structure that may be obscured by even small amounts of noise). Smith (1991) has discussed the problem of dimension estimation for systems with this type of observational noise.

A different type of noise results when the system under evolution undergoes perturbations due to some external force or change. The perturbations are then propagated through the system. This also describes the situation of rounding error in numerical simulations of chaotic systems; the original rounding error is repeatedly magnified by the "stretching" behavior of the map and the computed numerical trajectory [called a pseudo-orbit by Hammel et al. (1988)] diverges far from the true path. Hammel, Yorke and Grebogi pointed out that often pseudo-orbits are in fact true orbits corresponding to different initial conditions, but even in the ergodic case, this is not necessarily reassuring. The dyadic map $T(x) = 2x \pmod 1$ on the unit interval is ergodic and chaotic with the uniform distribution as invariant measure. However, all orbits of this map quickly iterate to zero on the computer. These are true orbits of the system; unfortunately, they correspond to initial conditions (dyadic rationals) that are attracted to the fixed point at zero and do not exhibit "typical" system behavior. Thus, in numerical simulations, it is not always easy to determine whether observed behavior is "real" or an artifact of the simulation procedure.

Corless (1991) has looked at the related problem of approximating solutions to differential equations by numerical methods (here again the computed solution may not resemble the intended system; see Hockett, 1990 and Corless, Essex and Nerenberg, 1991) and has proposed an "operational" definition of chaos. He suggests that a system should be considered chaotic if all "nearby" solutions are chaotic (regardless of the actual properties of the system itself). The reasoning here is that perturbations will cause any physical system to be pushed into neighboring states and these should be the real objects of study.

Comment: Inference and Prediction in the Presence of Uncertainty and Determinism

John Geweke

1. INTRODUCTION

The discovery of nonlinear determinism and chaos in physical systems, the study of these phenomena by physicists and mathematicians and their consideration by investigators in a wide array of disciplines have been ably surveyed from the perspective of statistics and probability in these two articles. The authors have indicated clearly that the contributions and relevance of statistical science are still unresolved, and some basic questions are open. Because chaotic dynamics generate realizations that can be characterized as purely random, what is role of stochastic modeling? If observed deterministic nonlinear processes always interact with stochastic processes, then are the conventional tools of statistical inference any less adequate here than elsewhere? The resolution of these issues will take time, and these surveys will contribute to this process by having brought the statistically relevant aspects of nonlinear determinism and chaos to a wider audience.

Independent of how these questions are answered, the models discussed bring to the practical level latent questions about the implications of determinism for the fundamental role that randomness seems to play in so much of statistics. Berliner has discussed these matters in the final section of his contribution. I have found deterministic models an enlightening vehicle for taking up these questions on a practical level, and in these brief remarks I will provide a few illustrations. The next section provides an approach to inference and prediction in the nonstochastic world of the models these authors have discussed. The likelihood function is presented for two simple models in Section 3, and the construction of predictive densities (for the past, as well as the future) is illustrated in Section 4.

John Geweke is Professor of Economics, University of Minnesota, Minneapolis, Minnesota 55455.

2. LIKELIHOOD FUNCTION AND PREDICTIVE DENSITIES

We address a deterministic system evolving in discrete time,

$$x_t = f(x_{t-1};\theta) \quad (t = 1, 2, \ldots)$$

and the implied iterates

$$x_t = f^{(n)}(x_{t-n};\theta) \quad (t = n+1, n+2, \ldots)$$

defined recursively as $f^{(1)} = f$, $f^{(n)} = f[f^{(n-1)}]$ for $n = 2, 3, \ldots$. The functional form of f is known, but the initial condition x_1, or the parameter θ, or both, are unknown; in all that follows, we shall take both to be unknown, and the specialization to other cases will be obvious. Both are vectors of finite order. The process $\{x_t\}$ is not observed directly. Rather, the deterministic function $d_t = m(x_t)$ is recorded for $t = 1, 2 \ldots T$. The function m is a partitioning of the phase space into a finite set of regimes $\mathbf{D}$, as assumed in symbolic dynamics (Destexhe, 1990). This formulation is sufficiently general to include missing observations, symbols depending only on a subvector of x_t and a wide variety of other problems that can arise in data records; but, it explicitly rules out the possibility of stochastic measurement error that is taken up in Geweke (1989).

Let the function $J(\cdot,\cdot)$ be defined on $\mathbf{D} \times \mathbf{D}$, with $J(a, b) = 1$, if $a = b$ and $J(a, b) = 0$ otherwise. Then, the likelihood function for the unknown initial condition x_1 and parameter θ, given the data set $\{d_1, d_2, \ldots, d_T)$ is

$$L(x_1, \theta \mid d_1, d_2, \ldots, d_T) = \prod_{t=1}^{T} J\{m[f^{(t-1)}(x_1, \theta)], d_t\}.$$

The likelihood function takes on only two values, 0 and 1. If $L(x_1, \theta) = 0$ for all x_1 and θ, then the model is inappropriate regardless of the prior. If $L(x_1, \theta) = 1$ for some values of x_1 and θ, then posterior and predictive densities may be constructed. Let $\Pi(x_1, \theta)$ be a prior cdf for x_1 and θ. (We return, in Section 4, to aspects of the construction of coherent priors specific to chaotic models.) Let $g(x_t)$ be a function of x_t, whose expected value under the posterior is of interest. Then,

$$E(g(x_t) \mid d_1, d_2, \ldots, d_T; \Pi) = \int\int g[f^{(t-1)}(x_1;\theta)]\, L(x_1, \theta)\, d\Pi(x_1, \theta) \quad (t = 1, 2, \ldots).$$

Posterior cdf's and pdf's may be constructed by letting $g(\cdot)$ rotate within the appropriate set of indicator functions. The predictive density for future recorded values is

$$P_t(d_t) = \int\int J\{m[f^{(t-1)}(x_1, \theta)], d_t\} \cdot L(x_1, \theta)\, d\Pi(x_1, \theta) \quad (t = T+1, T+2, \ldots).$$

Consistent with Berliner's discussion, we see that the principles of Bayesian inference are not changed in any important way by the complete absence of any stochastic terms in the model. Rather, models of this kind bring to the fore subjective uncertainty rather than random variables as the generator of the problems addressed in statistical inference and prediction. To make these points more concrete, we next take up the construction of the likelihood function and predictive pdf's in two simple models. These applications show that the procedures just outlined can be made practicable, and they also uncover some unusual likelihood functions.

3. THE LIKELIHOOD FUNCTION: EXAMPLES

To illustrate the construction of the likelihood function, consider two specific maps. The first is a tent map,

$$x_t = \alpha(1 - 2\,|\,x_{t-1} - 0.5\,|),$$

with parameter value $\alpha = \alpha^* = 0.8$ and initial condition $x_1 = 0.35$. This is not quite the same as the tent or hut map described in the foregoing article by Chatterjee and Yilmaz; both are special cases of the tent maps discussed by Jackson (1989, Section 4.2). This map generates chaotic behavior for $\alpha \in (0.5, 1.]$, with basin of attraction $[2\alpha(1-\alpha), \alpha]$ (Ott, 1981). The second map is logistic,

$$x_t = \alpha x_{t-1}(1 - x_{t-1}),$$

with parameter value $\alpha = \alpha^* = 4.0$ and initial condition $x_1 = 0.35$. The behavior of this process is described in the foregoing article by Berliner (Section 2.1) and Chatterjee and Yilmaz (Section 1.1). Each process is recorded using three symbols, $d_t = 3[x_t] + 1$. For each process, $\{x_t\}$ is indicated for $t = 1, \ldots, 21$, in Table 1, and the corresponding $\{d_t\}$ is indicated in the column headed "Symbol A."

The likelihood function corresponding to the tent map is displayed in Figures 1 and 2. The shaded areas correspond to the value 1, the unshaded areas to the value 0. Because the points x_1 and $1 - x_1$ each generate the same $\{x_t, t > 1\}$, and because the recorded symbol for x_1 is consistent with values of x_1 in the interval (1/3, 2/3) that is centered on the point 0.5, the likelihood function is symmetric in x_1. It immediately follows that x_1 cannot be consistently estimated. The figures also reflect the

TABLE 1
Generated series and associated symbols

t	x_t	Symbol A	Symbol B	x_t	Symbol A	Symbol B
1	0.350000	2	3	0.350000	2	2
2	0.560000	2	4	0.910000	3	5
3	0.704000	3	5	0.327600	1	2
4	0.473600	2	3	0.881113	3	5
5	0.757760	3	5	0.419012	2	3
6	0.387584	2	3	0.973763	3	5
7	0.620134	2	4	0.102192	1	1
8	0.607785	2	4	0.366996	2	2
9	0.627544	2	4	0.929240	3	5
10	0.595929	2	4	0.263011	2	2
11	0.646513	2	4	0.775345	3	4
12	0.565580	2	4	0.696740	3	4
13	0.695073	3	5	0.845174	3	5
14	0.487884	2	3	0.523421	2	3
15	0.780614	3	5	0.997806	3	5
16	0.351018	2	3	0.008757	1	1
17	0.561629	2	4	0.034722	1	1
18	0.701394	3	5	0.134065	1	1
19	0.477770	2	3	0.464367	2	3
20	0.764432	3	5	0.994921	3	5
21	0.376909	2	3	0.020213	1	1

fact that $\{(x_1, \theta): L(x_1, \theta \mid d_1, d_2, \ldots, d_T) = 1\} \supseteq \{(x_1, \theta): L(x_1, \theta \mid d_1, d_2, \ldots, d_{T+s}) = 1\}$ for all $s > 0$. Figure 1 demonstrates two of the bifurcations of the likelihood function that are characteristic of this map and symbolic representation, the first occurring between $T = 2$ and $T = 3$, and the second between $T = 9$ and $T = 10$. Figure 2 illustrates the behavior of the left lobe of the likelihood function as sample size increase from $T = 10$ to $T = 18$. All that is apparent in these figures is a steady shrinking in the size of two detached regions of positive likelihood. However, this is an artifact of the plotting algorithm, which shows insufficient detail. A further magnification, not presented here, reveals that the two lobes shown at $T = 18$ are each comprised of two disjoint, parallel strands.

The likelihood function corresponding to the logistic map is displayed in Figures 3 and 4. Symmetry about $x_1 = 0.5$ occurs again for the same reason, and bifurcation occurs at least twice, once between $T = 6$ and $T = 7$, and once between $T = 10$ and $T = 11$. However, no evidence of further bifurcation with increasing sample size turns up. Rather, the three strands persist for values of α close to

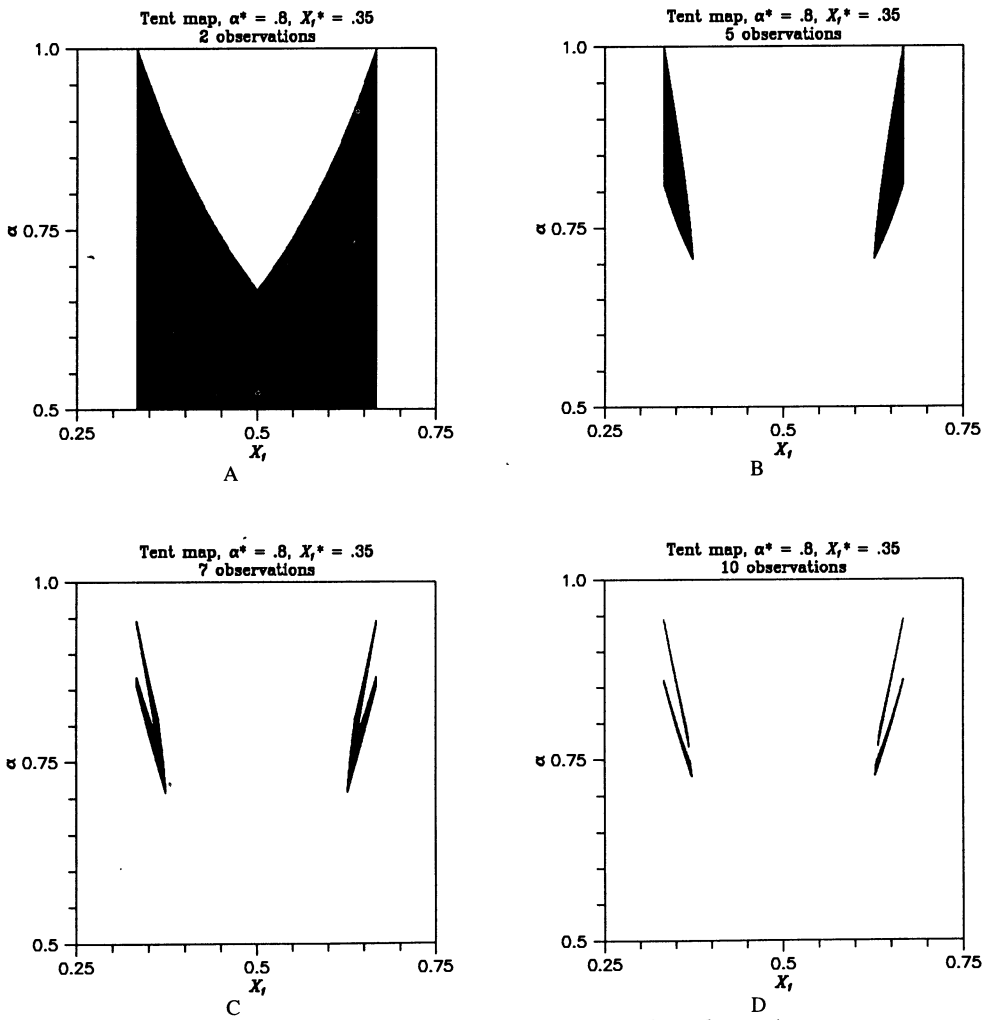

FIG. 1. *Likelihood surface, tent map*: 2, 5, 7 *and* 10 *observations.*

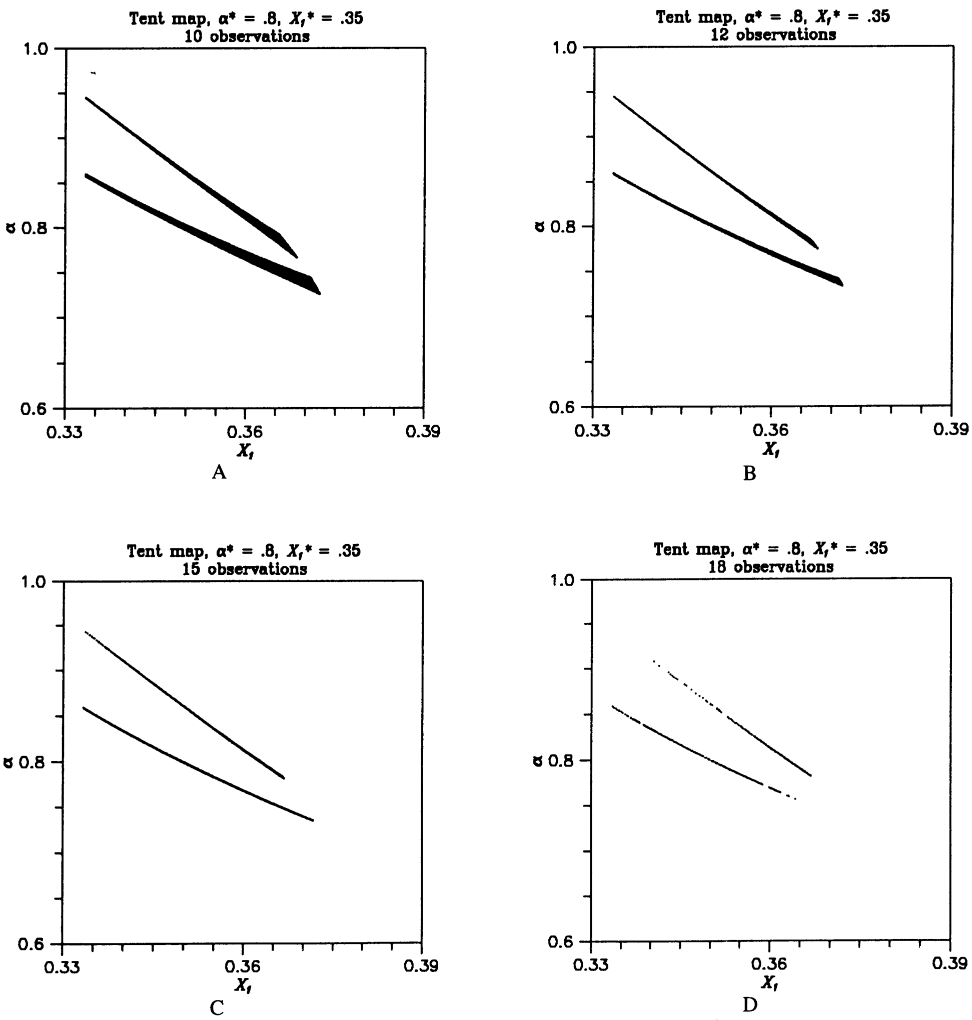

FIG. 2. *Likelihood surface, tent map*: 10, 12, 15 *and* 18 *observations.*

4.0, at least through sample size $T = 21$. The similarity of this likelihood surface, and that for the tent map, to the phase space for various chaotic maps is apparent. To explore the depth of this connection fully is beyond the scope of these comments, however.

4. THE PREDICTIVE DENSITY: EXAMPLES

No matter how great the sample size, uncertainty remains about $\{x_t\}$, both within and beyond the sample, and there will be uncertainty about $\{d_t\}$ for t sufficiently larger than T. The general procedures of Section 2 for Bayesian inference may be applied to express this uncertainty, using an appropriate family of priors and simple numerical integration. We present results here using a reference prior $d\Pi(\alpha)$ that places a uniform distribution on the parameter α, over the interval [0, 1] in the case of the tent map and over the interval [0, 4] in the case of the logistic map. Conditional on $d\Pi(\alpha)$, the prior $d\Pi(x_1 \mid \alpha)$ is taken to be the ergodic distribution for $\{x_t\}$, given α. More elaborate reference priors for α would also be interesting, especially for the logistic map in light of the sensitivity of properties of the trajectory to values of α as described by Berliner in Section 2.1.1.

This reference prior is easy to implement in an effective (if crude) scheme for numerical integration. Draws are made from each lattice of a grid of points on the spaces for α and the x_t. For the tent map, α ranges from 0 to 1 and the grid is 20,000 by 10,000, and for the logistic map α ranges from 0 to 4 and the grid is 20,000 by 20,000. In both cases, the space for x_t is the unit interval. For each grid point, the map is iterated 20 times. The motivating idea behind this procedure is that, for a given value of α, the 10,000 or 20,000 values at the twentieth iterate produce an approximation to the density $d\Pi(x_1 \mid \alpha)$ just described. The approximation would become exact if the 20 iterations were increased without limit; but, for an increase not too far beyond 20, the approximation of the true iterated mapping by double precision arithmetic breaks down (Geweke, 1989). I conjecture that the quality is better for nonperiodic mappings (the tent map

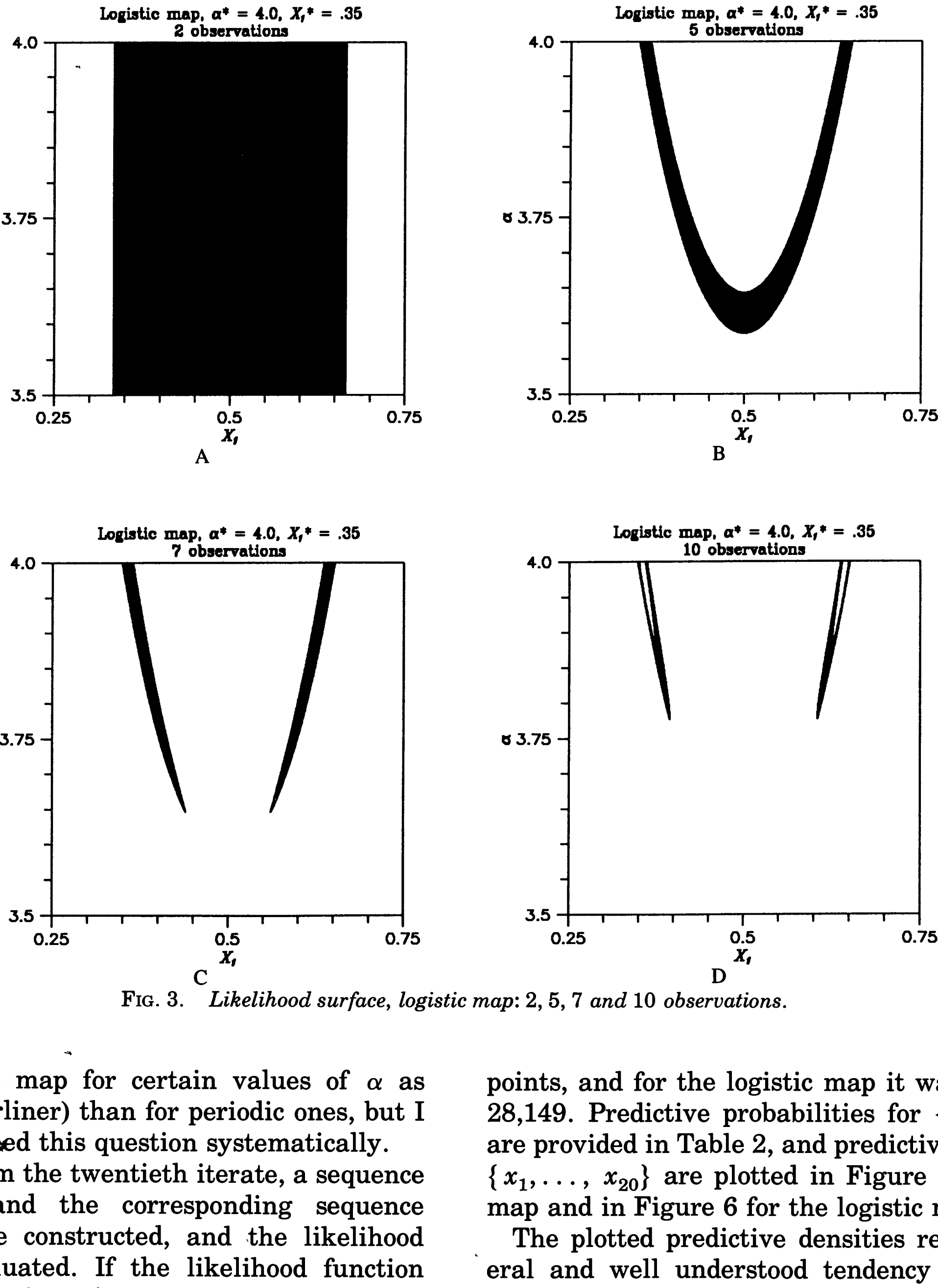

FIG. 3. *Likelihood surface, logistic map*: 2, 5, 7 *and* 10 *observations.*

and the logistic map for certain values of α as discussed by Berliner) than for periodic ones, but I have not examined this question systematically.

Beginning from the twentieth iterate, a sequence $\{x_1, \dots, x_T\}$ and the corresponding sequence $\{d_1, \dots, d_T\}$ are constructed, and the likelihood function is evaluated. If the likelihood function is 1, then the values $\{x_1, \dots, x_T\}$ are retained, and values $\{x_{T+1}, \dots, x_{T+s}\}$ beyond the sample are generated, as are the corresponding $\{d_{T+1}, \dots, d_{T+s}\}$. This procedure was applied for the tent and logistic maps described in the previous section, taking the first 10 observations presented in Table 1 as the sample in each case. The measurement process was different than the one described in Section 3 that underlies Figures 1 through 4. For the tent map, $d_t = m(x_t) = [6x_t] + 1$, and for the logistic map $d_t = [5x_t] + 1$. Because the basin of attraction is (0.32, 0.8) in the former case and $(0, 1)$ in the latter, the range of $\{d_t\}$ consists of five values in both cases. For the tent map, the likelihood function was nonzero for 51,508 points, and for the logistic map it was nonzero for 28,149. Predictive probabilities for $\{d_{11}, \dots, d_{20}\}$ are provided in Table 2, and predictive densities for $\{x_1, \dots, x_{20}\}$ are plotted in Figure 5 for the tent map and in Figure 6 for the logistic map.

The plotted predictive densities reflect the general and well understood tendency for increased certainty about earlier time periods inherent in chaotic maps (Geweke, 1989). However, there are notable local exceptions to this tendency within the sample that result from the discontinuity of the function $m(\cdot)$. For example, the density of x_6 is more concentrated than that of x_3 in the tent map, and in the case of the logistic map, the predictive density of x_6 is more concentrated than that of any earlier x_t. For the tent map, the dispersion of the predictive density increases monotonically for $t > 11$, essentially reaching the ergodic distribution by $t = 15$. The dispersion in the out-of-sample predictive density for the logistic distribution also increases with t, but variations from period-to-period are much greater and persist through $t = 20$.

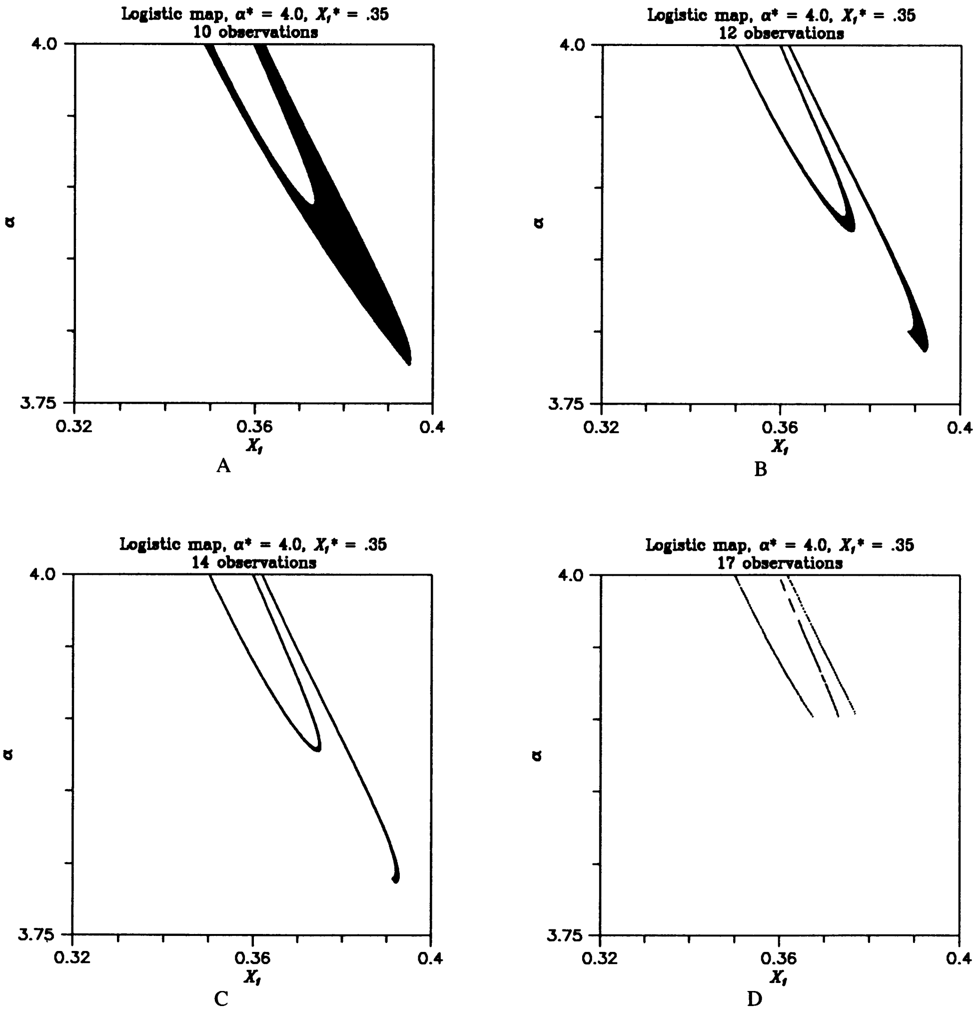

FIG. 4. *Likelihood surface, logistic map*: 10, 12, 14 *and* 17 *observations.*

TABLE 2
Predictive probabilities for $d_t = j$

	Tent map						Logistic map				
t	$j =$ 1	2	3	4	5	6	1	2	3	4	5
11	0.000	0.000	0.000	0.643	0.357	0.000	0.000	0.000	0.000	0.342	0.659
12	0.000	0.000	0.264	0.492	0.244	0.000	0.005	0.258	0.358	0.300	0.080
13	0.000	0.000	0.185	0.369	0.441	0.005	0.324	0.195	0.156	0.162	0.164
14	0.000	0.043	0.349	0.317	0.288	0.003	0.324	0.195	0.156	0.162	0.164
15	0.000	0.026	0.244	0.408	0.321	0.002	0.068	0.198	0.198	0.195	0.341
16	0.000	0.021	0.263	0.406	0.306	0.004	0.143	0.142	0.109	0.135	0.471
17	0.000	0.029	0.242	0.418	0.307	0.003	0.180	0.263	0.153	0.157	0.246
18	0.000	0.026	0.249	0.416	0.307	0.002	0.108	0.144	0.155	0.182	0.411
19	0.000	0.024	0.249	0.408	0.316	0.003	0.145	0.166	0.188	0.149	0.352
20	0.000	0.027	0.254	0.404	0.312	0.003	0.144	0.180	0.153	0.149	0.374

5. CONCLUSIONS

These simple examples show that Bayesian inference expresses subjective uncertainty in a deterministic world in the same way that it does in a stochastic world. Even the operational details differ in no important respects. On the other hand, as Berliner suggested would be the case in his article, there is not much for the frequentist to do here. In a stochastic world, ignoring subjective uncertainty often leads to an expression of the problem that is irrelevant, but to do so here would assume away the whole problem. That is why the comparison is so stark. As deterministic models (exhibiting chaos

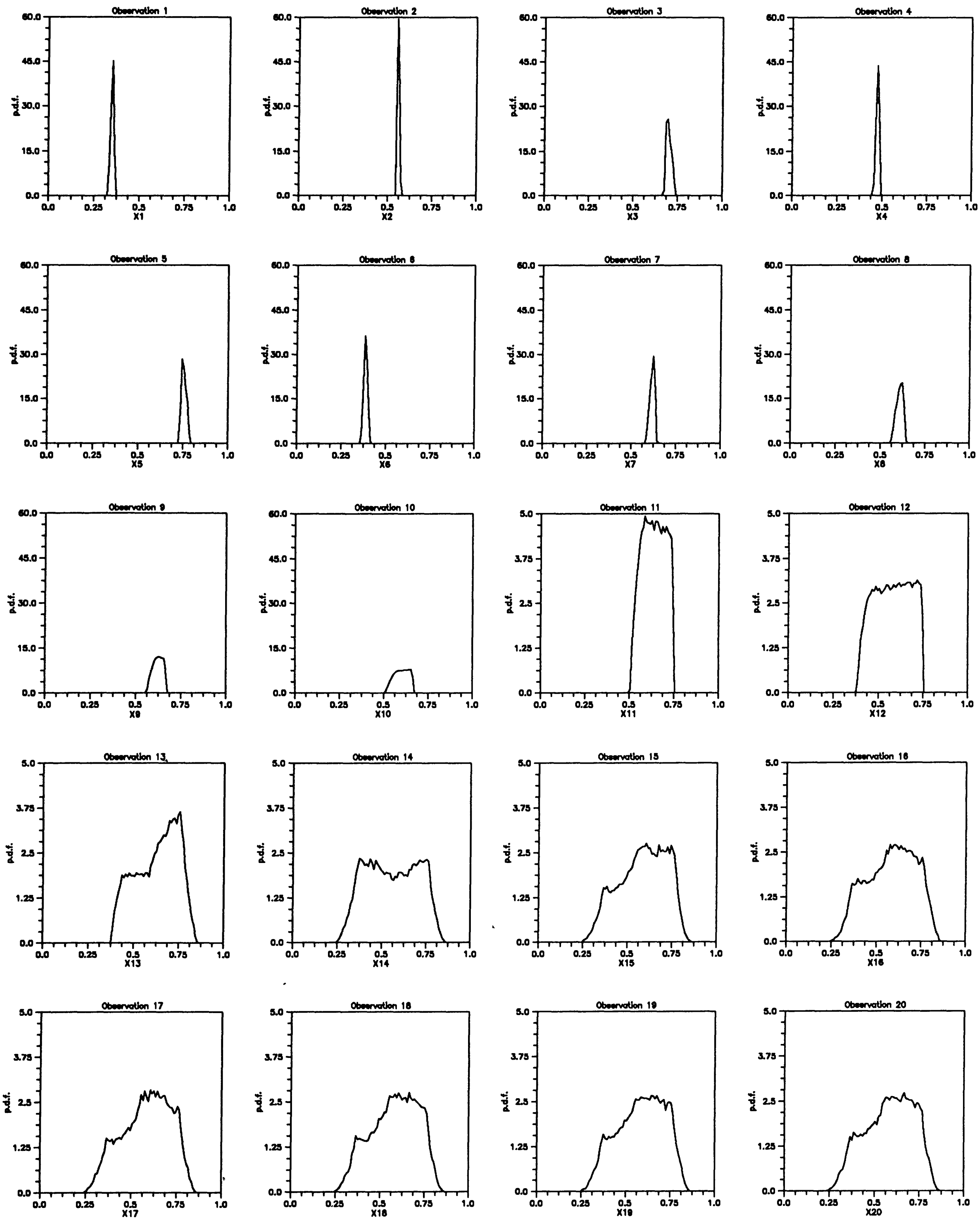

FIG. 5. *Tent map, predictive density.*

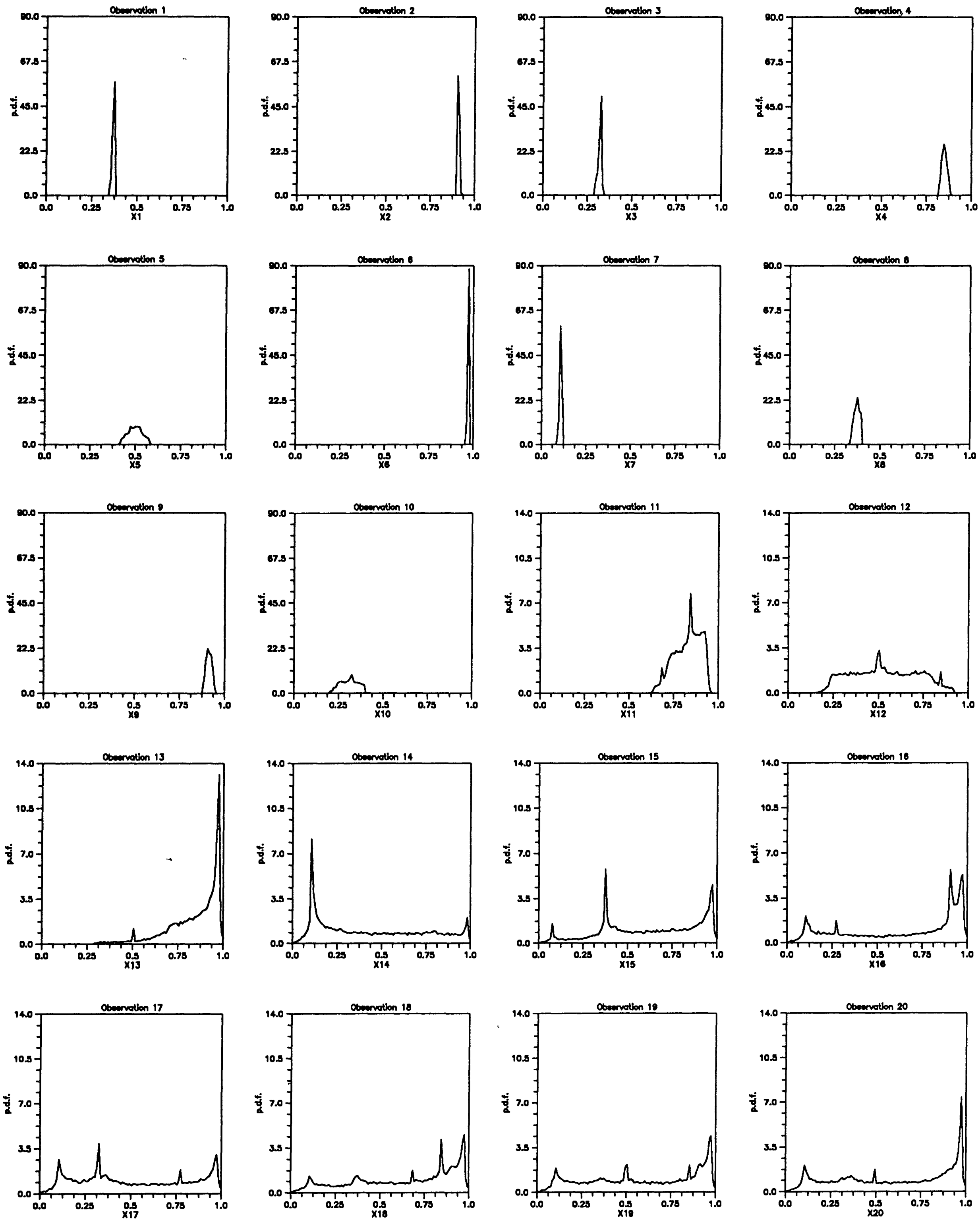

FIG. 6. *Logistic map, predictive density.*

or not) become more refined, the importance and relevance of Bayesian inference to decision making will become ever clearer. The development of practical tools in this environment should be a source of rich, rewarding challenges to Bayesian statistics for some time.

ACKNOWLEDGMENTS

Thanks to Zhenyu Wang for resourceful research assistance, Mark Berliner for helpful discussions and the National Science Foundation for financial support through Grant SES-89-08365.