# Appendix A Mathematical Notes and Computer Programs

This appendix lists a number of mathematical theorems and formulae that are used in the text of the book, but have proofs that do not contribute to the narrative, as such. The first section consists of mathematical notes and a following section lists all the computer programs that have been used in logistic-J theory research.

## Mathematical Notes

### A.1 Useful formulas

The following formulas are used in the mathematical developments that appear in this book. To search for a particular formula or expression, use the chapter and section number where it appears in order to find it here.

### A.1.1 Integration by substitution [ref 1.1.2, 9.1]

The integral below appears in the formula for the mean of the logistic-J distribution in Section 1.1.2.

$$c\int_{0}^{\Delta-\varepsilon} x/(x+\varepsilon)\,dx$$

Integrals that do not correspond to standard integration formulas may sometimes be solved by a substitution of variables, as here. Substituting for the variable x, using x' = x + ε, we obtain

$$\begin{aligned} c\int_{\varepsilon}^{\Delta} (x'-\varepsilon)/x')dx' &= c\int_{\varepsilon}^{\Delta}(1 - e/x')\,dx' \\ &= c[x' -\varepsilon\ln(x')]\,\Big|_{\varepsilon}^{\Delta} \\ &= c((D - \varepsilon) - \varepsilon\ln(\Delta/\varepsilon)) \end{aligned}$$

In both cases where this formula is used, the integrand also contains the term $\delta x$, but that is readily integrated separately as $\delta x^2/_2$.

### A.1.2 Summation of a power series [ref 9.2.1, A.6]

The infinite series $\sum x^k$ converges when $x < 1$, the sum being given in this case by the formula,

$$\sum x^k = 1/(1 - x)$$

The most elementary proof is obtained by simply carrying out the division implicit in the expression

$$1/(1 - x),$$

namely,

$$\begin{array}{ll} & \underline{1 + x + x^2 + \ \ldots\ } \\ 1/(1 - x): \quad 1 - x) & 1 \\ & 1 - x \\ & ---- \\ & x \\ & x - x^2 \\ & ------- \\ & \quad x^2 \ldots \end{array}$$

If $x = 1$, then the expression has the value $\infty$.

### A.2 Derivation of the formula for the variance [ref. 2.1.2]

For the sake of simplicity, we will employ the formula,

$$\mathrm{Var}(x) = E[(x - m)^2] = E\,[x^2] - m^2$$

$$= c\int(x^2/(x+e) - x^2d)dx - m^2$$

Transform the variable: $x' = x+e$, with $x'$ running from $e$ to $D$.

$$\mathrm{Var}(x) = c\int(x'-e)^2/x' - (x'-e)^2ddx' - m^2$$

$$= c\int(x'^2 - 2x'e + e^2)/x'dx - cd\int(x'-e)^2dx' - m^2$$

$$= c\int(x' - 2e + e^2/x')dx - cd(x'-e)^3/3)| - m^2$$

$$= c(x'^2/2 - 2x'e + e^2\ln(x')| - cd(x'-e)^3/3)| - m^2$$

$$= c(x'^2/2 - 2x'e + e^2\ln(x')| - cd((D'-e)^3/3) - (e'-e)^3/3)) - m^2$$

$$= c(x'^2/2 - 2x'e + e^2\ln(x')| - cd(D'-e)^3/3) - m^2$$

$$= c(D^2/2 - 2De + e^2(\ln(D) - (e^2/2 - 2e^2 + e^2\ln(e)) - cd(D'-e)^3/3) - m^2$$

$$= c[(D^2-e^2)/2 - 2(D-e)e + e^2(\ln(D)-\ln(e)) - d(D'-e)^3/3] - m^2$$

$$= c[(D^2-e^2)/2 - 2(D-e)e + e^2(\ln(D/e)) - d(D'-e)^3/3] - m^2$$

with $\ln(D/e) = 1/c + 1$ we have,

$$\mathrm{Var}(x) = c[(D^2-e^2)/2 - 2(D-e)e + e^2(1/c+1) - d(D'-e)^3/3] - m^2$$

$$= c[(D^2-e^2)/2 - 2(D-e)e + e^2 - d(D'-e)^3/3] + e^2 - m^2$$

$$= c[(D^2-e^2)/2 - 2De + 3e^2 - d(D-e)^3/3] + e^2 - m^2$$

$$= c[(D^2/2 - 2De + (2.5)e^2 - d(D-e)^3/3] + e^2 - m^2$$

The mean m has the formula,

$$m = c((D^2 - e^2)/2D \ - e\ln(D/e))$$

So that $m2 = c^2[(D^2 - e^2)/2D \ - e\ln(D/e)]\cdot[(D^2 - e^2)/2D \ - e\ln(D/e)]$

$$\approx c^2[D^2/4 - (1/c + 1)eD + e^2(1/c + 1)^2]$$

### A.3 Derivation of the canonical formula [ref. 2.2.2]

The discrete version of the logistic-J distribution described in Section 1.1.1 summarizes the expected numbers of species per abundance category. Another description of the distribution lists the expected abundances of the species themselves. The canonical abundance $a_k$ of the kth species (in abundance rank) is readily found via the integral equation,

$$Rc\int_0^{a_k} (1/(x+\varepsilon) - \delta)dx = k,$$

By integrating over the logistic-J function up to the expected kth abundance, a, we will have "captured" k species.

Thus, the expected abundance $a_1$ of the first species may be found by solving for a in

$$Rc(\ln(x + \varepsilon) - \delta x)\Big|_0^a = 1$$

The solution is easy:

$$Rc(\ln((a + \varepsilon)/\varepsilon) - \delta a) = 1$$

Letting $T = 1/Rc$, we have

$$\ln((a + \varepsilon)/\varepsilon) - \delta a = T$$

This equation can be solved for a only by iterative methods. Call the solution $a_1$.

The position of the next canonical abundance, $a_2$, is the solution for a in the equation,

$$Rc\int_0^a (1/(x + \varepsilon) - \delta)dx = 2$$

In this case,

$$\ln((a + \varepsilon)/\varepsilon) - \delta a = 2T$$

The general equation is clearly

$$\ln((a + \varepsilon)/\varepsilon) - \delta a = kT,$$

where k indicates the kth species in the sequence.

If we remove the term $\delta a$ (being negligible for small values of k), the equation becomes directly solvable, yielding

$$(a + \varepsilon)/\varepsilon = \exp(kT),$$

or $$a_k = \varepsilon(\exp(kT) - 1) \qquad \text{(i)}$$

Loss of the term $\delta a$ has little effect on the solution, since the factor $\varepsilon^{-\delta c}$ that would otherwise appear in equation (i) is very close to unity when $\delta c$ is small (typically < 0.01)

## A.4 Proof of the random hierarchy [ref. 3.4]

**Theorem:** If a procedure is k-random, then it is also (k-1)-random, for $k > 1$.

Proof: Let P be an effectively k-random procedure. Then every sequence of length k has an equal probability of being generated by P. This probability must be $p = 2^{-k}$. Given an arbitrary sequence S of length k-1, the sequences {S, 0} and {S, 1} have length k and each must appear with probability p. Since the sequences are different, the probability that either sequence will appear must be 2p, using the additive property of independent events. But this is just the probability that S (by itself) will be generated by P, namely,

$$2p = 2^{-(k-1)}$$

## A.4 A uniform random time series follows the exponential distribution [ref 6.1.1]

Let A be a time series that satisfies the two following conditions:

1. a long term average of $\lambda$ time units between events prevails,
2. events are statistically independent; the probability of an event at any point in time depends only on the time since the last event.

**Theorem:** The distribution of inter-event times follows the negative exponential distribution.

Proof: Consider a small time increment dt and note that the probability of an arrival during this small period is $dt/\lambda$. As seen already in Section 6.1.1, this is only an approximation, but very accurate for small intervals like dt. The probability that no event will occur during this interval is

$$1 - (dt/\lambda)$$

Let f(t) be the probability that the next event will occur at least t time units after the previous one. The expression $f(t + dt)$ represents the probability that the next event will occur at least t + dt time units after the previous one. The latter probability compounds two independent possibilities, namely that a) an event does not occur during the first t time units and b) an event does not occur in the subsequent period of duration dt.

Since the next event must occur still later that t + dt time units, it follows that

$$f(t + dt) = f(t)(1 -0\ dt/\lambda),$$

by the product rule for independent probabilities. By the definition of the differential,

$$df/dt = (f(t+dt) - f(t))/dt$$

$$= (f(t)(1-dt/\lambda) - f(t))/dt$$

$$= (f(t)/dt)(1 - dt/\lambda -1)$$

$$= - f(t)/\lambda$$

taking the limit, we have

$$df/dt = \lim_{dt \dashrightarrow 0} (-f(t)/\lambda)(1 - dt/l -1)$$

$$= - f(t)/\lambda$$

Using the boundary condition f(0) = 1, the solution is

$$f(t) = e^{-t/\lambda}$$

Now f(t)dt represents the probability of an event between time t and time t+dt. The density function is simply this probability divided by dt. We now have two independent probabilities:

$e^{-t/\lambda}$, the probability of no event before t time units, and

$dt/\lambda$, the probability of an event within dt time units

Applying the product rule once more reveals the form of the density function:

$$(e^{-t/\lambda}dt/\lambda)/dt = 1/\lambda e^{-t/\lambda},$$

as required.

## A.6 The logistic theorem for strongly stochastic communities [ref 6.1.2]

In this section I derive the mathematical theory that confirms the presence of the logistic-J distribution in strongly stochastic communities. The theorem uses a timeline of events in each of which the population of a species of the community increases or decreases by unity. It makes no difference if the event times are themselves regular (e.g. 1.2, 2.4, 3.6, etc.) or random (1.2, 3,2, 3,3, 7,5, etc.,), as long as they are monotonic increasing.

**Theorem:** A strongly stochastic community has the logistic-J distribution.

Proof: I derive the theorem first for uniform communities in which the probabilities of increase and decrease are not only equal for each species, but equal across species. In such a case the probability p that a species of abundance k will lose an individual at the next tick of the clock is k/N. Under the strong hypothesis, this is also the probability q that a species will gain an individual. The probability that a population will change (either way) is therefore $p + q = 2k/N$ and the probability r that a population of size k will not change at all is therefore

$$r(k) = 1 - 2k/N.$$

I may now apply the expectation formula for the time it takes for the population size to change for this species. This formula says that if a phenomenon happens with probability r(k) over 1 time unit, then it takes, on average, E(k) time units for the phenomenon to happen, where

$$E(k) = \sum t \cdot r^t(k),$$

the summation being taken over an index t ( = 0, 1, 2, . . .∞) that represents successive time units on the clock. If we write the ratio 2/N as ν, we have a simpler expression,

$$E(k) = \sum t \cdot (1 - \nu k)^t$$

Treating k as a continuous variable for the moment, one may apply the differential operator Dk:

$$D_k(1 - \nu k)t = (-\nu)^t(1 - \nu k)^{t-1}, \qquad \text{(ii)}$$

Thus,

$$E(k) = \sum t(1 - \nu k)^{t-1}(1 - \nu k)$$

$$= (-1/\nu)\sum t(1-\nu k)^{t-1}(-\nu)(1-\nu k)$$

$$= (-1/\nu)[D_k\sum((1 - \nu k)^t](1 - \nu k),$$

by equation (ii) and by interchanging Dk and the summation. The operand is an infinite sum,

$$E(k) = (-1/\nu)\cdot(1 - \nu k)D_k\sum(1 - \nu k)^t$$

and the summation may be put in closed form by invoking the identity from Appendix A.1.2.

$$\sum^{\infty} x^t = 1/(1 - x) -1,$$

$t = 1$

where I have replaced x in the original identity by (1-x) throughout. The summation converges to the right hand side, provided that (1-x) < 1, true when $x = 1 - \nu k$. We therefore have,

$$E(k) = (-1/\nu)\cdot(1 - \nu k)D_k(1/\nu k - 1)$$

Taking the derivative $D_k$ yields

$$E(k) = (1 - nk)/k^2$$

This represents the expected number of time units a species will retain the abundance k from its first arrival at k. However, species are arriving at abundance k and departing from it more or less continually, there being, on average, F(k) species of abundance k by definition of F. Now the average proportion of time spent by a species at k is E(k) multiplied by the rate of arrival which, because it also equals the departure rate at equilibrium, represents the turnover rate. This rate is simply the probability $\nu k$.

Suppose now that R species currently inhabit the dynamical system. Then F(k)/R is the proportion of time spent by the average species in the kth abundance category. The two expressions for the proportion of time spent by an arbitrary species in the kth abundance category may be equated:

$$\begin{aligned} F(k)/R &= \nu k\cdot E(k) \\ &= \nu.(1 - \nu k)/k \end{aligned}$$

Thus $$F(k) = R\nu(1 - \nu k)/k$$

and the frequency function may be turned into a probability distribution function by dividing through by R, so that

$$f(k) = \nu(1 - \nu k)/k$$

$$= \nu(1/k - \nu).$$

This completes the first portion of the proof. I now turn to a nonuniform community in which the ith species has an associated probability $p_i$ of increase.

I define the *reproductive rate* of a species as the average time between the first appearance of an individual of that species and the appearance of a new individual, offspring of the first. In the abstract community of the foregoing theorem, all species had the same reproductive rate since individuals of different species were equally likely to be chosen for an event. In a community with nonuniform reproductive rates the same theorem holds. In this case, one invokes probabilities that are modified from the form k/N, as in the foregoing theorem. The density function associated with the rate νi is clearly

$$f_i(k) = \nu_i(1/k - \nu_i)$$

Species with different reproductive rates will simply undergo stochastic vibration at different speeds. In doing so, they will obviously follow the same distribution. A useful analogy would be the relatively uniform random distribution of pedestrians on a busy city street. If you speed some of them up (like a movie from the 1920s), even as others enter a state of slow motion, the uniform nature of the distribution is unaffected.

Corollary: A nonuniform strongly stochastic community also follows the logistic-J distribution

### **A.7 The log-series distribution** [ref. 7.1]

The general term of the log-series distribution is

$$F(k) = ac^k/k,$$

where a is a constant defined by

$$a = N(1 - c)/c,$$

N being the total number of individuals in the sample and c being a free parameter that normally takes values that are less than unity, but close to it.

Since the sum ∑F(k) yields the number of species, R, and since the series sums to a(-ln(1-c)), we have,

$$R = a(-\ln(1 - c))$$

It now follows that c may be found by solving the equation,

$$(1 - c)(-\ln(1 - c)/c = R/N \; (= 1/\mu)$$

The probability density function for the log-series distribution is obtained simply by substituting for the mean abundance μ in a specific instance, then solving for the constant c.

$$f(k) = ac^k/k$$

Thus the log-series distribution is discrete and has just one parameter, namely c, the pseudo-parameter a being actually a joint function of c and $\mu$. The total number of degrees of freedom for a chi square test of an n-category histogram with the log-series would therefore be n - 1. In the case of the logistic-J distribution the corresponding number of degrees of freedom is n - 2.

No optimization is possible for the log-series distribution, since $\alpha$ and c are determined completely by $\mu$.

### **A.8 The transfer equations** [ref. 7.4]

The transfer equation permits us to express one set of parameters and/or constants in terms of another. The application developed here permits us to infer a value for epsilon from values for $\mu$ and $F_1$, the number of species of abundance 1, as well as delta, the maximum expected abundance. The starting point for developing the transfer equation is our formula (Section 2.1.2) for the mean. (Section 2.1.2)

$$\mu = c((\Delta - \varepsilon)(\Delta + \varepsilon)/2\Delta) - \varepsilon \ln(\Delta/\varepsilon)) \qquad \text{(iii)}$$

Since $1/c = (\ln(\Delta/\varepsilon) - 1 + \varepsilon/\Delta)$, we may write

$$\ln(\Delta/\varepsilon) = 1/c + 1 - \varepsilon/\Delta$$

and substitute this value into (iii) to obtain,

$$\mu = c((\Delta - \varepsilon)(\Delta + \varepsilon)/2\Delta - \varepsilon\cdot(1/c + 1 - \varepsilon/\Delta))$$

$$= c((\Delta - \varepsilon)(\Delta + \varepsilon)/2\Delta - \varepsilon\cdot(1 - \varepsilon/\Delta)) - \varepsilon$$

So $$\mu + \varepsilon = c((\Delta - \varepsilon)(\Delta + \varepsilon)/2\Delta - \varepsilon\cdot + \varepsilon^2/\Delta)$$

Thus, we have eliminated the logarithmic term and may now write,

$$c = (\mu + \varepsilon)((\Delta - \varepsilon)(\Delta + \varepsilon)/2\Delta - \varepsilon\cdot + \varepsilon^2/\Delta)^{-1} \qquad \text{(iv)}$$

I will call this equation *system one*.

Consider now the height of the first abundance category in the discrete form of the logistic-J pdf. The height (number of species) in the first abundance category is given by the integral,

$$F(1) = Rc\int_{0}^{1} (1/(x+\varepsilon) - \delta)dx$$

$$= Rc(\ln(x+\varepsilon) - \delta x)\Big|_{0}^{1}$$

Simplifying F(1) notationally to $F_1$, we may write,

$$F_1 = Rc(\ln((1+\varepsilon)/\varepsilon) - 1/\Delta)$$

or, $$c = (F_1/R)(\ln((1+\varepsilon)/\varepsilon) - 1/\Delta)^{-1} \qquad \text{(v)}$$

The two expressions for c (equations (iv) and (v)) may now be equated, yielding

$$(\mu + \varepsilon)((\Delta - \varepsilon)(\Delta + \varepsilon)/2\Delta - \varepsilon\cdot + \varepsilon^2/\Delta)^{-1} = (F_1/R)(\ln((1 + \varepsilon)/\varepsilon) - 1/\Delta)^{-1}$$

Let $F_1/R = T$ and rearrange the terms of the previous equation to yield the following:

$$(\mu + \varepsilon)(\ln((1 + \varepsilon)/\varepsilon) - 1/\Delta) = T((\Delta - \varepsilon)(\Delta + \varepsilon)/2\Delta - \varepsilon\cdot + \varepsilon^2/\Delta)$$

Multiplying through by Δ and simplifying further, one obtains,

$$(\mu + \varepsilon)(\Delta\ln((1 + \varepsilon)/\varepsilon) - 1) = T((\Delta - \varepsilon)(\Delta + \varepsilon)/2 - \varepsilon\cdot\Delta + \varepsilon^2)$$

The next step in the development of the transfer equation is to expand the expressions above, then to collect like terms in powers of Δ:

$$(\mu + \varepsilon)\Delta\ln((1 + \varepsilon)/\varepsilon) - (\mu + \varepsilon) = T((\Delta^2 - \varepsilon\Delta + \varepsilon\Delta - \varepsilon^2)/2 - \varepsilon\cdot\Delta + \varepsilon^2),$$

$$(\mu + \varepsilon)\Delta\ln((1 + \varepsilon)/\varepsilon) - (\mu + \varepsilon) = T\Delta^2/2 - T\varepsilon^2/2 - T\varepsilon\cdot\Delta + T\varepsilon^2,$$

and $$(T/2)\Delta^2 - (T\varepsilon + (\mu + \varepsilon)\ln((1 + \varepsilon)/\varepsilon))\Delta + (\mu + \varepsilon + T\varepsilon^2/2) = 0$$

Now let $A = T/2$,
$B = T\varepsilon + (\mu + \varepsilon)\ln((1 + \varepsilon)/\varepsilon)$,
and $C = \mu + \varepsilon + T\varepsilon^2/2$

and note that the equation $A\Delta^2 - B\Delta + C = 0$ is a quadratic in Δ and therefore has a solution of the form,

$$\Delta = (B +/- \sqrt{B^2 - 4AC} )/2A \qquad \text{(vi)}$$

Note also that Δ can be computed in one step, via equation (vi), since none of the three expressions involving A, B, and C, contain Δ. I will call equation (vi) *system two*.

Solutions of the transfer equations may be found by using the program SolveIt.PAS, as described in Appendix B.4). The underlying algorithm proceeds iteratively by choosing a starting value for ε, then substituting the values on hand for R, $F_1$, and μ. into system two. This produces a value for Δ that is then substituted, along with the values for ε and μ, into system one. The resulting value for ε is then recycled back into system two, where the process begins anew. Over a number of iterations, the value of ε (as well as that of Δ) converges to a stable value. These values are the ones implied (in the mathematical sense) by the observed values of R, $F_1$ and μ.

A more general form of the transfer equation can be developed from the foregoing systems by replacing the integration limit 1 in system two by the parameter a, which can be any abundance, whether integral or fractional. This yields the equation,

$$c = T(\ln((a + \varepsilon)/\varepsilon) - a/\Delta)^{-1}$$

on the way to a more general version of system two, whereas system one remains the same.

$$A = T/2$$
$$B = T\varepsilon + (\mu + \varepsilon)\ln((a + \varepsilon)/\varepsilon)$$
$$C = a(\mu + \varepsilon) + T\varepsilon^2/2$$

where T = Fa/R.

### A.9 Behaviour of strongly stochastic communities [ref. 8.2.1]

In a strongly stochastic community, each individual in the system has an equal chance of reproducing or dying over an average time period of l units of time. For a collection of k individuals the average period is λ/k. A species of abundance k will therefore have a *stochastic frequency* of k/λ. This is the average number of events affecting that species per unit time.

Over λ time units , the population of this species could increase or decrease by 0 up to k individuals. Suppose that exactly m individuals reproduce during the time λ. Then during the same period k-m individuals die. The net gain (or loss) in individuals is

$$p(2m-k) = C(k, m)\, p^{m} \cdot p^{(k - m)}]$$
$$= C(k, m)\, p^{k} \qquad \text{(vii)}$$

This is clearly the binomial distribution. The expression (vii) may be simplified by substituting r for 2m - k and replacing m by (k+r)/2:

$$p(r) = C(k,(k+r)/2)\ p^{k}$$

It is readily verified that

$$C(k,\ k+r)/2) = C(k,\ k-r)/2)$$

and this implies the following equalities:

$$p(+r) = p(-r) \text{ for } r = 0, 1, \ldots k.$$

The distribution is clearly symmetric. The stochastic frequency k/λ for abundant species is greater than that of less abundant ones. The frequency increases linearly with k. The distribution for all F(k) species of abundance k may be derived by "pooling" all the k' individuals within these species, where k' = k.F(k):

$$p(r) = k'C(k',\ (k'+r)/2\cdot p^{k'} \qquad \text{(viii)}$$

This distribution is also binomial with mean zero and the stochastic frequency is now

$$k\cdot F(k)/\lambda$$

Indeed, we may also pool all N individuals of the community and conclude that formula (viii) continues to hold, this time with k' = N. The frequency is now

$$N/\lambda,$$

so that over the relatively small period λ/N, we may again apply the probability of a unit increase and find that it equals the probability of a unit decrease.

### A.10 Sampling the univoltine distribution [ref. 9.2.1]

In this note an example is used to illustrate the role played by recurrence relations in the solution of the accumulation problem for species as a function of sample size. The example will involve a community with the (distinctly unnatural) univoltine distribution U[8] x 10, with 10 species, each with 8 individuals.

Let R(k) represent the expected number of species to have appeared in the sample at the time the kth individual is observed. Given a sample of size k, what is the probability of a new species appearing at the next observation? The probability that the next individual will belong to a new species will depend on the number of individuals belonging to the set of 10 - R(k) species that have not yet been sampled. The number is, of course, 8 x (10 - R(k)) and the associated probability must be

$$8(10 - R(k))/80 \quad \text{or} \quad 1 - 0.1R(k)$$

The expected number of new species at the (k+1)st drawing is simply this probability. When added to R(k) we obtain the following recurrence relation:

$$R(k + 1) = 0.9R(k) + 1$$

In such relations, the value of a function at k is stated as a function of previous values, often more than one, such as k-1, k-2, and so on. It would be convenient to have a closed-form version of the formula that would be easier to calculate. Before solving the recurrence relation, however, I will show how the recurrence may be used to generate values of R(k) for any value of k.

Starting at R(1) = 1.0, for example, the relation tells me that R(2) = 1.9 and applying the relation once again, I obtain R(3) = 2.71 Table A.1 gives the first 20 values of the function R, as calculated by this method.

| k | R(k) | k | R(k) |
|---|---|---|---|
| 1 | 1.000 | 11 | 6.862 |
| 2 | 1.900 | 12 | 7.176 |
| 3 | 2.710 | 13 | 7.458 |
| 4 | 3.349 | 14 | 7.712 |
| 5 | 4.095 | 15 | 7.941 |
| 6 | 4.685 | 16 | 8.147 |
| 7 | 5.216 | 17 | 8.332 |
| 8 | 5.695 | 18 | 8.499 |
| 9 | 6.126 | 19 | 8.649 |
| 10 | 6.513 | 20 | 8.784 |

**Table A.1.** Expected number R(k) of species as a function of sample size

A simple empirical test reveals the accuracy of the formula when compared with the results of a simulated sampling. Table A.2 shows the comparison at a few evenly spaced points.

| **k** | 4 | 8 | 12 | 16 | 20 |
|---|---|---|---|---|---|
| **R(k)** | 3.35 | 5.69 | 7.18 | 8.15 | 8.78 |
| **S(k)** | 3.45 | 5.50 | 7.00 | 8.20 | 8.80 |

**Table A.2.** Comparison of theoretical **R** and empirical **S** sample richness

The match between empirical and theoretical values gives us confidence that no mistake was made in deriving the recurrence relation. To obtain a closed form solution, we must solve the recurrence.

$$R(k) = 0.9R(k-1) + 1 \qquad \text{(ix)}$$

One way to solve recurrence relations in general is to back-substitute, replacing R(k-1), for example, by

$$0.9R(k-2) + 1.0$$

in formula (ix), then continuing this process down to the last replacement by R(1). The general result for the example follows:

$$R(k) = \sum_{i=0}^{k} 0.9^i \qquad \text{(x)}$$

This expression for the formula is slightly more useful than the earlier recurrence relation, but the sum cannot be simplified. In the limit, however, we may apply the well-known formula for an infinite geometric series (Appendix A.1.2 ) to this component of the formula, yielding

$$\lim_{k \to \infty} R(k) = 0 + 1.0/(1 - 0.9) = 10.0, \text{ as expected.}$$

Can formula (xi) be generalized to any univoltine distribution? Obviously it can. We cut to the chase by replacing the constant 0.9 in formula (x) by (1 - 1/R)

$$R(k) = \sum_{i=0}^{k} (1-1/R)^i$$

This completes the solution of the accumulation curve for the univoltine distribution with replacement sampling.

## Computer Research Tools

It seem fair to say that no successful theory of population or community dynamics is possible without the aid of computer software that is capable of simulating the sampling process accurately in a statistical sense or emulating dynamical processes in 1-1 fashion. Armchair insights will often founder on the rocks of a decisive computer test.

Each program is listed by its filename, followed by a description of its operation. Several of the key programs used in this research program are listed on the author's website in their entirety. The remainder are listed by title, along with a brief description.

### A.11 Systems emulators

Scramble1:

The most important and basic program for understanding logistic systems and their relationship to communities of real organisms is the program (actually a series) I call "Scramble," in which an arbitrary number of species undergo increase and decrease based on pair interactions. The most general view is that the death of one organism is accompanied by the "birth" of another. (A source code listing of Scramble1 appears on my university website, along with a working version that demonstrates its operation.)

Scramble 2: sets up two separate communities with migrations permitted between them. Initial conditions are specified as with Scramble.

Scramble3: sets up three separate communities with cross-migration permitted. Initial conditions are specified as with Scramble.

ScramWeb: establishes four separate trophic groups analogous to plants, animals, fungi and omnitrophs. Trophism is directed: animals eat plants; plants are autotrophic; fungi eat plants and animals; omnitrophs eat anything. Initial conditions involves number of species in each group.

StoComm: emulates a stochastic community using a probabilistic random walk with continually varying probabilities. Initial conditions are specified as in Scramble1.

### A.12 Sampling and estimation programs

Commrich:

The program CommRich is a richness-estimation program based directly on logistic-J theory.

The remaining programs listed below are all sampling programs of various kinds. Using either the raw sample histogram or the best-fit version of it, this program computes the expected sample from a community and compares it with the sample data via the least squares measure. To do this, it needs a sampling intensity estimate from the user. The program structure is reentrant, allowing the user to refine parameter estimates, converging on a final value for richness R. The user has the option of inputting raw sample data or best fit to sample. }

Comm2: is the same as CommRich, but uses the hypergeometric distribution instead of the Poisson distribution. In certain cases this completely accurate version of the Pielou transform is preferable to the (close) approximation to the hypergeometric employed by CommRich.

SampleSim1: computes multiple samples from the community LJ[e, D] x R at intensity r. It does this as many times as the user specifies, then computes an average sample.

SampleSim2: operates exactly like SampleSim, but samples an empirical community specified by the user.

LJTestor: produces perturbations of a given LJ distribution in ten categories.

PTransform: calculates the expected sample from the community LJ[e, D] x R This program is embedded in CommRich, described above.

## A.13 Statistical programs

ChiSquare:

This program takes an empirical (field) distribution as input and compares it with either the logistic-J or log-series distributions, as specified by the user. It uses the standard rule-of-five that, until a category has accumulated a total of at least 5.0 of the theoretical distribution, the category is extended until the criterion is met. The program outputs the chi square score, as well as the number of degrees of freedom used by the fitting process.

BestFit: employs the chi square test in reentrant fashion, allowing the user to adjust parameters as the computation proceeds. The user must find the combination of parameter values that produces the minimum chi square score.

ScanFit: In this version of the score minimization program, the user specifies a grid of test-points in parameter space. The program computes the score for each combination, ultimately reporting on the combination that produced the minimum score.

KSTest: calculates the Kolmogoroff-Smirnov statistic for an empirical distribution (specified by the user) against a particular logistic-J distribution (also specified by the user).

## A.14 Special utility programs

Solveit1:

This program solves the logistic equation iteratively, starting with an initial guess by the user. This guess may be critical in the sense that if it is too large or too small, the iterative process may diverge. Other inputs include richness R, mean μ, category width a (usually unity), and $F_a$, the number of lowest-abundance species. The program outputs the value for ε that solves the system, as well as values for Δ, δ, and the coefficient c.

SolveIt2: solves another set of transfer equations that carry ε as an implicit function of μ and Δ. instead of μ and Fa.

Canon: calculates the canonical sequence for the specified distribution LJ[ε, Δ] x R. Presently, this program operates in interactive mode. It can be rewritten to find each canonical abundance without human intervention.

LJGen: generates theoretical values of a given logistic-J distribution with parameters, ε, Δ and richness R. To do this it uses the formula for the logistic-J pdf, then multiplies by R.

Overlap: computes the expected overlap of two samples based on empirical distribution for the source community.

overlap: computes the expected overlap of two samples of given sizes from a logistic-J source, LJ[ε, Δ] x R.

Interpol: Given a chi square score at n degrees of freedom, Interpol finds the equivalent score at 10 degrees of freedom by using linear interpolation on the entries of a chi square table.