

Irreproducible Results and the Breeding of Pigs

(or Nondegenerate Limit Random Variables in Biology)

Joel E. Cohen

Many people believe that a coin in ordinary currency will come up heads nearly half the time it is tossed. Few people have reported a systematic experimental test of that belief.

During World War II an English statistician, J. E. Kerrich, was in Denmark when the Germans overran it. Interned under benevolent Danish supervision, he performed and recorded (Kerrich 1946) 10,000 spins of an ordinary coin. The proportions of trials which came up heads after ten, a hundred, a thousand, and ten thousand trials were, respectively, 0.400, 0.440, 0.502, and 0.507. If the ten thousand trials are broken into ten blocks of a thousand trials each, then the proportions of heads after each of the ten blocks were 0.502, 0.511, 0.497, 0.519, 0.504, 0.476, 0.507, 0.518, 0.504, and 0.529.

Viewed as a single long series, the data show that the proportion of heads tended toward and remained near one-half as the number of trials (tosses or spins) increased. Viewed as ten shorter series, the data suggest that the proportions of heads in independent experiments under the same conditions tended toward a single common value.

In 1713, Bernoulli constructed a mathematical idealization of the coin-tossing experiment as a sequence of independent trials each with a fixed probability p of coming up heads. Here p is some fraction near one-half. Imagine a very large number of copies of Kerrich all tossing copies of the same coin under the same conditions in perfect synchrony, but with the outcome of each coin toss independent of every other outcome. Bernoulli showed that,

as the number of tosses increases, the proportion of all the copies of Kerrich for each of whom the fraction of his trials coming up heads differs from p by less than some arbitrarily small fixed amount approaches 100%. Mathematicians call this phenomenon convergence in probability to the constant limit p .

Two centuries later, in 1909, Borel proved that the same imaginary situation is even more lawful than Bernoulli had supposed. Bernoulli's result does not rule out the possibility that the proportion of heads in the trials of a particular copy of Kerrich could continue indefinitely to wander away from p by at least some fixed nonzero amount. Borel ruled out this possibility: for 100% of the copies of Kerrich, as the number of each man's tosses increases, the proportion of his trials coming up heads must approach and remain arbitrarily near the value p . Mathematicians call this phenomenon convergence with probability 1 or almost sure convergence to the constant limit p (see Loève 1963, pp. 14 and 19 for short proofs). Few people find these results a shock to their intuition.

ONE GOOD URN

Now consider an equally simple experiment. Suppose a very large box (whose capacity can be extended indefinitely by adjunction of similar boxes) initially contains one green ball and one blue ball. Choose one ball at random, look at its color, replace the ball in the box, and add to the box another ball of the same color as the one chosen. At each successive point in time, say once every second, choose one ball at random and then repeat exactly the above. The precise meaning of "at random" is that if there are n balls in the box when a

drawing is made, each ball has an equal chance $1/n$ of being drawn.

The proportion of green balls in the box is the number of green balls divided by the total number of balls, whether blue or green. What will happen to the proportion of green balls as time increases?

Before reading further, please make a serious effort to guess. You have three guesses. When I proposed this problem to a very august mathematical ecologist in the course of a country march, he gave up after two wrong guesses. When I first heard the answer myself, I was astonished both by the general phenomenon it exemplifies and by the particular details. On reflection, I think the general phenomenon permeates population biology. My purpose here is to describe the phenomenon, give some biological examples of it, and suggest its consequences for the interpretation of biological data.

The experiment just described is a special case of what is known as "Polya's urn scheme." Eggenberger and Polya (1923) introduced the scheme in 1923 to model the spread of infection in a population. David Blackwell and David Kendall (1964) studied another generalization of this experiment and even mentioned its implications for stochastic population growth. But an overgrowth of related mathematical results obscured their message for biologists.

So, suppose a single Kerrich performs the experiment with blue and green balls. As time goes on, the proportion of green balls will converge to some limit p . (This fact alone is not obvious.) As in the Bernoulli model, "converge" means that if you pick some nonzero tolerance interval around p , then there is some time at which the proportion of green balls will be in that tolerance

The author is with Rockefeller University, New York, NY 10021.

interval and after which it will never leave it; the proportion approaches and remains near p .

But what is p ? For the first copy of Kerrich, call him Kerrich₁, all one can say is that his value of p , call it p_1 , lies between 0 and 1 inclusive. The chance that his p_1 is exactly equal to any particular fixed p between 0 and 1 is zero! However, the chance that his p_1 falls between 0.2 and 0.3 inclusive is exactly $0.3 - 0.2 = 0.1$.

At the next desk, Kerrich₂ is finding that the proportion₂ of green balls in his box₂ is getting and remaining closer and closer to a fixed number p_2 . But whereas Kerrich₁'s proportion seems to be approaching $p_1 = 0.2435871 \dots$, his proportion₂ is approaching $p_2 = 0.9342265 \dots$. And on his other side, Kerrich₃'s proportion is approaching $p_3 = 0.59943312 \dots$. Each man's proportion of green balls converges to a limit, which is constant for each particular man but which varies from one man to another, even though all are performing exactly the same experiment. In this case, the limiting proportion p is uniformly distributed over the interval from 0 to 1: that is, the chance that Kerrich₁₇'s limit p_{17} falls between a and b , where $0 \leq a \leq b \leq 1$, is $b - a$.

Blackwell and Kendall (1964) proved that if the box starts out with one ball of each of k different colors, where k may exceed 2, then the limiting distribution of proportions of each of the k colors is uniform over the set of all possible ways of dividing 100% into k proportions.

The behavior of this hypothetical experiment exemplifies what mathematicians call almost sure convergence to a nondegenerate limit random variable. "Nondegenerate" means that the limiting value of the proportion of green balls is not restricted to a single point. In Bernoulli's model of the real experiment which the real Kerrich performed, the limit random variable is degenerate because every such Kerrich would (in theory) have obtained the same limiting proportion p of heads.

This hypothetical experiment behaves identically to an apparently quite different experiment. Suppose each Kerrich has a box with one green and one blue ball. He receives a coin; one side is green, one blue. Though the coin looks fair, the real probability p that the coin will come up green is distributed uniformly between 0 and 1. For any given coin, p is constant in time. No man has any reason to suspect that his coin

differs from any of the others; in particular, he does not know his coin's value of p . Once a second, each man flips his coin and adds to his box a ball of the color indicated. Then (since this experiment is just Bernoulli's model and Borel's theorem applies) each man's proportion of green balls converges with probability 1 to his coin's p .

Here each man's limiting proportion p is assigned first. The color of the next ball is chosen by an independent trial with probability p of green. In the previous hypothetical experiment, each new ball's color is determined by random choice among the colors which have occurred so far. To an observer of the balls deposited in the boxes, the two experiments are indistinguishable.¹

Before proceeding to biological examples, I want to emphasize what almost sure convergence to a nondegenerate limit random variable looks like to people participating in an experiment with this property. With increasing time, each man's experiment settles down to systematic, regular, and lawful behavior; his graphical plot of proportion green as a function of time wiggles at first but smooths out gradually to a steady flat line. However, if he repeats the experiment or gets a friend to do so under identical conditions, where the curve of the replicate experiment levels out seems to bear no relation to the original.

It is only after a change in the level of analysis—only after considering an ensemble of replications—that regularity and simplicity reappear. It is the law of the limit random variable that is simple.

In retrospect, Bernoulli himself made just such a change in the level of analysis. If each copy of Kerrich were to toss a coin just once, then Kerrich₁ might get heads, Kerrich₂ tails, Kerrich₃ again heads, and so on without apparent pattern. As the size of the ensemble of copies of Kerrich increases, however, the proportion of copies whose single trial results in heads approaches the limit p near one-half. Already Bernoulli knew that life is lawful to the ensemble though chaos to the individual.

BACK ON THE FARM

Now suppose that a breeding stock on a pig farm is maintained by mating a

¹Violet Cane (1973) has discovered an equally surprising, and closely connected, observational equivalence of models for negative binomially distributed counts, such as accident statistics.

boar and a sow each generation. One male and one female from the offspring are chosen to mate in the next generation. Suppose there is a single gene locus at which, in the initial generation, both parents are heterozygotes. For example, each has genotype Aa. Since each offspring receives one allele chosen at random from each of its parents, there is positive probability that both offspring will have the genotype aa. If this happens, all future offspring will have the same genotype at that locus. There is an equal positive probability that both offspring will have the same genotype AA, with the same consequence.

Sooner or later both offspring must become homozygous for the same allele, and geneticists have calculated the rate of approach to homozygosity under the regime of inbreeding just described. The offspring of a particular pair of heterozygous parents will fixate on the genotype aa with probability one-half and on the genotype AA with probability one-half.

Aside from their good looks and intelligence, pigs are bred for characteristics of commercial interest such as quantity of edible meat. These quantitative characters are believed to be controlled by the additive effects of genes at several loci. Suppose, for the sake of illustration, that weight is controlled by five independently assorting loci with alleles A, a; B, b; and so on up to E, e. Let homozygosity for the capital letter at a locus correspond to an increase in one kilogram over the heterozygote and homozygosity for the small letter at a locus correspond to a decrease in one kilogram below the heterozygote.

If a breeding line is started with parents both of genotype AaBbCcDdEe, then eventually the descendants in that line are certain to drift to homozygosity, at each locus the same for both male and female. The weight of pairs in successive generations will cease fluctuating eventually, all else being equal, and will be the same for both members of the pair. Their weight at fixation will be 5 kilograms above those of their initial ancestors if all five loci fixate at capital letters, 3 kilograms above if four of the five loci fixate at capitals, 1 kilogram above if three of the five fixate at capitals, or symmetrically below the weights of their ancestors. The weight at fixation of another line of descent might differ. As the size of an ensemble of lines of descent increases, the proportions of lines at each weight approach the probabilities calculated from a

binomial distribution with parameters 5 and $1/2$ (roughly, a bell-shaped histogram with its highest values symmetrically placed on either side of the ancestral weight).

When a selective breeding program uses a finite stock of pigs (and infinite numbers of pigs have not yet been observed), this underlying drift, due to random sampling of genes, sets limits to what selection can accomplish. Moreover, drift to a nondegenerate limit random variable sets different limits in different replications of an identical breeding program. As Robertson (1960, p. 244) observed: "If $u(q)$ [the chance of fixation of an allele whose frequency at the beginning of a breeding program is q] is very different from unity for many genes, we will notice that replicate lines from the same initial population will be very different in the limit they reach." In our example, $q = 1/2$ and $u(q) = 1/2$ for all five loci, so Robertson's warning applies.

A failure to recognize the nondegeneracy of the limit random variable to which polygenic characters drift has practical consequences. Hill (1971, p. 294) points out that some authors estimate realized heritability in a single selection program "by fitting a linear regression to cumulative response and cumulative selection differential each generation. But with genetic sampling (drift) the variance of the population mean increases each generation, and these means become correlated. In standard regression analysis the observations are assumed to have equal variance and be uncorrelated, so that the estimates of variance of realised heritability obtained by... using standard regression techniques are biased downwards. In other words, the observed variance among heritability estimates from a replicated experiment would exceed the variance predicted from a single replicate." Hill gives an explicit quantitative analysis of what nondegenerate drift does and what to do about it in an important series of papers (most recently, 1974).

While geneticists have long known of genetic drift and have recently assayed its practical impact on breeding programs, other areas of population biology seem to have remained in bliss. Suppose two bacteria, say a wild type and a mutant, are distinguishable by some marker but are absolutely identical with respect to growth in a particular culture, which is sufficiently favorable to growth that no deaths occur. After a while one or the other of the bacteria will divide,

giving three bacteria. Then, one of those three will divide, each one being equally likely, and so on. If we ignore the interval between divisions and advance an artificial clock by one unit at each division in the culture, we obtain exactly the Polya urn model. If we identify the wild type with green balls, and the mutant with blue, then after a long time, since the proportion of green balls converges to a limit, so will the proportion of wild type bacteria in the culture, and to the same limit. The chance that this limit is exactly one-half is zero. If the limit is p , the culture would behave as if each new bacterium added were wild type with probability p . Blackwell and Kendall (1964, p. 295) state succinctly: "This might lead the incautious observer to attribute a real difference to the... clones in respect of their growth mechanism, although in fact they are in all ways identical." The same phenomenon might lead incautious observers to infer that a genetic change affecting growth had occurred if they attempted to replicate the experiment and found, as they must, a different limiting proportion of the wild type.

Similarly, suppose that individuals of a growing population fall into one of k age categories, where k may exceed 2. Under certain assumptions (Athreya and Ney 1972, p. 206), which may even be defensible in some real situations, the proportions in each class will approach proportions which depend only on the fertility and mortality, but not on the initial numbers of individuals, of each age class. Moreover, the population will (with positive probability) eventually grow exponentially at a rate which also depends only on the fertility and mortality of each age class. If total population size is plotted on a logarithmically scaled ordinate against time on the abscissa, the graph will eventually fall along a straight line. The point at which this straight line intersects the time axis is where a deterministically growing population with the same growth rate would have had to begin growing exponentially in order to fall into step alongside the stochastically modeled population. Call this point the lag time. It is a nondegenerate random variable. Though the laws of growth are the same, the lag times, or times to apparent exponential take-off of growth, of initially identical populations obeying this stochastic model are different.

No deterministic interpretation of such differences in limiting proportions

or lag times could possibly be right, though the differences are real enough. The variation in an ensemble of replicates must become the object of study when the limit random variable of an individual replicate is nondegenerate. Luria and Delbrück (1943) practiced this precept in their classic experimental proof that phage-resistant mutants arise randomly.

But population biologists who study macroscopic populations seem less inclined to this view of nature. Here are a few heretical possibilities. Is it possible that differences in successional changes and in so-called climax state in apparently similar habitats are not to be explained as due to any causal difference between the habitats, but should be interpreted as variation in an ensemble of such habitats? Is it possible that the differences in species composition of apparently similar islands result from the operation of identical forces which produce regularity only in an ensemble of islands? Could observations that one animal population cycles with a period of 4 years and another with a period of 3 or 13 or 17 years become intelligible if the ensemble of periods of cycling animal populations were examined? Could the differences in the sizes of prides of lions or in the social organization of troops of Japanese macaques reflect the inherently but lawfully variable outcome of identical underlying stochastic forces, rather than deterministic ecological differences?

In behavior, for example, is it possible that some of the significant differences among mother-child interactions, which are obvious by the time a child reaches five years of age, are due neither to inherent differences among individuals nor to environmental differences, but to sequentially dependent random forces applying equally to all mother-infant pairs? (Which hand the mother first holds the infant with is random, perhaps, but that choice affects the skill with which she performs tasks with her remaining hand, which affects the infant's response, which affects...) Could nondegenerate stochastic limits provide useful models of what students of plant development (Evans 1972) call "ontogenetic drift"?

The possibilities I raise will leave cold or enrage people who believe they *know* that deterministic factors explain some of the differences I cite. They may well be right in part. All I suggest is that there may be variation which deterministic factors do not usefully explain and

that the possibility of understanding phenomena is preserved by redirecting attention to a lawful-looking ensemble.

CHAOS AND COSMOLOGY

May (1974, p. 645) has emphasized the ecological interpretation of the mathematical fact that very simple deterministic difference equations can have astonishingly messy trajectories, including "cycles of any period, or even totally aperiodic but bounded... fluctuations." The recent, still unpublished work of several people shows that many (though not necessarily all) difference equations studied by ecologists can act so weirdly. Implicit is the suggestion that the apparent variability of population fluctuations may represent the working of a simple deterministic mechanism. The behavior predicted by this mechanism is so sensitive to the values of the parameters in at least some ranges that it will probably be necessary to compare observations with a probabilistic approximation. Thus, the apparent variability of population fluctuations, for example, is interpreted at two levels in the models May and others consider:

INDUSTRY, GOVERNMENT, ACADEME ARE YOU SEEKING BIOLOGISTS?

Your job description can be listed without charge in the *AIBS Employment Newsletter*, which is distributed to approximately 300 job-seekers.

The *Newsletter* is issued six times annually (February, April, June, August, October, and December) to qualified persons trained in all disciplines of the life sciences.

To take advantage of this opportunity to recruit your personnel from the widest possible selection, send your job descriptions to the *AIBS Employment Newsletter*, 1401 Wilson Blvd., Arlington, VA 22209. Telephone: (703) 527-6776.

first, in the complexity of the trajectories predicted with fixed parameters (including initial values); and second, in the impossibility of estimating exactly, and the likely actual fluctuation of, the parameter values. These models do not attempt to account for uncertainty or fluctuations in parameters but assume, at the kernel of phenomena, a simple determinism.

It seems impossible to reject with any data an affirmation of faith that a deterministic mechanism could supply sufficient apparent variability to describe a real population whose parameters were known and constant. The preceding biological examples, and others which could be cited, suggest an alternate view: At least some biological processes incorporate stochastic elements that can cause long-term behavior which appears lawful only in an ensemble of replicates. The empirical program suggested by this view is to examine such ensembles.

Worn exclusively, the deterministic glasses of Laplace and the stochastic glasses of Charles Sanders Peirce give equally roseate views of the world. In the interest of fair advertising, I have to admit that the strategy of moving from the individual to the ensemble to find order in variability will not always work. There are stochastic processes which approach a limit (any kind of limit, degenerate or not) only with a probability 0. Some misanthropes claim experience is like that, too: Some parts of nature simply change more slowly than others, they say, and those parts that change slowly compared to the time scale we are interested in serve us as points of reference, or limits built on sand. Such misanthropes may be right.

Having speculated thus far, let me raise and answer a metaphysical question prompted by Polya's urn which would, I hope, have amused Peirce as it amuses me on dark nights. If you and I had been born in another universe which had started from exactly the same initial conditions as our present one and which had been subject to the same dynamics, would we necessarily infer the same laws of nature as we (in the collective sense of civilized thought) infer for this universe? I take the existence of genetic drift on pig farms as establishing a stochastic element in the dynamics of the universe, and therefore

²Anne Whittaker points out that Ray Bradbury has dramatized this possibility (see Bradbury 1962).

have no guarantee that the apparent lawfulness in this copy of the universe would take the same form in any other. The order of this universe may be an irreproducible result.²

ACKNOWLEDGMENTS

This paper is based in part on a talk given in the Provost's Lodge, King's College, Cambridge, 25 April 1975. For extraordinary hospitality during 1974-75, I thank members of King's College Cambridge and the University of Cambridge Statistical Laboratory. I thank Robin Sibson for excellent teaching; and John Hajnal, David Kendall, and the referees for helpful comments. This work was supported by King's College Research Centre and the U.S. National Science Foundation.

REFERENCES CITED

- Athreya, K. B., and P. E. Ney. 1972. *Branching Processes*. Springer, New York.
- Blackwell, D. and D. G. Kendall. 1964. The Martin boundary for Polya's urn scheme and an application to stochastic population growth. *J. Appl. Probab.* 1(2): 284-296.
- Bradbury, R. 1962. A sound of thunder. Reprinted in *R Is for Rocket*. Doubleday, Garden City, New York, 1952.
- Cane, V. R. 1973. The concept of accident proneness. *Académie Bulgare de Sciences Bulletin de l'Institut de Mathématiques* 15: 183-188.
- Eggenberger, F., and G. Polya. 1923. Über die Statistik verketteter Vorgänge. *Zell angew. Math. Mech.* 3: 279-289.
- Evans, G. C. 1972. *The Quantitative Analysis of Plant Growth*. Blackwell Scientific, Oxford.
- Hill, W. G. 1971. Design and efficiency of selection experiments for estimating genetic parameters. *Biometrics* 27: 293-311.
- _____. 1974. Variability of response to selection in genetic experiments. *Biometrics* 30: 363-366.
- Kerrick, J. E. 1946. *An Experimental Introduction to the Theory of Probability*. Einar Munksgaard, Copenhagen.
- Loève, M. 1963. *Probability Theory*, 3rd ed. D. Van Nostrand, Princeton.
- Luria, S. E., and M. Delbrück. 1943. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* 28: 491-511.
- May, R. M. 1974. Biological populations with nonoverlapping generations: stable points, stable cycles, and chaos. *Science* 186: 645-647.
- Robertson, A. 1960. A theory of limits in artificial selection. *Proc. R. Soc. Lond. B Biol. Sci.* 153: 234-249.