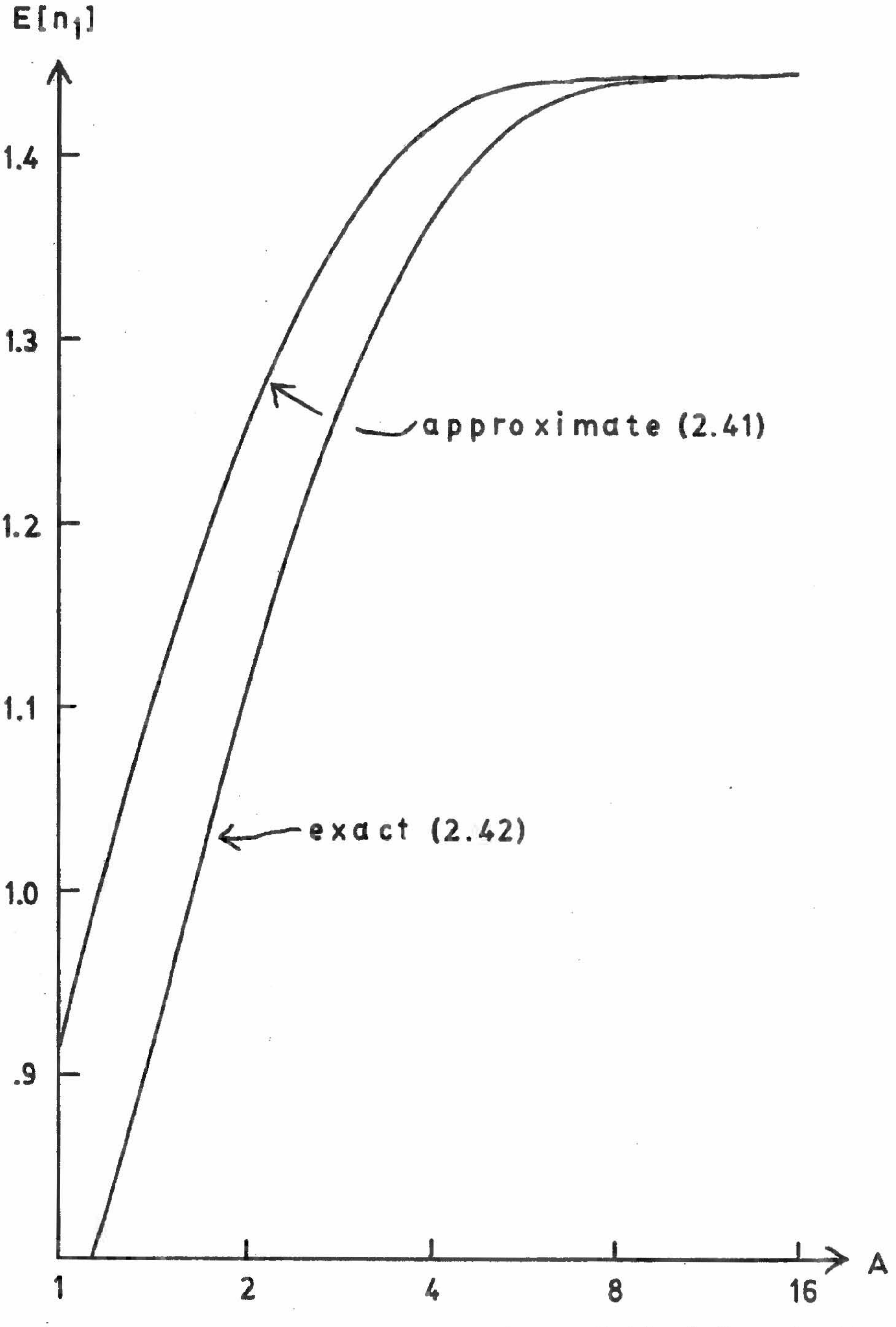


Figure 2.3 Exact and Approximate Values of $E[n_1]$ for the Geometric Series Model ($\theta = 1/2$ and $\alpha = 1$)

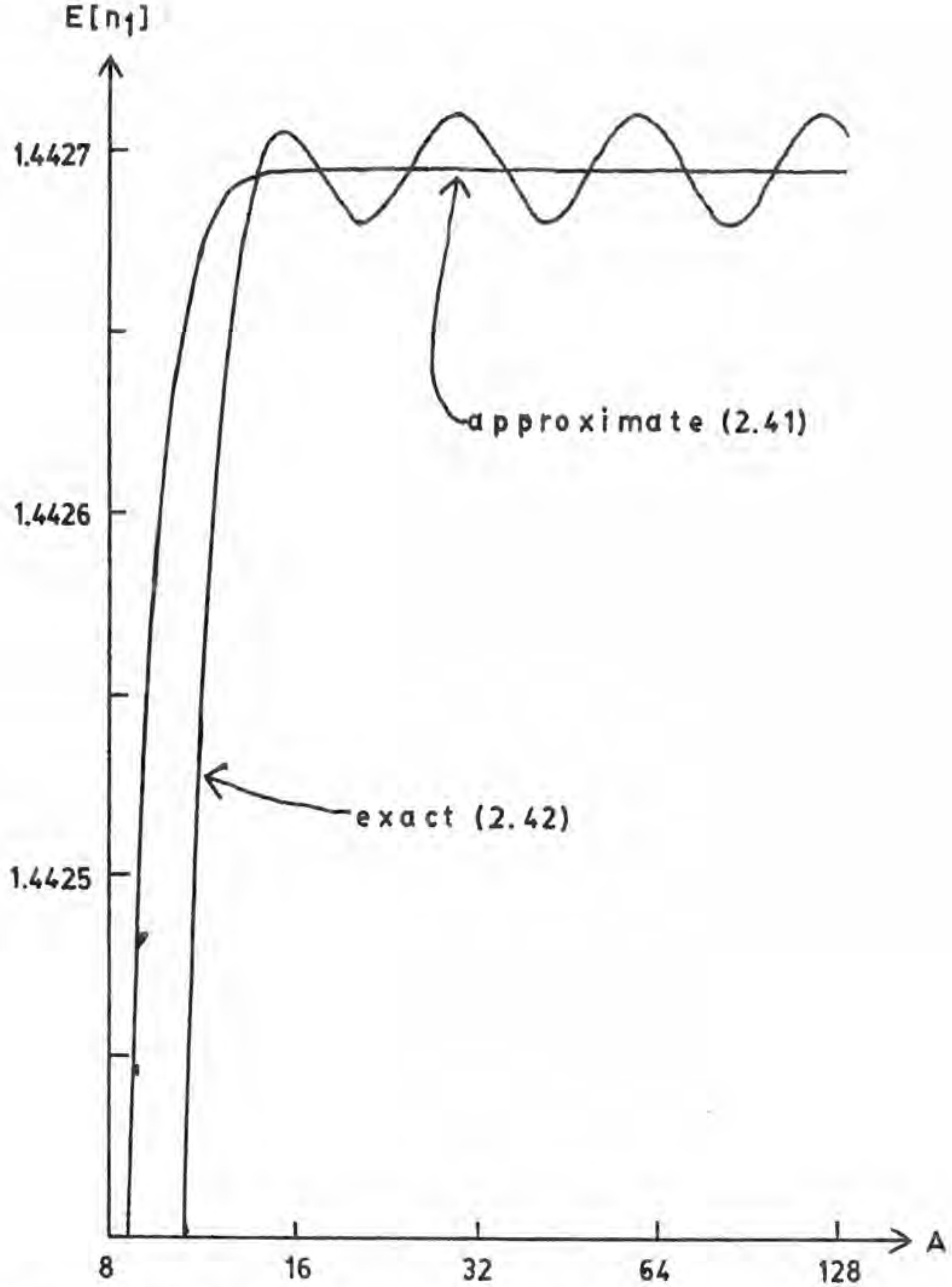


Figure 2.4 Limiting Periodicity of $E[n_1]$ for the Geometric Series Model ($\theta = 1/2$ and $\alpha = 1$)

In the following we provide a brief summary of the material in this section:

1. Two requirements are imposed on any fixation $\underset{\sim}{\nu}$ of the limiting Dirichlet: (i) $\underset{\sim}{\nu}$ should be asymptotically geometric and (ii) the mean sampling frequencies $E[n_x]$ should be approximately proportional to the terms of the log series distribution.

2. Watterson's [90] exponential integral model meets both requirements although the asymptotic geometricity is smoother than would be observed in practice.

3. For given α, Engen's [84] geometric series model is intrinsically more diverse than the exponential integral model. The geometric series model satisfies the first requirement but not the second; the mean sampling frequencies are a smooth truncation of the rectangular hyperbola. The geometric series model will seldom provide an adequate fit to species frequency data.

4. Any fixed model which is asymptotically geometric in the strong sense has a multimodal likelihood function.

2.4 Stochastic Diversity

This section extends the intrinsic diversity ordering to random communities. As was the case for fixed communities, the ordering is only partial and two random communities need not be comparable. In fact, comparability is the exception rather than the rule. However, some comparability relations are established for Dirichlet communities and these relations shed further light on the role of Fisher's "alpha" as a diversity parameter.

There are several stochastic analogues of the intrinsic diversity ordering. Four such analogues are considered here. Let $\underset{\sim}{\pi}$ and $\underset{\sim}{\nu}$ be two random communities and define $\underset{\sim}{\nu}$ to be *stochastically more diverse* than $\underset{\sim}{\pi}$ in the sense (SD2, SD3, SD4, SD5) if:

SD2: $E[\Delta(\underset{\sim}{\nu})] \geq E[\Delta(\underset{\sim}{\pi})]$ whenever Δ satisfies Criterion C2.

SD3: $E[\Delta(\underset{\sim}{\nu})] \geq E[\Delta(\underset{\sim}{\pi})]$ whenever Δ satisfies Criterion C3.

SD4: For every positive integer m, the random variable $\sum_{i \geq m} \nu_i^*$ is stochastically greater than the random variable $\sum_{i \geq m} \pi_i^*$.

SD5: The fixed community $E[\underset{\sim}{\nu}^*]$ is intrinsically more diverse than the fixed community $E[\underset{\sim}{\pi}^*]$.

As a technical point, the above diversity indices should be required to be Borel measurable. It will be seen below that these four definitions are equivalent for fixed, but not for random, communities. Notice also that each definition is invariant to permutations of $\underset{\sim}{\pi}$ and of $\underset{\sim}{\nu}$.

Stochastic diversity in the sense SD2 is the dual of stochastic majorization as defined by Nevius, Proschan & Sethuraman [96]. In the same paper, these authors have shown (Theorem 2.2) that the following statements are equivalent:

i) $\underset{\sim}{\nu}$ is stochastically more diverse than $\underset{\sim}{\pi}$ in the sense SD2.

ii) The random variable $\Delta(\underset{\sim}{\nu})$ is stochastically greater than $\Delta(\underset{\sim}{\pi})$ whenever Δ satisfies Criterion C2.

iii) $P(\underset{\sim}{\nu} \in A) \geq P(\underset{\sim}{\pi} \in A)$ whenever A is a Schur convex set.

Statement (iii) is a direct requirement on the probability distributions of $\underset{\sim}{\pi}$ and $\underset{\sim}{\nu}$ and is similar to the definition of stochastic ordering of random variables. For this reason SD2 is the most elegant and, perhaps, the most appropriate formulation of stochastic diversity. However, SD2 is a very stringent requirement. A few of the results established below make use of Jensen's inequality and apply to the definition SD3; some of these results are definitely false for SD2 while others are unknown.

First the logical relations that exist among the alternative formulations of stochastic diversity will be given.

<u>Theorem 2.15</u>: Each of the following implications is true:

$$\begin{tikzcd}
 & \text{SD3} \arrow[dr, Rightarrow] & \\
\text{SD2} \arrow[ur, Rightarrow] \arrow[dr, Rightarrow] & & \text{SD5}. \\
 & \text{SD4} \arrow[ur, Rightarrow] &
\end{tikzcd}$$

In general, none of these implications can be reversed; however, for fixed communities the four definitions are equivalent.

<u>Proof</u>: Clearly SD2 implies SD3 since any index satisfying Criterion C3 also satisfies Criterion C2. The proof of the remaining three implications uses the rank type index $\Delta = T_m(\underset{\sim}{\pi})$ defined by $T_m(\underset{\sim}{\pi}) = T_m(\underset{\sim}{\pi}^*) = \sum_{i \geq m} \pi_i^*$ $(m = 1,2,3,\ldots)$, which satisfies both Criterion C2 and C3 by Theorem 1.9. Applying statement (ii) to the index T_m shows that SD2 implies SD4. Next, if $\underset{\sim}{\nu}$ is stochastically more diverse than $\underset{\sim}{\pi}$ in either of the senses SD3 or SD4, then $E[T_m(\underset{\sim}{\nu}^*)] \geq E[T_m(\underset{\sim}{\pi}^*)]$. Since T_m is linear, this implies that $T_m(E[\underset{\sim}{\nu}]) \geq T_m(E[\underset{\sim}{\pi}^*])$ and, consequently, that $E[\underset{\sim}{\nu}^*]$ is intrinsically more diverse that $E[\underset{\sim}{\pi}^*]$. To prove that none of the implications can be reversed, it will suffice to show that neither of the two definitions SD3 and SD4 implies the other. For a counterexample to SD4 => SD3, let $\underset{\sim}{\pi}$ and $\underset{\sim}{\nu}$ be defined as in Example 2.3 and Example 2.4, respectively. A counterexample to SD3 => SD4 is given below in Example 2.15. The last assertion

of the theorem is clear since SD5 implies SD2 for fixed communities.

Example 2.14: This is the first application of Jensen's inequality. Let $\underset{\sim}{\pi}$ be a random community and Δ an index satisfying Criterion C3. Then $E[\Delta(\underset{\sim}{\pi})] = E[\Delta(\underset{\sim}{\pi}^*)] \leq \Delta(E[\underset{\sim}{\pi}^*])$ which shows that:

(a) $E[\underset{\sim}{\pi}^*]$ is stochastically more diverse than $\underset{\sim}{\pi}$ in the sense SD3.

Now let $\underset{\sim}{\nu}$ be any random permutation of $\underset{\sim}{\pi}$ for which $E[\underset{\sim}{\nu}]$ is ranked, i.e., $E[\nu_1] \geq E[\nu_2] \geq E[\nu_3] \geq \ldots$; for example, $\underset{\sim}{\nu}$ might be the size-biased permutation $\underset{\sim}{\pi}^{\#}$. Clearly $\pi_1^* + \pi_2^* + \ldots + \pi_m^* \geq \nu_1 + \nu_2 + \ldots + \nu_m$ for any positive integer m. Taking expectations shows that:

(b) $E[\underset{\sim}{\nu}]$ is intrinsically more diverse than $E[\underset{\sim}{\pi}^*]$, and

(c) $E[\underset{\sim}{\pi}^{\#}]$ is intrinsically more diverse than $E[\underset{\sim}{\pi}^*]$.

Taken together, these relations indicate that fixing a random community by the method of expectations will exaggerate the community's diversity. Also, (a) and (c) help explain the earlier result that Engen's fixation of the limiting Dirichlet is intrinsically more diverse than Watterson's fixation.

Example 2.15: Let $\underset{\sim}{\pi}$ take the two values (1/2, 1/2, 0) and (1/3, 1/3, 1/3) each with probability 1/2 and let $\underset{\sim}{\nu}$ take the value (5/12, 5/12, 2/12) with probability 1. It is easy to see that $\underset{\sim}{\pi}$ and $\underset{\sim}{\nu}$ are not comparable in either of the senses SD2 or SD4. On the other hand, $\underset{\sim}{\nu} = E[\underset{\sim}{\pi}] = E[\underset{\sim}{\pi}^*]$ and, by Example 2.14(a), $\underset{\sim}{\nu}$ is stochastically more diverse than $\underset{\sim}{\pi}$ in the sense SD3. Finally, in the sense SD5, each of $\underset{\sim}{\nu}$ and $\underset{\sim}{\pi}$ is stochastically greater than the other. This example indicates that weakening the definition of stochastic diversity enriches the class of communities that are comparable with one another.

Example 2.16: Draw a random sample of size N from the fixed community $\underset{\sim}{\nu}$ and let Y_i $(i = 1,2,3,\ldots)$ be the number of times the ith ranked species is represented in the sample. Since $E[\underset{\sim}{Y}/N] = \underset{\sim}{\nu}^*$ is already ranked, Example 2.14(a,b) may be applied to show that $E[\underset{\sim}{Y}/N]$ is stochastically more diverse than $\underset{\sim}{Y}/N$ in the sense SD3. Using the permutation invariance, $\underset{\sim}{\nu}$ is then stochastically more diverse than the sample $\underset{\sim}{Y}/N$. In practice, only the order statistics $X_1 \geq X_2 \geq X_3 \geq \ldots$ of $\underset{\sim}{Y}$ can be observed; but, again by permutation invariance $\underset{\sim}{\nu}$ is stochastically more diverse than $\underset{\sim}{X}/N$ in the sense SD3. Thus $\Delta(\underset{\sim}{\nu}) \geq E[\Delta(\underset{\sim}{X}/N)]$ whenever Δ satisfies Criterion C3. In other words, $\Delta(\underset{\sim}{X}/N)$ is always a negatively biased estimator of $\Delta(\underset{\sim}{\nu})$. Simple counterexamples can be constructed to show that this is not generally true for indices satisfying Criterion C2.

The assumption that $\underset{\sim}{\nu}$ is a fixed community is not essential in the last example. In fact, the next theorem states that stochastic diversity in any of the four senses is preserved under mixing. The straightforward proof using conditional expectations is omitted. The case SD2 has been previously noted by Nevius, Proschan & Sethuraman [96, Theorem 3.1].

Theorem 2.16: Let $\underset{\sim}{\pi}$ and $\underset{\sim}{\nu}$ be random communities and let U be a random variable. Assume the conditional distribution of $\underset{\sim}{\nu}$ given $U = u$ is stochastically more diverse in any one of the four senses than the conditional distribution of $\underset{\sim}{\pi}$ given $U = u$ for each u. Then $\underset{\sim}{\nu}$ is stochastically more diverse than $\underset{\sim}{\pi}$ in the same sense.

Most of the results in Nevius, Proschan & Sethuraman [96] and in Proschan & Sethuraman [97] are stated under an assumption that certain parameters λ and λ_i are positive. In the present terminology, this requires that both communities have the same number of species which is too restrictive for applications to diversity. Fortunately, this assumption can be relaxed to the requirement that λ and λ_i be nonnegative. Simply note that the proof of the preservation theorem as given by Proschan & Sethuraman [97] remains valid for $\lambda = 0$ provided $\phi(0, x)$ is interpreted as the Dirac delta function with unit mass at $x = 0$. With this extension, Application 4.2(a) of Nevius, Proschan & Sethuraman [96] implies that random samples reflect the intrinsic diversity ordering in the following sense:

Theorem 2.17: Let $\underset{\sim}{Y}$ and $\underset{\sim}{Y}'$ be random samples of size N from the fixed communities $\underset{\sim}{\pi}$ and $\underset{\sim}{\pi}'$ as in Example 2.16. Then $\underset{\sim}{Y}'/N$ is stochastically more diverse than $\underset{\sim}{Y}/N$ in the sense SD2 when $\underset{\sim}{\pi}'$ is intrinsically more diverse than $\underset{\sim}{\pi}$.

The rest of this section is devoted to diversity comparisons of Dirichlet communities. The Dirichlet model has the two parameters s and k as well as the derived parameter $\alpha = sk$ and any diversity relations will have to involve two of these parameters. Necessary and sufficient conditions for the comparability of two Dirichlet communities have not been obtained but the next theorem is a close approximation.

Theorem 2.18: Let $\underset{\sim}{\pi}$ and $\underset{\sim}{\pi}'$ be Dirichlet communities with parameters (s, k) and (s', k'), respectively. In order that $\underset{\sim}{\pi}'$ be stochastically more diverse than $\underset{\sim}{\pi}$ in the sense SD2:

a) the pair of conditions $s' \geq s$ and $(s' - 1)\, k' \geq (s - 1)k$ are necessary, and

b) the pair of conditions $s' \geq s$ and $s'k' \geq sk$ are sufficient but not necessary.

Proof: a) Necessary conditions are obtained by requiring that

that $E[\Delta(\underset{\sim}{\pi}')] \geq E[\Delta(\underset{\sim}{\pi})]$ for appropriate choices of the index Δ. The choices should include indices sensitive to rare species as well as indices sensitive to abundant species. The choice of $\Delta_{-1} = s - 1$ establishes the necessity of $s' \geq s$. Next consider Δ_β for large β. Using Theorem 2.1(a), this choice requires that

$$A[\Gamma(k' + \beta + 1)/\Gamma(k + \beta + 1)][\Gamma(sk + \beta + 1)/\Gamma(s'k' + \beta + 1)] \leq 1, \qquad (2.43)$$

where $A = [\Gamma(k + 1)/\Gamma(k' + 1)][\Gamma(s'k' + 1)/\Gamma(sk + 1)]$. But as $\beta \to \infty$, the left hand side of (2.43) is asymptotic to $A\ \beta^{(s-1)k-(s'-1)k'}$ so that $(s - 1)k - (s' - 1)k'$ must be less than or equal to zero. Note: if $(s - 1)k = (s' - 1)k'$, one must also have $A \leq 1$. However, Lemma 2.1, below, shows that this is automatically satisfied when $s' \geq s$.

b) Consider separately the two cases: (i) $s' > s$ but $s'k' = sk$, and (ii) $s' = s$ but $s'k' > sk$. Case (i) follows from Application 4.2(d) of Nevius, Proschan & Sethuraman [96]. Recall that the λ_i need only be nonnegative. For case (ii), the s-component symmetric Dirichlet family has monotone likelihood ratio in the index in the sense that the likelihood ratio is a Schur concave function of $\underset{\sim}{\pi}$. Now proceed just as in the usual proof that a monotone likelihood ratio family is stochastically increasing. That b) is not necessary is demonstrated in Theorem 2.19, below.

It seems likely that the conditions given in Theorem 2.18(a) are both necessary and sufficient. This will be proved when $s = 2$ and $s' = 3$. We need a preliminary lemma whose proof is given in Appendix A.8.

Lemma 2.1: Let $0 < A_1 < B_1 < B_2 < A_2$ with $A_1 + A_2 = B_1 + B_2$. Then $\Gamma(B_1 + 1)\Gamma(B_2 + 1) < \Gamma(A_1 + 1)\Gamma(A_2 + 1)$.

<u>Theorem 2.19</u>: Let $\underset{\sim}{\pi} \sim D(2, k)$ and $\underset{\sim}{\pi}' \sim D(3, k')$. Then $\underset{\sim}{\pi}'$ is stochastically more diverse than $\underset{\sim}{\pi}$ in the sense SD2 $\Leftrightarrow 2k' \geq k$.

<u>Proof</u>: Necessity has already been proved. Since the stochastic diversity of $\underset{\sim}{\pi}$ is increasing in k, we may suppose that $k = 2k'$ and show that $P(\underset{\sim}{\pi}' \in T) - P(\underset{\sim}{\pi} \in T) \geq 0$ for T any Schur convex subset of the two-dimensional simplex. It may also be assumed that T has a nonempty intersection with the one-dimensional faces. Referring to Figure 2.5, the intersection of T with any one of these faces is an interval symmetric about the midpoint and of length $2t - 1$ where $1/2 \leq t \leq 1$. But

$$P(\underset{\sim}{\pi}' \in T) \geq 1 - 3\Gamma(3k')/[\Gamma(2k')\Gamma(k')]\int_t^1 x^{k'-1}(1 - x)^{2k'-1}dx,$$

$$P(\underset{\sim}{\pi} \in T) = 1 - 2\Gamma(2k)/[\Gamma(k)\Gamma(k)]\int_t^1 x^{k-1}(1 - x)^{k-1}dx.$$

Thus $P(\underset{\sim}{\pi}' \in T) - P(\underset{\sim}{\pi} \in T) \geq F(t)$ where $F(t)$ is given by

$$\begin{aligned}&2\Gamma(2k)/[\Gamma(k)\Gamma(k)] \int_t^1 x^{k-1}(1 - x)^{k-1}dx - \\ &3\Gamma(2k')/[\Gamma(2k')\Gamma(k')] \int_t^1 x^{k'-1}(1 - x)^{2k'-1}dx.\end{aligned}$$

The strategy is to show that (i) $F(1/2) > 0$, (ii) $F(1) = 0$, (iii) $F'(t)$ has at most one zero for $1/2 < t < 1$, and (iv) $F'(t) < 0$ for t close to 1. It then follows that $F(t) \geq 0$ for $1/2 \leq t \leq 1$. Now (i) and (ii) are obvious. Since $k = 2k'$, we have

$$\begin{aligned}F'(t) &= 3\Gamma(3k')/[\Gamma(2k')\Gamma(k')]\, t^{k'-1}(1 - t)^{2k'-1} - \\ &\qquad 2\Gamma(2k)/[\Gamma(k)\Gamma(k)]t^{k-1}(1 - t)^{k-1} \\ &= t^{k'-1}(1 - t)^{2k'-1}/\Gamma(2k')[3\Gamma(3k')/\Gamma(k') - 2t^{k'}\Gamma(4k')/\Gamma(2k')].\end{aligned}$$

The expression in brackets has at most one zero so that (iii) has been

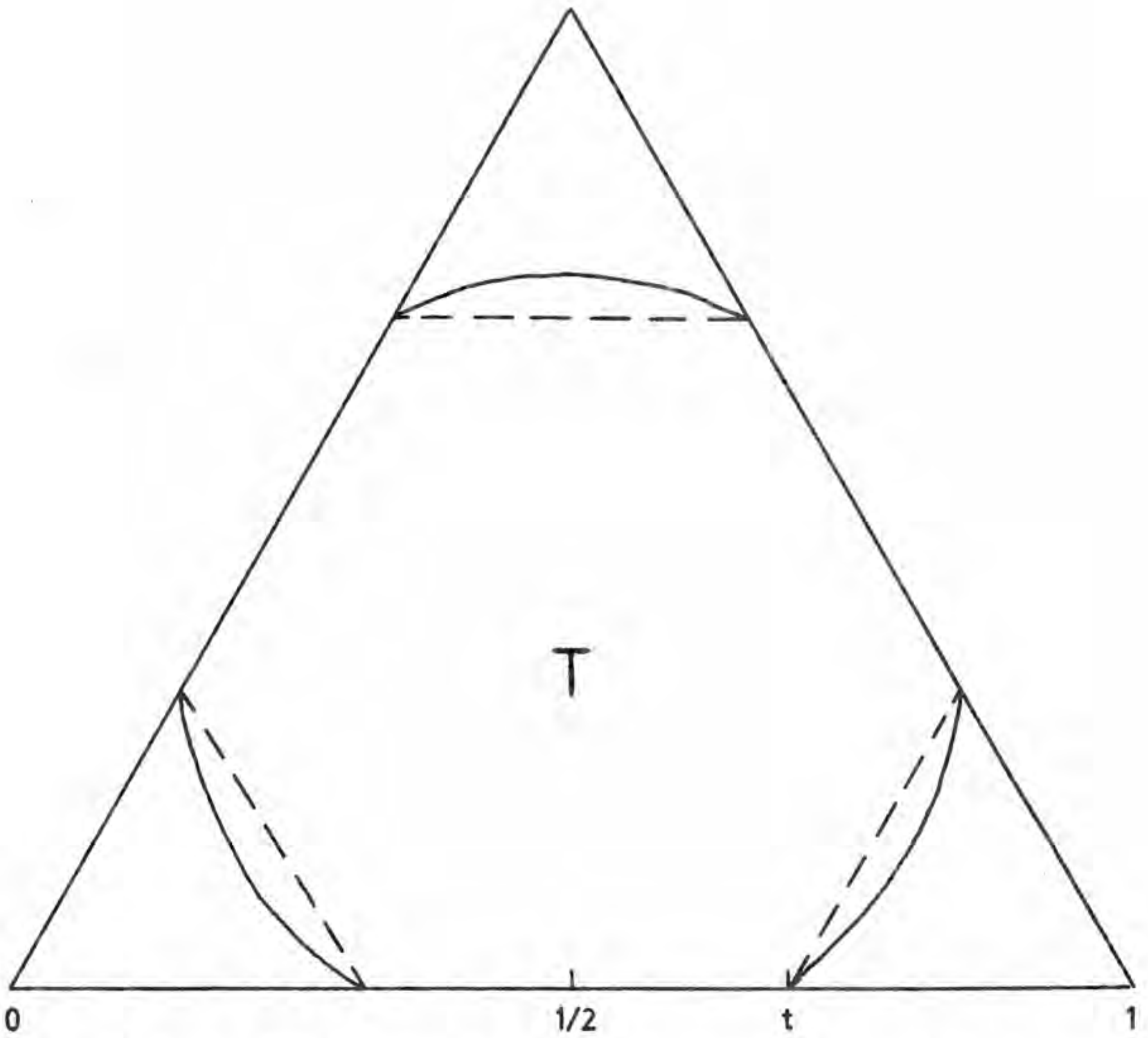


Figure 2.5 A Schur Convex Subset T of the Two-Dimensional Simplex

established. For (iv), it will suffice to show that $3\Gamma(3k')/\Gamma(k') - 2\Gamma(4k')/\Gamma(2k') < 0$ or, equivalently, that $3\Gamma(2k')\Gamma(3k') < 2\Gamma(k')\Gamma(2k')$. Multiplying both sides by $2k'k'$, this is the same as $\Gamma(2k' + 1)\Gamma(3k' + 1) < \Gamma(k' + 1)\Gamma(4k' + 1)$ which is true by Lemma 2.1.

Remark 2.11: If the conditions given in Theorem 2.18(a) are always sufficient, then SD2, SD3, and SD4 are equivalent for the Dirichlet model. This is so because the proof of necessity used only the index Δ_β which satisfies Criterion C3 and so these conditions are necessary for SD3. By arguments similar to the proof of Theorem 2.19, it may be shown that they are also necessary for SD4. On the other hand, SD5 is not equivalent to any of SD2, SD3, or SD4 for the Dirichlet model. In any event, the two sets of conditions given in Theorem 2.18 are numerically indistinguishable for realistic parameter values (s large and k small).

Corollary 2.3: Fisher's α is a complete diversity parameter for the Dirichlet model when either s or k is held fixed.

Proof: The two sets of conditions of Theorem 2.18 coincide in either of these cases.

If $X_1, X_2, X_3, \ldots, X_n$ is a random sample from a nonnegative distribution, it is easy to show that $E[\mathrm{Max}(X_1, X_2, \ldots, X_n)]/n$ is a decreasing function of the sample size. The next corollary generalizes this for the Gamma distribution.

Corollary 2.4: Let $X_1^* \geq X_2^* \geq \ldots \geq X_n^*$ and $Y_1^* \geq Y_2^* \geq \ldots \geq Y_{n'}^*$, be the order statistics from standard Gamma distributions with index k and k', respectively. Assume that $n' \geq n$ and $n'k' \geq nk$. Then one has $E[X_1^* + X_2^* + \ldots + X_m^*]/(nk) \geq E[Y_1^* + Y_2^* + \ldots + Y_m^*]/(n'k')$, $m = 1,2,\ldots,n$.

Proof: This follows from Theorem 2.18 since SD2 => SD5.

We conclude this section by showing that Fisher's α is a complete diversity parameter for the limiting Dirichlet, at least in the sense SD3. The proof uses the notion of random deletion of species as discussed in Example 2.13. The following theorem, which contains no new information, is included to clarify the proof.

Theorem 2.20: Let $\underset{\sim}{\nu} \sim D(3, k)$ and $\underset{\sim}{\pi} \sim D(2, k)$ (same k). Then $\underset{\sim}{\nu}$ is stochastically more diverse than $\underset{\sim}{\pi}$ in the sense SD3.

Proof: Obtain $\underset{\sim}{\nu}$ as $(\lambda_1, \lambda_2, \lambda_3)/\lambda$ where $\lambda = \lambda_1 + \lambda_2 + \lambda_3$ and the λ_i are independent realizations of the standard Gamma distribution with index k. Consider the following collection of communities and their associated weights:

weight	community
$w_1 = (\lambda_2 + \lambda_3)/(2\lambda)$	$\underset{\sim}{\nu}^{(1)} = (0, \lambda_2, \lambda_3)/(\lambda_2 + \lambda_3)$
$w_2 = (\lambda_1 + \lambda_3)/(2\lambda)$	$\underset{\sim}{\nu}^{(2)} = (\lambda_1, 0, \lambda_3)/(\lambda_1 + \lambda_3)$
$w_3 = (\lambda_1 + \lambda_2)/(2\lambda)$	$\underset{\sim}{\nu}^{(3)} = (\lambda_1, \lambda_2, 0)/(\lambda_1 + \lambda_2)$

Now, it is easy to check that (i) $w_1 + w_2 + w_3 = 1$, (ii) the mean $\underset{\sim}{\nu}^{(i)}$ is $w_1 \underset{\sim}{\nu}^{(1)} + w_2 \underset{\sim}{\nu}^{(2)} + w_3 \underset{\sim}{\nu}^{(3)} = \underset{\sim}{\nu}$, (iii) w_i is independent of $\underset{\sim}{\nu}^{(i)}$, and (iv) after omitting the zero components, each $\underset{\sim}{\nu}^{(i)}$ equals $\underset{\sim}{\pi}$ in distribution. Let Δ satisfy Criterion C3. By (i) and (ii), we find that $\Delta(\underset{\sim}{\nu}) \geq \Sigma\, w_i \Delta(\underset{\sim}{\nu}^{(i)})$ and taking expectations of both sides gives

$$E[\Delta(\underset{\sim}{\nu})] \geq \Sigma E[w_i \Delta(\underset{\sim}{\nu}^{(i)})]. \tag{2.44}$$

But, by (iii) and (iv), $E[w_i\ \Delta(\underset{\sim}{\nu}^{(i)})] = E[w_i]\ E[\Delta(\underset{\sim}{\nu}^{(i)})] = E[w_i]E[\Delta(\underset{\sim}{\pi})]$. Inserting this into (2.44) shows that $E[\Delta(\underset{\sim}{\nu})] \geq E[\Delta(\underset{\sim}{\pi})]$, as desired.

Theorem 2.21: Let $\underset{\sim}{\nu}$ and $\underset{\sim}{\pi}$ be limiting Dirichlet models with parameters α and α_0, respectively. Then $\underset{\sim}{\nu}$ is stochastically more diverse than $\underset{\sim}{\pi}$ in the sense SD3 $\Leftrightarrow \alpha \geq \alpha_0$.

Proof: ($\Rightarrow$). We must have $E[\Delta_1(\underset{\sim}{\nu})] \geq E[\Delta_1(\underset{\sim}{\pi})]$ or, by Remark 2.7, $\alpha/(\alpha + 1) \geq \alpha_0/(\alpha_0 + 1)$. This implies that $\alpha \geq \alpha_0$. ($\Leftarrow$). Without loss of generality, assume that $\alpha > \alpha_0$ and write $\alpha_0 = p\alpha$ where $0 < p < 1$. Now $\underset{\sim}{\pi}$ may be obtained from $\underset{\sim}{\nu}$ by random deletion of species with zeros omitted, as in Theorem 2.11. Let $\underset{\sim}{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots)$ be the selection variables (see Example 2.13). The sample space Ω of $\underset{\sim}{\varepsilon}$ consists of all infinite strings of zeros and ones. Let $d\mu(\underset{\sim}{\varepsilon})$ be the probability measure on this space which specifies the distribution of $\underset{\sim}{\varepsilon}$. For a given realization of $\underset{\sim}{\nu}$, let $\underset{\sim}{\nu} \times \underset{\sim}{\varepsilon}$ be the vector whose ith component is $\nu_i\varepsilon_i$ and put $S(\underset{\sim}{\nu}, \underset{\sim}{\varepsilon}) = \Sigma\, \nu_i\varepsilon_i$. Then $\underset{\sim}{\nu} \times \underset{\sim}{\varepsilon}/S(\underset{\sim}{\nu}, \underset{\sim}{\varepsilon})$ is the same as $\underset{\sim}{\pi}$ but without the zeros omitted. Since $\int_\Omega \varepsilon_i d\mu(\underset{\sim}{\varepsilon}) = p$, one easily checks that, for a given realization of $\underset{\sim}{\nu}$, $S(\underset{\sim}{\nu}, \underset{\sim}{\varepsilon})d\mu(\underset{\sim}{\varepsilon})/p$ is a probability measure on Ω and

$$\underset{\sim}{\nu} = \int_\Omega (\underset{\sim}{\nu} \times \underset{\sim}{\varepsilon}/S(\underset{\sim}{\nu}, \underset{\sim}{\varepsilon}))\ S(\underset{\sim}{\nu}, \underset{\sim}{\varepsilon})d\mu(\underset{\sim}{\varepsilon})/p. \tag{2.45}$$

Now let Δ satisfy Criterion C3. By Jensen's inequality, (2.45) gives

$$\begin{aligned}\Delta(\underset{\sim}{\nu}) &\geq \int_\Omega \Delta(\underset{\sim}{\nu} \times \underset{\sim}{\varepsilon}/S(\underset{\sim}{\nu}, \underset{\sim}{\varepsilon}))\ S(\underset{\sim}{\nu}, \underset{\sim}{\varepsilon})d\mu(\underset{\sim}{\varepsilon})/p \\ &= E[\Delta(\underset{\sim}{\nu} \times \underset{\sim}{\varepsilon}/S(\underset{\sim}{\nu}, \underset{\sim}{\varepsilon}))\ S(\underset{\sim}{\nu}, \underset{\sim}{\varepsilon})|\underset{\sim}{\nu}]/p.\end{aligned}$$

Take expectations with respect to $\underset{\sim}{\nu}$ to find that

$$E[\Delta(\underset{\sim}{\nu})] \geq E[\Delta(\underset{\sim}{\nu} \times \underset{\sim}{\varepsilon}/S(\underset{\sim}{\nu}, \underset{\sim}{\varepsilon}))S(\underset{\sim}{\nu}, \underset{\sim}{\varepsilon})]/p.$$

But $\Delta(\underset{\sim}{\nu} \times \underset{\sim}{\varepsilon}/S\ (\underset{\sim}{\nu}, \underset{\sim}{\varepsilon})) = \Delta(\underset{\sim}{\pi})$ since omitting the zero components does not change the value of Δ. By Theorem 2.11, $\underset{\sim}{\pi}$ is independent of $S(\underset{\sim}{\nu}, \underset{\sim}{\varepsilon})$ so that $E[\Delta(\underset{\sim}{\nu})] \geq E[\Delta(\underset{\sim}{\pi})]\ E[S(\underset{\sim}{\nu}, \underset{\sim}{\varepsilon})]/p = E[\Delta(\underset{\sim}{\pi})]$.

<u>Open Problem:</u> Is Theorem 2.21 true in the sense SD2?

CHAPTER 3

SIZE-BIASED SAMPLING AND THE EXPONENTIAL INTEGRAL

3.1 Introduction

In developing species abundance models, it is often assumed that the unnormalized abundances $X_1, X_2, \ldots, X_s$ of the various species are independent realizations of some nonnegative random variable X with density f(x), say. In Chapter 2, for example, X was taken to have a Gamma distribution.

Now X is the abundance of a randomly chosen species. But in practice sampling is by individuals and we will be interested in the individuals abundance distribution,[1] i.e., the distribution of the abundance of the species to which a randomly chosen individual belongs. Sampling by individuals has the effect of shifting the probability density f(x) to the right since the more abundant species now have a greater propensity to be observed. In fact, the probability of choosing a given species is proportional to the abundance of that species, which has led several authors [11, 98, 99, 100] to adopt the first moment distribution $x\ f(x)/E[X]$ as the individuals abundance distribution. This choice cannot be strictly correct since the first moment distribution does not involve the parameter s. The exact distribution is derived in Section 3.2 and is shown to converge to the first moment distribution when $s \to \infty$. The techniques used in the proof are then extended in Section 3.3 to cover Rao's [101] weighted distributions.

---

[1] Preston [98, 99] uses the term "individuals curve."

The exact individuals distribution is quite complicated although, in Section 3.4, we are able to express it in terms of the exponential integral function when X has a Gamma distribution. Even so, one will usually have to adopt the first moment distribution as an approximation and the rate of convergence becomes of interest. This question is considered in Section 3.5 where it is shown that the rate of convergence is of order 1/s, at least for sampling from the exponential distribution.

3.2 Convergence to the First Moment Distribution

Let X be a nonnegative random variable having (right continuous) distribution function F. It is not necessarily assumed that $F(0) = 0$. In this section, we define a sequence $X^{\#s}$, $s = 1,2,3,\ldots$, of size-biased versions of X, derive a general expression for the distribution function $F^{\#s}$ of $X^{\#s}$, and obtain the first moment distribution of F as the limit when $s \to \infty$.

Take a random sample $X_1, X_2, X_3, \ldots$ from the distribution F and regard these variates as the successive gaps in a renewal process. For a fixed positive integer s, randomly select a point from the union of the first s gaps and record the length of the gap that covers this point as $X^{\#s}$. Note that $X^{\#s}$ assumes the value zero when and (with probability one) only when the first s gaps are all of zero length. The distribution function of $X^{\#s}$ will be written as $F^{\#s}$. Clearly $F^{\#s} = F$ when $s = 1$.

<u>Theorem 3.1</u>: The distribution of $X^{\#s}$ is given by

$$F^{\#s} = \begin{cases} 0 & \text{if } x < 0 \\ [F(0)]^s & \text{if } x = 0 \\ [F(0)]^s + s\int_{0+}^{x} uE[1/(u + S_{s-1})]dF(u) & \text{if } x > 0, \end{cases}$$

where $S_s = X_1 + X_2 + \ldots + X_s$ and $S_0 = 0$.

Proof: Let $x \geq 0$ and define $h(u)$ to be u for $u \leq x$ and zero otherwise. Conditional on the realized values of $X_1, X_2, \ldots, X_s$, the probability of the event $X^{\#s} \leq x$ is $[\Sigma h(X_j)]/S_s$ where the sum ranges from $j = 1$ to $j = s$ and where 0/0 is taken to be one. Consistent with the 0/0 convention, this conditional probability may be written as

$$[sh(X_1)/S_s + sh(X_2)/S_s + \ldots + sh(X_s)/S_s]/s.$$

Since the terms $sh(X_j)/S_s$ are identically distributed, we obtain

$$\begin{aligned} P(X^{\#s} \leq x) &= E[sh(X_s)/S_s] \\ &= \int_0^\infty E[sh(u)/(u + S_{s-1}) \mid X_s = u] \, dF(u). \end{aligned}$$

By the independence of S_{s-1} and X_s, the condition $X_s = u$ may be dropped from the expectation. The proof is completed by decomposing the integral as

$$\int_{0-}^{\infty} = \int_{\{0\}} + \int_{0+}^{\infty} ,$$

and noting that the integral over $\{0\}$ is $[F(0)]^s$.

Corollary 3.1: Assume that X is absolutely continuous with density $f(x)$. Then $X^{\#s}$ is absolutely continuous with density $xf(x)E[s/(x + S_{s-1})]$.

When s is large, $F(0)^s = F^{\#s}(0)$ is approximately zero unless $F(0) = 1$. In effect, any probability at the origin disappears in the limit. Also for large s, $s/(x + S_{s-1})$ is approximately $1/E[X]$ by the law of large numbers. Corollary 2.1 thus suggests that $F^{\#s}$ converges to the first moment distribution of F. The next theorem makes this rigorous.

Theorem 3.2: Assume that $\mu = E[X]$ is finite and nonzero. Then $F^{\#s}$ converges in distribution to $F^{\#}$ where $dF^{\#}(x) = \mu^{-1}x\,dF(x)$ is the first moment distribution of F.

Proof: Conditional on the realized values of $X_1, X_2, \ldots, X_s$, the characteristic function of $X^{\#s}$ is $T_s = [\Sigma X_j \exp(itX_j)]/S_s$ where the sum ranges from $j = 1$ to $j = s$ and where $0/0 = 1$. Divide numerator and denominator of T_s by s and apply the strong law of large numbers to find that, with probability one, T_s converges to $\phi(t) = \mu^{-1}E[X \exp(itX)]$ which is the characteristic function of $F^{\#}$. But $|T_s| \leq 1$ so that the dominated convergence theorem gives $E[T_s] \to \phi(t)$. Since $E[T_s]$ is the characteristic function of $X^{\#s}$, the proof is complete.

The above results assume that the renewal process is replicated for each realization of $X^{\#s}$. In practice, it will be more realistic to consider repeated size-biased sampling from a single realization of the renewal process. Suppose that $r \leq s$ such selections are made without replacement from among the first s gaps. Let $\underset{\sim}{X}^{r\#s} = (X_1^{r\#s}, X_2^{r\#s}, \ldots, X_r^{r\#s})$ be the vector of selected gaps and let $Y = S_{s-1} - \sum_{i=1}^{r} X_i^{r\#s}$ be the residual.

Theorem 3.3: Assume that X is absolutely continuous with density f(x) and let f_{s-r} be the $(s - r)$-fold convolution of f. The joint density of $\underset{\sim}{X}^{r\#s}$ and Y is

$$s(s - 1)\ldots(s - r + 1)f_{s-r}(y) \prod_{i=1}^{r} [x_i f(x_i)/(x_i + x_{i+1} + \ldots + x_r + y)]$$

and the marginal density of $\underset{\sim}{X}^{r\#s}$ is

$$s(s - 1)\ldots(s - r + 1)\cdot \prod_{i=1}^{r} [x_i f(x_i)]\cdot E[1/ \prod_{i=1}^{r} (x_i + x_{i+1} + \ldots + x_r + S_{s-r})].$$

Proof: Let $H_x(u)$ be the indicator function of the interval $(-\infty, x]$. For notational simplicity, take $r = 2$ and let $0 \le x_1, x_2, y$. As in the proof of Theorem 2.1, $P(X_1^{2\#s} \le x_1, X_2^{2\#s} \le x_2, Y \le y)$ is given by

$$E[\sum_{\substack{i,j=1 \\ i\neq j}}^{r} (X_i/S_s)(X_j/(S_s - X_i))H_{x_1}(X_i)H_{x_2}(X_j)H_y(S_s - X_i - X_j)] =$$

$$s(s - 1)E[(X_1/S_s)(X_2/(S_s - X_1))H_{x_1}(X_1)H_{x_2}(X_2)H_y(S_s - S_1 - S_2)] =$$

$$s(s - 1) \int_{u_1=0}^{x_1} \int_{u_2=0}^{x_2} \int_{v=0}^{y} [u_1/(u_1 + u_2 + v)][u_2/(u_2 + v)] \times$$

$$f(u_1)f(u_2)f_{s-2}(y)du_1du_2dv.$$

Differentiating with respect to x_1, x_2, and y now gives the result.

When $r = 1$, $X^{\#s}$ and $Y = S_s - X^{\#s}$ have joint density $sx_1f(x_1)f_{s-1}(y)/(x_1 + y)$. Since the convolution of this joint density must give the density of S_s, we obtain the functional equation

$$zf_s(z) = \int_{x=0}^{z} sxf(x)f_{s-1}(z-x)dx. \qquad (3.1)$$

After normalizing, (3.1) states that the first moment distribution of the s-fold convolution of X is obtained by convolving the first moment distribution of X with the (s - 1)-fold convolution of X.

Theorem 3.4: Let r be a fixed positive integer. In the limit as $s \to \infty$, $\underset{\sim}{X}^{r\#s}$ is distributed as a random sample of size r from the first moment distribution of X provided $0 < E[X] < \infty$.

Proof: Similar to the proof of Theorem 3.2.

In the same way, one can show that Theorem 3.4 holds for sampling with replacement. This has been stated without proof by Patil & Rao [102].

3.3 Weighted Distributions

As a generalization of the first moment distribution, Rao [101] has introduced the concept of weighted distributions. These have density $f^w(x) = w(x)f(x)/\int w(x)f(x)dx$ where $f(x)$ is the original density and $w(x)$ is a nonnegative weight function. Distributions of this form occur in many applications (see Patil & Rao [103] for a survey). Here we show how weighted distributions fall into the framework of Section 3.2.

Suppose we have a pair $(X, \underset{\sim}{Z})$ where X is a nonnegative random variable called the concomitant variable and where $\underset{\sim}{Z}$ is a random vector called the recorded variable. The components of $\underset{\sim}{Z}$ do not have to be nonnegative and one of these components may be X. An important special case occurs when $X = w(\underset{\sim}{Z})$ is a nonnegative function of $\underset{\sim}{Z}$. We assume that $(X, \underset{\sim}{Z})$ has some naturally occurring joint distribution, but the method of sampling is such that sampling units with larger values of X are more likely to be observed, in fact with probability proportional to X.

As in Section 3.2, let $(X_1, \underset{\sim}{Z}_1)$, $(X_2, \underset{\sim}{Z}_2)$, $(X_3, \underset{\sim}{Z}_3)$,... be independent realizations of $(X, \underset{\sim}{Z})$ and regard $X_1, X_2, X_3, \ldots$ as the successive gaps in a renewal process. Randomly select a point from the union of the first s gaps and let j be the index of the covering gap. Record $\underset{\sim}{Z}_j$ as $\underset{\sim}{Z}^{\#s}$. Let $F(\underset{\sim}{z})$ and $F^{\#s}(\underset{\sim}{z})$ be the distribution functions of $\underset{\sim}{Z}$ and $\underset{\sim}{Z}^{\#s}$, respectively.

<u>Theorem 3.5</u>: Assume $\mu = E[X]$ is finite and nonzero. Then $F^{\#s}$ converges in distribution to the weighted distribution F^w given by $dF^w(\underset{\sim}{z}) = \mu^{-1}w(\underset{\sim}{z})dF(\underset{\sim}{z})$ where $w(\underset{\sim}{z}) = E[X|\underset{\sim}{Z} = \underset{\sim}{z}]$.

<u>Proof</u>: Conditional on $(X_1, \underset{\sim}{Z}_1)$, $(X_2, \underset{\sim}{Z}_2)$,...,$(X_s, \underset{\sim}{Z}_s)$, the characteristic function of $\underset{\sim}{Z}^{\#s}$ is $T_s = [\Sigma X_j \exp(i\underset{\sim}{t}\,\underset{\sim}{Z}_j)]/[\Sigma X_j]$ where

both sums range from $j = 1$ to $j = s$ and where $0/0 = 1$. The rest of the proof is the same as that of Theorem 3.2.

Theorem 3.5 can also be extended to repeated sampling, with or without replacement, as in Theorem 3.4.

3.4 The Exponential Integral Distribution

Closed form expressions for the factor $E[s/(x + S_{s-1})]$ in Corollary 3.1 will be difficult to obtain unless the distribution of X behaves well under convolutions. When X has a Gamma distribution, we are able to express this factor in terms of the exponential integral functions $E_n(x)$, $x > 0$, which are defined by

$$E_n(x) = \int_1^\infty \exp(-xt)/t^n \, dt, \; n \geq 0. \tag{3.2}$$

Alternatively,

$$E_n(x) = \begin{cases} \exp(-x)/x & \text{if } n = 0 \\ \int_x^\infty (t-x)^{n-1} \exp(-t)/t \, dt/\Gamma(n) & \text{if } n > 0. \end{cases} \tag{3.3}$$

Properties of $E_n(x)$ for integral values of n can be found in Abramowitz & Stegun [89].

Let $k, n \geq 0$ with $k + n > 0$. From (3.2) it follows that

$$(k + n)x^k E_n(x)/\Gamma(k + 1), \; x > 0, \tag{3.4}$$

is a probability density function. The corresponding distribution will be referred to as the exponential integral distribution with index k and order n (EID(k, n)). This distribution results when the Gamma distribution with index $k + 1$ and mean $(k + 1)/a$ is randomized by letting parameter a have a standard Pareto distribution with index $k + n$.

Theorem 3.6: a) $EID(k, 0)$ is the same as the standard Gamma distribution with index k. b) As $n \to \infty$, $EID(k, n)$ converges to the standard Gamma distribution with index $k + 1$. c) $EID(0, n)$ has density $nE_n(x)$, $x > 0$, and converges to the standard exponential distribution when $n \to \infty$.

Proof: Use the fact that $\exp(-x)/(x + n) \leq E_n(x) \leq \exp(-x)/(x + n - 1)$ for $n \geq 1$ (Abramowitz & Stegun [89, equation 5.1.19]).

Theorem 3.7: Let X have a standard Gamma distribution with index $k > 0$ and let $X^{\#s}$ be the sth order size-biased version of X as defined in Section 3.2. Then $X^{\#s} \sim EID(k, (s - 1)k)$. Note: there is no loss of generality in assuming that X has scale parameter equal to one since $(aX)^{\#s} = a(X^{\#s})$, $a > 0$.

Proof: Let $f(x) = x^{k-1}\exp(-x)/\Gamma(k)$ be the Gamma density with index k. Since S_{s-1} has a Gamma distribution with index $(s - 1)k$, Corollary 3.1 may be applied to show that $X^{\#s}$ has density given by

$$sxf(x)E[1/(x + S_{s-1})] = sxf(x)\int_0^\infty (x + u)^{-1}u^{(s-1)k-1}e^{-u}du/\Gamma((s - 1)k).$$

After making the change of variable $t = x + u$ and using (3.3), this reduces to

$$sxf(x)\exp(x)\int_x^\infty (t - x)^{(s-1)k-1}\exp(-t)/t\,dt/\Gamma((s - 1)k) =$$

$$sxf(x)\exp(x)E_{(s-1)k}(x) = sx^kE_{(s-1)k}(x)/\Gamma(k).$$

Multiplying and dividing by k gives

$$(sk)x^kE_{(s-1)k}(x)/\Gamma(k + 1), \tag{3.5}$$

which is the same as (3.4) with $n = (s - 1)k$.

If we apply Fisher's limit ($s \to \infty$, $k \to 0$, $sk = \alpha$) to (3.5), the limiting density of $X^{\#s}$ becomes $\alpha E_{\alpha}(x)$. In Chapter 4, it will be shown that this limiting density can be interpreted in terms of size-biased sampling even though the Gamma density $f(x) = x^{-1}\exp(-x)$ cannot be normalized to unity when $k = 0$. Instead of considering samples of size s as in Section 3.2, the "sample" size will be Poisson random variable with mean α. By Theorem 3.6(c), when $\alpha \to \infty$ the density $\alpha E_{\alpha}(x)$ converges to $\exp(-x)$ which is the first moment distribution of $f(x)$. This is the formal analogue of Theorem 3.2.

Theorem 3.3 can also be applied to obtain the joint density of $\underset{\sim}{X}^{r\#s}$. This density is very complicated and will not be given. However, it should be noted that $\underset{\sim}{X}^{s\#s}/S_s$ is equal in distribution to the size-biased permutation of the Dirichlet community with s components and with index k.

3.5 Rate of Convergence

In most cases, the sth order size-biased version of F is intractable and one will use the first moment distribution as an approximation. It then becomes important to determine the rate at which $F^{\#s}$ converges to $F^{\#}$. Under suitable regularity assumptions, the rate of convergence appears to be of order $1/s$.[2] This may be compared with the central limit theorem where the rate of convergence is only of order $1/\sqrt{s}$ (see Feller [82, p. 533]).

---

[2] The author has "proved" this statement by formal manipulations but has not identified the regularity assumptions needed to justify the manipulations.

For concreteness, only size-biased sampling from the exponential distribution is considered here.

Let X have a standard exponential distribution with density $f(x) = \exp(-x)$. By (3.5), $X^{\#s}$ has density $f^{\#s}(x) = sxE_{s-1}(x)$ and converges to the density $f^{\#\infty}(x) = f^{\#}(x) = x \exp(-x)$ when $s \to \infty$. The density $f^{\#s}$ is plotted in Figure 3.1 for $s = 1,2,3,11,\infty$. Examination of these plots suggests that $X^{\#s}$ is stochastically increasing in s.

Theorem 3.8: The distribution $F^{\#s+1}$ has monotone likelihood ratio with respect to $F^{\#s}$ so that $X^{\#s+1}$ is stochastically greater than $X^{\#s}$ and $F^{\#1}(x) \geq F^{\#2}(x) \geq F^{\#3}(x) \geq \ldots \geq F^{\#}(x)$. Also, $F^{\#}$ has monotone likelihood ratio with respect to $F^{\#s}$.

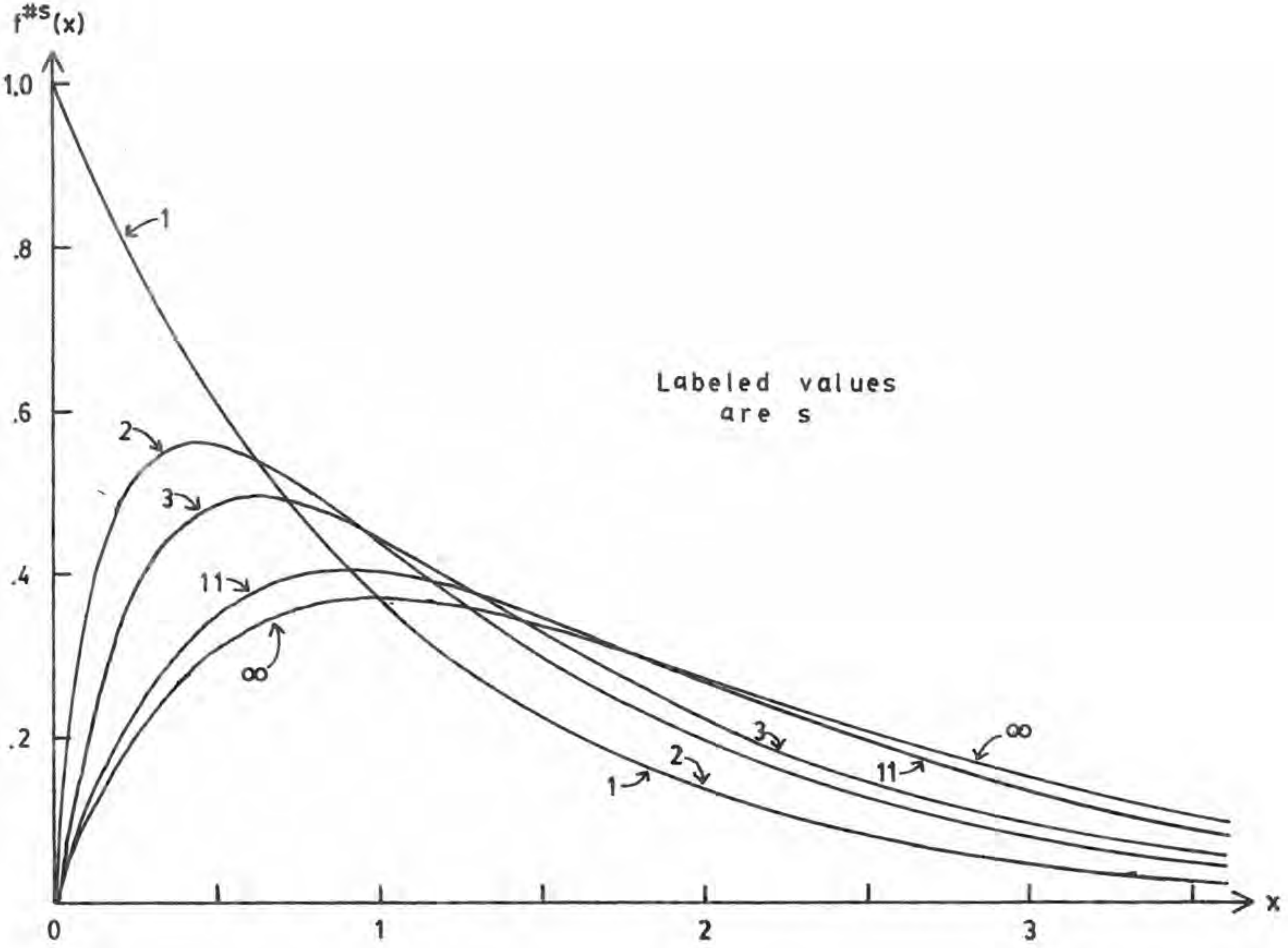
Proof: The ratio $f^{\#s+1}(x)/f^{\#s}(x)$ equals $(s + 1)E_s(x)/[sE_{s-1}(x)]$. But the derivative of $E_s(x)/E_{s-1}(x)$ is positive by 5.1.21 of Abramowitz & Stegun [89]. The ratio $r(x) = f^{\#s}(x)/f^{\#}(x)$ equals $sE_{s-1}(x)\exp(x)$. Since $dE_{s-1}(x)/dx = - E_{s-2}(x)$, we obtain $r'(x) = s \exp(x)[E_{s-1}(x) - E_{s-2}(x)]$ and this is negative by 5.1.17 of Abramowitz & Stegun [89].

From Theorem 3.8, it follows that $F^{\#s}(x) - F^{\#}(x) \geq 0$. An explicit expression for this difference will now be derived. Since $X^{\#s}$ has density $sxE_{s-1}(x)$, we obtain

$$
\begin{aligned}
1 - F^{\#s}(x) &= \int_x^\infty suE_{s-1}(u)du \\
&= \int_{u=x}^\infty su \int_{t=1}^\infty \exp(-tu)/t^{s-1}\, dt\, du \\
&= \int_{u=x}^\infty \int_{v=u}^\infty s\, u^{s-1}\exp(-v)/v^{s-1}\, dv\, du,
\end{aligned}
$$

where the change of variable $v = ut$ has been made in the last step.

Figure 3.1 Probability Density Function $f^{\#s}(x)$ for Size-Biased Sampling of Order s from the Exponential Distribution ($s = 1, 2, 3, 11, \infty$)


$f^{\#s}(x)$
1.0
.8
.6
.4
.2
0
1
2
3
x
Labeled values are s
1
2
3
11
∞
1
2
3
11
∞

Changing the order of integration gives

$$1 - F^{\#s}(x) = \int_{v=x}^{\infty} \int_{u=x}^{v} s\ u^{s-1} \exp(-v)/v^{s-1} du\ dv$$

$$= \int_{x}^{\infty} [\exp(-v)/v^{s-1}][v^s - x^s]\ dv$$

$$= \int_{x}^{\infty} v\ \exp(-v)dv - \int_{x}^{\infty} x^s \exp(-v)/v^{s-1} dv$$

$$= 1 - F^{\#}(x) - x^2 \int_{1}^{\infty} \exp(-xt)/t^{s-1}\ dt,$$

where the change of variables $v = xt$ has been made in the last step. Since $\int_{1}^{\infty} \exp(-xt)/t^{s-k} dt = E_{s-1}(x)$, we obtain

$$F^{\#s}(x) - F^{\#}(x) = x^2 E_{s-1}(x). \tag{3.6}$$

It will now be proved that the rate of convergence of $F^{\#s}$ to the first moment distribution has order $1/s$.

<u>Theorem 3.9</u>: Let $x_0 = s_0(s)$ be the solution of the equation $\exp(-x_0) = sE_{s-1}(x_0$. Then $1 \leq x_0 \leq 2$ and

$$\sup_{s>0} \{F^{\#s}(x) - F^{\#}(x)\} = x_0^2 \exp(-x_0)/s \leq 4 \exp(-2)/s = .541/s.$$

<u>Proof</u>: The bounds $1 \leq x_0 \leq 2$ are easily established using 5.1.19 of Abramowitz & Stegun [89]. Now $F^{\#}$ has monotone likelihood ratio with respect to $F^{\#s}(x)$ so that the maximum difference $F^{\#s}(x) - F^{\#}(x)$ occurs when $f^{\#s}(x) = f^{\#}(x)$, i.e., when $sE_{s-1}(x) = \exp(-x)$. The maximum difference itself is obtained by substituting $E_{s-1}(x_0 = \exp(-x_0)/s$ into (3.6). Finally $x^2\exp(-x)$ is increasing for $1 \leq x \leq 2$ so that $\exp(-1)/s \leq x_0^2\exp(-x_0)/s \leq 4 \exp(-2)/s$ which proves that the maximum difference is of order $1/s$.

Remark 3.1: Using 5.1.22 of Abramowitz & Stegun [89], it can be shown that x_0 is an increasing function of s and $x_0 \rightarrow 2$ as $s \rightarrow \infty$. Further $2 - 2/(s - 2) \leq x_0 \leq 2 - 2/(s + 1)$. Some values of $\sup\{F^{\#s}(x) - F^{\#}(x)\}$, $4 \exp(-2)/s$, and x_0 are given in Table 3.1.

Table 3.1: Rate of Convergence for Size-Biased Sampling from the Exponential Distribution

s	$\sup\{F^{\#s}(x) - F^{\#}(x)\}$	$4 \exp(-2)/s$	x_0
1	.368	.541	1.00
2	.229	.271	1.29
3	.165	.180	1.45
4	.128	.135	1.56
5	.104	.108	1.63
8	.067	.068	1.76
10	.054	.054	1.80
20	.027	.027	1.90
50	.011	.011	1.96
100	.005	.005	1.98

CHAPTER 4

ABUNDANCE MODELS AND SUBORDINATORS

4.1 Introduction

The classical approach to abundance models considers the unnormalized abundances $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_s$ to be iid random variables. In relation to the sample size, the number of species in the community is often so large that efficient and robust estimation of s becomes impossible and one is led to reduce the number of parameters by letting s go to infinity. But can the abundances still be treated as iid? The answer is no because the number of individuals in the sample is Poisson with parameter proportional to the total abundance $\Sigma\lambda_i$. When the abundances are iid, this sum becomes infinite with probability one, thus implying that the sample contains infinitely many individuals. Since this is clearly impossible, models must be developed for which the infinite series $\Sigma\lambda_i$ converges.

Fisher's limiting scheme provides a clue as to the nature of such models. Recall that the index k of the Gamma distribution goes to zero simultaneously as s goes to infinity and, ignoring normalizing constants, the Gamma density converges to $f(u) = \exp(-u)/u$. Near the origin, this function becomes so large that it cannot be normalized to unity. Heuristically, an infinite random sample from f(u) contains such a preponderance of small values that the sample sum converges.

Given an improper density such as f(u), there exists a stochastic process, known as a subordinator, whose sample paths are increasing and increase only in jumps. Moreover, a countable infinity of such jumps occur in any time interval. The sample paths are difficult to visualize and impossible to draw, but a subordinator can be thought of intuitively as a compound Poisson process in which the intensity of jumps goes to infinity while at the same time the jump distribution converges to an improper density like f(u). The various jumps λ_i occurring during a fixed time interval $0 \leq t \leq \alpha$ will be taken as the species abundances.

Three particular subordinators, called the Hyperbolic process, Gamma process, and Beta process are introduced in Section 4.2. Each of the first two processes is a stochastic analogue of the deterministic geometric series. It is not widely recognized that the geometric series can be generated by two different mechanisms. The preemption mechanism successively assigns to each species a fixed fraction of the remaining available abundance. Under the dominance mechanism, each species receives a fixed fraction of what has been assigned to the preceding species. Both mechanisms produce a geometric series when the fractions are constant numbers, but this is not so when the fractions are iid random variables. Stochastic dominance gives rise to the Hyperbolic process while stochastic preemption results in the Gamma process.

It was a failure to recognize that the two mechanisms are distinct at the stochastic level which led Engen [84] to mistakenly propose the geometric series as a fixed version of the limiting Dirichlet. The limiting Dirichlet is associated with the Gamma process but it is the Hyperbolic process that has the geometric series as a fixed version.

Section 4.3 develops a random dispersal model in which the members of each species are initially clustered with some intrinsic abundance at focal points that are Poisson distributed in the plane. The individual organisms then disperse about their focal points in accord with a circular Gaussian distribution having variance σ^2. Depending on the intrinsic abundance distribution and the distribution of σ^2, the effective abundances with which the different species are observed at a fixed sampling site can be realized as the jumps of either the Hyperbolic process, the Gamma process, or the Beta process. Further, Fisher's α has the interpretation of the mean number of focal points, i.e. species, to be found in a standardized unit of area. This reinforces the claim made in Chapter 2 that α is best regarded as a species richness parameter.

Taken individually, the jumps λ_i of a subordinator are not valid random variables. However, both the ranked permutation $\lambda_i^{\#}$ and the size-biased permutation $\lambda_i^{\#}$ of the jumps are random variables, whose joint densities are derived in Section 4.4. As $\alpha \to 0$, the first size-biased component $\lambda_1^{\#}$ is shown to have a limiting distribution whose probability density function is the first moment distribution of f(u). The λ_i^* and the $\lambda_i^{\#}$ are permutations of one another and provide equivalent descriptions of the same underlying community structure. The choice of which to use can be made on the grounds of mathematical convenience. For the Hyperbolic process the ranked permutation is more manageable while the size-biased permutation is best for the Gamma process.

Section 4.4 also describes a standardized method for replacing the random abundances generated by a subordinator with a typical set of fixed abundances. The fixed versions of the Hyperbolic process and the Gamma process are the geometric series model and the exponential integral model. Under a fairly broad set of conditions, the fixed version of a subordinator is asymptotically geometric in the strong sense.

Section 4.5 considers Poisson sampling from a subordinator. The jumps λ_i are replaced with independent observations from the Poisson $(A\lambda_i)$ distributions. The resulting sample process is compound Poisson and both the jump intensity and the jump distribution are derived. A general expression for the species-area curve is also obtained from the sample process. The sample can alternatively be described by the statistics n_x. These are shown to have independent Poisson distributions and general expressions for their expectations are given. Both moment and maximum likelihood estimation are briefly discussed.

Treating the species abundances as random introduces an additional component of variability into the sampling distribution. A method is given for calculating the pure sampling variability of a statistic (the mean conditional variance of the statistic given the realized abundances). The results of this method are compared with the approximate expressions obtained by replacing the random abundances with their fixed versions. An explanation is also given of why Watterson's [90] formal generating function fails to be a valid probability generating function.

When samples are taken at different points in time or in space, it has been observed that the pooled sample may fail to fit the log series distribution even though the individual samples do fit. The random dispersal model is used to explain this effect. It is found that the pooled sample tends to have an excess of singletons when the various sampling sites are widely separated.

4.2 Subordinators

This section is a reference compendium of basic facts about subordinators. The most readable source of information is Kingman [79, 104]. More detailed treatments may be found in Feller [82], Breiman [105], Kallenberg [106], and Grandell [107]. Application of the Gamma process to species abundance models has previously been considered by McCloskey [108] and Shorrock [109].

Let $d\mu(u)$ be a measure concentrated on the positive real axis such that $\int_0^\infty d\mu(u)$ is infinite but $M(A) = \int_0^\infty [1 - \exp(-Au)]d\mu(u)$ is finite for all nonnegative A. We call μ a <u>Levy measure</u> and M(A) the <u>Levy transform</u> of μ. Necessary and sufficient conditions for $M(A)$, $A \geq 0$, to be finite are (i) $\int_\varepsilon^\infty d\mu(u) < \infty$ for all $\varepsilon > 0$ and (simultaneously) (ii) $\int_0^\varepsilon u d\mu(u) < \infty$ for some $\varepsilon > 0$. A sufficient condition is $\int_0^\infty u d\mu(u) < \infty$. It will always be supposed that μ is absolutely continuous with density $f(u)$, i.e., $d\mu(u) = f(u)du$.

Associated with each Levy measure is a stochastic process $\{\xi_t : t \geq 0\}$ having stationary independent increments and whose sample paths are increasing. Moreover, ξ_t increases only in jumps and the times t at which these jumps occur form a random countable dense subset of the positive real axis. The process is often called a subordinator. With A as the generating symbol, ξ_t has Laplace transform and expectation given by:

$$E[\exp(-A\xi_t)] = \exp[-tM(A)], \tag{4.1}$$

$$E[\xi_t] = t \int_0^\infty u d\mu(u) = t \int_0^\infty u f(u)du. \tag{4.2}$$

The right hand side of (4.2) may possibly be infinite.

For $\alpha > 0$, let the jumps occurring during the time interval $0 \leq t \leq \alpha$ be enumerated in some arbitrary manner as $\lambda_1, \lambda_2, \lambda_3, \ldots$. The number α will correspond to Fisher's "alpha" and the λ_i will be the abundances of the various species. Note that $\xi_\alpha = \Sigma\lambda_i$ is the total abundance. Properly speaking, the λ_i should be embellished with an α, but a fixed value for α will always be understood. While it is permissible to regard the entire collection $\{\lambda_i\}$ of abundances as a random entity, the individual members of this collection (λ_1, say) are <u>not</u> random variables. This is because there is no way to establish a correspondence from one realization of the process to a second of the jumps labeled λ_1.[1] For the same reason, species identity cannot be compared across realizations.

For a given realization of $\{\xi_t : 0 \leq t \leq \alpha\}$, let λ_i^* be the jumps arranged in descending order and let $\lambda_i^{\#}$ be the size-biased permutation of the jumps (defined essentially as in Section 2.2 but with proper account being taken of the fact that $\Sigma\lambda_i = \xi_\alpha \neq 1$). It will be seen later that the λ_i^* as well as the $\lambda_i^{\#}$ are legitimate random variables, although they are neither independent nor identically distributed. Associated with $\{\xi_t | 0 \leq t \leq \alpha\}$ are the two random relative abundance vectors $\underset{\sim}{\pi}^* = (\lambda_1^*, \lambda_2^*, \lambda_3^*, \ldots)/\xi_\alpha$ and $\underset{\sim}{\pi}^{\#} = (\lambda_1^{\#}, \lambda_2^{\#}, \lambda_3^{\#}, \ldots)/\xi_\alpha$. These are permutations of one another and are simply different ways of describing the same underlying community structure. Depending on the process, one or the other of these descriptions may be mathematically more convenient.

---

[1] In a compound Poisson process one can take λ_1 to be the first jump. A subordinator has no first jump, however.

The probability density function of ξ_α will be written as $f_\alpha(u)$. In principle, $f_\alpha(u)$ can be obtained by inverting the Laplace transform (4.1) but, in practice, this is seldom feasible. Also define the right tail integral to be $H(u) = \int_u^\infty f(u)du$. This is a decreasing function whose inverse function is denoted H^{-1}. Let $0 < u_1 < u_2$. The number of species whose abundances lie between u_1 and u_2 is then a Poisson random variable with mean

$$\alpha[H(u_1) - H(u_2)] = \alpha \int_{u_1}^{u_2} f(x)dx. \tag{4.3}$$

Further, the Poisson variates corresponding to disjoint intervals are independent.

Three additional facts about subordinators will be needed. The first two are trivial; the third is no doubt well-known but, lacking a reference, a proof is given.

1. Effect of a scale change. Let the Levy density $f(u)$ be replaced with $c^{-1}f(u/c)$, $c > 0$. Then α and t are unchanged but λ_i becomes $c\lambda_i$ and ξ_t becomes $c\xi_t$.
2. Effect of renormalization. Let $f(u)$ be replaced with $cf(u)$. Then the jumps are unchanged but t becomes ct and ξ_t becomes ξ_{ct}.
3. Strong law of large numbers for subordinators. When $t \to \infty$, ξ_t/t converges to $\int uf(u)du$ with probability one.

Proof: For simplicity, suppose that t goes to infinity through integral values n. Since the process has stationary independent increments, ξ_n is the n-fold sum of copies of ξ_1. The result now follows from (4.2) and the strong law of large numbers.

The first two facts permit any convenient scaling and normalization of the Levy density without altering the essential features of the process.

For example, one can either take the Levy density to be $\alpha f(u)$ and observe the process for $0 \leq t \leq 1$ or take the Levy density to be $f(u)$ and observe the process for $0 \leq t \leq \alpha$. The second approach is more convenient for our purposes.

The rest of this chapter focuses on three particular processes which we call the Hyperbolic process, the Gamma process, and the Beta process. Their main features are summarized in Table 4.1.

Example 4.1: The Levy density for the Hyperbolic process is $f(u) = u^{-1}$, $0 < u < 1$. By direct integration, the tail integral is $H(u) = -\log(u)$, $0 < u < 1$. Using (5.1.39) of Abramowitz & Stegun [89], the Levy transform becomes $M(A) = E_1(A) + \log(A) + \gamma$, where $\gamma = .5772\ldots$ is Euler's constant. Explicit expressions for the probability density function of ξ_α are not known but some partial results are obtained in Appendix A.9. It will be seen later that Engen's geometric series model is an appropriate fixed version of this model.

Example 4.2: The well-known Gamma process has $f(u) = u^{-1}\exp(-u)$, $0 < u < \infty$, as the Levy density and $H(u) = E_1(u)$, $0 < u < \infty$, as the tail integral. Using (5.1.32) of Abramowitz & Stegun [89], the Levy transform is $M(A) = \log(1 + A)$. The Gamma process has the pleasant feature that the Laplace transform (4.1) can be inverted to find that ξ_α has density $f_\alpha(u) = u^{\alpha-1} \exp(-u)/\Gamma(\alpha)$, i.e., ξ_α is a standard Gamma random variable with index α. It will be seen later that the Gamma process leads to the limiting Dirichlet after normalizing the jumps to unity.

Both the Hyperbolic process and the Gamma process are stochastic analogues of the deterministic geometric series. Ecologists (e.g., Pielou [33] and May [100]) do not seem to realize that the geometric

Table 4.1: Summary of the Processes[a]

	General	Hyperbolic[b]	Gamma[c]	Beta[c]
Levy density:	$f(u)$	u^{-1}	$u^{-1}\exp(-u)$	$u^{-1}(1+a^{-1}u)^{-q}$
tail integral:	$H(u)$	$-\log(u)$	$E_1(u)$	---
Levy transform:	$M(A)$	$E_1(A)+\log(A)+\gamma$	$\log(1+A)$	---
density of ξ_α:	$f_\alpha(u)$	---	$u^{\alpha-1}\exp(-u)/\Gamma(\alpha)$	---
density of λ_1^*:	$\alpha f(u)\exp(-\alpha H(u))$	$\alpha u^{\alpha-1}$	$\alpha u^{-1}\exp[-u-\alpha E_1(u)]$	---
density of $\lambda_1^\#$:	$uf(u)E[\alpha/(u+\xi_\alpha)]$	---	$\alpha E_\alpha(u)$	---
mean sampling frequencies $E[n_x]$:	$\alpha\int_0^\infty \frac{(Au)^x e^{-Au}}{x!} \frac{e^{-Au}}{M(A)} f(u)\, du$	$(\alpha/x)\Gamma(x;A)$	$(\alpha/x)[A/(1+A)]^x$	---
fixed version:	$H^{-1}(i/\alpha)$	geometric series model	exponential integral model	---
intrinsic abundances:	not applicable	Degenerate	Exponential	Exponential
dispersion parameter $(2\pi\sigma^2)$:	not applicable	Degenerate	Degenerate	Gamma

[a] A missing entry indicates that no simplification of the general expression is known.

[b] The range of u is $0 < u < 1$. In the Levy transform, γ is Euler's constant.

[c] The range of u is $0 < u < \infty$.

series may be generated by two essentially different mechanisms : preemption and dominance. Preemption supposes that the first species preempts a fraction Q_1 of the available resource, the second a fraction Q_2 of the remainder, etc. In dominance, the first species receives an amount of resource equal to P_1, the second species receives a fraction P_2 as much as the first, the third species receives a fraction P_3 as much as the second, etc. Both preemption and dominance result in a geometric series of species abundances when $P_i = 1 - Q_i$ are all equal to some fixed number. The stochastic analogues are obtained by letting $P_i = 1 - Q_i$ be iid random variables. Here preemption and dominance are quite different. Due to randomness, the successive abundances assigned by the preemption mechanism need not be decreasing. Hence, the expected ranked abundance vector may not be a geometric series. In dominance, however, the successive abundances are decreasing and have geometric expectations.[2]

When $Q_i \sim \text{Beta}_1(1, \alpha)$, the preemption mechanism leads to the size-biased version of the limiting Dirichlet. Hence, the successive absolute abundances may be realized as the size-biased jumps from the Gamma process. Similarly, when $P_i = (1 - Q_i) \sim \text{Beta}_1(\alpha, 1)$, the successive absolute abundances generated by the dominance mechanism can be realized as the ranked jumps from the Hyperbolic process. This is proved in Section 4.4 but can be seen heuristically via a Fisher-like limit. Consider the classical framework of s species having iid

---

[2] It is the absolute abundances that have geometric expectations. The expected relative abundances are difficult to calculate since the total abundance is not independent of the relative abundances.

abundances $\lambda_1, \lambda_2, \ldots, \lambda_s$. Motivated by the dominance mechanism, we would like the expected descending order statistics to form a geometric series, say, $E[\lambda_i^*] = \theta^i$, $i = 1,2,\ldots,s$, $0 < \theta < 1$. Kadane [110] has shown that, for each value of s, there exists a probability distribution having this property. He did not explicitly give the distribution, but, using his methods, the probability density function can be obtained as

$$\theta(1-\theta)^{-1}(s+1)^{-1}/u^{1+1/(s+1)}, \ \theta^{s+1} < u < 1. \tag{4.4}$$

Note that the functional form of this density as well as the range, depends explicitly on s. Letting $s \to \infty$ and ignoring the normalizing constant, (4.4) converges to the Levy density of the Hyperbolic process.

Remark 4.1: By refining some arguments of May [100], Pielou [33, p. 21] has suggested that the expected descending order statistics $E[\lambda_i^*]$ are approximately a geometric series when $\lambda_1, \lambda_2, \ldots, \lambda_s$ have probability density function

$$(-\log(c))^{-1}/u, \ 0 < c < u < 1. \tag{4.5}$$

Note that (4.5) compares favorably with (4.4) for large s provided the truncation parameter c is approximately θ^{s+1}. The exact expected order statistics from (4.5) are

$$E[\lambda_i^*] = cM(s+1-i, s+1, -\log(c)), \ i = 1,2,\ldots,s,$$

where $M(\cdot, \cdot, \cdot)$ is Kummer's confluent hypergeometric function. Numerical calculations using the tables in Abramowitz & Stegun [89] indicate that these expectations are very close to a geometric series when s is large and c is small. More precisely, using the recurrence (13.4.1) of Abramowitz & Stegun [89], one finds that

$$iE[\lambda^*_{i+1}] + (s + 1 - 2i - \log(c))E[\lambda^*_i] - (s + 1 - i)E[\lambda^*_{i-1}] = 0.$$

From this, it follows that the expected order statistics converge to a geometric series with ratio $\alpha/(\alpha + 1)$ when $s \to \infty$ and $c \to 0$ such that $\log(c)/s \to -1/\alpha$.

The preceding calculations are of limited practical interest since, as Pielou [33, p. 23] points out, the dominance model does not often fit empirical data and then only when the number of species is small. This is consistent with the observations made in Chapter 2 concerning Engen's geometric series model. The hyperbolic process is nonetheless of some theoretical interest; it will be seen in the next section that the Gamma process is a randomized version of the Hyperbolic process.

Example 4.3: Let $0 < a, q < \infty$. The Beta process (of the second kind) has Levy density $f(u) = u^{-1}(1 + u/a)^{-q}$, $0 < u < \infty$. The Gamma process can be obtained as a limiting case by letting $a, q \to \infty$ with $q/a \to 1$. Closed form expressions for $M(A)$ and $H(u)$ are not known.

A Levy density and its associated subordinator process are said to have index k provided $0 < \lim u^{1-k}f(u) < \infty$ as $u \to 0$. Necessarily $k > -1$. The three processes considered above have index zero and their Levy densities have been normalized so that $uf(u) \to 1$ as $u \to 0$. As we shall see, this normalization implies that the ranked abundances are asymptotically geometric in the weak sense with asymptotic ratio θ given by $\log(\theta) = -1/\alpha$. In addition, it is convenient to scale the Levy density so that $\int_0^\infty uf(u)du = 1$ whenever the integral is finite. With this scaling, $E[\xi_\alpha] = \alpha$ by (4.2), so that Fisher's α may be regarded as the mean total abundance. The Levy densities for the Hyperbolic process and the Gamma process have been scaled in this way. For the Beta process, one should take $a = q - 1$ but this is possible only when $q > 1$.

4.3 A Random Dispersal Model

Neyman [111] has described an insect dispersal model in which egg clusters are Poisson distributed throughout the plane. The number of insects to emerge from any cluster is a Poisson random variable with mean λ'. After emergence, the insects are assumed to distribute themselves across the plane according to some dispersal distribution centered at the original egg cluster. Given λ' and given the location of the original cluster, the number of insects (per unit area) to arrive at a small fixed sampling site is also Poisson but whose intensity λ differs from λ'. Clearly λ depends on the distance between the sampling site and the egg cluster and is small when this distance is large. Since each cluster contributes a value for λ, the model contains all the ingredients of a subordinator: a countable collection of positive numbers λ (the jumps), most of which are small, and, if the model is to work, whose sum is finite. Neyman [111] assumed that the dispersal distribution had compact support so that the issue of subordinators did not arise in his treatment.

A species abundance model may be constructed along the same lines by regarding the cluster locations as the focal points of the various species. Call λ' and λ the intrinsic and the effective species abundances, respectively. Note that the intrinsic abundances are dimensionless numbers while the effective abundances have units of 1/area. The dispersal distribution for each species is assumed to be circular Gaussian with center at the species focal point and with variances σ^2. It will be shown that this model results in the Hyperbolic process, the Gamma process, or the Beta process depending on the distributions given λ' and σ^2.

Let the focal points be Poisson distributed in the plane with intensity α' where $\alpha' > 0$ has the units of 1/area. With the sampling site as the origin of coordinates, a given species has effective abundance

$$\lambda = \lambda' \, (2\pi\sigma^2)^{-1}\exp[-\pi(x^2 + y^2)/(2\pi\sigma^2)], \tag{4.6}$$

where (x, y) are the coordinates of the focal point and where λ' and σ^2 are the intrinsic abundance and the variance, respectively. Let $u > 0$ and define $\mathrm{Log}(z)$ to be $\log(z)$ when $0 < z < 1$ and zero otherwise. Then, the requirement $\lambda > u$ is equivalent to

$$\pi(x^2 + y^2) < -2\pi\sigma^2\mathrm{Log}(2\pi\sigma^2 u/\lambda').$$

It follows that the number of species with effective abundance greater than u is a Poisson random variable with mean

$$\alpha' \, E[-2\pi\sigma^2\mathrm{Log}(2\pi\sigma^2 u/\lambda')],$$

where the expectation is taken with respect to the distribution of λ' and σ^2. By (4.3), this expression should be equated with $\alpha H(u)$ where $H(u)$ is the tail integral of the Levy density.

First we suppose that λ' and σ^2 have degenerate distributions so that

$$\alpha H(u) = \begin{cases} -\alpha'(2\pi\sigma^2)\log(2\pi\sigma^2 u/\lambda') & \text{if } 0 < u < \lambda'/(2\pi\sigma^2) \\ 0 & \text{otherwise.} \end{cases}$$

Differentiating with respect to u, the Levy density becomes

$$\alpha f(u) = \begin{cases} \alpha'(2\pi\sigma^2)/u & \text{if } 0 < u < \lambda'/(2\pi\sigma^2) \\ 0 & \text{otherwise.} \end{cases} \qquad (4.7)$$

Up to a scale transformation, (4.7) is the same as the Levy density of the hyperbolic process provided we take $\alpha = \alpha'(2\pi\sigma^2)$. Note that $2\pi\sigma^2$ is the mean dispersal area and the dimensionless number α is the mean number of species to be found in such a standardized area. This is consistent with our earlier contention that α is best thought of as a species richness parameter.

Next, let σ^2 continue to be degenerate but give λ' an exponential distribution. Since λ' enters only the range of (4.7), there is no loss of generality in taking the scale parameter of the exponential distribution as unity. Randomizing (4.7), the Levy density is

$$\alpha f(u) = \alpha'(2\pi\sigma^2)u^{-1}\exp(-2\pi\sigma^2 u),\ 0 < u < \infty. \qquad (4.8)$$

Again taking $\alpha = \alpha'(2\pi\sigma^2)$, (4.8) is the same as the Levy density of the Gamma process, up to a scale transformation.

Now, the MacArthur model assumed an exponential abundance distribution and is often found to provide a reasonable fit when the sample contains only a few species. But the fit usually deteriorates with increasing sample size. This may possibly be explained by noticing that, when the sampling intensity is small, only those species with focal point close to the sampling site will be observed. But the effective abundances of these species are approximately proportional to their intrinsic abundances, by (4.6). With increased sampling intensity, the intrinsic abundances become seriously distorted.

Finally, let λ' have a standard exponential distribution and $2\pi\sigma^2$ a Gamma distribution with index b and mean b/a. Randomizing (4.8) with respect to the Gamma distribution gives

$$\alpha f(u) = (\alpha' b/a)u^{-1}(1 + a^{-1}u)^{-b-1}.$$

Here, we take $\alpha = \alpha' b/a = \alpha' E[2\pi\sigma^2]$ and $q = b + 1$ to obtain the Levy density of the Beta process. Note that $q > 1$ in this model.

4.4 Constructing the Process

Three methods of constructing the subordinator process with Levy density f(u) will be described.

Limit of compound Poisson processes. Suppose that $0 < \ldots < x_3 < x_2 < x_1 < x_0 = \infty$ is a partition of the positive real axis where $x_i \to 0$ as $i \to \infty$. For $i = 1,2,3,\ldots$, let the probability d.f. $f^{(i)}(u) = f(u)/[H(x_i) - H(x_{i-1})]$, $x_i \leq u < x_{i-1}$, be the truncation of the Levy density to the interval $[x_i, x_{i-1})$. Let $\xi_t^{(i)}$, $0 \leq t < \infty$, $i = 1,2,3,\ldots$, be independent compound Poisson processes whose intensity is $H(x_i) - H(x_{i-1})$ and whose jump distribution is $f^{(i)}(u)$. Then ξ_t may be realized as the superposition $\sum_i \xi_t^{(i)}$ of the processes $\xi_t^{(i)}$. Alternatively,

$$\xi_t = \lim_{n\to\infty} \sum_{i=1}^{n} \xi_t^{(i)}$$

is the limit of compound Poisson processes with intensity $H(x_n)$ and with the truncation of f(u) to the interval $[x_n, \infty)$ as jump distribution. For full details concerning this construction, see Breiman [105, p. 310ff.]. As an application, we prove the following fundamental lemma:

Lemma 4.1: Suppose the nonnegative function $B(u_1, u_2, \ldots, u_r; y)$ is defined for $0 < u_1, u_2, \ldots, u_r$, $0 < y < \infty$, $u_1 + u_2 + \ldots + u_r \leq y$. Let ξ_t be a subordinator with Levy density $f(u)$ and with jumps $\lambda_1, \lambda_2, \lambda_3, \ldots$ over the time interval $0 \leq t \leq \alpha$, $\alpha > 0$. Then

$$E[\Sigma B(\lambda_{i(1)}, \lambda_{i(2)}, \ldots, \lambda_{i(r)}; \xi_\alpha)] =$$

$$\alpha^r \int_R B(u_1, u_2, \ldots, u_r; V + u_1 + u_2 + \ldots + u_r) f(u_1) f(u_2) \ldots f(u_r) \times$$

$$f_\alpha(V) du_1 du_2 \ldots du_r dV, \qquad (4.9)$$

where the sum ranges over all r-tuples $(i_{(1)}, i_{(2)}, \ldots, i_{(r)})$ of distinct positive integers and where the range of integration is the region R defined by $0 < u_1, u_2, \ldots, u_r, V < \infty$.

Proof: Obtain ξ_t as the limit of compound Poisson process with intensity $H(x_n)$ and jump distribution $f(u)/H(x_n)$, $x_n \leq u < \infty$. Let g_m, $m = 0, 1, 2, \ldots$, be the m-fold convolution of this jump density. For simplicity, take $r = 3$. Conditional on the number of jumps s, the expectation in (4.9) for the compound Poisson process is

$$s(s - 1)(s - 2) \int_{R_n} B(u_1, u_2, u_3; V + u_1 + u_2 + u_3) f(u_1) f(u_2) f(u_3) \times$$

$$[H(x_n)]^{-3} g_{s-3}(V) du_1 du_2 du_3 dV, \qquad (4.10)$$

where the range of integration is the region R_n defined by $x_n \leq u_1, u_2, u_3 < \infty$, $0 < V < \infty$. But s has a Poisson distribution with mean $\alpha H(x_n)$ so that

$$E[s(s - 1)(s - 2) g_{s-3}(V)] = [\alpha H(x_n)]^3 \sum_{j=0}^{\infty} [\alpha H(x_n)]^j \exp(-\alpha H(x_n)) g_j(V)/j!.$$

Inserting this into (4.10) and letting $n \to \infty$ gives the result since

$$\sum_{j=0}^{\infty} [\alpha H(x_n)]^j \exp(-\alpha H(x_n)) g_j(V)/j! \to f_\alpha(V).$$

Uniform arrangement of ranked jumps. This construction yields the process ξ_t only for $0 \le t \le \alpha$. The procedure is to first obtain realizations $\lambda_1^*, \lambda_2^*, \lambda_3^*, \ldots$ of the ranked jumps and then to distribute these ranked jumps uniformly across the interval $0 \le t \le \alpha$. Formally, let $U_1, U_2, U_3, \ldots$ be a random sample from the Uniform $(0, \alpha)$ distribution independent of the λ_i^*. Then ξ_t, $0 \le t \le \alpha$, is realized by placing a a jump of magnitude λ_i^* at time $t = U_i$, i.e., $\xi_t = \sum_i \{\lambda_i^* : U_i \le t\}$. According to Kingman [79], the ranked jumps can be obtained by putting

$$\lambda_i^* = H^{-1}(T_i/\alpha), \quad i = 1,2,3,\ldots, \tag{4.11}$$

where $0 < T_1 < T_2 < T_3 < \ldots$ are the successive epochs of a Poisson process with unit intensity.

This construction shows that the ranked jumps are random variables having a transformed Gamma distribution. In particular, the largest jump λ_1^* has cumulative distribution function $\exp(-\alpha H(u))$ and probability density function

$$\alpha f(u) \exp(-\alpha H(u)), \quad 0 < u < \infty. \tag{4.12}$$

Notice that $H(u)$, and hence the process, is uniquely determined by the distribution of λ_1^* for even a single value of α. This is the analogue of the well-known theorem that a probability distribution is uniquely determined by its largest expected order statistic for all sample sizes.

Next we show that the ranked abundances are asymptotically geometric in the weak sense whenever the Levy density has index zero.

Theorem 4.1: Assume the Levy density has index zero and is normalized so that $uf(u) \to 1$ as $u \to 0$. Then, with probability one,

$$\lim_{i\to\infty} \log(\lambda_i^*)/i = \lim_{i\to\infty} \log(\lambda_i^*/\xi_\alpha)/i = -1/\alpha.$$

Proof: Given $0 < \varepsilon < 1$, there is a $u_0 > 0$ such that $1 - \varepsilon \leq uf(u) \leq 1 + \varepsilon$ for $0 < u < u_0$. Dividing by u and integrating from u to u_0 gives $H(u_0) + (1 - \varepsilon)\log(u_0/u) \leq H(u) \leq H(u_0) + (1 + \varepsilon)\log(u_0/u)$ for $0 < u < u_0$. But this implies that

$$u_0 \exp[-(x - H(u_0))/(1 - \varepsilon)] \leq H^{-1}(x) \leq u_0 \exp[-(x - H(u_0))/(1 + \varepsilon)]$$

for $x > H(u_0)$. From (4.11),

$$\text{constant}\cdot\exp[-\alpha^{-1}T_i/(1 - \varepsilon)] \leq \lambda_i^* \leq \text{constant}\cdot\exp[-\alpha^{-1}T_i/(1 + \varepsilon)].$$

By the strong law of large numbers, $T_i/i \to 1$. Therefore,

$$-\alpha^{-1}/(1 - \varepsilon) \leq \lim \log(\lambda_i^*)/i \leq -\alpha^{-1}/(1 + \varepsilon),$$

and the proof is completed by letting $\varepsilon \to 0$.

In view of the loss of memory property of the Poisson process, one can also obtain the ordered jumps via successive truncation: first obtain λ_1^* from the distribution (4.12), then truncate this distribution to the interval $(0, \lambda_1^*)$ and obtain λ_2^* from the truncated distribution, then truncate (4.12) to the interval $(0, \lambda_2^*)$ and obtain λ_3^* from the newly truncated distribution, etc. This enables us to write down the joint density of $\lambda_1^*, \lambda_2^*, \ldots, \lambda_r^*$ as

$$\alpha^r f(u_1)f(u_2)\ldots f(u_r)\exp(-\alpha\, H(u_r)),\quad u_1 > u_2 > \ldots > u_r > 0. \qquad (4.13)$$

From (4.13), it follows that the ranked jumps from the Hyperbolic process can also be obtained from a stochastic dominance mechanism.

Theorem 4.2: Let P_1, P_2, $P_3, \ldots$ be a random sample from the Beta_1 $(\alpha, 1)$ distribution and put $\lambda_1^* = P_1$, $\lambda_2^* = P_1 P_2$, $\lambda_3^* = P_1 P_2 P_3, \ldots$. Then λ_1^*, λ_2^*, $\lambda_3^*, \ldots$ are equal in distribution to the ordered jumps from the Hyperbolic process. In particular, $E[\lambda_i^*] = [\alpha/(\alpha + 1)]^i$.

Proof: For the Hyperbolic process, (4.13) simplifies to $\alpha^r u_r^{\alpha-1}/[u_1 u_2 \ldots u_{r-1}]$. But this is also the joint density of P_1, $P_1 P_2, \ldots, P_1 P_2 \ldots P_r$.

It follows that λ_1^* is independent of λ_2^*/λ_1^*. It can be shown that this characterizes the Hyperbolic process, up to a scale transformation.

Uniform arrangement of size-biased jumps. Generate the ranked jumps λ_i^* and the uniform random variables U_i as in the previous construction and let $\lambda_i^{\#}$ be the size-biased permutation of the ranked jumps. Since the λ_i^* are random variables, so are the $\lambda_i^{\#}$. Further, the uniform variables are independent of the λ_i^* as well as the permutation so that ξ_t, $0 \leq t \leq \alpha$, can be realized as $\xi_t = \sum_i \{\lambda_i^{\#} : U_i \leq t\}$. Since $P(U_i \leq p\alpha) = p$, $0 < p < 1$, it follows that the size-biased jumps over the time interval $0 \leq t \leq p\alpha$ are obtained by random deletion of the size-biased jumps over the time interval $0 \leq t \leq \alpha$. The same applies to the ranked jumps. This together with Corollary 4.2, below, proves Theorem 2.11.

Theorem 4.3: The joint density of $\lambda_1^{\#}$, $\lambda_2^{\#}, \ldots, \lambda_r^{\#}$ and the residual $Y = \xi_\alpha - \lambda_1^{\#} - \lambda_2^{\#} - \ldots - \lambda_r^{\#}$ is

$$\alpha^n f_\alpha(y) \prod_{i=1}^{r} [u_i f(u_i)/(u_i + u_{i+1} + \ldots + u_r + y)].$$

The density of $\lambda_1^{\#}$ is $uf(u)E[\alpha/(u + \xi_\alpha)]$.

Proof: Similar to the proof of Theorem 3.3 but using Lemma 4.1.

Corollary 4.1: For the Gamma process, $\lambda_1^{\#}$ has density $\alpha E_\alpha(u)$, $0 < u < \infty$.

Corollary 4.2: For the Gamma process, the normalized abundance $\underset{\sim}{\pi}^{\#} = (\lambda_1^{\#}, \lambda_2^{\#}, \lambda_3^{\#}, \ldots)/\xi_\alpha$ is independent of ξ_α and is equal in distribution to Engen's model with parameter α.

Proof: Apply Theorem 4.3 with a change of variable.

Since Kingman's version of the limiting Dirichlet is $\underset{\sim}{\pi}^{*} = (\lambda_1^{*}, \lambda_2^{*}, \lambda_3^{*}, \ldots)/\xi_\alpha$, we obtain,

Corollary 4.3: Engen's model is the same as the size-biased permutation of Kingman's limiting Dirichlet and Kingman's limiting Dirichlet is the same as the ranked permutation of Engen's model.

Corollary 4.4: Engen's model is invariant under size-biased permutation.

Proof: Immediate from Corollary 4.3 since $\underset{\sim}{\pi}^{\#}$ and $\underset{\sim}{\pi}^{\#\#}$ are equal in distribution.

Corollary 4.5: For the Gamma process, the ranked abundance $\underset{\sim}{\pi}^{*} = (\lambda_1^{*}, \lambda_2^{*}, \lambda_3^{*}, \ldots)/\xi_\alpha$ is independent of ξ_α.

Proof: This follows from Corollary 4.2 and Corollary 4.3.

Finally, we give the promised proof of Theorem 2.13(f).

Corollary 4.6: For the Gamma process, $E[\Sigma(i - 1)\lambda_i^{*}/\xi_\alpha] = \alpha \log (2)$.

Proof: By Corollary 4.5, $E[\Sigma(i - 1)\lambda_i^{*}/\xi_\alpha] = E[\Sigma(i - 1)\lambda_i^{*}]/E[\xi_\alpha] = E[\Sigma(i - 1)\lambda_i^{*}]/\alpha$. Define $B(u_1, u_2) = u_1$ if $0 < u_1 < u_2$ and zero otherwise. Using an argument of Kingman [79], the expression $\Sigma(i - 1)\lambda_i^{*}$ can now be recast in terms of the function $B(u_1, u_2)$ in the following manner:

$$\sum_i (i-1)\lambda_i^* = \sum_i \lambda_i \cdot (\text{number of } \lambda_j \text{ such that } \lambda_j > \lambda_i)$$

$$= \sum_{i \neq j} B(\lambda_i, \lambda_j).$$

By Lemma 4.1,

$$E[\sum_i (i-1)\lambda_i^*] = \alpha^2 \int B(u_1, u_2) f(u_1) f(u_2) du_1 du_2$$

$$= \alpha^2 \int_{u_1<u_2} u_1 f(u_1) f(u_2) du_1 du_2$$

$$= \alpha^2 \int_{u_2=0}^{\infty} u_2^{-1} e^{-u_2} \int_{u_1=0}^{u_2} e^{-u_1} du_1 du_2$$

$$= \alpha^2 \int_0^{\infty} (e^{-u_2} - e^{-2u_2})/u_2 \, du_2.$$

But this is $\alpha^2 \log(2)$ by (5.1.32) of Abramowitz & Stegun [89].

Next, we prove an analogue of the first moment convergence result given in Theorem 3.2.

Theorem 4.4: Consider a subordinator with Levy density $f(u)$ for which $\int uf(u)du < \infty$. As $\alpha \to \infty$, $\lambda_1^{\#}$ has a limiting distribution for which the probability density function is $uf(u)/\int uf(u)du$.

Proof: Same as the proof of Theorem 3.2 but using Lemma 4.1 and the strong law of large numbers for subordinators. Complex numbers can be avoided by using the Laplace transform instead of the characteristic function.

Applying this result to Corollary 4.1, it follows that the density $\alpha E_{\alpha}(u)$ converges to the standard exponential density, as pointed out in

Section 3.4. In terms of the random dispersal model, large α means that the species focal points are densely distributed throughout the plane. A size-biased selection is then likely to produce a species with focal point close to the sampling site and an effective abundance roughly proportional to the intrinsic abundance. So it is intuitively clear for the Gamma process that $\lambda_1^{\#}$ converges to an exponential distribution as $\alpha \to \infty$.

Corollary 4.7: The distribution of $\lambda_1^{\#}$ for all α uniquely determines the process provided the Levy density has finite first moment.

Proof: The Levy density is determined by its first moment distribution.

Open Problem: Does the distribution of $\lambda_1^{\#}$ for just a single α determine the Process?

Open Problem: Develop an extreme value theory for the limiting distribution of λ_1^*.

Open Problem: Is α a complete diversity parameter when the Levy density has no unknown parameters? A major obstacle in attacking this problem is the lack of independence of ξ_α and $(\lambda_1^*, \lambda_2^*, \ldots)/\xi_\alpha$. Such independence is well-known to characterize the Gamma process.

Consider a subordinator process which generates a random set $\{\lambda_i\}$ of abundances. How might we replace $\{\lambda_i\}$ with a typical fixed set of abundances? The choices $E[\lambda_i^*]$ and $E[\lambda_i^{\#}]$ are perhaps natural, but, as pointed out in Section 2.4, these will exaggerate community diversity. Another possibility is suggested by (4.11): replace T_i by its expectation $E[T_i] = i$ and adopt

$$\mu_i = H^{-1}(i/\alpha), \ i = 1,2,3,\ldots \tag{4.14}$$

as the fixed abundances. Note that $\mu_1 \geq \mu_2 \geq \mu_3 \geq \ldots$. Applied to the Hyperbolic process, (4.14) gives the geometric series model $\mu_i = \exp(-i/\alpha)$ with ratio $\exp(-1/\alpha)$. The Gamma process results in Watterson's [90] exponential integral model $\mu_i = E_1^{-1}(i/\alpha)$.

We wish to show that the fixed abundances are asymptotically geometric with asymptotic ratio $\exp(-1/\alpha)$ when the Levy density has index zero and is normalized as described in Section 4.2. This is true for weak asymptotic geometricity and can be proved just as in Theorem 4.1. To obtain strong asymptotic geometricity, an additional regularity assumption will be imposed on the Levy density. Suppose that

$$\left|\lim_{u\to 0} [uf(u) - 1]/u\right| < \infty. \tag{4.15}$$

Equivalently, the function $uf(u)$ should take the value one at the origin and should have a finite derivative at the origin. The three processes under consideration satisfy (4.15).

Theorem 4.5: The set of fixed abundances $\mu_i = H^{-1}(i/\alpha)$ is strong AG with asymptotic ratio $\exp(-1/\alpha)$ when the Levy density satisfies (4.15).

Proof: We have to prove that $\mu_i/\exp(-i/\alpha) = \mu_i\exp(i/\alpha)$ has a finite nonzero limit as $i \to \infty$. Since $i/\alpha = H(\mu_i)$ and $\mu_i \to 0$, it suffices to show that $u\exp(H(u))$ has a finite nonzero limit as $u \to 0$. By (4.15), there are finite numbers A, B such that $A \leq [uf(u) - 1]/u \leq B$ or $u^{-1} + A \leq f(u) \leq u^{-1} + B$ for all sufficiently small u, say $0 < u < \varepsilon$. Integrating from u to ε gives

$$\log(\varepsilon/u) + A(\varepsilon - u) \leq H(u) - H(\varepsilon) \leq \log(\varepsilon/u) + B(\varepsilon - u).$$

Transpose $H(\varepsilon)$ and exponentiate to find that

$$(\varepsilon/u)\exp(H(\varepsilon))\exp[A(\varepsilon - u)] \leq \exp(H(u)) \leq (\varepsilon/u)\exp(H(\varepsilon))\exp[B(\varepsilon - u)].$$

After multiplying by u and letting $u \to 0$, this gives

$$\limsup_{u\to 0} u\exp(H(u)) \leq \varepsilon \exp(H(\varepsilon)) \exp(B\varepsilon) \tag{4.16}$$

and

$$\varepsilon \exp(H(\varepsilon))\exp(A\varepsilon) \leq \liminf_{u\to 0} u\exp(H(u)).$$

These inequalities show that the upper and lower limits are both finite and nonzero. To show that the limit actually exits, let $\varepsilon \to 0$ in (4.16) to find that the upper limit is less than or equal to the lower limit.

The preceding proof only requires that 9uf(u) - 1)/u be bounded as $u \to 0$.

4.5 Poisson Sampling from a Subordinator

Let the species abundances λ_i be realized as the jumps of a subordinator during the time interval $0 \leq t \leq \alpha$. It will be assumed that the number of individuals in the sample from species i is a Poisson random variable with mean $A\lambda_i$. In effect, the jump λ_i is replaced with an observation on the Poisson $(A\lambda_i)$ distribution, i.e., with the number of times species i appears in the sample. All but finitely many of these replacements are zero since most λ_i are small, and the resulting jump process is compound Poisson with a positive integer valued jump distribution. Call this compound Poisson process the <u>sample process</u>.

Each jump of the sample process represents a species that is physically present in the sample. Moreover, it is proved in Appendix A.10 that (i) the sample process has jump intensity (number of jumps per unit time) equal to M(A) where M is the Levy transform and (ii) the jump distribution for the sample process is

$$p_x = \int_0^\infty (Au)^x \exp(-Au) f(u) du / [x! M(A)], \; x = 1,2,3,\ldots \; . \qquad (4.17)$$

From (i) it follows that the number T of species in the sample is a Poisson random variable whose mean is the product of the jump intensity and the observation time. Thus

$$E[T] = \alpha \, M(A). \qquad (4.18)$$

When A can be interpreted as area, (4.18) is the species-area curve. It is a concave function of A which uniquely determines the Levy transform and, hence, the process. Using Table 4.1, the species-area curves for the Hyperbolic process and the Gamma process are, respectively,

$$E[T] = \alpha[E_1(A) + \log(A) + \gamma]$$

and

$$E[T] = \alpha \log(1 + A).$$

Both are asymptotic to $\alpha \log(A)$ for large A.

Conditional on the subordinator, the number N of individuals in the sample has a Poisson distribution with mean $A\Sigma\lambda_i = A\xi_\alpha$. After randomizing with respect to ξ_α, the distribution of N is compound Poisson with mean

$$E[N] = AE[\xi_\alpha] = \alpha A\int_0^\infty uf(u)du. \tag{4.19}$$

For the Hyperbolic process and the Gamma process, (4.19) simplifies to $E[N] = \alpha A$. Moment estimates of α and A can be obtained for these two processes by equating the observed values of T and N to their expected values. Moment estimation is not generally possible for the Beta process since $E[N] = \infty$ when $0 < q \leq 1$. The maximum likelihood estimates of q often fall into this range (Kempton [95]).

Consider the sample process and let $n_x(t)$, $x = 1,2,3,\ldots$, be the number of jumps through time t of size exactly equal to x. The $n_x(t)$ are obtained by screening the sample process and are themselves independent Poisson processes with intensities $M(A)\ p_x$, where p_x is given by (4.17). But $n_x \equiv n_x(\alpha)$ is the number of species with x representatives in the sample. Thus the n_x are independent Poisson random variables with means

$$\begin{aligned} E[n_x] &= \alpha M(A)\ p_x \\ &= (\alpha/x!)\int_0^\infty (Au)^x \exp(-Au)f(u)du. \end{aligned} \tag{4.20}$$

For the Hyperbolic, Gamma, and Beta process, these expected frequencies become, respectively,

$$E[n_x] = (\alpha/x)\ \Gamma(x;\ A),$$

$$E[n_x] = (\alpha/x)[A/(1 + A)]^x,$$

and

$$E[n_x] = (\alpha/x!)\int_0^\infty u^{x-1}[1 + (Aa)^{-1}u]^{-q}\exp(-u)du. \tag{4.21}$$

Note the confounding between a and A in this last equation.

Because the n_x are independent Poisson variates, the likelihood equations are easily shown to be

$$\sum_{x=1}^{\infty} n_x(\partial/\partial\theta)(\log E[n_x]) = (\partial/\partial\theta)(\alpha M(A)),$$

where θ represents any of the various parameters of the model. The first likelihood equation ($\theta = \alpha$) coincides with the first moment equation: $T = E[T] = \alpha M(A)$.

For the Hyperbolic model, the second likelihood equation ($\theta = A$) is

$$\begin{aligned} \sum_{x=1}^{\infty} x n_x \left[\frac{A^x e^{-A}/x!}{\Gamma(x;A)}\right] &= \alpha(1 - e^{-A}) \\ &= T(1 - e^{-A})/[E_1(A) + \log(A) + \gamma]. \qquad (4.22) \end{aligned}$$

Both sides of (4.22) are decreasing functions of A. When $A \to 0$, the left hand side converges to $\Sigma x n_x = N$ while the right side converges to $T \leq N$. On the other hand, when $A \to \infty$, the right hand side goes to zero much more slowly than the left hand side. Thus the likelihood equations always have a solution when $0 < T < N$. We have not investigated the uniqueness of this solution.

For the Gamma process, the two likelihood equations are the same as the moment equations. Estimation for this model has been extensively studied by Fisher [85], Bliss [53], Rao [112], and Watterson [90]. We have generally found the fit of the Gamma process to be far superior to that of the Hyperbolic process. Since both models have the same number of parameters, the Gamma process is to be preferred.

Kempton [95] has discussed likelihood estimation for the Beta process. He obtained the expected frequencies (4.21) by supposing that the community contains s species whose abundances follow a Beta distribution of the second kind with parameters k and q. After determining the expected sample frequencies for this model, he let $k \to 0$ and $s \to \infty$ with $sk = \alpha$.

Remark 4.2: The jump distribution (4.17) of the sample process may be rewritten as

$$\int_0^\infty \left[\frac{(Au)^x e^{-Au}}{x!(1 - e^{-Au})}\right] \cdot \left[\frac{(1 - e^{-Au})f(u)}{M(A)}\right] du, \quad x = 1,2,3,\ldots .$$

The first term in brackets is a zero-truncated Poisson probability while the second term is a probability density function. The jumps of the sample process thus have a compound zero-truncated Poisson distribution. Boswell & Patil [113] have noticed that improper distributions like f(u) can be handled at the sampling level in the above manner.

Remark 4.3: For the Gamma process, the jump distribution (4.17) is the log series distribution with parameter $A/(1 + A)$. Since ξ_α has a Gamma distribution, the number of individuals in the sample is a negative binomial random variable (see the remarks preceding (4.19)). We have thus obtained the well-known result that a Poisson sum from the log series distribution follows the negative binomial law.

4.6 Conditional Sampling Variance

The variance of a sample statistic T contains two components,

$$\mathrm{Var}\,[T] = E[\mathrm{Var}(T|\underset{\sim}{\lambda})] + \mathrm{Var}[E(T|\underset{\sim}{\lambda})]. \tag{4.23}$$

The first term--the conditional variance -- represents the pure sampling effect while the second is the contribution due to the variability of species abundances. We wish to indicate how Lemma 4.1 can be used to calculate the conditional variance.

For simplicity, take T to be the number of species in the sample so that Var $(T|\underset{\sim}{\lambda})$ is

$$\sum_i \exp(-A\lambda_i)(1 - \exp(-A\lambda_i)) = \sum_i [\exp(-A\lambda_i) - \exp(-2A\lambda_i)]. \qquad (4.24)$$

Taking expectations and using Lemma 4.1, the conditional variance becomes

$$\begin{aligned} E[\text{Var}(T|\underset{\sim}{\lambda})] &= \alpha \int [\exp(-Au) - \exp(-2Au)]f(u)du \\ &= \alpha[M(2A) - M(A)]. \end{aligned}$$

For the Gamma process, this simplifies to $\alpha \log[(1 + 2A)/(1 + A)]$ which is the same as (2.42) of Watterson [90]. Watterson obtains this expression by replacing the abundances λ_i in (4.24) with their fixed versions $E_1^{-1}(i/\alpha)$ and then approximating the infinite sum by an integral. This method will yield the correct conditional variance of any sample statistic provided the fixed abundances are taken to be $H^{-1}(i/\alpha)$ as in (4.14).

Watterson [90, equation (2.34)] has also attempted to derive the joint probability generating function of the n_x by formally applying Fisher's limit. As he notes, the result is not a valid probability generating function, but it does yield the correct expectations as well as the correct conditional variances. In taking the limit $s \to \infty$, Watterson treated s as though it were a constant. But in obtaining a subordinator as the limit of compound Poisson processes, s has a Poisson distribution. Watterson thus overlooked a component of variability.

It is unclear to the present author why Watterson's generating function gives precisely the correct conditional variances.

Since the n_x are independent Poisson random variables, their joint probability generating function is easily written down and results from (2.32) of Watterson by letting s have a Poisson distribution with mean α/k, then randomizing with respect to the Gamma distribution, and, finally, letting $k \to 0$.

4.7 Multiple Sampling Sites

When samples are taken at different points in time or in space, it has been observed that the pooled sample may fail to fit the log series distribution although the individual samples do fit (Bliss [53]). The random dispersal model of Section 4.3 provides a framework for explaining this effect.

Consider two sites with Poisson sampling intensities A_1 and A_2. It is reasonable to assume that the two samples are conditionally independent given the intrinsic abundances λ' and the dispersal parameters σ^2. However independence is lost after randomizing with respect to λ' and σ^2.

Let n_{xy}, $x, y = 0,1,2,\ldots,(x,y) \neq (0,0)$, be the number of species with x representatives at the first site and y representatives at the second site. The n_{xy} are independent Poisson random variables, but their expected values are very complicated integrals which will not be given here. Clearly the distance d between the two sites enters these integrals as a parameter.

Two limiting cases are of interest. First, when $d \to 0$, so that the samples are taken at the same site, the expected values become

$$E[n_{xy}] = [\alpha/(x!y!] \int_0^\infty (A_1u)^x(A_2u)^y\exp(-A_1u - A_2u)f(u)du.$$

The expected frequencies $E[n'_z]$ for the pooled sample are obtained by summing over $x + y = z$. Using the binomial theorem, this gives

$$E[n'_z] = (\alpha/z!) \int_0^\infty [(A_1 + A_2)u]^z\exp[-(A_1 + A_2)u]f(u)du, \qquad (4.25)$$

which are the same as the expected frequencies for a single sample of intensity $A_1 + A_2$. Thus, pooling two samples from the same site does not alter the model.

When $d \to \infty$, the expected joint frequencies are zero unless $x = 0$ or $y = 0$, i.e., the two samples have no species in common. Pooling the two samples in this case gives

$$E[n'_z] = (\alpha/z!)\int_0^\infty[(A_1u)^z\exp(-A_1u) + (A_2u)^z\exp(-A_2u)]f(u)du. \qquad (4.26)$$

Now (4.26) exceeds (4.25) for small z and vice versa for large z. Thus the pooled sample will have more singletons than are predicted by the fitted model. An excess of singletons is frequently observed even for a single sample which, in reality, is a composite of many adjacent samples.

APPENDIX

A.1 Proof of Lemma 1.2

Repeated use is made of the following consequence of Abel's partial summation formula:

Abel's inequalities. Let $\underset{\sim}{f}$ be a nonzero transfer vector and $\{a_i : 1 \leq i < \infty\}$ a monotone sequence. Then

(i) $\Sigma f_i a_i \leq 0$ if $\{a_i\}$ is decreasing,

(ii) $\Sigma f_i a_i \geq 0$ if $\{a_i\}$ is increasing,

(iii) $\Sigma f_i a_i = 0 \Leftrightarrow \underset{\sim}{f}$ vanishes over tied sets of $\{a_i\}$,

(iv) $\Sigma f_i a_i \neq 0$ if $f_i \geq f_{i+1}$ whenever $a_i = a_{i+1}$.

For the proof of Lemma 1.2, let f_n be the last nonzero component of $\underset{\sim}{f}$ and let λ_s be the last finite term of $\{\lambda_i\}$. Using Abel's inequalities, the signs of the numerator and the denominator of $\sigma(x)$ are determined as in the following tabulation:

Range of x	Numerator		Denominator	
	$n \leq s$	$s < n$	a)	b) or c)
$x<0$	$+$	$+\ \infty$	$+\ 0$	$+$
$x=0$	$-$	$-\ \infty$	$-\ 0$	$-$
$x>0$	$-$	$-$	$-\ 0$	$-$

.

All assertions of the lemma, other than the monotonicity of $\sigma(x)$, are clear from the tabulation. For the proof that $\sigma(x)$ is strictly decreasing, take $x \neq 0$ in case b) and $x > 0$ in case c). Now $\sigma'(x)$ has the same sign as the expression $\sum_{j=1}^{m} b_j A(\lambda_j)$ where

$$A(\lambda) = \sum_{i=m}^{n} f_i(\lambda_i - \lambda)\exp[(\lambda_i + \lambda)x].$$

We show that $A(\lambda)$ is strictly increasing for $\lambda > \lambda_m$; it then follows by Abel's inequalities that $\sigma'(x) < 0$. But $A'(\lambda) = \sum_{i=m}^{n} f_i B(\lambda_i)$ where

$$B(\mu) = (\mu x - \lambda x - 1) \exp[(\mu + \lambda)x].$$

Again by Abel's inequalities, it suffices to show that $B(\mu)$ is strictly decreasing for $\lambda_m > \mu$. But $B'(\mu) = (\mu x^2 - \lambda x^2)\exp[(\mu + \lambda)x] < 0$ since $\lambda > \lambda_m > \mu$ and $x \neq 0$.

A.2 The MacArthur Frequencies

MacArthur's fixed model is the expected ranked abundance vector $\underset{\sim}{\pi}$ that results when the unit interval is randomly (uniformly) apportioned among the s different species. As shown in Chapter 2, $\underset{\sim}{\pi}$ is given by

$$\pi_i = \begin{cases} (1/i + 1/(i+1) + \ldots + 1/s)/s & \text{if } 1 \leq i \leq s \\ 0 & \text{if } s < i. \end{cases}$$

These frequencies can be arrived at directly by supposing that the total available resource is represented by the unit square which is divided into s vertical slices of equal width. Refer to Figure A.1 when s = 4. The first slice is apportioned equally among the s different species after which one species is removed from competition and the second slice apportioned equally among the remaining s - 1 species. The procedure is now iterated with an additional species being removed after the apportionment of each slice. The procedure is deterministic and is intended to yield only MacArthur's fixed model, not the random model.

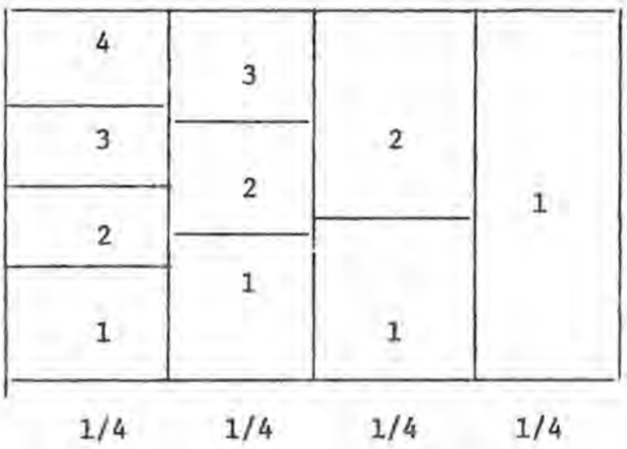


Figure A.1 Apportionment of the Unit Square $(s = 4)$

The equiprobable model and MacArthur's model have been employed by Joanes & Gill [114] as two series of communities of increasing species richness upon which the bias reducing effectiveness of flattening can be assessed. They further remark that it would be desirable to have available additional series of test communities. A simple generalization of the preceding apportionment scheme will yield such additional series.

Let $q_1, q_2, \dots, q_s$ be a given sequence of real numbers with $q_1 = 1$ and $0 \leq q_i \leq 1$. Take the width of the ith slice to be $(1 - q_s) \times (1 - q_{s-1}) \dots (1 - q_{i+1}) q_i$, i.e., the width of the first slice is q_s, the width of the second slice is a fraction q_{s-1} of the remaining width, etc. Let $\underset{\sim}{\pi}(q_1, q_2, \dots, q_s)$ be the community generated when the individual slices are equally apportioned as in the MacArthur model.

<u>Theorem A.1</u>: $\underset{\sim}{\pi}(1, 1/2, 1/3, \dots, 1/s)$ is the MacArthur model and $\underset{\sim}{\pi}(1, 1, \dots, 1)$ is the completely even community. Finally, the single-species community is obtained as $\underset{\sim}{\pi}(1, 0, 0, \dots, 0)$.

Theorem A.2: The community $\underset{\sim}{\pi}(q_1,\ldots,q_s)$ can be generated iteratively as follows:

$$\underset{\sim}{\pi}(q_1) = \underset{\sim}{\pi}(1) = (1,\ 0,\ 0,\ 0,\ldots)$$

$$\underset{\sim}{\pi}(q_1,\ q_2,\ldots,q_s) = (1 - q_s)\underset{\sim}{\pi}(q_1,\ q_2,\ldots,q_{s-1}) + q_s\ \underset{\sim}{\varepsilon}(s)$$

where $\underset{\sim}{\varepsilon}(s) = (1/s,\ 1/s,.,\ 1/s,\ 0,\ 0,\ 0,\ldots)$ is the completely even community with s species.

Theorem A.3: The community $\underset{\sim}{\pi}(q_1,\ q_2,\ldots,q_{s+1})$ is intrinsically more diverse than $\underset{\sim}{\pi}(q_1,\ q_2,\ldots,q_s)$.

Theorem A.4: $\underset{\sim}{\pi}(q_1',\ q_2',\ldots,q_s')$ is intrinsically more diverse than $\underset{\sim}{\pi}(q_1,\ q_2,\ldots,q_s)$ when $q_i' \geq q_i$, $i = 1,2,\ldots,s$.

A.3 Measure Theoretic Framework for Infinitely Many Species

Let $I^{(\infty)} = \{(x_1,\ x_2,\ x_3,\ldots): 0 \leq x_i \leq 1\}$ be the Hilbert cube equipped with the product Borel structure and $SP^{(\infty)} = \{\underset{\sim}{x} \in I^{(\infty)}: \Sigma x_i = 1\}$ the infinite dimensional simplex.

Theorem A.5: $SP^{(\infty)}$ is a Borel subset of $I^{(\infty)}$.

Proof: Define $f : I^{\infty} \to [0,\ \infty]$ by $f(x) = \Sigma x_i$. Now f is measurable (by the monotone convergence theorem since the partial sums are measurable) and $SP^{(\infty)} = f^{-1}(1)$.

Equip $SP^{(\infty)}$ with the subspace Borel structure. The inclusion map $SP^{(\infty)} \to I^{(\infty)}$ and the coordinate maps $SP^{(\infty)} \to [0,1]$: $\underset{\sim}{x} \to x_i$ are then measurable.

Theorem A.6: Suppose $\Delta = \Sigma\pi_i R(\pi_i)$ a dichotomous index with a Borel measurable rarity measure R. Then $\Delta : SP^{(\infty)} \to [-\infty,\ \infty]$ is measurable whenever it is defined, in particular, whenever R is nonnegative.

Theorem A.7: The mapping $SP^{\infty} \to SP^{\infty} : \underset{\sim}{x} \to \underset{\sim}{x}^{*}$ which rearranges the components of $\underset{\sim}{x}$ in descending order is measurable.

Proof: For example, $\{x_1^* \le t_1, x_2^* \le t_2, x_3^* \le t_3\}$ is the intersection of the following events:

$$\text{(i)} \quad \bigcap_{i=1}^{\infty} \{x_i \le t_1\},$$

$$\text{(ii)} \quad \bigcup_{i=1}^{\infty} \bigcap_{\substack{j=1 \\ j \ne i}}^{\infty} \{x_j \le t_2\},$$

$$\text{(iii)} \quad \bigcup_{\substack{i,j=1 \\ i \ne j}}^{\infty} \bigcap_{\substack{k=1 \\ k \ne i,j}}^{\infty} \{x_k \le t_3\}.$$

Theorem A.8: Let $\Delta = \Sigma \pi_i^* R(i)$ be a rank type index. Then $\Delta : SP^{(\infty)} \to [-\infty, \infty]$ is measurable whenever it is defined, in particular, whenever R is nonnegative.

The size-biased permutation requires the auxiliary uniform random variables as described in Example 2.6. The sample space of these uniform random variables is the Hilbert cube equipped with the product of uniform measures. The sample space of the size-biased permutation is thus the product $SP^{(\infty)} \times I^{(\infty)}$. It can be shown that the size-biased permutation is measurable but we do not give the details. Similarly, random deletion of species is a measurable operation.

A.4 Asymptotic Geometricity

Let $0 < \theta < 1$ and $a_1, a_2, a_3, \ldots$ a sequence of positive numbers. Write $a_i = b_i \theta^i$ where the b_i are positive.

Theorem A.9: The sequence a_i is

(i) strong AG with asymptotic ratio $\theta \Leftrightarrow \lim_{i\to\infty} b_i$ exists and

$0 < \lim b_i < \infty$,

(ii) intermediate AG with asymptotic ratio $\theta \Leftrightarrow \lim_{i\to\infty} b_i/b_{i+1} = 1$,

(iii) weak AG with asymptotic ratio $\theta \Leftrightarrow \lim_{i\to\infty} \log(b_i)/i = 0$.

Theorem A.10: Strong AG => intermediate AG => weak AG.

Proof: The first implication is clear from Theorem A.9. For the second implication, take logarithms of Theorem A.9 (ii) to find that $\log(b_i) - \log(b_{i+1})$ converges to zero which implies that the Caesaro means of these differences also converge to zero. Telescoping then shows that $\log(b_i)/i$ goes to zero.

Theorem A.11: The series Σa_i converges whenever a_i is weak AG.

Proof: Write $\log(a_i)/i = \log(\theta) + \varepsilon_i$ where ε_i converges to zero. Then $a_i = \theta^i \exp(i\cdot\varepsilon_i)$ and $\sqrt[i]{a_i} = \theta \exp(\varepsilon_i)$. Thus $\lim \sqrt[i]{a_i} = \theta < 1$ and Σa_i converges by the root test.

Example A.1: The negative binomial series

$$a_i = \binom{k + i - 1}{i}$$

is intermediate AG for all k but is strong AG <=> k = 1. Use Theorem A.9.

Example A.2: The log series θ^i/i is intermediate AG but not strong AG.

Example A.3: The Poisson series $\theta^i/i!$ is not AG in any sense. Use Theorem A.9 (iii) and Stirling's formula.

Theorem A.12: The sequence $a_i = b_i\theta^i$ is weak AG whenever the b_i are bounded or, more generally, whenever $b_i \le i^p$ where p is a constant.

Proof: Use Theorem A.9 (iii).

Example A.4: The sequence a_i is weak AG but not intermediate AG when $a_1 = \theta$, $a_2 = 10\ \theta^2$, $a_3 = \theta^3$, $a_4 = 10\ \theta^4$, $a_5 = \theta^5$, $a_6 = 10\ \theta^6, \ldots$.

Example A.5: Let $U_1, U_2, U_3, \ldots$ be a random sample from the Uniform (0,1) distribution and put $a_i = U_i\theta^i$. With probability one, a_i is weak AG but also with probability one, a_i is not intermediate AG.

Example A.6: The zeta series $1/i^p$ is not AG in any sense.

A.5 Screening the Geometric Series Model

Let $X_1, X_2, X_3, \ldots$ be independent geometric random variables with $P(X_i = x) = p(1 - p)^{x-1}$, $x = 1,2,3,\ldots$. Suppose $\underset{\sim}{\nu}$ is a fixed community with infinitely many species and define $\tau_i = \nu_{X_1+X_2+\ldots X_i}$, $i = 1,2,3,\ldots$ and $\tau = \tau_1 + \tau_2 + \tau_3 + \ldots$. Let $\underset{\sim}{\pi}$ be the fixed community with $\pi_i = E[\tau_i]/E[\tau] = E[\tau_i]/p$. It might be more natural to take π_i as $E[\tau_i/\tau]$ but this expectation is intractable since τ is not independent of the ratios τ_i/τ. To see this, suppose $\underset{\sim}{\nu}$ is a geometric progression with ratio 1/2. The dyadic expansion of τ is a sequence of zeros and ones and $X_1 + X_2 + \ldots + X_i$ is the waiting time to the occurrence of the ith one in this sequence. Thus τ uniquely determines the realized values of $X_1, X_2, \ldots$ and hence of $\tau_1/\tau, \tau_2/\tau, \ldots$. The second part of Theorem 2.11 does have an analogue, however. Let $G_{\underset{\sim}{\pi}}(t) = \Sigma\pi_i t^{i-1}$ and $G_{\underset{\sim}{\nu}}(t) = \Sigma\nu_i t^{i-1}$ be the generating functions of $\underset{\sim}{\pi}$ and $\underset{\sim}{\nu}$.

Theorem A.13: a) $G_{\underset{\sim}{\pi}}(t) = G_{\underset{\sim}{\nu}}(1 - p + pt)$, b) $\underset{\sim}{\nu}$ is a geometric progression with ratio θ and odds ratio $\alpha' = \theta/(1 - \theta)$ if and only if $\underset{\sim}{\pi}$ is a geometric progession with ratio $p\theta/(1 - \theta + p\theta)$ and odds ratio $p\alpha'$, and c) $\underset{\sim}{\nu}$ is strong AG with asymptotic odds ratio α' if and only if $\underset{\sim}{\pi}$ is strong AG with asymptotic odds ratio $p\alpha'$.

Proof: a) The equation $G_{\underset{\sim}{\pi}}(t) = G_{\underset{\sim}{\nu}}(1 - p + pt)$ can be rewritten as

$$E[\Sigma\ \tau_i t^{i-1}] = p\ G_{\underset{\sim}{\nu}}(1 - p + pt), \tag{A.1}$$

and this will be proved by the method of marks. Let $\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots$ be the sequence of selection variables as described in Example 2.13 and let $0 \leq t \leq 1$. The statement "choose a component of $\underset{\sim}{\nu}$" will mean that component i is chosen with probability ν_i. Now generate a sequence of observations on the selection variables and execute the following set of instructions:

```
choose a component of $\underset{\sim}{\nu}$, component i, say;
if $\varepsilon_i = 0$, make a mark and stop; otherwise;
do for $j = 1$ to $i - 1$;
    read the value of $\varepsilon_j$; if
    $\varepsilon_j = 0$, next j; otherwise;
    make a mark with probability
    $1 - t$; next j; stop.
```

Given the values of the selection variables, the conditional probability of no marks is $\Sigma\tau_i t^{i-1}$. Thus the left hand side of (A.1) is the probability of no marks for the compound experiment of first observing the selection variables and then executing the instructions. But the right hand side of (A.1) is also the probability of no marks when, in the above set of instructions, the phrase "read the value of" is changed to "generate an observation on." This proves a). Next suppose $\underset{\sim}{\nu}$ is a geometric progression with ratio θ so that $G_{\underset{\sim}{\nu}}(t) = (1 - \theta)/(1 - \theta t)$. Then a routine calculation shows that the generating function of the vector $\underset{\sim}{\pi}$ is

$$G_{\underset{\sim}{\pi}}(t) = G_{\underset{\sim}{\nu}}(1 - p + pt) = (1 - \theta)/(1 - \theta + \theta p - \theta pt)$$
$$= (1 - \overline{\theta})/(1 - \overline{\theta}t),$$

where $\overline{\theta} = \theta p/(1 - \theta + \theta p)$. Thus $\underset{\sim}{\pi}$ is a geometric progression with ratio $\overline{\theta}$. The converse of b) is proved in the same way. Finally, assume $\underset{\sim}{\nu}$ is strong AG with ratio θ so that ν_i/θ^i converges to a finite positive constant A. Given $0 < \delta < 1$, there is an i_0 such that $A(1 - \delta)\theta^i \leq \nu_i \leq A(1 + \delta)\theta^i$ whenever $i > i_0$. It follows that

$$A(1 - \delta)\theta^i/(1 - \theta t)^i \leq G_{\underset{\sim}{\nu}}^{(i-1)}(t)/(i - 1)! \leq A(1 + \delta)\theta^i/(1 - \theta t)^i,$$

whenever $i > i_0$. But

$$\pi_i = G_{\underset{\sim}{\pi}}^{(i-1)}(0)/(i - 1)! = p^{i-1}G_{\underset{\sim}{\nu}}^{(i-1)}(1 - p)/(i - 1)!$$

which implies that

$$Ap^{-1}(1 - \delta)\overline{\theta}^i \leq \pi_i \leq Ap^{-1}(1 + \delta)\overline{\theta}^i,$$

whenever $i > i_0$. Thus $\underset{\sim}{\pi}$ is strong AG with asymptotic ratio $\overline{\theta}$. The converse of c) is proved in the same way. This completes the proof of the theorem.

The above results assert that, on the average, randomly deleting a fraction 1 - p of the species from an asymptotically geometric series still leaves an asymptotically geometric series but whose asymptotic odds ratio is reduced to a fraction p of its former value. Thus the asymptotic odds ratio may be regarded as a species richness parameter for fixed models that are asymptotically geometric in the strong sense, just as Fisher's α is a species richness parameter for the limiting Dirichlet model. However, it is only for the rigidly geometric series that the odds ratio is a complete diversity parameter.

A.6 Engen's Jackknife Procedure

Consider an infinite multinomial population $\underset{\sim}{\pi} = (\pi_1, \pi_2, \pi_3, \ldots)$ where the π_i form a geometric progression with $\pi_i = \theta^{i-1}(1 - \theta)$, $i = 1,2,3,\ldots, 0 < \theta < 1$. Put $\alpha' = \Sigma i\pi_i - 1 = \theta/(1 - \theta)$. Let $\underset{\sim}{Y} = (Y_1, Y_2, Y_3, \ldots)$ be a random sample of size N drawn from $\underset{\sim}{\pi}$. When the species ranking is known, the MLE of α' is $\tilde{\alpha}' = \Sigma i(Y_i/N) - 1$ and is unbiased. Usually the ranking is unknown, in which case the MLE is $\hat{\alpha}' = \Sigma i(X^{(i)}/N) - 1$, where $X^{(1)} \geq X^{(2)} \geq \ldots$ are the descending order statistics of $\underset{\sim}{Y}$. Now $\hat{\alpha}' \leq \tilde{\alpha}'$ so that $\hat{\alpha}'$ is negatively biased. Let $b(N) = b(N, \theta) = \alpha' - E[\hat{\alpha}'] \geq 0$ be the magnitude of the bias. Engen [84] has made two assertions:

(i) the asymptotic bias is of order 1/N in the sense that $\limsup Nb(N) < \infty$ when $N \to \infty$,

and therefore,

(ii) the first order term in the bias can be corrected for by jackknifing.

The purpose of this appendix is to point out that (i) does not justify the jackknife procedure and that, in any case, Engen's proof of (i) is invalid.

The usual justification for jackknifing assumes that the bias has a Taylor expansion in powers of 1/N, which is a stronger requirement than (i). For an example of the consequences of jackknifing in the presence of (i), assume the bias has the form

$$b(N) = N^{-1}(1/\sqrt{1} - 1/\sqrt{2} + 1/\sqrt{3} - \ldots \pm 1/\sqrt{N}). \qquad (A.2)$$

Now Nb(N) is of order 1 since the alternating series converges. After jackknifing the bias becomes

$$\begin{aligned} b_J(N) &= Nb(N) - (N - 1)b(N - 1) \\ &= \pm 1/\sqrt{N}. \end{aligned}$$

Thus $|Nb_J(N)| \to \infty$ and jackknifing has increased the order of the bias. In fairness it has to be pointed out that the magnitude of the bias of $\hat{\alpha}'$ is monotone decreasing in N and could never have the form given in (A.2).

Next consider Engen's proof of (i). Letting $C(N) = Nb(N)$, Engen makes a clever use of the loss of memory property of the geometric distribution to show that $\lim E[C(N) - C(N - Y_1)] = 0$ as $N \to \infty$; but he then concludes that "C(N) is at least a periodic function of log N in the limit which is sufficient to conclude that C(N) is of order 1" (Engen [84, p. 699]). Apparently he has in mind replacing $N - Y_1$ by its asymptotic value $N(1 - \pi_1) = N\theta$ and using the (fallacious) result that

$$C(N) - C(N(1 - \pi_1)) \to 0 \Rightarrow \limsup C(N) < \infty.$$

That this line of argument is incorrect can be seen by taking $C(N) = \log\log(N + 2)$. Then, as $N \to \infty$, one has

a) $E[C(N) - C(N - Y_1)] \to 0,$

b) $C(N) - C(N(1 - \pi_1)) \to 0,$

but,

c) $C(N) \to \infty.$

Part (b) can be established by a routine application of L'Hospital's rule while (c) is obvious. For the proof of (a), first notice that

$$\begin{aligned} 0 \leq C(N) - C(N - Y_1) &= \log[\log(N + 2)/\log(N - Y_1 + 2)] \\ &\leq \log\,[\log(N + 2)/\log(2)] \\ &\leq N. \end{aligned}$$

To find an upper bound for the expectation, let $0 < \varepsilon < 1 - \pi_1$ and partition the sample space into the two regions $\{Y_1 \leq N\pi_1 + N\varepsilon\}$ and $\{Y_1 > N\pi_1 + N\varepsilon\}$. Over the first region, the integrand is bounded above by $\log[\log(N + 2)/\log(N(1 - \pi_1 - \varepsilon) + 2)]$ which goes to zero as $N \to \infty$. Over the second region the integrand is bounded by $N\,P(Y_1/N - \pi_1 > \varepsilon)$. But this expression is well-known to go to zero as $N \to \infty$. For completeness, a proof is sketched in Appendix A.11.

Remark A.1: The estimator $\hat{\alpha}'$ can be shown to take the form

$$N\,\hat{\alpha}' = \sum_{x<y} x n_x n_y + (1/2) \sum_x x n_x (n_x - 1), \tag{A.3}$$

where n_x, $x = 1,2,3,\ldots$, is the number of species represented by x individuals in the sample. Similarly, the jackknifed version of $\hat{\alpha}'$ is

$$N\,\hat{\alpha}'_J = \sum_{x<y} x n_x n_y + \sum_x x n_x (n_x - 1).$$

Clearly, $\hat{\alpha}' \leq \hat{\alpha}'_J$ which implies the signed bias of $\hat{\alpha}'_J$ exceeds the signed bias of $\hat{\alpha}'$. From this one easily proves a previous comment that $b(N)$ is decreasing in N.

Remark A.2: In light of Section 2.3, especially Figure 2.4, it is not unreasonable to think that $C(N)$ has the asymptotic form $a + D(N)$ where a is a positive finite constant and $D(N)$ is a periodic function of $-\log_\theta N$ for large N. Presumably this is what Engen is attempting to show. In principle, it should be possible to determine the value of a, and prove that it is finite, by taking expectations of (A.3), expressing

$E[n_x n_y]$ and $E[n_x(n_x - 1)]$ as infinite sums, replacing these infinite sums by integrals, and letting $N \to \infty$. Unfortunately, attempts to carry out this program have proved unsuccessful. As an indication of the difficulties, consider the problem of approximating $E[n_x]$. The exact expression is

$$E[n_x] = \sum_{i=1}^{\infty} \binom{N}{x} \pi_i^x (1 - \pi_i)^{N-x}.$$

Replacing the sum by an integral with respect to i and making a change of variable gives the approximation

$$E[n_x] \simeq \alpha\binom{N}{x} \int_0^{1/\alpha'} (1 - u)^{N-x} u^x du/u,$$

where $\alpha = -1/\log \theta$ is Fisher's α and $\alpha' = \theta/(1 - \theta)$ is the odds ratio. Now for any given x, the approximation gives the correct asymptotic value α/x as $N \to \infty$. But the approximation is not uniformly accurate in x. To see this, notice that $\Sigma\, x\, E[n_x]$ must be exactly N while the approximation gives $N\alpha/\alpha' > N$. This suggests using $1/\alpha$ as the upper limit of integration instead of $1/\alpha'$. In carrying out the above program, the asymptotic behavior of C(N) is found to depend critically upon the choice of the upper limit of integration and it will be difficult to justify rigorously any particular choice. Most likely, the "correct" choice depends on both x and N. It may also be remarked that formally, if not realistically, both $1/\alpha$ and $1/\alpha'$ can exceed 1 and thus be outside the usual range of the Beta integral.

At present it is unknown if jackknifing reduces the asymptotic bias. One must also be concerned with how large N should be before the asymptotic theory can be confidently applied.

A.7 The Geometric Series Model is More Diverse than the Exponential Integral Model

Theorem A.14: Let $\nu_i = E_1^{-1}(i/\alpha)$ and $\pi_i = \theta^i$, $i = 1,2,3,\ldots$, where $\alpha > 0$ and $\log(\theta) = -1/\alpha$. Then ν_i/π_i is decreasing in i which implies that $(\pi_1, \pi_2, \pi_3, \ldots)/\Sigma\pi_i$ is intrinsically more diverse than $(\nu_1, \nu_2, \ldots)/\Sigma\nu_i$.

Proof: Since $\nu_i/\pi_i = E_1^{-1}(i/\alpha)\exp(i/\alpha)$, it will suffice to show that $E_1^{-1}(t)\exp(t)$ is decreasing in t. Write $y = E_1^{-1}(t)$, $E_1(y) = t$ and put $f(y) = y \exp(E_1(y))$. But

$$\begin{aligned} f'(y) &= \exp(E_1(y)) + y \exp(E_1(y))dE_1(y)/dy \\ &= \exp(E_1(y))[1 - \exp(-y)] \geq 0. \end{aligned}$$

Thus f is increasing in y. Since t is a decreasing function of y, the proof is complete.

A.8 Proof of Lemma 2.1

We use a well known series expansion of the digamma function (Abramowitz & Stegun [89, equation 6.3.16]):

$$\psi(z + 1) = \Gamma'(z + 1)/\Gamma(z + 1) = -\gamma + \sum_{n=1}^{\infty} z/[n(n + z)],$$

where $\gamma = .5772\ldots$ is Euler's constant. Define $f(t) = \Gamma(B_1t + 1)\Gamma(B_2t + 1)/[\Gamma(A_1t + 1)\Gamma(A_2t + 1)]$, $t \geq 0$. Take logarithm derivatives to find that

$$\begin{aligned} f'(t)/f(t) &= B_1\psi(B_1t + 1) + B_2\psi(B_2t + 1) - A_1\psi(A_1t + 1) - A_2\psi(A_2t + 1) \\ &= \sum_{n=1}^{\infty} (t/n)[B_1^2/(n+B_1t) + B_2^2/(n+B_2t) - A_1^2/(n+A_1t) - A_2^2/(n+A_2t)]. \end{aligned}$$

After a lengthy calculation, this becomes

$$f'(t)/f(t) = -(B_1 - A_1)t \sum_{n=1}^{\infty} [n^2(A_2 + B_2 - A_1 - B_1) + nt(A_2B_2 - A_1B_1)] \div$$

$$[(n + A_1t)(n + A_2t)(n + B_1t)(n + B_2t)].$$

Since this expression is strictly negative for $t > 0$, $f(t)$ is strictly decreasing and $f(1) < f(0) = 1$.

A.9 The Distribution of Total Abundance for the Hyperbolic Process

A method is given which can, in principle, be used to determine the probability density function $f_\alpha(z)$, $0 < z < \infty$, of the total abundance ξ_α for the Hyperbolic process. This density function exhibits certain pecularities. For example, $f_\alpha(z)$ may fail to be continuously differentiable when z is a positive integer, although $f(z)$ is analytic on each of the intervals (0, 1), (1, 2), (2, 3),... .

Using Theorem 4.3, the joint density of $\lambda_1^{\#}$ and ξ_α is

$$f_{\lambda_1^{\#},\xi_\alpha}(u, z) = \begin{cases} \alpha u f(u) f_\alpha(z - u)/z & \text{if } 0 < u < z \\ 0 & \text{otherwise.} \end{cases}$$

Integrating out u shows that $f_\alpha(z)$ satisfies the functional equation

$$z f_\alpha(z) = \alpha \int_0^z u f(u) f_\alpha(z - u) du. \tag{A.4}$$

But $f(u) = 1/u$, $0 < u < 1$. Thus, if $0 < z \leq 1$,

$$z f_\alpha(z) = \alpha \int_0^z f_\alpha(z - u) du = \alpha \int_0^z f_\alpha(V) dV. \tag{A.5}$$

Now (A.5) implies that f_α is continuous for $0 < z < 1$ and hence that f_α is differentiable for $0 < z < 1$. Taking the derivative of (A.5) with respect to z gives the differential equation $zf_\alpha'(z) + f_\alpha(z) = \alpha f_\alpha(z)$, whose general solution is

$$f_\alpha(z) = c\ z^{\alpha-1},\ 0 < z \le 1,\ c > 0.$$

When $1 \le z \le 2$ (A.4) gives

$$zf_\alpha(z) = \alpha \int_0^1 f_\alpha(z - u)du = \alpha \int_{z-1}^{z} f_\alpha(V)dV. \tag{A.6}$$

Again $f_\alpha(z)$ is differentiable for $1 < z < 2$ and comparison of (A.5) with (A.6) shows that $f_\alpha(z)$ is continuous at $z = 1$. Differentiating (A.6) gives the differential equation $zf_\alpha'(z) + f_\alpha(z) = \alpha f(z) - \alpha f(z - 1) = \alpha f(z) - \alpha c(z - 1)^{\alpha-1}$. Since $f(1) = c$, the solution of this differential equation is

$$f_\alpha(z) = cz^{\alpha-1}[1 - \int_1^z u^{-\alpha}(u - 1)^{\alpha - 1}du],\ 1 \le z \le 2.$$

Note that (A.6) holds for $1 \le 2 < \infty$, so that the method can be iterated to determine $f_\alpha(z)$ for all z. The required integrals become increasingly complex, however.

A.10 Proof that the Sample Process is Compound Poisson

The compound Poisson process with intensity I and with positive integer valued jump distribution p_x, $x = 1,2,3,\ldots$, has probability generating function

$$\exp[-tI(1 - \sum_{x=1}^{\infty} g^x p_x)], \tag{A.7}$$

where t is the time and g is the generating symbol.

Let N_t be a Poisson process with intensity A. The sample process may be obtained as N_{ξ_t}, $0 \leq t \leq \alpha$. The interpretation is as follows: the events of the process N_t represent the individuals entering the sample; these individuals are then grouped into species by the subordinator ξ_t, $0 \leq t \leq \alpha$.

The probability generating function of the sample process is

$$\begin{aligned} E[g^{N_{\xi_t}}] &= E[E(g^{N_T}|\xi_t = T)] \\ &= E[\exp(-TA(1 - g))|\xi_t = T] \\ &= E[\exp(-A(1 - g)\xi_t]= \exp[-tM(A(1 - g))], \end{aligned} \qquad (A.8)$$

where M is the Levy transform (refer to (4.1)). But

$$M(A(1 - g)) = \int_0^\infty [1 - \exp(-Au + Agu)]\ f(u)du.$$

Adding and subtracting exp(-Au) under the integral gives

$$\begin{aligned} M(A(1 - g)) &= M(A) - \int_0^\infty [\exp(Agu) - 1]\ \exp(-Au)f(u)du \\ &= M(A)[1 - \sum_{x=0}^{\infty} g^x \int_0^\infty A^x u^x \exp(-Au)f(u)du/(x!M(A))]. \end{aligned} \qquad (A.9)$$

Comparison of (A.8) and (A.9) with (A.7) shows that the sample process is compound Poisson with intensity M(A) and with jump distribution given by (4.17).

A.11 Asymptotic Nullity of the Binomial Tail

<u>Theorem A.15</u>: Let $Y \sim$ Binomail (N, p) and let $\varepsilon > 0$. Then, for any real number r,

$$\lim_{N \to \infty} N^r P(Y/N - p > \varepsilon) = 0.$$

This theorem can be (and often is) proved by employing the asymptotic approximation to the normal integral. The following proof uses only Stirling's formula. Now $P(Y/N - p > \varepsilon) = P(Y > N(p + \varepsilon))$ is a sum of binomial probabilities and, since $N(p + \varepsilon)$ exceeds the mean, these probabilities are decreasing. Furthermore, there are no more than N such probabilities. Write $q = p + \varepsilon > p$ and put y equal to the integral part of Nq. Then $N^r P(Y/N - p > \varepsilon) \leq N^{r+1} \binom{N}{y} p^y (1 - p)^{N-y}$. By Stirling's formula, this upper bound is asymptotic to a constant times

$$N^{r+.5}\{(p/q)^q[(1 - p)/(1 - q)]^{1-q}\}^N. \qquad \text{(A.10)}$$

With q fixed, let f(p) be the expression in braces. Taking the logarithmic derivative, f(p) is found to be strictly increasing for $p \leq q$. Thus $f(p) < f(q) = 1$ and (A.10) is a polynomial infinity times an exponential zero, which goes to zero.

BIBLIOGRAPHY

[1] Wallace, A. R., *Tropical Nature and Other Essays*, New York: AMS Press Inc., 1975 (Reprint of the 1875 ed. published by Macmillan and Co., London).

[2] Aczél, J., *Lectures on Functional Equations and Their Applications*, New York: Academic Press, Inc., 1966.

[3] Simpson, E. H., "Measurement of Diversity," *Nature*, 163 (April 30, 1949), 688.

[4] MacArthur, R. H., "Patterns of Species Diversity," *Biological Reviews of the Cambridge Philosophical Society*, 40, No. 4 (1965), 510-533.

[5] Hurlbert, S. H., "The Nonconcept of Species Diversity: A Critique and Alternative Parameters," *Ecology*, 52, No. 4 (1971), 577-586.

[6] Hardy, G. H., Littlewood, J. E. and Polya, G., *Inequalities*, London: Cambridge University Press, 1934.

[7] Smith, W., "Sampling Properties of a Family of Diversity Measures," Contribution No. 3558, Woods Hole Oceanographic Institution, 1976.

[8] Adelman, M. A., "Comment on the H Concentration Measure as a Numbers-Equivalent," *The Review of Economics and Statistics*, 51, No. 1 (1969), 99-101.

[9] Hart, P. E., "Entropy and Other Measures of Concentration," *Journal of the Royal Statistical Society*, *Ser. A.*, 134, Part I (1971), 73-85.

[10] Kemp, A. W., "On Entropic Measures of Industrial Diversification,"

Statistics Reports and Preprints No. 21, School of Mathematics, University of Bradford, 1975.

[11] Renyi, A., "On Measures of Entropy and Information," in J. Neyman, ed., Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. I, Berkeley: The University of California Press, 1961, 547-561.

[12] Hill, M. O., "Diversity and Evenness: A Unifying Notation and its Consequences," Ecology, 54, No. 2 (1973), 427-432.

[13] Dalton, H., "The Measurement of the Inequality of Incomes," Economic Journal, 30 (September 1920), 348-361.

[14] Good, I. J., "The Population Frequencies of Species and the Estimation of Population Parameters," Biometrika, 40 (December 1953), 237-264.

[15] Solomon, D. L., "A Mathematical Foundation for Ecological Diversity," Annual Report, Oak Ridge National Laboratory, 1975, 24-25.

[16] Marshall, A. W. and Olkin, I., "Majorization in Multivariate Distributions," The Annals of Statistics, 2, No. 6 (1974), 1189-1200.

[17] Dasgupta, P., Sen, A. and Starrett, D., "Notes on the Measurement of Inequality," Journal of Economic Theory, 6, No. 2 (1973), 180-187.

[18] Mirsky, L., "Results and Problems in the Theory of Doubly Stochastic Matrices," Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete, 1 (1963), 319-334.

[19] Rado, R., "An Inequality," Journal of the London Mathematical Society, 27, No. 105 (1952), 1-6.

[20] Rota, G. C. and Harper, L. H., "Matching Theory, An Introduction," in P. Ney, ed., Advances in Probability and Related Topics, Vol. 1, New York: Marcel Dekker, Inc., 1971, 169-215.

[21] Lewontin, R. C., "The Apportionment of Human Diversity," Evolutionary Biology, 6 (1972), 381-398.

[22] Fager, E. W., "Diversity: A Sampling Study," The American Naturalist, 106, No. 949 (1972), 293-310.

[23] Peet, R. K., "Relative Diversity Indices," Ecology, 56, No. 2 (1975), 496-498.

[24] Lehmann, E. L., Testing Statistical Hypotheses, New York: John Wiley & Sons, Inc., 1959.

[25] Peet, R. K., "The Measurement of Species Diversity," in R. F. Johnston, ed., Annual Review of Ecology and Systematics, Vol. 5, Palo Alto, Calif.: Annual Reviews, Inc., 1974, 285-307.

[26] Allan, J. D., "Components of Diversity," Oecologia, 18 (1975), 359-367.

[27] Pielou, E. C., "The Use of Information Theory in the Study of the Diversity of Biological Populations," in L. M. LeCam and J. Neyman, eds., Proceedings of the Fifth Berkeley Symposium on Mathamatical Statistics and Probability, Vol. IV, Berkeley: The University of California Press, 1967, 163-177.

[28] Lloyd, M., Inger, R. F. and King, F. W., "On the Diversity of Reptile and Amphibian Species in a Bornean Rain Forest," The American Naturalist, 102, No. 928 (1968) 497-515.

[29] Wilson, J. W., "Analytical Zoogeography of North American Mammals," Evolution, 28, No. 1 (1974), 124-140.

[30] Theil, H., Economics and Information Theory, Amsterdam: North-Holland Publishing Co., 1967.

[31] Theil, H., Statistical Decomposition Analysis, Amsterdam: North Holland Publishing Co., 1972.

[32] Nei, M., "Analysis of Gene Diversity in Subdivided Populations," Proceedings of the National Academy of Sciences, USA, 70 (December 1973), 3321-3323.

[33] Pielou, E. C., Ecological Diversity, New York: John Wiley & Sons, Inc., 1975.

[34] Aczél, J. and Daroczy, Z., On Measures of Information and Their Characterizations, New York: Academic Press, Inc., 1975.

[35] Colwell, R. K. and Futuyma, D. J., "On the Measurement of Niche Breadth and Overlap," Ecology, 52, No. 4 (1971), 567-576.

[36] Horn, H. S., "Measurement of Overlap in Comparative Ecological Studies," The American Naturalist, 100, No. 914 (1966), 419-424.

[37] Pielou, E. C., "Niche Width and Niche Overlap: A Method for Measuring Them," Ecology, 53, No. 4 (1972), 687-692.

[38] MacArthur, R. H., Geographical Ecology, New York: Harper & Row, Publishers, Inc., 1972.

[39] Lee, P. M., "On the Axioms of Information Theory," The Annals of Mathematical Statistics, 35, No. 1 (1964), 415-418.

[40] Sibson, R., "Information Radius," Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 14 (1969), 149-160.

[41] Gove, P. B. (ed), Webster's Third New International Dictionary

of the English Language, Springfield, Mass.: G. C. Merriam Co., 1967, 759.

[42] Elandt-Johnson, R. C., Probability Models and Statistical Methods in Genetics, New York: John Wiley & Sons, Inc., 1971.

[43] Dobzhansky, T., Genetics of the Evolutionary Process, New York: Columbia University Press, 1970.

[44] Kempthorne, O., An Introduction to Genetical Statistics, New York: John Wiley & Sons, 1957.

[45] Hoeffding, W., "Masstabinvariante Korrelationstheorie," Schriften des Mathematischen Instituts der Universität Berlin, 5 (1940), 181-233.

[46] Tchen, A. H., "Inequalities for Distributions with Given Marginals," Technical Report No. 19, Department of Statistics, Stanford University, 1976.

[47] Cairns, J., "The Sequential Comparison Index - a Simplified Method for Non-Biologists to Estimate Relative Differences," Journal of the Water Pollution Control Federation, 40, No. 9 (1968), 1607-1613.

[48] Cairns, J., and Dickson, K. L., "A Simple Method for the Biological Assessment of the Effects of Waste Discharges on Aquatic Bottom Dwelling Organisms," Journal of the Water Pollution Control Federation, 43, No. 5 (1971), 755-772.

[49] Noether, G. E., "A Central Limit Theorem with Nonparametric Applications," Annals of Mathematical Statistics, 41, No. 5 (1970), 1753-1755.

[50] Mood, A. M., "The Distribution Theory of Runs," The Annals of Mathematical Statistics, 11, No. 4 (1940), 367-392.

[51] Bowman, K. O., Hutcheson, K., Odum, E. P. and Shenton, L. R., "Comments on the Distribution of Indices of Diversity," in G. P. Patil, E. C. Pielou and W. E. Waters, eds., Statistical Ecology, Vol. 3, University Park: The Pennsylvania State University Press, 1971, 315-366.

[52] Nei, M. and Roychoudhury, A. K., "Sampling Variances of Heterozygosity and Genetic Distance," Genetics, 76 (February 1974), 379-390.

[53] Bliss, C. I., "An Analysis of Some Insect Trap Records," in G. P. Patil, ed., Classical and Contagious Discrete Distributions, London: Pergamon Press, 1963, 385-397.

[54] Bulmer, M. G., "On Fitting the Poisson Lognormal Distribution to Species-Abundance Data," Biometrics, 30 (March 1974), 101-110.

[55] Robbins, H., "Estimating the Total Probability of the Unobserved Outcomes of an Experiment," The Annals of Mathematical Statistics, 39, No. 1 (1968), 256-257.

[56] Greenberg, J. H., "The Measurement of Linguistic Diversity," Language, 32, No. 1 (1956), 109-115.

[57] Lieberson, S., "An Extension of Greenberg's Linguistic Diversity Measures," Language, 40, No. 4 (1964), 526-631.

[58] Lieberson, S., "Measuring Population Diversity," American Sociological Review, 34, No. 6 (1969), 850-862.

[59] Guiraud, P., Problèmes et Méthodes de la Statistique Linguistique, Dordrecht: D. Reidel Publishing Co., 1959.

[60] Herdan, G., Quantitative Linguistics, Washington, D. C.: Butterworth, Inc., 1964.

[61] Herdan, G., The Advanced Theory of Language as Choice and Chance, New York: Springer-Verlag, 1966.

[62] Yule, G. U., The Statistical Study of Literary Vocabulary, London: Cambridge University Press, 1944.

[63] Brainerd, B., "On the Relation Between Types and Tokens in Literary Text," Journal of Applied Probability, 9, No. 3 (1972), 507-518.

[64] Horowitz, I., "Employment Concentration in the Common Market: an Entropy Approach," Journal of the Royal Statistical Society, Ser. A, 133, Part 3 (1970), 463-475.

[65] Finkelstein, M. O. and Freidberg, R. M., "The Application of an Entropy Theory of Concentration to the Clayton Act," Yale Law Journal, 76, No. 4 (1967), 677-717.

[66] Hexter, J. L. and Snow, J. W., "An Entropy Measure of Relative Aggregate Concentration," Southern Economic Journal, 36, No. 3 (1970), 239-243.

[67] Horowitz, A. and Horowitz, I., "Entropy, Markov Processes and Competition in the Brewing Industry," Journal of Industrial Economics, 16, No. 3 (1968), 196-211.

[68] Hall, M. and Tideman, N., "Measures of Concentration," Journal of the American Statistical Association, 62 (March 1967), 162-168.

[69] Hart, P. E., "Moment Distributions in Economics: An Exposition," Journal of the Royal Statistical Society, Ser. A, 138, Part 3 (1975), 423-434.

[70] Kendall, M. G. and Stuart, A., The Advanced Theory of Statistics, Vol. 1, London: Charles Griffin & Co., Ltd., 1958.

[71] Thompson, W. A., Jr., "Fisherman's Luck," Biometrics, 32 (June 1976), 265-271.

[72] Kolm, S. Ch., "The Optimal Production of Social Justice," in J. Margolis and H. Guitton, eds., Public Economics, London: Macmillan and Co., Ltd., 1969, 145-200.

[73] Sen, A., "Poverty, Inequality and Unemployment: Some Conceptual Issues in Measurement," Sankhyā, Series C, 36, Parts 2 & 4 (1974), 67-82.

[74] Rathie, P. N. and Kannappan, Pl., "A Directed Divergence Function of Type β," Information and Control, 20, No. 1 (1972), 38-45.

[75] Bhattacharyya, A., "On a Measure of Divergence Between Two Multinomial Populations," Sankhyā, 7, Part 4 (1946), 401-406.

[76] Chiswick, C. U., "Application of The Theil Index to Income Inequality," Working Paper Series B-2, Development Research Center, The World Bank, 1976.

[77] Atkinson, A. B., "On the Measurement of Inequality," Journal of Economic Theory, 2, No. 3 (1970), 244-263.

[78] Jardine, N. and Sibson, R., Mathematical Taxonomy, London: John Wiley & Sons, Ltd., 1971.

[79] Kingman, J. F. C., "Random Discrete Distributions," Journal of the Royal Statistical Society, Ser. B, 37, No. 1 (1975), 1-15.

[80] MacArthur, R. H., "On the Relative Abundance of Bird Species," Proceedings of the National Academy of Sciences, USA, 43 (March 1957), 293-295.

[81] Cohen, J. E., "Alternate Derivations of a Species-Abundance Relation," The American Naturalist, 102, No. 924 (1968), 165-172.

[82] Feller, W., An Introduction to Probability Theory and its Applications, Vol. 2, New York: John Wiley & Sons, Inc., 1971.

[83] Kotlarsky, I., "On Groups of n Independent Random Variables Whose Product Follows the Beta Distribution," Colloquium Mathematicum, 9, Fasc. 2 (1962), 325-332.

[84] Engen, S., "A Note on the Geometric Series as a Species Frequency Model," Biometrika, 62, No. 3 (1975), 697-699.

[85] Fisher, R. A., "The Relation Between the Number of Species and the Number of Individuals in a Random Sample From an Animal Population: Part 3. A Theoretical Distribution for the Apparent Abundance of Different Species," Journal of Animal Ecology, 12, No. 1 (1943), 54-57.

[86] Kempton, R. A. and Taylor, L. R., "Log-series and Log-normal Parameters as Diversity Discriminants," Journal of Animal Ecology, 43, No. 2 (1974), 381-399.

[87] Cox, D. R., "Discussion of Professor Kingman's Paper," Journal of the Royal Statistical Society, Ser. B, 37, No. 1 (1975), 18.

[88] Engen, S., "On Species Frequency Models," Biometrika, 61, No. 2 (1974), 263-270.

[89] Abramowitz, M. and Stegun, I. A., Handbook of Mathematical Functions, New York: Dover Publications, Inc., 1965.

[90] Watterson, G. A., "Models for the Logarithmic Species Abundance Distributions," Theoretical Population Biology, 6, No. 2 (1974), 217-250.

[91] Holgate, P., "Species Frequency Distributions," Biometrika, 56, No. 3 (1969), 651-660.

[92] Anscombe, F. J., "Sampling Theory of the Negative Binomial and the Logarithmic Series Distribution," Biometrika, 37, Parts 3 & 4 (1950) 358-382.

[93] Willis, J. C., Age and Area, London: Cambridge University Press, 1922.

[94] Chamberlin, J. C., "The Hollow Curve of Distribution," The American Naturalist, 58, No. 698 (1924), 350-374.

[95] Kempton, R. A., "A Generalized Form of Fisher's Logarithmic Series," Biometrika, 62, No. 1 (1975), 29-38.

[96] Nevius, S. E., Proschan, F. and Sethuraman, J., "Schur Functions in Statistics: Stochastic Majorization," The Annals of Statistics, 5, No. 2 (1977), 263-273.

[97] Proschan, F. and Sethuraman, J., "Schur Functions in Statistics," The Annals of Statistics, 5, No. 2 (1977), 256-262.

[98] Preston, F. W., "The Commonness, and Rarity of Species," Ecology, 29, No. 3 (1948), 254-283.

[99] Preston, F. W., "The Canonical Distribution of Commonness and Rarity," Ecology, 43, No. 2 & 3 (1962), 185-215, 410-432.

[100] May, R. M., "Patterns of Species Abundance and Diversity," in M. L. Cody and J. M. Diamond, eds., Ecology and Evolution of Communities, Cambridge, Mass.: The Belknap Press of Harvard University Press, 1975, 81-120.

[101] Rao, C. R., "On Discrete Distributions Arising Out of Methods of Ascertainment," in G. P. Patil, ed., Classical and Contagious Discrete Distributions, London: Pergamon Press, 1965, 320-332.

[102] Patil, G. P. and Rao, C. R., "Weighted Distributions and Size Biased Sampling with Applications to Wildlife Populations and Human Families," submitted to Biometrics.

[103] Patil, G. P. and Rao, C. R., "The Weighted Distributions: A Survey of Their Applications," in P. R. Krishnaiah, ed., Proceedings of the International Conference on Applications of Statistics, Amsterdam: North Holland Publishing Co., to appear.

[104] Kingman, J. F. C., "Completely Random Measures," Pacific Journal of Mathematics, 21, No. 1 (1967), 59-78.

[105] Breiman, L., Probability, Reading, Mass.: Addison-Wesley Publishing Co., 1968.

[106] Kallenberg, O., Random Measures, Berlin: Akademia-Verlag, 1975.

[107] Grandell, J., Doubly Stochastic Poisson Processes, New York: Springer-Verlag, 1976.

[108] McCloskey, J. W., "A Model for the Distribution of Species by Individual in an Environment," Technical Report No. 7, Department of Statistics, Michigan State University, 1965.

[109] Shorrock, R. W., "Extremal Processes and Random Measures," Journal of Applied Probability, 12, No. 2 (1975), 316-323.

[110] Kadane, J. B., "A Moment Problem for Order Statistics," The Annals of Mathematical Statistics, 42, No. 2 (1971), 745-751.

[111] Neyman, J., "On a New Class of 'Contagious' Distributions, Applicable in Entomology and Bacteriology," The Annals of Mathematical Statistics, 10, No. 1 (1939), 35-57.

[112] Rao, C. R., "Some Comments on the Logarithmic Series Distribution in the Analysis of Insect Trap Data," in G. P. Patil, E. C. Pielou

and W. E. Waters, eds., Statistical Ecology, Vol. 1, University Park: The Pennsylvania State University Press, 1971, 131-142.

[113] Boswell, M. T. and Patil, G. P., "Chance Mechanisms Generating the Logarithmic Series Distributions Used in the Analysis of Numbers of Species and Individuals," in G. P. Patil, E. C. Pielou and W. E. Waters, eds., Statistical Ecology, Vol. 1, University Park: The Pennsylvania State University Press, 1971, 99-127.

[114] Joanes, D. and Gill, C., "The Sampling Properties of an Estimate of the Diversity of an Ecological Community," Research Report, Department of Statistics, University of Leeds, 1976.

VITA

Name: Charles Taillie Jr.

Birthplace: Great Falls, Montana, U.S.A. 1943

Education: A. B., with Highest Distinction, in Mathematics and Physics, University of Rochester, 1965.

Professional and Academic Experience:

1975-1977	Graduate Fellow, The Pennsylvania State University
1974-1975	Graduate Teaching Assistant, Department of Statistics, The Pennsylvania State University
1970-1974	Assistant Professor, Department of Mathematics, The State University of New York at Oneonta
1968-1970	Graduate Teaching Assistant, Department of Mathematics, Cornell University
1965-1968	Andrew S. White Fellow, Cornell University
1962-1965	General Motors Scholarship, University of Rochester
1961-1962	New York State Regents Scholarship, University of Rochester
1961	National Science Foundation Summer Traineeship in Mathematics, Cornell University.

Professional and Honorary Affiliations:

American Mathematical Society

American Statistical Association

Institute of Mathematical Statistics

Mathematical Association of America

Society of Naturalists

Phi Beta Kappa