Chapter 2. The Logistic-J Distribution and Variations

In the introductory chapter, the logistic-J distribution appeared only in the form of its probability density function (pdf). In this chapter it appears as a distribution populated by a number R of species. The distribution comes in a continuous version and in a corresponding discrete version that translates the continuous form into a prediction of the number of species to appear at each possible abundance. Carrying this process one step further, it is possible to predict specific abundances for each species as a function of rank order. The canonical sequence developed for this role is mathematically equivalent to the other formulations.

This chapter begins with two developments based on the probability density function.

2.1 Properties of the probability density function

Recall the density function in Equation (i) of Section 1.4:

$$f(x) = c(1/(x+\varepsilon) - \delta); \quad 0 \leq x \leq \Delta - \varepsilon,$$

$$= 0; \; x \geq \Delta - \varepsilon$$

It may cause some consternation to have a density function with a finite domain. Can there be no abundances greater than $\Delta - \varepsilon$? It will be natural to address this question in the context of its distribution form. (See Section 2.2 below.)

The normalizing coefficient c is a function of the two parameters, ε and Δ, and does not constitute a new parameter, being merely a notational simplification:

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

In what follows, the simplified (closely approximate) form of the formula is adequate to the task of portraying the role of c as a shape parameter.

$$c \approx (\ln((\Delta/\varepsilon) - 1)^{-1} \qquad \text{(i)}$$

In most cases the approximation error would be less than 0.0001.

2.1.1. The capacity constant

The constant c yields a closely related constant C, the *capacity* of the community, defined as the inverse of c. Inverting formula (i) produces the following expression:

$$C = \ln(\Delta/\varepsilon) - 1$$

$$= \ln(1/\delta\varepsilon) - 1 \text{ (since } \Delta = 1/\delta)$$

Clearly, C is inversely proportional to the (log of) the twin parameters. It gets larger when the parameters ε and δ become smaller. The capacity C is smaller, on the other hand, when the same parameters are larger. The difference between the two possibility is illustrated in Figure 2.1 a..

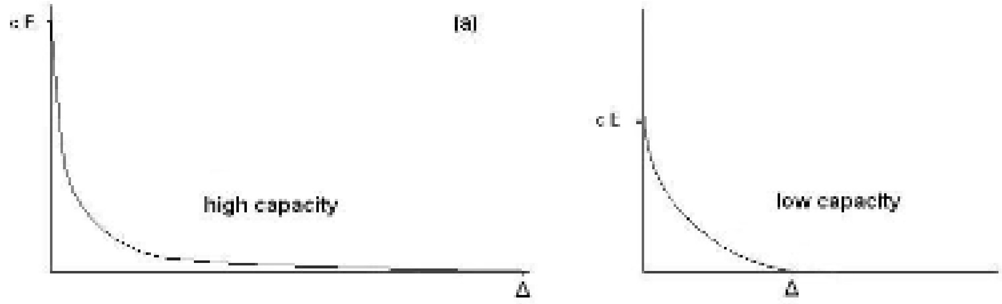

Figure 2.1a. Examples of high (left) and low (right) capacities

If, on the other hand, the capacity is kept constant, the parameters ε and δ may nevertheless vary, as long as the product εδ remains constant, as in Figure 2.1b below.

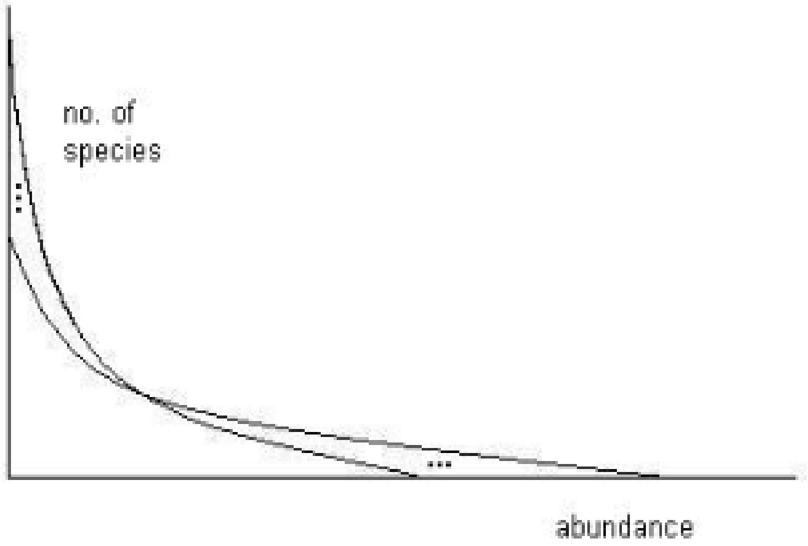

Figure 2.1b. Different shapes may share the same c-value

For each fixed value of c, the range of density functions produced by varying ε and δ in this manner all belong to the same capacity class. The concept of capacity will be revisited in Section 2.2.2, where it plays a role in the spacing between consecutive abundances; the greater the capacity C of a logistic distribution, the greater is the gap between the abundance of the kth species and that of the k+1st. In distributions having a larger capacity, there is more "room", abundance-wise, for additional species. Apart from appearing in Section 2.2.2, the concept plays a relatively minor role in the rest of this book. However, it strikes one as having promise for an important role in later developments.

2.1.2 The mean and variance of the logistic-J distribution

According to the standard definition of the mean as the first moment of an (arbitrary) probability density function g, we have,

$$\mu = \int xg(x)dx$$

Applied to the logistic-J distribution (pdf), the expression for the mean becomes

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

The first term of the integrand is not in standard form but is readily integrated as shown in Appendix A1.1. In this case and in others to follow I will drop the epsilon from the limit of integration as it makes very little difference to do so.

$$\mu = c(x - \varepsilon\ln(x) - \delta(x-\varepsilon)^2/2)\Big|_{\varepsilon}^{\Delta}$$

$$= c((\Delta - \varepsilon) - \varepsilon\ln(\Delta/\varepsilon) - (\Delta - \varepsilon)^2/2\Delta),$$

which simplifies to the following expression

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

Expanding the constant c in terms of Δ and ε does not result in a simpler expression, so it will be left in this form. In order to find the number N of individuals in a given theoretical distribution, simply multiply the value for μ just obtained by R, the richness of the sample. Thus,

$$N = R\mu$$

The formula for the mean will reappear in Section 6.1.4, where it plays a key role in the explanation of oscillations in parameter values over time. It turns out that if one fixes a value for μ, equation (ii) has zero, one, or two solutions in terms of ε and Δ, depending on the value of μ.

The variance of the logistic-J distribution is based on the second moment about the mean,

$$V = = c\int_0^{\Delta-\varepsilon} ((x-m)^2/(x+\varepsilon) - \delta x)dx$$

The resulting formula, as derived in Appendix A.2, is not especially elegant:

$$V = c[(\Delta^2/2 - 2\Delta\varepsilon + (2.5)\varepsilon^2 - \delta(\Delta-\varepsilon)^3/3] + \varepsilon^2 - \mu^2$$

The variance of a distribution function F(x) = Rf(x) is obtained by dividing this expression through by R-1. Because of its non-central nature and long tail, the logistic distribution has high variance. This is true to varying extents of all the proposals for species-abundance distribution, as described in Section 1.2, 4.2 and elsewhere.

2.2. The logistic-J distribution: species and individuals

The distribution function F of the logistic-J probability density function f is obtained by multiplying it by R, the number of species in a sample -- or in a community, depending on the focus of one's investigation.

$$F(x) = Rf(x)$$

Owing to the finite nature of the domain, one must interpret the function F carefully. In general, the integral of the continuous form over a given interval of abundances yields the expected number of species having abundances in that interval. Statistical fluctuations guarantee that some species in both samples and communities will, from time to time, wander rather far from their expected positions, resulting in overruns even on the logistic limit D, which we take as a "soft" limit in this sense. We can only say that Δ bounds the average maximum abundance, if not all maximum abundances. Under the circumstances, it seems a reasonable practice to accept abundances where the formula says there should be none. Operationally, it makes no difference. For example, in calculating the chi square statistic for such sample data, an abundance a beyond Δ will be compared with 0.000, in effect, instead of, say 0.002.

The relationship between the value of the parameter ε and the number F_1 of species in the minimum abundance category, as expressed in the continuous versions of the logistic-J distribution, hints at the somewhat more subtle nature of the parameter ε:

$$F_1 = Rc(1/\varepsilon - \delta) \qquad \text{or} \qquad Rc(E - \delta)$$

Of course, c is already a function of epsilon. If the right hand side of this expression is expanded in terms of the constant c, a mixed log/linear equation results that has no closed-form solution. The equation can be solved by numerical methods on a computer using the largest abundance as an estimate for Δ (or $\delta = 1/\Delta$) and the number of lowest abundance species as an estimate for F_1. this approach will yield a good approximation, in general, for the corresponding value of ε.

2.2.1 The discrete logistic-J distribution

To produce a discrete version of the logistic-J distribution function one first selects appropriate abundance categories, depending on the data on hand. Most commonly, the field worker will use categories 1, 2, 3, etc., representing simple counts: how many species appeared once in the sample? How many twice? And so on. But observations may also be grouped, so that, for example, the categories might be 1 to 3 for the first category, 4 to 6 for the next, and so on. Another common abundance format is density. What is the average number of robber flies observed per hectare? One species might have a density of 3.7,.while another has density 42.6. Densities are normally grouped. Thus the first (lowest) abundance category might run from 0 to 1.5, the next from 1.5 to 3.0, etc. Examples of such grouping will be found on the author's website (Dewdney, 2011).

Suppose that the category width is a real number, a, with categories (0, a], (a, 2a], (2a, 3a], and so on. With this notation one may produce a discrete distribution function by integrating over each range of abundances, as follows:

$$F(ka) = \int_{(k-1)a}^{ka} F(x)\, dx = Rc[\ln((ka+\varepsilon)/((k-1)a+\varepsilon) - \delta a];\ k = 1, 2, 3, \ldots$$

Given values for R, c, k, and a, the function to the right is readily computed using a hand calculator. Of course, if one is adept at programming a computer, it would take only a matter of minutes to write a program to perform the same calculation much more rapidly. (See the program LJGen in Appendix A.10.) When this is done for each category, a histogram like the following will emerge.

The first abundance category (k =1) has the greatest number of species. The numbers decline with increasing abundance, abruptly at first, more gradually later, to zero. In this idealized situation the number of species in a given category might well be fractional. The interpretation of this number is simple. If, for example, the number of species in a given category is 5.6, it would mean that over all instances of samples which this particular theoretical curve fit best, there would be an average of 5.6 species in this category. Most of the time, six species would inhabit

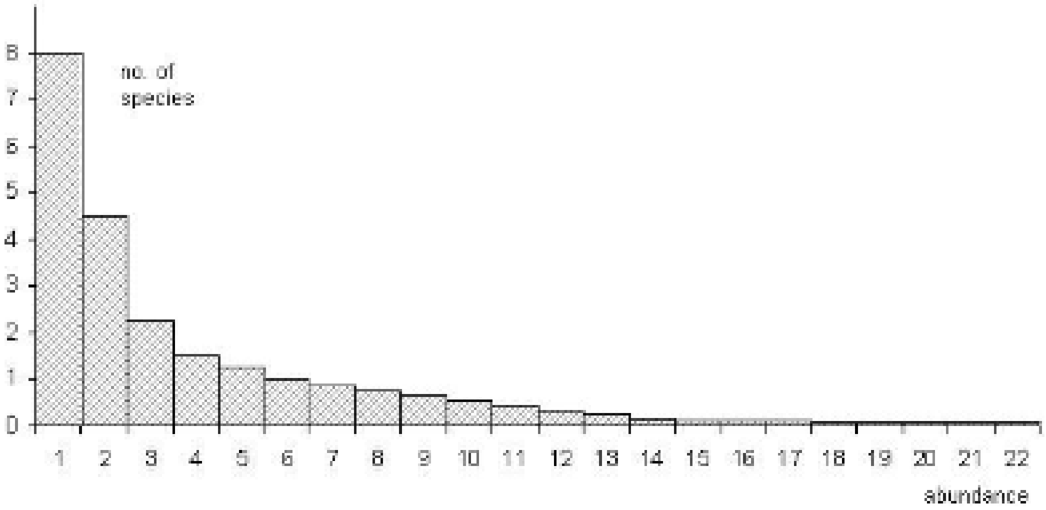

Figure 2.2. A discrete version of the logistic-J distribution.

it, much of the time five, occasionally four or seven. (Chapter 4 explains how to derive this theoretical curve and to compare it statistically with a field sample.)

We can rarely, if ever, know the abundances of the species in a community *per se*. According to logistic-J theory, the general shape of the histogram in communities would resemble Figure 2.3, except that some species would appear displaced in one direction or the other. Samples of such a distribution, however ragged, would resemble what is typically seen in field histograms.

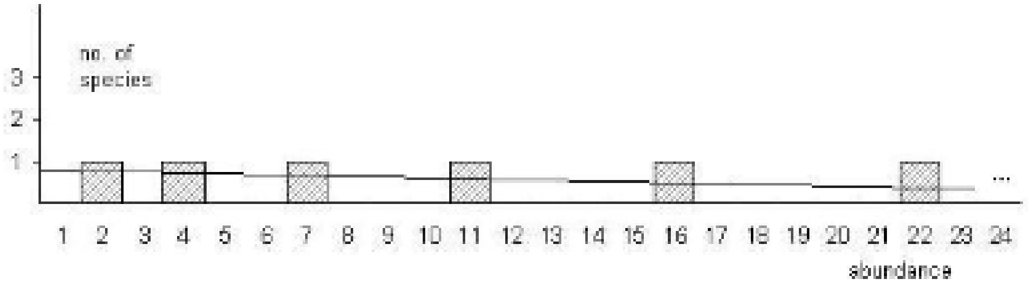

Figure 2.3. The logistic-J distribution for a community

In the example of Figure 2.3 the continuous version of the logistic-J distribution (thin, flat curve) is superimposed upon the histogram to show the relationship between the two versions of the distribution in the context of a community, rather than a sample. Each bar (a block of unit height) represents the area under the continuous curve between limits that are derived from its formula. This subject will be revisited in the next section.

Those who worry that the logistic-J shape in samples indicates a large number of species about to be extirpated may take some comfort from Figure 2.3. Not only are species that appear only once in a sample likely to be more abundant in the source community, but the lowest abundance there is not necessarily 1, as the figure makes clear. Over time, species come and go from communities, now appearing by immigration or, now disappearing by emigration or extirpation. By the same token, at the scale of larger landscapes, species may also appear by speciation. Chapter 8 examines the role of the logistic-J distribution in evolution.

2.2.2. The canonical sequence

The discrete version of the logistic-J distribution described in Section 2.2.1 describes the expected numbers of species per abundance category. Another description of the distribution lists the expected abundances themselves, as in Figure 2.3 above. The canonical abundance a_k of the kth species (in abundance rank) is readily found via the integral equation,

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

Appendix A.2 contains the full derivation of the resulting formula:

$$a_k = \varepsilon(\exp(kT) - 1), \qquad \text{(iv)}$$

where a_k represents the abundance of the kth species and $T = c/R$. Recall that C is the capacity of the distribution, defined as the inverse of the constant c.

Example:

Let LJ(2.00, 150.0) X 50 be a logistic-J distribution. Then c = 0.3003 and T = 1/15.015 = 0.06660, so that

$$a_k = 2.0(e^{0.0666k} -1.0)$$

Table 2.1 displays the first five canonical abundances corresponding to this distribution, as well as the last five. The formula is approximate, being highly accurate for low abundances and slightly less so for the highest ones.

As the example shows, more than one species can have an expected abundance that is less than one. At the discrete distribution level, this number merely reflects the expected number of species to show up once in the sample. In this case, the number is 6. As it happens, some 4 species would be expected to show up with abundance 2 in the sample, and so on. At some point we must switch from formula (iv) to the exact or “general” expression referred to in Appendix A.2: The last five canonical positions were calculated by this method.

k	1	2	3	4	5	. . .	46	47	48	49	50
abund.	0.138	0.285	0.442	0.611	0.790		63.4	72.0	83.4	99.8	150.0

Table 2.1. Canonical abundances for a community of 50 species

Plotted on an abstract abundance axis, the species in a canonical sequence can be displayed as individual markers, as in Figure 2.4 below.

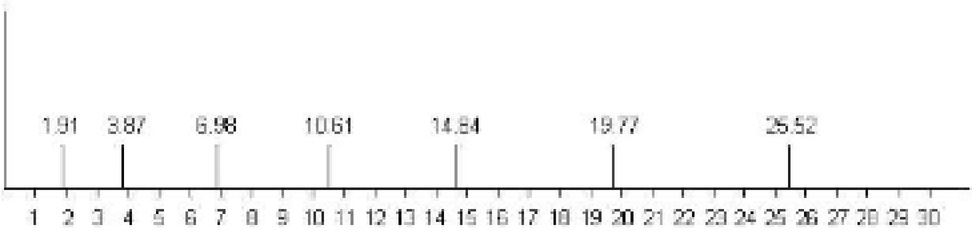

Figure 2.4. Canonical abundances of species in a small community

Finally, it should be noted that the canonical sequence formulation is simply another way of presenting the logistic-J distribution, being mathematically equivalent to it. The approximate formula used here is essentially a geometric progression, similar to the proposal of (May 1975) discussed earlier in Section 1.2. I must also insert the following caveat: defining the canonical abundance of the kth species (in order) as occurring when the integral of Equation (iii) has the value k may not be quite correct. The resulting mean abundance of the kth species may be slightly lower than the value given by formula (iv) but subsequent developments based on the integral values k chosen in Equation (iii) will not be altered enough to invalidate any of the principal conclusions arrived at in this chapter.

The canonical sequence for a given logistic-J distribution, LJ[e, D] x R illuminates the use of the word “capacity” for constant C. Recall that

$$C = \ln(\Delta/\varepsilon) - 1$$

According to formula (iii) for the canonical abundances, we may replace the factor T by 1/Rc to obtain the formula,

$$a_k = \varepsilon\ (\exp(kC/R) - 1.0),$$

which readily yields an expression for the space between consecutive abundances, as follows.

$$a_{k+1} - a_k = \varepsilon(\exp((k+1)C/R) - 1.0) - \varepsilon(\exp(kC/R) - 1.0)$$

$$= \varepsilon(\exp(kC/R))(\exp(C/R) - 1)$$

With ε and R both fixed, the expression is obviously larger when C is larger and, conversely, smaller when C is smaller. In other words, there is more “room” between successive abundances in high capacity communities.

2.2.3. An implicit formula for rank abundance

Much of the prior work on species abundances has been framed in the context of the rank abundance diagram, a method of representing sample data that holds the same information as the standard representation, but in a rather different manner. (See Section 4.2.) One arranges all the species in one’s sample in order of decreasing abundance (the rank order), then plots their logarithms as vertical bars. This section notes, in passing, that the rank abundance formula is implicit in the expression from which the canonical sequence is derived in Appendix A.2. Unfortunately, the formula has no closed-form solution, so must be solved by numerical methods.

$$\ln((a + \varepsilon)/\varepsilon) - \delta a = kT, \qquad \text{(v)}$$

If one represents the implicit solution by a(k), the order of the variable k must be reversed since, in the present instance, the diagram begins not with the smallest abundances, but the largest. Let A(k) represent the resulting function. Then,

$$A(k) = a(k)(R - k + 1)$$

The resulting curve can be at least be illustrated through a worked example, as shown in Figure 2.5. Using the earlier example of LJ[2.0, 150.0] x 50, formula (v) results in the plot of Figure 2.5, where, for the sake of simplicity only multiples of 5 are plotted as values for k. Although it is based on a rather small example, the resulting rank abundance diagram has the sinusoidal shape that is typical of such diagrams when based on actual samples.

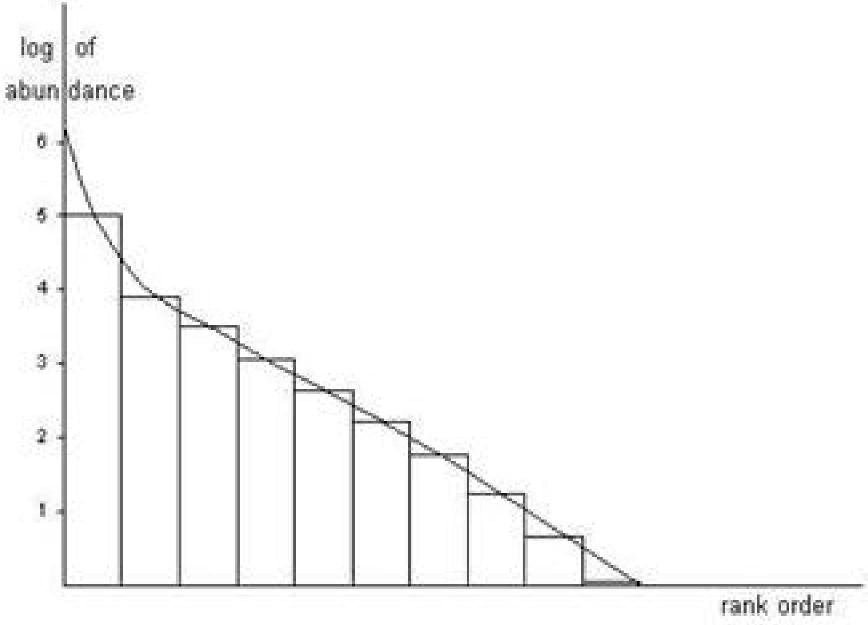

Figure 2.5. A simplified example of a logistic-J rank abundance diagram

2.3. Equilibrium dynamics of the logistic-J distribution

Starting in the abstract, consider a system of particles vibrating in one dimension. The particles occupy discrete positions, numbered 1, 2, 3, etc., while a clock ticks off time units. At any given tick of the clock, each particle has an equal probability p of moving to the next higher position or the next lower one. Moreover, the probability is proportional (symbol α) to the position k of the particle,

$$p_k \; \alpha \; k. \qquad \text{(vi)}$$

2.3.1 The Equilibrium Theorem

Theorem: The system just described must attain equilibrium and produces a hyperbolic distribution of species over abundances.

Proof: First the formula is derived on the basis of the assumption that the system reaches equilibrium and then it is shown that the system must tend toward equilibrium.

Given that the system is in equilibrium, the number of particles moving from position k to k+1 would equal the number of particles moving in the other direction, on average. Suppose now that the number of particles at each position is given by the function F. The number of particles moving to the right at each tick of the clock is $p_k \cdot F(k)$, while the number of particles moving to the left is $p_{k+1} \cdot F(k+1)$. These numbers are fractional, of course, since they corresponds to probabilities.

At equilibrium the quantities are equal;

$$p_k \cdot F(k) = p_{k+1}.F(k+1). \qquad \text{(vi)}$$

Now let ν be a proportionality constant that corresponds to the proportionality α in equation (vi), so that, for example, $p_k = \nu k$. Substituting for p_k and p_{k+1}, we have,

$$\nu k \cdot F(k) = \nu(k+1) \cdot F(k+1)$$

Canceling the νs, we obtain,

$$k \cdot F(k) = (k+1) \cdot F(k+1)$$

This equation has a unique solution, namely $F(k) = 1/k$, a hyperbola.

To see how equilibtrum occurs, suppose that

$$k \cdot F(k) > (k+1) \cdot F(k+1)$$

This inequality implies a corresponding iequality replacing equation (vi), so that it is now more probable that a species of abundance k will increase than that a species of abundance k+1 will increase. It is therefore more probable that at the next iteration the left hand quantity will be smaller by unity than that F(k+1) will be smaller. Departures from hyperbolic values in F will therefore result in a compensatory increase or decrease in the corresponding F-values. This completes the proof.

The theorem implies that the “mechanism” that produces the logistic-J distribution has nothing to do with specific ecological interactions or forms of ecological interaction, except insofar as they contribute collectively to the probabilities of increase and decrease. The hyperbolic shape just adduced is not the full story, of course. In the logistic version of this system, there is a limit, as shown in Appendix A.5. The dynamical system just examined has an interesting physical interpretation; if a thin cylinder of gas is heated at one end, the temperature gradient will ensure that molecules vibrate at different speeds, moving more quickly at the hot end and more slowly at the cool end. According to thermodynamic theory, the motions will be random. Molecules will nevertheless “prefer” to be at the cool end, since that yields a minimum overall energy for the system. Indeed, if the temperature increases in linear fashion along the tube, the density of molecules in the cylinder will fall off toward the hot end in inverse linear fashion.