Digital Appendix E. More on measuring deviation.

How might fluctuations best be quantified, assuming that we have determined comparable intervals of time and a curve against which to compare the species involved in the lottery we have described? Let μ_i be the number expected in bin *i* according to, for example, an exponential distribution and n_i the number actually observed. One measure of deviation from the expected smooth curve is

$$\chi_p^2 = \sum_i \frac{(n_i - \mu_i)^2}{\mu_i} \qquad \text{(E.1)}$$

This quantity is well defined but will not have a χ^2 distribution unless n_i is normally distributed about μ_i. In Kelly et al. (2001), the measure

$$\sum_i \left\{ 2(\mu_i - n_i) + 2 n_i \ln \frac{n_i}{\mu_i} \right\}$$

was employed, which is an analog of χ^2 for Poisson distributed n_i (Baker and Cousins 1984).

A measure of the fractional deviation from an expected (or hypothesised) smooth curve is

$$\xi^2 = \sum_i \left(\frac{n_i}{\mu_i} - 1 \right)^2 \qquad \text{(E.2)}$$

Both (E.1) and (E.2) can be divided into interesting and uninteresting pieces, as follows:

Let $\qquad n_i = \mu_i + e_i + \gamma_i$

where e_i is an interesting deviation (in this paper, a fluctuation in recruitment caused by an environmental effect) and γ_i is a statistical fluctuation (which is not interesting). Then to the extent that e_i and γ_i are uncorrelated

$$\chi_p^2 = \sum_i \left(\frac{e_i^2}{\mu_i} + \frac{\gamma_i^2}{\mu_i} \right) \qquad \text{(E.3)}$$

$$\xi^2 = \sum_i \left(\frac{e_i}{\mu_i} \right)^2 + \sum_i \left(\frac{\gamma_i}{\mu_i} \right)^2 \qquad \text{(E.4)}$$

If γ_i is Poisson distributed about $\mu_i + e_i$, $\overline{\gamma_i^2} = \mu_i + e_i$ and e_i fluctuates.

Thus $\dfrac{\overline{\gamma_i^2}}{\mu_i} \cong 1$ and $\dfrac{\overline{\gamma_i^2}}{\mu^2{}_i} \cong \dfrac{1}{\mu_i}$.

In (E.3) the second (statistical) term is independent of the sample size or shape of the expected distribution (but depends on the number of bins) but the first (interesting) term depends on the shape and the sample size. In (E.4) the situation is reversed: the first (interesting) term is independent of the shape or sample size, whereas the second (statistical) term depends on both – but can be calculated for a given expected distribution.

A measure of the χ^2 type can be used to establish deviation from some hypothesized distribution beyond the merely statistical, but care may be necessary in interpreting differences in the estimator of $\sum_i \left(\dfrac{e_i^2}{\mu_i}\right)$ which is

$$\chi_p^2 - \sum_i 1. \qquad \text{(E.5)}$$

The measure we are after for comparison with the lottery model is the measure of fractional deviation $\sum_i \left(\dfrac{e_i}{\mu_i}\right)^2$ for which we took the estimator

$$d^2 = \xi^2 - \sum_i \frac{1}{\mu_i}. \qquad \text{(E.6)}$$

There are two caveats that should be borne in mind. The first is that the statistical terms have a distribution and the corrections in (E.5) and (E.6) are the mean values of the statistical terms. The second caveat is that the populations of individual bins may not be Poisson distributed.

The variance on (E.5) due to statistical fluctuations about $\mu_i + e_i$ which are Poisson distributed is

$$\sum_i \left(\frac{\overline{\gamma_i^4}}{\mu_i^2} - \frac{\left(\overline{\gamma_i^2}\right)^2}{\mu_i^2} \right) \cong \sum_i \left\{ 2 + \frac{1}{\mu_i} + 2\left(\frac{e_i}{\mu_i}\right)^2 \right\}$$

and the variance on (E.6) due to such statistical fluctuations is

$$\sum_i \left(\frac{\overline{\gamma_i^4}}{\mu_i^4} - \frac{\left(\overline{\gamma_i^2}\right)^2}{\mu_i^4} \right) \cong \sum_i \left\{ \frac{2}{\mu_i^2} + \frac{1}{\mu_i^3} + 2\left(\frac{e_i}{\mu_i}\right)^2 \frac{1}{\mu_i^2} \right\}$$

We do not think it possible (or necessary) to extract a better estimate of the variance **V** on d^2 than

$$\mathbf{V}\left(d^2\right) \cong (2-4)\sum \frac{1}{\mu_i^2} \qquad \text{(E.7)}$$

without a rather complete understanding of the details of each two component system and the larger forest environment within which they operate.

Literature Cited

Baker, S., and R. D. Cousins. 1984. Clarification of the use of chi-square and likelihood functions in fits to histograms. Nuclear Instruments and Methods in Physics Research **221**:437-442.

Kelly, C. K., H. Banyard Smith, Y. M. Buckley, R. Carter, M. Franco, W. Johnson, T. Jones, B. May, R. Perez Ishiwara, A. Perez-Jimenez, A. Solis Magallanes, H. Steers, and C. Waterman. 2001. Investigations in commonness and rarity: a comparative analysis of co-occuring, congeneric Mexican trees. Ecology Letters **4**:618-627.