# BAYESIAN NUMERICAL ANALYSIS

PERSI DIACONIS

Department of Statistics
Stanford University
Stanford, California 94305, U.S.A.

## 1. INTRODUCTION

Consider a given function $f: [0,1] \to \mathbf{R}$ such as

$$f(x) = \exp\left\{\cosh\left(\frac{x + 2x^2 + \cos x}{3 + \sin x^3}\right)\right\}. \tag{1}$$

If you require $\int_0^1 f(x)dx$, a formula such as (1) isn't of much use and leads to questions like "What does it mean to 'know' a function?" The formula says some things (e.g. $f$ is smooth, positive, and bounded by 20 on $[0,1]$) but there are many other facts about $f$ that we don't know (e.g., is $f$ monotone, unimodal, or convex?).

Once we allow that we don't know $f$, but do know some things, it becomes natural to take a Bayesian approach to the quadrature problem:

- Put a prior on continuous functions $C[0,1]$
- Calculate $f$ at $x_1, x_2, \ldots, x_n$
- Compute a posterior
- Estimate $\int_0^1 f$ by the Bayes rule

Most people, even Bayesians, think this sounds crazy when they first hear about it. The following examples may help.

EXAMPLE 1. Brownian Motion. The easiest prior to work with is Brownian motion on $C[0,1]$ with a normal prior on $f(0)$. The Markov property implies that

$$E\{f(t) | f(x_i) = y_i, \quad 0 \le i \le n\}$$

is the straight line interpolent of $(x_i, y_i)$, assuming for simplicity that $x_0 = 0$, $x_n = 1$. This is the Bayes rule for interpolation if squared error is used as loss. The associated quadrature rule is the well known trapezoid rule.

EXAMPLE 2. (Splines). If the prior is taken as once integrated Brownian motion $f(t) \sim \int_0^t B_x\, dx$, an easy computation shows that the Bayes rule is the cubic spline interpolent. Integrating $k$ times yields splines of order $2k + 1$.

Seeing standard procedures emerge from the Bayesian approach may convince readers the argument isn't so crazy after all. The examples suggest the following program: Take standard numerical analysis procedures and see if they are Bayes (or admissible, or minimax).

This program leads to interesting problems and results: Is there a prior giving even order splines? (I don't know.) Is Simpson's rule Bayes? (Only for priors concentrated on cubic polynomials.) Among the possible benefits:

*Benefit (A)* It offers a clear approach to the design problem: Where should we sample $f$ to get the best estimate of the integral? This can have surprising answers, even in the simplest cases. For example, in one-dimension, with Brownian motion as prior the best sampling points are at $x_i = \frac{2i}{2n+1}, 1 \leq i \leq n$, not at $i/n$. Thus, for $n = 2$, it's best not to sample at the ends, but at $\left(\frac{2}{5}, \frac{4}{5}\right)$. Jerry Sacks and Don Ylvisaker have made profound contributions to this program. A recent survey is Sacks and Ylvisaker (1985). An early article that explicitly brings out the connections between time series design problems and Bayesian numerical analysis is Sacks and Ylvisaker (1970).

*Benefit (B)* In truth, we don't really need a lot of new machinery to deal with quadrature in one-dimension. This is a well understood problem. The Bayesian approach generalizes in a straightforward way to $p$-dimensions, where most of us feel we can use all the help we can get. O'Hagan (1985) and Shaw (1985) have recently reported useful results for quadrature in high dimensions using this approach. They worked on the problem of computing posterior expectations. Stewart has reported quite a number of successful applications in real problems. His article in Bernardo et al. (1985) is a convenient reference.

*Benefit (C)* The Bayesian approach yields more than the Bayes rule; it yields a posterior distribution. This can be used to give confidence sets as in Wahba (1983). See Cox (1986) for a more careful study.

*Benefit (D)* The Bayesian approach throws fresh light on the choice of numerical procedures. Consider the Brownian motion prior. In many contexts, the underlying $f$ is known to be smooth, so Brownian motion is not even a rough approximation to the prior knowledge. This suggests smoother priors, and more smoothing in the numerical procedure.

The list of benefits A–D above suggests a mathematics problem. In the special case of Brownian motion we may ask: Is Brownian motion the only prior yielding linear interpolation? A little reflection turns up the following.

(1) If $B_t$ is standard Brownian motion, then $\sigma B_t$ also interpolates linearly, for any fixed $\sigma$, and so for $\sigma$ a randomly distributed scale factor independent of $B_t$.

(2) A Poisson process, or any independent increments process, also interpolates

like $B_t$. This suggests that the underlying space of paths $C[0,1]$ is crucial.

It turns out that (1) and (2) are the only conditions required to get a theorem: the following result has essentially been proved by David Williams (1980).

**Theorem.** *1. If $\pi$ is a measure on $C[0,1]$ that predicts like Brownian motion $B_t$:*

$$E_\pi(f(t)|f(t_i) = y_i,\ 0 \le i \le n) = E_B(f(t)|f(t_i) = y_i,\ 0 \le i \le n)$$

*for every finite sequence $t_1, t_2, \ldots, t_n$, then $\pi$ has the distribution of a scale mixture of Brownian motion: $\pi \sim \sigma B_t$, with $\sigma$ independent of $B_t$.*

REMARK 1. Under $\pi$, the process $f(t)$ is a Martingale. The proof uses Levy's characterization of Brownian motion as the only Martingale with a linear compensator. The article of Dozzi (1981) gives details for a similar characterization of the Brownian sheet.

REMARK 2. The theorem suggests a rich class of open problems: to what extent do Bayes rules determine the prior. Diaconis and Ylvisaker (1979, 1983) survey known results in a finite-dimensional setting. In the present setting, almost all questions are open. One knows that posterior linearity characterizes mixture of Poisson processes on $D[0,1]$, but nothing about cubic splines or other linear operators of Brownian motion.

REMARK 3. The characterization of Brownian motion reported in Theorem 1 above is closely connected to versions of de Finetti's theorem. See Buhlman (1963) or Freedman (1963).

Section 2 of this paper sets out some history in modern language. Bayesian numerical analysis dates back (at least) to Poincare (1896). Section 3 provides entry to modern applications, work of the Gemans and Julian besag on image restoration, work of Steffen Lauritzen on geodetic interpolation, and work of Don Ylvisaker on a tractable class of priors called G-Maps.

The final section states a collection of open problems and conjectures.

Two other surveys detail work not described here: Smale (1985) or Lee and Woznakowski (1986) describe recent work of a group of numerical analysts with a similar flavor. Ylvisaker (1986) gives a comprehensive survey.

## 2. SOME HISTORY

### A. POINCARE'S APPROACH TO INTERPOLATION

An early, clear example of Bayesian numerical analysis appears in Poincare (1896). In Chapter 20, Poincare discusses the theory of interpolation. He states the problem thus: we measure certain values of an unknown function

$f(x)$: $f(a_1), f(a_2), \ldots, f(a_n)$. Construct the curve $y = f(x)$.

Poincare discusses several approaches to this problem: one can always find a function that passes exactly through the observed points, and then seek a function "as continuous as possible". A second approach "also arbitrary to a degree" is to find an approximation of the form $f(x) \doteq c_0 + c_1 x + \ldots + c_q x^q$ with $q$ fixed smaller than $n-1$. Poincare recognizes the difficulties in specifying $q$, and goes on to discuss algorithms involving continued fractions when $q$ is specified.

Poincare's final approach, in Chapter 21 is of particular interest. He begins by supposing that $f(x)$ has a power series expansion

$$f(x) = A_0 + A_1 x + \ldots ,$$

where the $A_i$ are unknown. He puts a prior distribution on the $A_i$, supposing them to be independent, mean zero Gaussian, with variance $\sigma_i^2$. Given $f(a_i)$, he searches for the "most probable value of $f(x)$ for another value $x$".

Poincare treats the problem using the language of Bayesian statistics ("the probability of causes"). He sets up and solves the following problem: $u_1, \ldots, u_p$ are unknown parameters with a Gaussian prior, $y_1, \ldots, y_n$ are $n$ observed linear functions of the $u_i$, find the posterior mean of another fixed linear combination of the $u_i$.

Formally passing to the limits as $p \to \infty$, Poincare arrives at the following solution to the original interpolation problem. Let

$$g(x) = \sigma_0^2 + \sigma_1^2 x + \sigma_2^2 x^2 + \ldots .$$

Then the posterior mean of $f(x)$ given $f(a_i)$ equals

$$\hat{f}(x) = \epsilon_1 g(a_1 x) + \epsilon_2 g(a_2 x) + \ldots + \epsilon_n g(a_n x)$$

with $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ chosen to exactly match the given data when $x = a_1, a_2, \ldots, a_n$. This requires solving $n$ linear equations in the unknowns $\epsilon_1, \ldots, \epsilon_n$.

This solution depends critically on the variances $\sigma_i^2$, and Poincare notes that if the high order variances are set to zero, one gets a new scheme for polynomial interpolation.

All of this took place many years before there was a rigorous theory of Gaussian processes, yet the language and spirit of Poincare's approach are in perfect accord with the modern approach.

## B. Palasti-Renyi's Minimax Result

I. Palasti and A. Renyi (1956) give an early game theoretic formulation of the quadrature problem and prove that the Monte Carlo strategy is minimax. In their setup, the admissible functions are continuous and satisfy

$$\int_0^1 \left\{ f(x) - \int_0^1 f(t)dt \right\}^2 dx = s^2. \qquad (1)$$

One desires to approximate the integral $I = \int_0^1 f(x)dx$ by the sum $S = \frac{1}{n}\sum_{k=1}^{n} f(x_k)$. A pure strategy for Player I consists in the choice of a function $f$. A pure strategy for Player II consists of an $n$-tuple of points $(x_1, \ldots, x_n)$ of (0,1). The loss for Player II is defined as $\Delta = (S-I)^2$. A mixed strategy for II is defined by a measure $\mu$ on the subsets of the $n$-dimensional unit cube. Then the loss is defined as $\Delta(f,\mu) = \int (S-I)^2 d\mu$.

They show that the uniform distribution $\mu_0$ is minimax in the sense that $\sup_f \Delta(f,\mu_0) \leq \sup_f \Delta(f,\mu)$ for any measure $\mu$.

The same type of result is shown to hold for $p$-dimensional integrals. Analogous results hold for the problem of estimating the sum $Y = \sum_{k=1}^{N} y_k$, by $\eta = \frac{N}{n}\sum_{j=1}^{n} y_{k_j}$ where

$$\sum_{k=1}^{N}(y_k - \frac{Y}{N})^2 = s^2 \tag{2}$$

and $(k_1, k_2, \ldots, k_n)$ is some subset of $(1, \ldots, N)$: simple random sampling without replacement is minimax.

Their paper gives rise to many questions: Does Player I have a minimax strategy? If so, to what prior does it correspond? How do minimax strategies change if conditions like (1) and (2) change?

A recent treatment of a similar problem deserves mention: Aldous (1983) considers finding the minimum of function $f$ defined on the vertices of a $d$-dimensional cube. He restricts attention to functions such that the only local minimum is a global minimum. He sets the problem up as a two-person game and finds its value – roughly, the minimax solutions require $2^{d/2}$ steps for any algorithm. Nature's prior on $f$ can be taken as the first hitting times of simple random walk on the cube.

There has also been some work on admissibility of quadrature rules. Ermakov (1968) is an early reference. Granovsky and Knoh (1981) or Granovsky (1986) are recent references.

### C. On Eberlein's Approach to Quadrature

W. F. Eberlein and his students developed an approach to quadrature that is usefully viewed as Bayesian numerical analysis. I will describe the theory in one dimension, although some striking results demonstrating the efficacy of Monte Carlo integration as the dimension increases will be cited.

Eberlein works with the set of all real functions on $[-1,1]$ which have con-

vergent power series expansions.

$$\mathcal{F} = \{f(t) = \sum_{n=0}^{\infty} x_n t^n : \Sigma|s_n| < \infty\}.$$

This space can be identified with elements $x = (x_0, x_1, \ldots)$ of the sequence space $\ell_1$. A prior is put on $\mathcal{F}$ by putting a prior on the coefficients. This in turn is done by a simple "stick breaking procedure".

Choose a sequence of random variables $X_0, X_1, X_2, \ldots$, as follows.

- Choose $X_0$ uniformly in [-1,1].
- Given $X_0$; choose $X_1$ uniformly in $[-(1 - |X_0|), 1 - |X_0|]$.
- Given $X_0, X_1, \ldots, X_n$, choose $X_{n+1}$ uniformly in $[-(1 - |X_0| - |X_n|), (1 - |X_0| - |X_n|)]$.

This puts a measure on $\ell_1$, and so on the space of functions $\mathcal{F}$. Integrals with respect to this measure are denoted

$$\int_{\mathcal{F}} T(f) df.$$

If $T$ is a function that only depends on a finite number of coordinates, then the integrals are computed as finite dimensional integrals. For example, the marginal distribution of $X_n$ is the law of the product of $n+1$ uniforms on $[0, 1]$, with a random $\pm$ sign attached (see Feller (1971, pg. 25)).

Eberlein (1962) shows that this measure is concentrated on functions with radius of convergence precisely $e$, so they are all analytic on $[-1, 1]$. Eberlein showed that $\Sigma|X_i| = 1$, so the integral concentrates on the unit ball. It is natural to multiply by an independent scale factor in applications.

This measure was independently described in Freedman (1963). Freedman put a probability on the set of all probabilities on $\{0, 1, 2, \ldots\}$ by repeated stick breaking. Freedman allowed more general distributions than uniform (his tale free measures). To get from Freedman's construction to Eberlein's, pick a random probability $P_0, P_1, P_2, \ldots$, by uniform stick breaking and then let $X_i = \pm P_i$ with the $\pm$ chosen independently at random. Diaconis and Freedman (1986) contains a survey of applications of this construction to problems in Bayesian statistics.

Eberlein and his students made two uses of this integral: It can be used to compare procedures by looking at mean square error between a quadrature rule and the true function. It can be used to generate quadrature formulas via Bayes theorem.

Here is an example of the first use, drawn from V. L. N. Sarma (1968). Consider two point Gaussian quadrature on $[-1, 1]$. This approximates $I(f) =$

$\int_{-1}^{1} f(x)dx$ by $\hat{I}(f) = a_1 f(t_1) + a_2 f(t_2)$ where $a_1 = a_2 = 1$, $t_1 = -1/\sqrt{3}$, $t_2 = 1/\sqrt{3}$. Sarma shows

$$\int (I(f) - \hat{I}(f))^2 df < (.019)3^{-7}.$$

Sarma also evaluates the accuracy of naive Monte Carlo, with $n$ points sampled at random in $[-1,1]$, and $\hat{\hat{I}}(f) = \frac{1}{n}\Sigma f(x_i)$. He shows $E((I(f) - \hat{\hat{I}}(f))^2) = \gamma/n$ for an explicit constant $\gamma$. His calculations show that to match the expected accuracy of two point Gaussian quadrature requires $n > 1,000$.

Of course, one must be careful in interpreting such results. The measure $df$ lives on functions having power series expansions with coefficients that tend rapidly to zero. Presumably, such functions are well approximated by a few terms of the series, and two-point Gaussian quadrature is exact for polynomials of degree 4.

Sarma (1968) shows that the accuracy of Monte Carlo improves with dimension. Sarma (1968), give two results for higher dimensional integrals.

Many natural questions remain. It would be worthwhile to think about the posterior distribution given the type of information usually computed about $f$. If $f(0), f'(0), \ldots, f^m(0)$ are given, the posterior is easy to compute. Presumably, the Bayes rules are consistent, though a proof is clearly in order. Finally, as far as I can judge, no serious work was done on the design problem for this prior.

## 3. RECENT APPLICATIONS

### A. Random Surfaces

There has been a great deal of interest in applying Bayesian methods to problems associated to tomography, computer vision, or image restoration. Consider an unknown two dimensional picture, coded up on a discrete grid, e.g. $256 \times 256$ in black and white "pixels". One receives some data about the picture. This could be a noisy version, or the average density along the rows and columns, or some combination of these.

It is desired to get a good version of the picture. The Bayesian approach begins with a prior distribution on the set of pictures, computes posteriors given the information, and then uses a Bayes rule such as the posterior mean or mode as an estimate.

Workers in this field have reported good results based on "Markov random field" priors similar to the Ising model of statistical mechanics. Useful surveys of applications and techniques are given by Besag (1987), Geman and Geman (1984), and Poggio (1985).

The main problems at the moment seem to be computational. With so many sites, updating is difficult as is computation of Bayes rules. Workers have found special classes of priors (e.g. nearest neighbor Markov random fields) and models (e.g. iid noise) where updating is possible. A variety of clever ad hoc procedures (e.g. simulated annealing, ICM) are currently being compared for the computation of Bayes rules. It is a lively world, which can almost certainly benefit from some more mathematics.

## B. Statistical Geodesy

A practical application of the methods outlined before has evolved in geodesists efforts to get good estimates of the force of gravity. We are all used to the usual approximations — 32 ft/(sec)$^2$ or 9.8 m./(sec)$^2$. These are useful only at the earth's surface, and even then are based on assumptions of a homogeneous earth which are sufficiently false to demand correction for engineering projects such as building a bridge (not to mention charting the course of a satellite).

There are a lot of data (millions of measurements) mostly taken at the surface, and near large population centers. This raises the problem of interpolating the potential field. The discussion below is based on work of Steffen Lauritzen (1972, 1973, 1976).

Lauritzen approximates the earth as a sphere, and bases his analysis on the physical assumption that the gravity potential $U$ at a given longitude, latitude and height is a harmonic function $\nabla^2 U = 0$. There is a well known basis for the harmonic functions in terms of harmonic polynomials. If the process is assumed orthogonally invariant, there is an expansion in terms of spherical functions:

$$U = \sum_{j=0}^{\infty} c_j s_j.$$

For an unknown function $U$, it is natural to take the coefficients $c_j$ as Gaussian with mean 0 and variance $\sigma_j^2$, just as Poincare did. This puts a prior on the harmonic functions.

Data consists of functionals of $U$ such as the direction and length of grad $U$ at various points. The covariance function of the process $U$ can be used to build a reproducing kernel space. Interpolation at an unobserved value now reduces to a well studied least squares problem.

Geodesists used a similar approach, without the probabilistic interpretation. They used classical kernels such as Poissons, and were not particularly pleased with the results.

Thinking probabilistically, and looking at the data, Lauritzen found error variances $\sigma_j^2$ which matched observations, but allowed the covariance to be summed into a closed form. This kernel has come into practical use among Geodesists in Denmark, Germany, and the United States.

The story has evolved since Lauritzen's work. Geodesists have enough data to approximate the projection of the process onto the space spanned by the first 30 or so spherical harmonics non-parametrically. The "tail" is handled using Lauritzen's approach.

C. YLVISAKER'S G-MAPS

Don Ylvisaker (1986) has developed a fascinating class of Gaussian priors for use in very general problems such as combinatorial optimization problems. To describe the construction, consider a finite set $X$. Let $P(x,y) = P(y,x)$ be the transition probability associated to a symmetric Markov chain on $X$. Think of the chain running in continuous time (with exponential holds between jumps). Let $T$ be an independent exponential variable, say the chain is killed at time $T$.

To such a chain we can associate the Greens function — the expected number of times the chain hits $y$ starting from $x$

$$G(x,y) = \int_0^\infty P^t(x,y)dt.$$

The Greens function is easily shown to be positive definite and thus can serve as a covariance function for a Gaussian process $Y_x$ indexed by $x \in X$. There are many curious links between the original chain and the process $Y$. One useful one: observing $\{Y_s\}$ $s \in S$, the best predictor of $Y_t$ is a linear combination

$$E\{Y_t|Y_s,\ s \in S\} = \Sigma w_t^S(s)Y_s.$$

The weights $w_t^S(s)$ can be computed as the hitting probabilities of the various sites in $S$, starting from $t$.

Ylvisaker has proposed simulating the original chain to calculate weights, or to give solutions to the design problem. For simple Markov chains such as random walk on groups or homogeneous spaces, analytic expressions and approximations for first hit probabilities are obtainable.

The Bayesian interpretation goes as follows: $Y_x$ is taken as a prior distribution for an unknown function on $X$. To estimate the total sum, interpolate at a point, or search for the maximum, a Bayes approach can be based on this prior.

Of course many things remain to be tried. The set of ideas, based on previous work of Hammersley and Dynkin, is so rich it should keep many of us busy for quite a while.

## 4. SOME OPEN PROBLEMS

### A. Stein Estimation

By now, the decision theoretic community has accumulated a good deal of wisdom on the properties of Bayes estimates. I have in mind the work of Berger, Efron, Morris and many collaborators on Stein estimation. We have all kinds of checks and balances — limited translation and various notions of robustness as ways of thinking of Bayes estimates based on priors chosen for computational convenience.

Many of these ideas can be carried to the infinite dimensional setting typical of Bayesian numerical analysis. Lillestöl (1977) discusses Stein estimation for time series as do Wolpert and Berger (1982).

One idea which is clearly understood in the statistical community but tends to be lost to outsiders: there is a big difference between a prior distribution based on experience and introspection and one based on mathematical convenience. Integrals with respect to the first deserve to be taken seriously. Integrals with respect to the second deserve to be treated skeptically. Users outside statistics seem to accept the first measure they can compute with as "God given", with nary a mention that it might be supported on a tiny, strange, part of the space.

### B. Consistency

Another hard won statistical lesson deserves to be remembered: it is hard to think about putting priors on high-dimensional spaces. It is possible to construct parametric families and priors supported on the entire parameter space which yield inconsistent Bayes rules. Diaconis and Freedman (1983) do this in the context of Bayesian numerical analysis. Diaconis and Freedman (1986) give a survey of problems and available positive results for Bayes rules in infinite dimensional parameter space.

All of the usual notions of frequentist statistics and their Bayesian counterparts can be thought about in the present setting. It seems as if there *are* surprises, and payoffs for doing so.

### C. More Examples

What is really needed most of all is explicit computation with medium to large scale examples in real settings. The constraints and intuition available in real problems cannot be beat for suggesting new priors, and methods of approximation. All of the recently accumulated experience by Bayesians (see e.g. Bernardo et al. (1982) for a fine set of examples) should be useable and extendable.

### Note Added in Proof.

Since writing this article, several further papers which treat closely related

problems have come to my attention.

Ajne, B. and Dalenius, T. (1960). Några Tillämpningar Av Statistika Ideer Pånumerisk Interration. *Nordisk Math. Tidskrift Band* **8**, 145–152.

These authors derive the trapezoid rule and Gaussian quadrature starting from statistical ideas.

Mochus, J. B. (1984). The Bayesian approach to Global optimization in Statistics: Applications and New Directions. *Proc. Indian Statist. Institute Golden Jubilee International Conference.* J. K. Ghosh, J. Roy (ed). Ind. Statist. Institute, Calcutta, pp. 405–430.

A survey of Russian work on finding the minimum of a continuous function using Bayes Methods.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting (with Discussion). *Journ. Roy. Statist. Soc. B* **47**, pp. 1–52.

A recent contribution emphasizing practical aspects of the approach. The discussion points to further relevant literature.

## BIBLIOGRAPHY


Aldous, D. (1983). Minimization algorithms and random walk on the d-cube. *Ann. Prob.* **11**, 403–413.

Bernardo, J., DeGroot, M., Lindley, D., Smith, A. (1985). *Bayesian Statistics 2.* North Holland, Amsterdam.

Besag, J. (1987). The statistical analysis of dirty pictures. *Jour. Roy. Statist. Soc. B.* To appear (with discussion).

Bühlman, H. (1963). $L^2$-martingales and orthogonal decomposition. *Zeit. Wahr.* **1**, 394–414.

Cox, D. (1986). An Analysis of Bayesian Inference for Non-Parametric Regression. Technical Report 93, Dept. of Statistics, University of Washington, Seattle, Washington.

Diaconis, P. and Freedman, D. (1983). Frequency properties of Bayes rules. In *Scientific Inference, Data Analysis, and Robustness.* (G. E. P. Box, T. Leonard, and C. F. Wu, eds.). Academic Press, New York.

Diaconis, P. and Freedman, D. (1986). On the consistency of Bayes estimates. *Ann. Statist.* **14**, 1–67.

Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7**, 269–281.

Diaconis, P. and Ylvisaker, D. (1985). Quantifying prior opinion. In *Bayesian Statistics 2* (J. Bernardo, M. DeGroot, D. Lindley, A. Smith, eds.) North Holland, Amsterdam, pp. 133–156.

Dozzi, M. (1981). Two-parameter harnesses and the Wiener process. *Zeit. Wahr.* **56**, 507–514.

Eberlein, W. F. (1962). An integral over function space. *Canadian Jour. Math.* **14**, 379–384.

Ermakov, S. M. (1968). On the admissibility of Monte Carlo method procedures *Soviet Math. Dokl.* **8** 66–67.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd edition. Wiley, New York.

Freedman, D. (1963). On the asymptotic behavior of Bayes estimates in the discrete case I. *Ann. Math. Statist.* **34**, 1386–1403.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Analysis, Machine Intelligence* **6**, 721–741.

Granovsky, B. (1986). Admissibility of Monte Carlo procedures. Technical Report No. 65, Department of Mathematics, Case Western University.

Granovsky, B. and Knoh, V. (1981). Admissibility of quadrature formulas with random nodes. In *Approximation Theory and Applications* (Zvi Ziegler, ed.). Academic Press, New York, 135–146.

Kadane, J. and Wasdkowski, G. (1983). Average case $\epsilon$ complexity in computer science, a Bayesian view. In *Bayesian Statistics 2.* (J. Bernardo, M. DeGroot, D. Lindley, A. Smith, eds.) North Holland, Amsterdam.

Lauritzen, S. (1972). An application of continuous random functionals and unitary representations of groups in physical geodesy in "Open House for Probability 1971". Var. Publ. Series. No. 21, Aarhus Univ.

Lauritzen, S. (1973). *The Probabilistic Background of Some Statistical Methods in Physical Geology.* Publication No. 48. Geodaetisk Institut, Copenhagen, Denmark.

Lauritzen, S. (1975). Random orthogonal set functions and stochastic models for the gravity potential of the earth. *Stochastic Proc. Appl.* **3**, 65–72.

Lee, E. and Woznakowski, G. W. (1986). Approximation of linear functionals on a Banach space with a Gaussian measure. *Jour. Complexity* **2**, 12–43.

Lillestøl, J. (1977). Improved estimates for multivariate complex-normal regression with application to analysis of linear time-invariant relationships. *Jour. Multivariate Analysis* **7**, 512–534.

O'Hagan, A. (1985). Bayesian quadrature. Technical Report, Department of Statistics, University of Warwick.

Palasti, I. and Renyi, A. (1956). On interpolation theory and the theory of games. *MTA. Mat. Kot. Int. Kozl* **1**, 529–540.

Poggio, T., Torre, V. and Christof, K. (1985). Computational vision and regularization theory. *Nature 317*, No. 6035, pp. 314–319.

Poincare, H. (1896). *Calcul des Probabilities*, Georges Carré, Paris.

Sacks, J. and Ylvisaker, D. (1970). Statistical designs and integral approximation. *Proc. 12th Bien Sem. Canad. Math. Conq.* Canadian Math: Cong.

Montreal 115–136.

Sacks, J. and Ylvisaker, D. (1985). Model robust design in regression: Bayes theory. In *Proc. Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer.* (L. Le Cam, R. Olshen, eds.) Vol. II Wadsworth, Monterey, CA.

Sarma, V. L. N. (1966). A generalization of Eberlein's integral over function space. *Transac. Amer. Math. Soc.* **121**, 52–61.

Sarma, V. L. N. (1968). Eberlein measure and mechanical quadrature formula I. Basic Theory. *Math. Comp.* **22**, 607–616.

Shaw, J. E. H. (1985). A quasi-random approach to integration in Bayesian statistics. Technical Report, Department of Statistics, University of Warwick.

Smale, D. (1985). On the efficiency of algorithms of analysis. *Bull. Amer. Math. Soc.* **13**, 87–121.

Wahba, G. (1983). Bayesian confidence intervals for the cross validated smoothing spline. *Jour. Roy. Statist. Soc.* **45**, 133–150.

Williams, D. (1973). Some basic theorems on harnesses. In *Stochastic Analysis.* D. G. Kendall and E. F. Harding, eds. Wiley, New York. pp. 349–366.

Williams, D. (1980). Brownian motion as a harness. Unpublished manuscript.

Wolpert, R. and Berger, J. (1982). Incorporating prior information in minimax estimation of the mean of a Gaussian process. In *Statistical Decision Theory and Related Topics III*, Vol. 2, (S. Gupta, J. Berger, eds.) Academic Press, New York. pp. 451–464.

Ylvisaker, D. (1986). Prediction and design. To appear, *Ann. Statist.*