The Annals of Statistics
1989, Vol. 17, No. 4, 1419–1440

INVITED PAPER

GAIT ANALYSIS AND THE BOOTSTRAP[1]

BY RICHARD A. OLSHEN, EDMUND N. BIDEN, MARILYNN P. WYATT AND DAVID H. SUTHERLAND

University of California, San Diego; University of New Brunswick and Children's Hospital, San Diego; Children's Hospital, San Diego; and University of California, San Diego and Children's Hospital, San Diego

This paper is about random coefficient trigonometric regression models and their use in gait analysis. Here gait analysis means free-speed walking on a level surface. Our study focuses on bootstrap-based prediction regions for the angular rotation curves of test children, when the relevant training data are gathered from normal children of comparable ages. Considerations that led to our choice of model and use of the bootstrap are given. Prediction regions and empirical bootstrap distributions are displayed, as is their application to several test cases. Also included is a study of the almost sure asymptotic behavior of the theoretical bootstrap probability of the prediction regions.

1. Introduction. Human motion has been a subject of inquiry since nearly the beginning of history. For example, more than 2300 years ago Aristotle described the actions of muscles and analyzed their effects in producing motion. In fact, the way an individual walks is a sensitive measure of neuromuscular development or impairment. A sound descriptive model helps to define normal and abnormal parameters and offers some clues to underlying causes of abnormality. The primary goals of our Motion Analysis Laboratory at Children's Hospital, San Diego have been to expand the understanding of what constitutes normal walking and to study deviations from norms seen in children and adults with a variety of diseases. Pretreatment studies are vital to any decision-making process in choosing the proper course of treatment. Posttreatment analyses are used to assess effectiveness of treatment and to furnish useful information for planning further management.

The most frequent referrals to our Laboratory are patients with cerebral palsy. However, patients with limb deficiencies, poliomyelitis, arthritis, muscle disease, head and spinal cord injuries, peripheral nerve injuries, stroke, congenital joint deformities, joint implants and many other disease processes have been studied.

Received January 1987; revised May 1989.

[1]Research supported in part by NIH Grants 5 R01 HD08520, 5 R01 HD15801 and 5 R01 PHS CA41628, NSF Grants DMS-85-05609 and DMS-87-22306 and the John Simon Guggenheim Memorial Foundation.

AMS 1980 *subject classifications*. Primary 62G15, 62P10; secondary 62J05, 60F99.

Key words and phrases. Bootstrapped prediction regions, gait analysis, trigonometric regression, Vapnik–Chervonenkis classes.

Four basic types of data are gathered at the Laboratory and are used in evaluating walking:

1. Linear measurements, or time–distance parameters.
2. Movement measurements, that is, calculations of angular rotations (in three planes) for the pelvis, hip, knee and ankle.
3. Force plate measurements—measurements of forces exerted by the subject's foot as it contacts the floor.
4. Dynamic electromyography—measurements of the firing pattern of the subject's muscles when he walks.

Pretreatment studies for a patient with cerebral palsy bear upon surgery planned to help the patient walk in a more normal fashion. All four types of data are used. Suppose, for example, that at some point in the walking cycle, an angular rotation falls particularly far from its normal value. The physician then focuses on a possible explanation. Electromyographic measurements may give clues to abnormal firing patterns of particular muscles. Asymmetries between motions of the right and left sides of the patient might suggest a lack of balance or coordination, or the presence of fixed muscle contractures or excessive muscle tone in some muscles. These findings can then be corroborated by electromyographic or force plate data. After the abnormalities are understood, surgery is designed to lessen or, if possible, eliminate them. With or without an intervention such as surgery or other therapy, gait studies are used to monitor patient progress. A choice between two different leg braces for the patient can be made on the basis of which rendered his or her gait more normal in appearance. Most patients (and their parents and physicians) hope that their walking appears to be as close as possible in appearance to the walking of clinically normal children throughout the walking cycle. So techniques that serve to quantify deviation from normal walking *somewhere* in the cycle are of great clinical interest. We believe that our work provides a tool for clinical applications that is of substantial potential importance because it strengthens the ability of the physician to use motion analysis data in a systematic way.

Angular rotation curves and bootstrap-based prediction regions for them are the main topics of this paper. The raw data from which the computations are made are taken by first recording a subject's movements as he walks down a walkway and then taking measurements from the recording medium. We study what is called free-speed, level walking. From 1973 until a recent improvement in technology, our Laboratory employed the same photographic method, utilizing four 16 mm motion picture cameras and a system to take measurements from the film. For this film system, reference marks are placed on the subject's skin, and lightweight sticks are secured on the pelvis and tibiae as landmarks from which to make measurements (see Figure 1). Three fixed-position movie cameras record the subject from the left side, right side and front as he walks. The fourth movie camera is located beneath the force plates and photographs the bottom of the foot as it strikes the plate. The film is analyzed frame by frame to make the measurements.

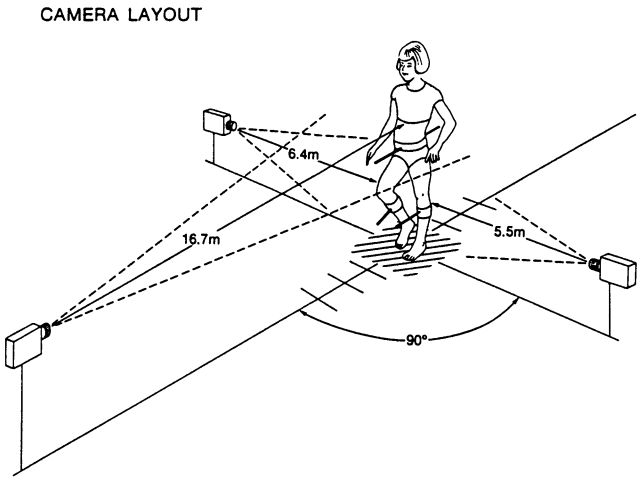


FIG. 1. *Camera layout for studies of joint rotations.*

More detailed descriptions of the way in which data were gathered are in the paper by Sutherland, Olshen, Cooper and Woo (1980) and especially in the book by Sutherland, Olshen, Biden and Wyatt (1988). Recently the Laboratory converted to a semiautomatic VICON (VICON is a trademark of Oxford Metrics, Inc.) movement measurement system originally developed by M. O. Jarrett. Now reflective markers are placed on the skin, and they are seen by television cameras; three-dimensional coordinates of the markers are used to calculate joint rotations.

The movement of subject may be described as a set of "joint rotations" for each side of the body. In our work, 12 rotations are measured and evaluated to describe the movement of each leg (see Figure 2). These include four side (or sagittal) rotations, pelvic tilt (how much the pelvis tilts forward or back), hip flexion–extension (how much the hip flexes), knee flexion–extension (measuring the bending of the knee) and ankle dorsi–plantarflexion—the moving of the foot up (dorsiflexion) or down (plantarflexion). Viewed from the front, the rotation of the pelvis to the right or left is measured as pelvic rotation, and pelvic obliquity measures whether one hip is higher than the other. The ab–adduction of the hip describes whether the thigh moves inward (adduction) or outward (abduction). Femoral rotation measures how much the whole leg turns in or out, while tibial rotation measures the rotation of the lower leg. Hip rotation is the difference between pelvic and femoral rotations, and knee rotation is the difference between femoral and tibial rotations. Finally, the angle the foot makes with the nominal direction in which the subject is walking is measured as the foot rotation angle. Figure 3 shows a plot of these values, averaged over subjects, for normal two-year-olds and seven-year-olds.

Though it is difficult to tell from Figure 3, the curves from two- and seven-year-olds are both similar and different in interesting ways. But for knee

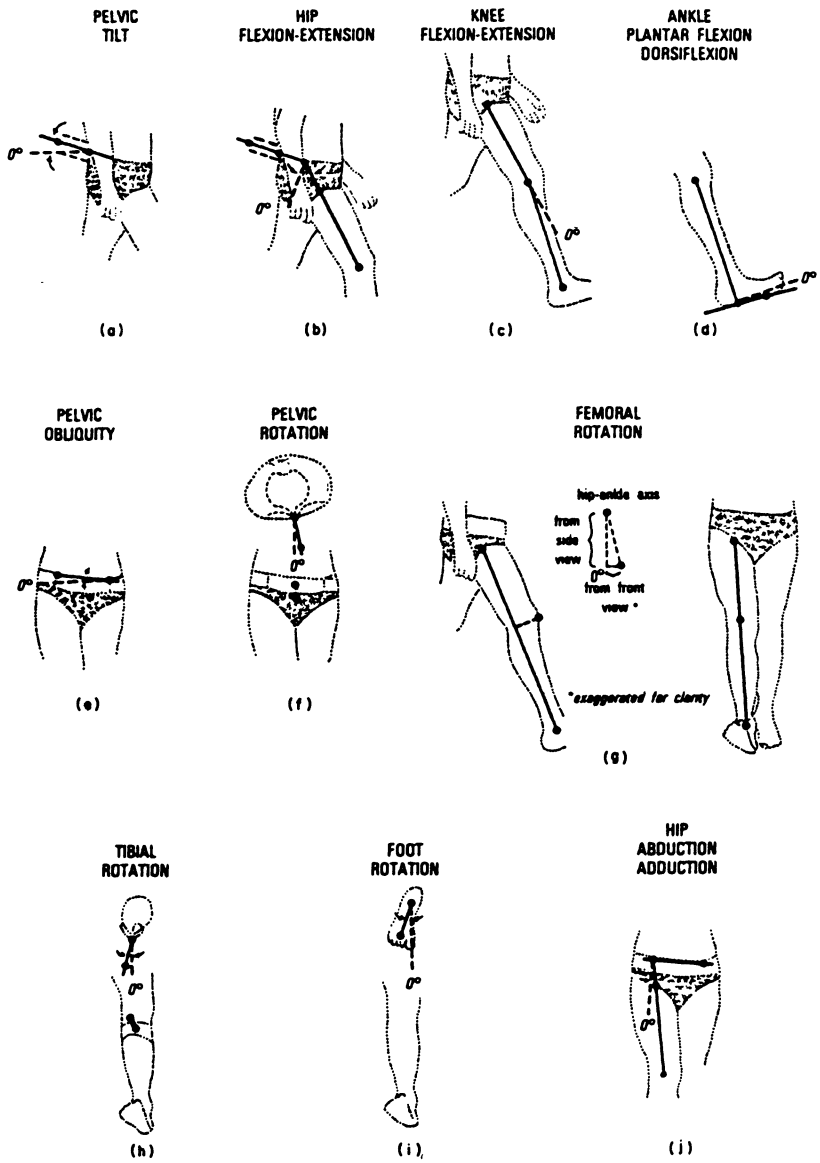


FIG. 2. *Figures depict the methods used in calculating the various joint rotations. Large dots indicate actual measurement points. Dashed lines show the zero degree reference for each angle.*

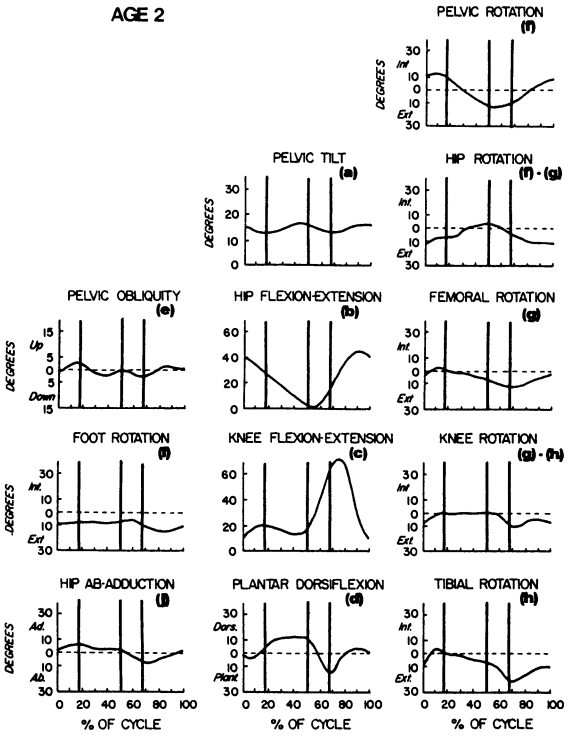
flexion–extension (c), the sagittal plane curves in developmentally normal two-year-olds closely resemble not only those of seven-year-olds, but also those of normal adults. One exception is during the early part of the knee flexion–extension curve. This portion of the curve is called an initial knee flexion wave, the result of an eccentric contraction of the quadriceps just after the foot hits the ground. The contraction, more evident in seven-year-olds than in two-year-olds,

is a shock absorber. We note that the sagittal plane curves are often the most interesting and informative ones in quantifying pathological gait [see Sutherland, Olshen, Cooper, Wyatt, Leach, Mubarak and Schultz (1981)]. Differences in rotations between normal two- and seven-year-olds can be found in the transverse plane. Both pelvic rotation (f) and knee rotation [(g)–(h)] are slightly exaggerated in the younger children. Hip rotation [(f)–(g)], femoral rotation (g) and tibial rotation (h) are all external for the two-year-olds relative to the seven-year-olds. When children begin to walk, they tend to walk with their feet and legs turned out, and they assume a wide stance to give maximum support. Their cadence is much higher than the cadence of adults and they spend less time supported by a single limb than do adults. These aspects of walking are still evident in two-year-olds, but not in seven-year-olds.

It is essential to assure that measurement systems are accurate and reproducible, especially when treatment is being evaluated or when decisions are being made about surgery. With the system used to gather our data, the mapping from digitizer screen to the actual laboratory is about 1 in. : 1 ft. In tests of positioning accuracy, a typical experienced technician could hit the screen surface within a circle of about 1/32 in. diameter. This corresponds to about 1 cm "error" on the subject. VICON has pixel resolution of about 1/1024 of the field of view. This implies that for a field of view of 2–3 m, one cannot resolve to less than about 3 mm. In practice, the error for a VICON study is approximately the same as that for digitizing film, since it is subject to errors of calibration and point reconstruction. From geometric and the above considerations it follows that with either system, an observer generally can reproduce his measurements to within 3°. Typically, between observers, measurements are reproducible to within 3°; the average absolute difference is 1 or 2°. For further information, see Sutherland, Olshen, Biden and Wyatt (1988).

At the time of this writing, there are many sophisticated motion analysis laboratories throughout the world. The accumulation of vast amounts of data is now relatively easy. We hope that our discussion has convinced the reader that the interpretation of these data presents an enormous challenge. Examples of clinical uses of gait data are given in the recent books by Sutherland (1984) and by Sutherland, Olshen, Biden and Wyatt (1988). The latter incorporates use of the bootstrap-based prediction regions studied here.

The remainder of the paper has four sections. In Section 2 we describe a training sample and a test sample whose data motivated this research. Also, further information is given on how the observations we utilize were gathered. Models and modeling are the subject matter of Section 3. We argue that a certain random coefficient trigonometric regression model is suitable for the rotation data and that the distributions of the coefficients need not be Gaussian. "Cleaning up" outlying data is mentioned, as is the use of least squares for estimating the expected Fourier coefficients in our model. Section 4 gives the algorithm for forming bootstrap predictions for the least squares fit to a test child's data. Some applications to test data are given. One noteworthy aspect of our approach is that prediction for the overall constant term is separated from that for the sum of harmonics. Graphical connection is made to recent

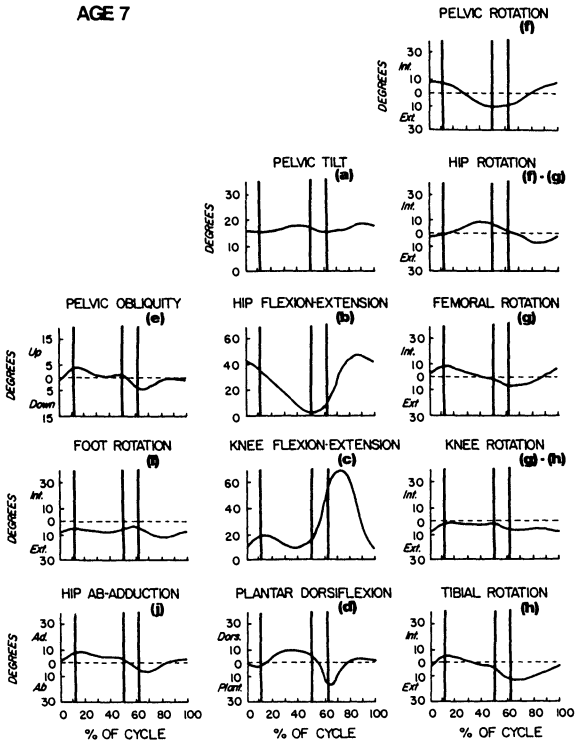


(a)

FIG. 3. *(a) Joint angles for two-year-olds. The curves are derived from the Fourier trigonometric model by least squares. The three vertical lines through the graphs as viewed from left to right indicate, respectively, percents of the gait cycle at which the opposite foot leaves the ground, the opposite foot strikes the ground and the foot being studied leaves the ground.*

theoretical work by Johansen and Johnstone (1990); their study, which builds on contributions by Hotelling (1939), was motivated in part by Figure 5. The last section is devoted to some mathematical aspects of the bootstrap prediction regions for the sum of harmonics. In particular, we study the almost sure behavior of theoretical bootstrap probabilities.

2. The data. The joint rotation data that are the subject matter of this paper come from two studies conducted at our Laboratory. The prediction regions by age, side and rotation were generated from the training sample data of 415 gait studies of normal children ages one through seven, by which age walking is nearly mature. It is natural to use age as a first variable on which to group subjects since the *development* of normal walking is of major interest. Subdividing by sex or dominant side, for example, is not useful; see Sutherland, Olshen, Biden and Wyatt (1988), Chapter 10.



(b)

FIG. 3. (*b*) *Joint angles for seven-year-olds.*

The age groups begin at one year and increase by one-half year to age four, and then by a full year to age seven, for a total of 10 groups. They range in size from a low of 36 (age $1\frac{1}{2}$) to a high of 49 (age 1). Each child was studied within 30 days of the date when his or her birthday corresponded to the indicated age group. The prediction regions and bootstrap percentiles have been utilized to assess departures from normal for data gathered from a few of the 240 children (in eight different age groups) from another study of prematurely born children. In one group, well over half the children fell outside the 95% prediction regions for knee flexion–extension. These are children who have had no orthopaedic intervention, but who may have neuromuscular impairment (cerebral palsy) or may merely be delayed in their neurologic development. Some clearly do have cerebral palsy.

A child being studied made at least three passes down the level walkway. Each pass included about three full walking cycles (say, right foot-strike to right foot-strike.) The data from the measurements of the three passes were averaged on seven simple but important descriptors of walking (cadence, step length, etc.). For each cycle a vector of standardized differences from the mean across cycles

was computed. Ideally, the cycle (from among the nine or so) for which the vector had the smallest Euclidean norm would be chosen for further analysis. In practice, the implementation was by eyeball. Thus, for the individuals studied we used representative gait, not average gait. The rationale for this is given by Sutherland, Olshen, Cooper and Woo (1980). In gathering our data we did not attempt to interpolate between frames to find the "real" beginning of the gait cycle. It was not obvious to us how to implement such a scheme, in principle or (in view of certain limitations of our equipment) in practice. Also, our choice of representative cycle for future study may have diminished problems with our analyses that owed to this shortcoming.

3. Models and modeling. Free-speed walking on a level surface by normal children is approximately periodic and the signals in the angular rotation data are of low frequency. Therefore, in keeping with tradition for such data, our analyses are based on Fourier models. The models are never exactly correct. Some biases are introduced by the deletion of high frequency harmonics; however, these biases do not appear to detract in important ways from the clinical uses of the technology we describe.

It is customary in gait analysis to parameterize the cycle by percent, from foot-strike to foot-strike. Thus, if $f = f(\theta)$ refers to some generic joint rotation at percent θ of the cycle (for a given age and side), then we assume that

$$f(\theta) = \alpha_0 + \sum_{j=1} \left[\alpha_j \cos\left(\frac{2\pi j\theta}{100} \right) + \beta_j \sin\left(\frac{2\pi j\theta}{100} \right) \right]. \tag{1}$$

The α's and β's are characteristic of a particular child and walk, and thus random. This simple observation is one key to what successes we have enjoyed in fitting prediction regions for angular rotations. The vector $(\alpha_0, \alpha_1, \beta_1, \alpha_2, \beta_2, \dots) = \mathscr{C}$ is assumed to have mean μ and covariance $\Gamma = [\gamma_{ij}]$. For a representative stride, data on f for a given child are digitized at $\theta_1 < \theta_2 < \cdots < \theta_d$, where always $d \geq 16$, and typically the θ_k are equally spaced:

$$\theta_k = \frac{100(k-1)}{d}. \tag{2}$$

Occasionally, isolated digitized values are greatly out of keeping with (1) for a moderate number of summands and any believable distribution of errors. For example, this can occur as the foot strikes the ground and the sticks shown in Figure 1 are shaken. (The problem is more acute with obese children than others.) If the θ_k's are as in (2) and, say, a digitized value at θ_l is aberrant, but those at θ_{l-2}, θ_{l-1}, θ_{l+1} and θ_{l+2} are not, then we replace the digitized value at θ_l as follows. One quadratic is fitted to values at θ_{l-2}, θ_{l-1} and θ_{l+1} and another to those at θ_{l-1}, θ_{l+1} and θ_{l+2}. The two fitted values at θ_l are averaged and that number replaces the digitized value at θ_l. It follows from the Lagrange interpolation formula [Powell (1981), Section 4.1] and (2) that this interpolation scheme is exact for cubics. Only a tiny fraction of the data are aberrant and therefore

corrected. That we have "cleaned up" our raw data is not incorporated into subsequent inferences drawn from least squares estimation and the bootstrap.

In practice, there are N children and therefore N independent, identically distributed (iid) random vectors of coefficients $\{(\alpha_0^{(i)}, \alpha_1^{(i)}, \beta_1^{(i)}, \alpha_2^{(i)}, \beta_2^{(i)}, \dots)\}_{i=1}^{N}$. Data from the ith child are assumed to be digitized at d_i points; the kth of these points is written $\theta_k^{(i)}$, where $\theta_k^{(i)} = 100(k-1)/d_i$ for $k = 1, \dots, d_i$. There are measurement errors ε_{ik} for the measurement of the ith child's data at $\theta_k^{(i)}$. The $\{\varepsilon_{ik}\}$ are assumed to be iid and independent of the set of random vectors of coefficients, with $E(\varepsilon_{ik}) \equiv 0$ and $\mathrm{Var}(\varepsilon_{ik}) \equiv \sigma^2$. Therefore, we write Y_{ik} for the observation on the ith child at $\theta_k^{(i)}$ as

$$Y_{ik} = \alpha_0^{(i)} + \sum_{j=1} \left[\alpha_j^{(i)} \cos\left(\frac{2\pi j \theta_k^{(i)}}{100} \right) + \beta_j^{(i)} \sin\left(\frac{2\pi j \theta_k^{(i)}}{100} \right) \right] + \varepsilon_{ik}. \tag{3}$$

If the number of summands in (3) is fixed at not more than the integer part of $(d_i - 1)/2$, then we can estimate μ from the $\{Y_{ik}\}_{k=1}^{d_i}$ (that is, from the data on the ith subject) by the method of least squares. We learned from M. L. Eaton, though it follows from results of Rao (1965), that for our model least squares and best linear unbiased estimation coincide. The ith separate estimate is substituted in (1) to form $\hat{f}^{(i)}(\theta)$, an estimate of $E(f(\theta))$. How to combine these estimates across subjects to form an overall estimate $\hat{f}(\theta)$ is not obvious since the optimal weights for a linear unbiased estimate depend on what θ is. Two extreme possibilities are:

1. Average with each subject weighted equally and thereby ignore the inequality in the d_i.
2. Weight each subject proportional to d_i and thereby ignore Γ.

The latter was our choice, though fortunately, for our data the two extreme methods of combining the $\hat{f}^{(i)}$ produce estimates whose graphs are visually indistinguishable.

It follows from the conditional variance formula that for any allowable assumed number of summands in (3), the unconditional covariance of the least squares estimator based on data of the ith child is

$$\Gamma + \mathrm{diag}\left[\sigma^2/d_i, 2\sigma^2/d_i, \dots, 2\sigma^2/d_i\right], \tag{4}$$

where diag[ ] is a diagonal matrix of the same dimension as Γ with the cited values on its diagonal. How many summands are utilized in (3) will be discussed. The parameter σ^2 is estimated unbiasedly from each least squares fit and the separate estimates are pooled across subjects to yield an overall estimate $\hat{\sigma}^2$. The expression (4) is estimable, at least approximately, from the sample covariance of the least squares estimates and thus Γ is, too, by subtraction. Virtually always, $\hat{\sigma}^2/(\hat{\sigma}^2 + \hat{\gamma}_{kk} d_i)$ was less than 0.05, typically much less.

Clearly, normal walking involves a variety of compensating actions of joints and limbs. Indeed, the α's and β's are not stochastically independent. Various ad hoc tests of the null hypothesis that Γ is diagonal resoundingly reject.

Ignoring this dependence is a major flaw in the analyses of Cappozzo, Leo and Pedotti (1975).

A first question that might be asked regarding the α's and β's concerns their marginal distributions. Are they Gaussian? We attempt to answer the question from the empirical distribution of their least squares estimates. In order to fix ideas, we focus on the estimation, for child i of, say $\alpha_1^{(i)}$. It is easy to compute that the least squares estimate is

$$\alpha_1^{(i)} + \frac{2}{d_i} \sum_{k=1}^{d_i} \varepsilon_{ik} \cos\left(\frac{2\pi\theta_k^{(i)}}{100}\right). \tag{5}$$

From first glance, it seems as if the assessment we require involves a deconvolution, since $\alpha_1^{(i)}$ is independent of the second term of (5). However, that sum is distributed approximately as $N(0, 2/d_i)$ according to the central limit theorem, so it is stochastically small. Also, from Cramér's theorem [Cramér (1936)], normality of (5) entails normality of $\alpha_1^{(i)}$. In any case, we proceed to make our assessment as outlined above. Plots of these estimates against their normal scores reveal occasional outliers, and correlations of data with their normal scores can be much lower than normal theory would predict. Generally speaking, for normal children the higher up the body the angular rotation to which f refers, the older the group of children from whom the data are gathered, and the lower the frequency of the harmonic, the more Gaussian are the marginal distributions of the Fourier coefficients. Part of our motivation for introducing the bootstrap in our prediction regions was that we do not always trust theories based on Gaussian assumptions regarding the coefficients. When we simulated data from (1) with Gaussian coefficients with the same means and covariance structure that we observed for the least squares estimates, the plots of rotation against θ typically looked very different from the data of observed children and from their averages in Figure 3.

We discuss now why, in practice, we take the upper limit of summation in (1) to be 6. There is precedent for this choice [Cappozzo, Leo and Pedotti (1975) and Sutherland, Olshen, Cooper and Woo (1980)]. Since setting standards and making comparisons are basic to our purpose, it was deemed appropriate to fit the same number of terms to each child for every rotation. One might guess that the process of fitting could be improved by first testing whether candidate Fourier coefficients are 0 and retaining in the model only those terms that pass some suitable test at, say the 5% level. This process is not reasonable from at least one decision-theoretic point of view [Sclove, Morris and Radhakrishnan (1972)]; in addition, it leads to fitted curves that obscure "blips" and other aspects of rotations believed to be important on the basis of clinical experience. Instead, we adapted Shibata's (1981) model selection procedure to the problem of choosing the upper limit to the summation. It seemed unreasonable to skip harmonics, so subsets of harmonics other than those formed in this way were not considered. We could not generally retain more than the first six sine and cosine terms because we wanted at least 3 degrees of freedom for residuals for each fit, and for some rotations $d_i = 16$ for some i. There would be only slight diagnostic value to

higher order terms anyway, since regardless of d_i, multiple correlations for our fits nearly always exceed 95%. (The extremely high values for multiple correlations persist for VICON data that are digitized more frequently than were the data on which this paper is based.) Two ideas that guided our application are these: There is some considerable spread to the distributions of $\mathscr{C}$ and so there should be some spread to the distributions of numbers of terms Shibata would retain; also, neurologic maturation is top to bottom (cephalocaudal), so the higher up the body the rotation, the fewer terms would likely be needed. For rotations such as the foot that involve "fine tuning" of the walk, older children should require more terms than younger. Both of these conjectures were validated by the data. For a given candidate value M to the upper limit of summation, Shibata's criterion is computed to be

$$S(d_i, M) = \left\{1 + \frac{4M + 1}{d_i}\right\} \{\text{residual sum of squares}\}. \tag{6}$$

The selection rule is to pick M for which (6) is minimized. The criterion is asymptotically optimal if the ε_{ik} are independent and identically distributed, $N(0, \sigma^2)$. Of course we do not believe the normality assumption in general. Even with the stated assumptions, however, Rice (1984) indicates that at least for some problems the technique tends to overfit in practice. This tendency is less of a problem in our situation with random coefficients than it might be in other contexts, because in deciding how many terms to retain it would seem unwise to be parsimonious at the expense of using models unable to describe the rotations of any sizable minority of children. Figure 4 gives histograms of the choice of M Shibata dictates for two rotations for our 38 normal five-year-olds. Clearly the ankle tends to require more terms than the hip, and yet even for the hip some children require $M = 6$. The choice of M and application of the Shibata criterion can be summarized thus. For each rotation, M was to be fixed. It could not be more than six. To have chosen it to be less for any rotation would have been inconsistent with the data of at least some children. So we feel that a uniform choice of $M = 6$ is reasonable.

4. Bootstrapping prediction regions and percentiles. While mean values are a *sine qua non* in our analyses, the process of comparing test children to normal children requires estimates of the variability of the rotations for the normal child. We formalize the use of training data to evaluate subsequent test subjects by forming *prediction regions*. Initial attempts to define such regions led to the discovery that for many rotations, the variability of the average across θ (that is, α_0) seems to be larger than the variability of the sum of harmonic terms. Therefore, for each rotation we separate the process of prediction into two (not independent) components: one for α_0 and one for the sum of harmonics. This is reasonable from a physiologic point of view, since the former may be primarily a function of body alignment and the latter largely dependent on muscle control. Another argument is that across motion analysis laboratories the α_0 term for a given walker is not necessarily reproducible since its value depends

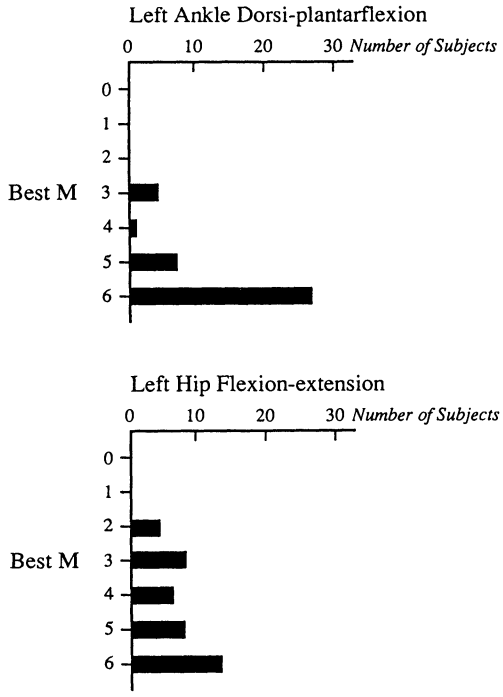


FIG. 4. *Choice of the upper limit of summation in* (1) *by the Shibata model selection criterion* (6) *for* 38 *normal five-year-old children.*

to a large extent on marker placement. On the other hand, the sum of harmonics for the same walker is closely reproducible. For now, discussion centers on prediction regions for the sum of harmonics, that is, for

$$f_h(\theta) = \sum_{j=1}^{6} \left[\alpha_j \cos\left(\frac{2\pi j\theta}{100} \right) + \beta_j \sin\left(\frac{2\pi j\theta}{100} \right) \right]. \tag{7}$$

As is clear from the discussion of the last section, variability of f_h is far from constant across the gait cycle. Typically, the range of data is wider during swing phase (when the foot is off the ground) than during stance phase. The variance of $f_h(\theta)$ is $\sigma_f^2(\theta) = \mathcal{T}'(\theta)\Gamma_h\mathcal{T}(\theta)$, where

$$\mathcal{T}'(\theta) = \left(\cos\left(\frac{2\pi\theta}{100} \right), \sin\left(\frac{2\pi\theta}{100} \right), \ldots, \sin\left(\frac{12\pi\theta}{100} \right) \right),$$

where the prime denotes transpose and Γ_h is the lower right 12×12 submatrix of Γ.

Denote by $\tilde{f}_h(\theta)$ the estimated $f_h(\theta)$ for a test subject for whom least squares estimates based on his data are plugged into (7). From (4) it follows that if d_0 is the number of points at which the subject's data are digitized, then

$$\mathrm{Var}\left(\tilde{f}_h(\theta) \right) = \mathcal{T}'(\theta)\Gamma_h\mathcal{T}(\theta) + \left(12\sigma^2/d_0 \right). \tag{8}$$

Because in practice, as was stated, σ^2/d_0 is always small relative to γ_{kk} and the d_i's from the training sample data are roughly equal, we simply estimate (8) by

$$\hat{\sigma}_f^2(\theta) = \mathscr{T}'(\theta)\hat{\mathrm{Cov}}_h\,\mathscr{T}(\theta), \tag{9}$$

where $\hat{\mathrm{Cov}}_h$ is the sample covariance of the collection of least squares estimates $\{(\hat{\alpha}_1^{(i)},\ldots,\hat{\beta}_6^{(i)})\}$. $E(f_h(\theta))$ is estimated by $\bar{f}_h(\theta)$, in which $f_h^{(i)}$ is weighted proportional to d_i. It is in keeping with clinical practice to say that a child departs from the normal range if for *some* θ, his $\tilde{f}_h$ and the population mean are too different, where the unit of discrepancy at θ is $\sigma_f(\theta)$. Thus, our estimated simultaneous prediction regions are of the form

$$\bigcap_\theta \left\{(\theta, \tilde{f}_h)\colon \bar{f}_h(\theta) - m\hat{\sigma}_f(\theta) \le \tilde{f}_h(\theta) \le \bar{f}_h(\theta) + m\hat{\sigma}_f(\theta)\right\} \tag{10}$$

$$= \left\{\tilde{f}_h\colon \max_\theta \left|\frac{\tilde{f}_h(\theta) - \bar{f}_h(\theta)}{\hat{\sigma}_f(\theta)}\right| \le m\right\}, \tag{11}$$

where m is a positive number. The bootstrap enables us to estimate the probability of the event (10).

The bootstrap process begins with the selection of a random sample of size N with replacement from our training sample of N subjects. Compute from this bootstrap sample N vectors of coefficients $\{(\hat{\alpha}_{1,B}^{(i)},\ldots,\hat{\beta}_{6,B}^{(i)})\colon i = 1,\ldots,N\}$ and from them an estimate $\bar{f}_{h,B} = \bar{f}_{h,B}(\theta)$ of $E(f_h(\theta))$ and $\hat{\sigma}_{f,B}^2(\theta)$ of $\mathrm{Var}(f_h(\theta))$. These vectors $\{(\hat{\alpha}_{1,B}^{(i)},\ldots,\hat{\beta}_{6,B}^{(i)})\colon i = 1,\ldots,N\}$, are N realizations of random vectors distributed according to the theoretical bootstrap distribution. In Section 5, where most of the discussion is restricted to the case $M = 1$, we have denoted by $\{(\hat{\alpha}_B^{(N,i)}, \hat{\beta}_B^{(N,i)})\colon i = 1,\ldots,N\}$ random vectors distributed according to the theoretical bootstrap distribution from which the N $\{(\hat{\alpha}_{1,B}^{(i)}, \hat{\beta}_{1,B}^{(i)})\colon i = 1,\ldots,N\}$ are sampled. The probability measure that gives rise to the theoretical bootstrap distribution is denoted in Section 5 by $P_B^{(N)}$.

For each positive m, let

$$\hat{F}_B(m) = N^{-1}\#\left\{\hat{f}_h\colon \max_\theta \left|\frac{\hat{f}_h(\theta) - \bar{f}_{h,B}(\theta)}{\hat{\sigma}_{f,B}(\theta)}\right| \le m\right\}. \tag{12}$$

[Here $\hat{f}_h(\theta)$ refers to the least squares estimate of $f_h(\theta)$ for an individual member of the training sample.] Now repeat the bootstrap sampling and computation of $\hat{F}_B(m)$, say R times. Then $\bar{F}_B(m)$ is the average of the R $\hat{F}_B(m)$. If $0 < p < 1$, then m_p is defined by

$$m_p = \min\left\{m\colon \bar{F}_B(m) \ge p\right\};$$

p is our estimate of the probability of the event (10) when $m = m_p$. That is, we predict that for approximately $100p\%$ of normal children, their $\tilde{f}_h$ will lie

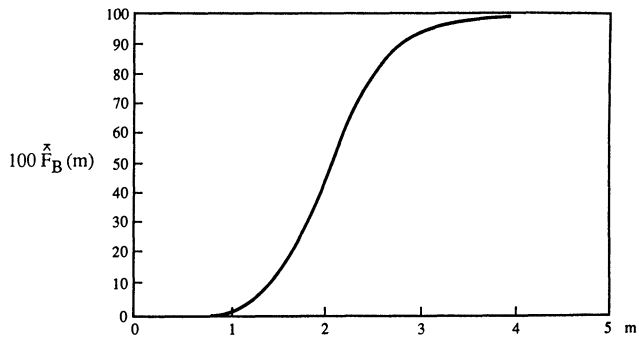


FIG. 5. *Plot of a bootstrap estimate of the probability of the event* (10) *as a function of m for left ankle dorsi–plantarflexion based on* $R = 10$ *bootstrap samples and* $N = 38$ *normal five-year-old children.*

between $\bar{\hat{f}}_h(\theta) - m_p\hat{\sigma}_f(\theta)$ and $\bar{\hat{f}}_h(\theta) + m_p\hat{\sigma}_f(\theta)$ for all θ. In practice the discreteness of $\hat{F}_B$ is not important since its typical jump is $1/NR$.

It follows from large deviation results [as in Babu and Singh (1983)], that if R is large enough, then $\hat{F}_B(m)$ arbitrarily closely approximates the corresponding theoretical bootstrap probability (uniformly in m) with arbitrarily high probability. In Section 5 it is demonstrated that subject to some conditions on the distribution of $\mathscr{C}$, the difference of the theoretical bootstrap probability and the true probability of (10) tends almost surely to 0 as N increases without bound.

Figure 5 is a plot of $100\hat{F}_B(m)$ versus m for the left sagittal plane ankle motion based on a training sample of 38 normal five-year-old children. Note the striking and typical sigmoidal shape to the curve. An explanation is given by Johansen and Johnstone (1990). They show that, at least approximately, the curve in Figure 5 is the distribution function of a product of two independent random variables, one the square root of a χ^2_{12} random variable. The other distribution is derived from work of Hotelling (1939). In constructing Figure 5, $R = 10$ was used. Further computing showed that the curves for $R = 10$ and $R = 25$ were, for any practical purposes, indistinguishable.

If we fix a p, say $p = 0.95$, then (10) and $\hat{F}_B$ combine to provide simultaneous prediction regions. The $\tilde{f}_h$ of any test subject can then be plotted with $\bar{\hat{f}}_h(\theta)$, $\bar{\hat{f}}_h(\theta) - m_{0.95}\hat{\sigma}_f(\theta)$ and $\bar{\hat{f}}_h(\theta) + m_{0.95}\hat{\sigma}_f(\theta)$ to see whether $\tilde{f}_h$ lies within the region. This has been done (for $R = 10$) in Figures 6 and 7. In the first instance, right foot rotation is plotted, with $\bar{\hat{f}}_h$ and the prediction region, for a clinically normal five-year-old who was actually a member of the training sample. The only known outlying aspect of this "test" child's walk was an unusually high ankle dorsi–plantarflexion $\hat{\alpha}_5$ and, of course, this $\tilde{f}_h$ lies within the bounds throughout. Figure 7 gives the prediction region and $\bar{\hat{f}}$ for five-year-old left ankle dorsi–plantarflexion. Here the test subject is a five-year-old child born 10 weeks prematurely. Obviously the child's walk is abnormal; he has cerebral palsy.

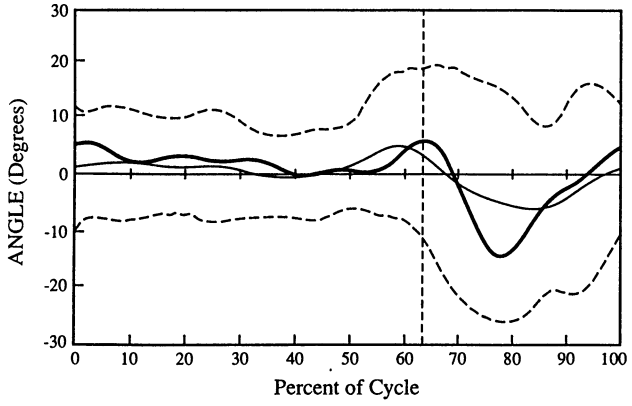


FIG. 6. *A 95% prediction region for five-year-old right foot rotation is bounded by the dotted lines; $\hat{f}_h$ is given by the solid line. The heavy line is $\tilde{f}_h$ for a clinically normal child who was actually a member of the training sample.*

In order to better understand the stability and robustness of our prediction regions, we ran a simulation study as follows. For a fixed f we simulated N Gaussian $\alpha_0, \ldots, \beta_6$ from a distribution with the same μ and Γ as were estimated. Both $\hat{f}_h$ and prediction regions were found for the simulated data. Recall from discussion of the previous section that individual f_h—and necessarily $\hat{f}_h$, too—did not look like observed f_h. However, the simulated prediction regions bear remarkable resemblance to the real regions and to those of Johansen and Johnstone. An example is given in Figure 8. In fact, $\hat{F}_B(m)$ was well approximated by the simulated version for all m.

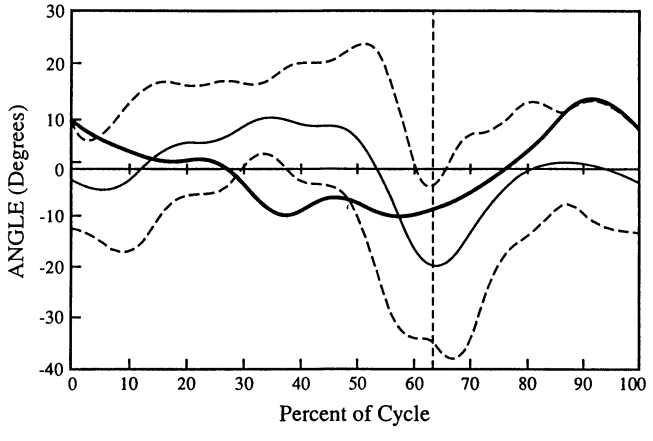


FIG. 7. *A 95% prediction region and $\bar{f}_h$ are given as in Figure 6, though here for five-year-old left ankle dorsi–plantarflexion. The test subject $\tilde{f}_h$ was computed from the data of a 10 week prematurely born child who has cerebral palsy.*

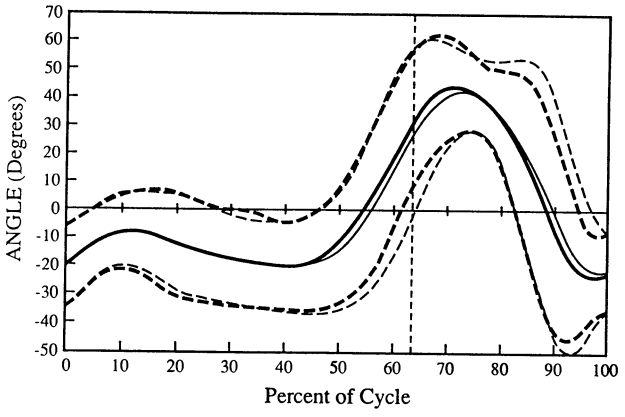


FIG. 8. *Prediction regions for five-year-old right knee flexion–extension, with* $R = 25$. *The lighter solid and dashed lines are based on real data, the others on Gaussian data with (true) mean and covariance that were observed.*

The cited theory of Johansen and Johnstone yields prediction regions that can in practice closely approximate what we compute from the bootstrap. Their theory as presently implemented depends on Gaussian assumptions on the distributions of Fourier coefficients and on the approximation $\hat{\sigma}_f(\theta) = \sigma_f(\theta)$. It follows from the comparison of Johansen and Johnstone with this paper that in the examples drawn from gait analysis, for the purpose of computing bootstrap prediction regions, the non-Gaussian distributions of the higher frequency Fourier coefficients do not bear heavily on $\hat{F}_B(m)$. We certainly did not know this when we began.

The bootstrap procedure that has been described for f_h can be carried out as well for α_0. In this case what was a prediction region becomes a prediction interval. The bootstrap process is used to estimate the probability of the event that corresponds to (10) (in an obvious notation),

$$\{\tilde{\alpha}_0 : \bar{\hat{\alpha}}_0 - m\hat{\sigma}_\alpha \le \tilde{\alpha}_0 \le \bar{\hat{\alpha}}_0 + m\hat{\sigma}_\alpha\}. \tag{13}$$

The theory of Johansen and Johnstone (1990) applies here, too. However, curves like the one depicted in Figure 9 are not expected to be sigmoidal in shape, because $\sqrt{\chi_1^2}$ is not approximately Gaussian like $\sqrt{\chi_{12}^2}$ is. We expect but are not certain that in practice the Johansen–Johnstone prediction intervals correspond closely to those that we compute from the bootstrap.

Table 1 gives results on bootstrap percentiles for α_0 for two prematurely born five-year-old children and 10 angular rotations. The subject denoted 8519 is the same child whose $\tilde{f}_h$ was displayed in Figure 7. Clearly 7 of his 10 rotations put him at no lower than the 99th percentile for normal children. The only three rotations for which he does better deal with movement toward the middle of the body and they are not associated with what one might think of as fine motor

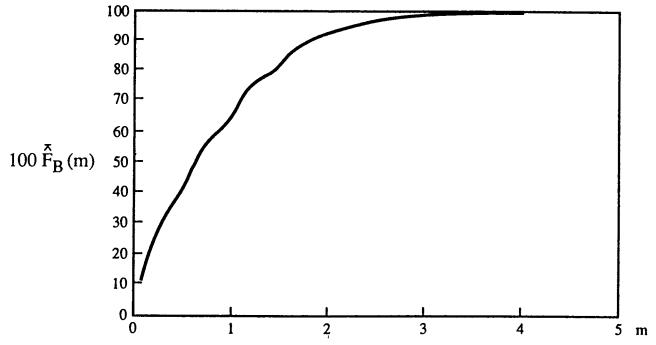


FIG. 9. *Plot of a bootstrap estimate of the probability of the event* (13). *As with Figure* 5, *the rotation is ankle dorsi–plantarflexion for five-year-old children with* $N = 38$ *and* $R = 10$.

control, as is the foot. Notice also that since most of the gait cycle is stance phase (when the foot is on the ground), the predicted average (across θ) ankle value could be -67.1, as it is, only if the child were walking on the balls of his feet. (A subject standing still with heels together has ankle dorsi–plantarflexion of approximately 0.) Subject 8506, two fewer weeks prematurely born than subject 8519, appears to be normal at a superficial glance. His bootstrap knee flexion–extension percentile is high and his left hip clearly does not rotate internally as he walks, quite like the hips of his full term peers. We once thought that this child's neuromuscular development was delayed because of his premature birth. However, for reasons not discussed here [see Sutherland, Olshen, Biden and Wyatt (1988), pages 191–200], we decided that the gait of this child more closely resembles that of an older, not younger, child.

TABLE 1

Left $\hat{\alpha}_0$ *averages and standard deviations and bootstrap distribution percentiles for two prematurely born children*

				Subject 8519 (10 weeks)			Subject 8506 (8 weeks)		
Plane	**Rotation**	$\bar{\hat{\alpha}}_0$	$\hat{\sigma}_{\hat{\alpha}}$	$\tilde{\alpha}_0$	$\#\hat{\sigma}_{\hat{\alpha}}$	**Bootstrap percentile**	$\tilde{\alpha}_0$	$\#\hat{\sigma}_{\hat{\alpha}}$	**Bootstrap percentile**
Sagittal	Knee F–E	30.9	3.2	47.9	5.3	99 +	23.8	2.2	98
Sagittal	Ankle D–P	1.5	3.4	−67.1	19.3	99 + +	0.8	0.2	19
Sagittal	Hip F–E	26.0	5.5	32.8	1.2	73	17.8	1.5	81
Transverse	Pelvic Rtn	−1.7	4.8	4.5	1.3	83	1.7	0.7	45
Transverse	Femoral Rtn	1.5	6.1	30.9	4.8	99 +	−5.7	1.2	77
Transverse	Tibial Rtn	−6.2	6.9	67.9	10.7	99 + +	−5.3	0.1	5
Transverse	Foot Rtn	−7.2	4.7	14.5	4.6	99 +	−13.7	1.4	86
Transverse	Hip Rtn	3.2	5.8	26.4	4.0	99 +	−7.4	1.8	90
Transverse	Knee Rtn	−7.6	6.7	37.0	6.7	99 +	0.3	1.2	80
Frontal	Pelvic Oblg	1.4	2.0	−2.7	2.1	96	−0.8	1.1	70

5. Some mathematical considerations. What follows are the beginnings of an asymptotic theory for the bootstrap prediction regions. Of course, one popular area of current research regarding the bootstrap is its hyperaccuracy for both critical points and coverage probability for certain confidence interval problems; see, for example, Babu and Singh (1983), Abramovitch and Singh (1985) and, especially, Hall (1988). One key to these results is an Edgeworth-like asymptotic expansion, difficulties with which have been discussed by Hartigan (1986) and Olshen (1986); see Hall (1988) for a definitive exposition. The presence of $\tilde{f}_h$ in the numerator of (11) precludes asymptotic normality of the statistic given there; rather, that statistic, asymptotically, is obviously a mixture of normals. In the special case of prediction intervals for α_0, it has been shown without Edgeworth expansions that bootstrap-based critical points are not hyperaccurate, but nonetheless coverage probabilities are!; see Bai (1988), Bai and Olshen (1988) and Bai, Bickel and Olshen (1990). (An estimator based on a learning sample of size N is said to be *hyperaccurate* if the difference between it and its putatively correct value tends to 0 faster than $N^{-1/2}$ as N tends to infinity. The sense in which the difference is computed depends on context. For bootstrap-based prediction intervals, the differences of critical values converge almost surely, while the convergence of differences of coverage probabilities is that of a sequence of real numbers. The name *hyperaccurate* is due to Wing Wong.) We concentrate here only on the sample path behavior of the theoretical bootstrap probability of the event given by (10) and (11). The arguments presented concentrate first on the case $M = 1$, that is, the upper limit of summation in (1) is 1.

To begin, we assume the existence of a sequence of random vectors $(\tilde{\alpha}, \tilde{\beta}), (\hat{\alpha}^{(1)}, \hat{\beta}^{(1)}), (\hat{\alpha}^{(2)}, \hat{\beta}^{(2)}), \ldots$ that are iid. These are the individual least squares estimates for, respectively, the test case and the learning sample. Next we introduce a triangular array of bootstrap random vectors $\{(\tilde{\alpha}_B^{(i)}, \tilde{\beta}_B^{(i)})\colon i = 1, 2, \ldots\}$ and $\{(\hat{\alpha}_B^{(i,j)}, \hat{\beta}_B^{(i,j)})\colon i = 1, 2, \ldots;\ j = 1, \ldots, i\}$. Their distributions are defined in terms of conditional distributions as follows. Let $\mathscr{B}_N$ be the σ-field generated by $\{(\hat{\alpha}^{(i)}, \hat{\beta}^{(i)})\colon i \le N\}$, $\{(\hat{\alpha}_B^{(i,j)}, \hat{\beta}_B^{(i,j)})\colon i \le N - 1,\ j \le i\}$ and $\{(\tilde{\alpha}_B^{(i)}, \tilde{\beta}_B^{(i)})\colon i \le N - 1\}$. $(\tilde{\alpha}, \tilde{\beta})$ is independent of $\bigvee_N \mathscr{B}_N$, the span of the $\mathscr{B}_N$. Given $\mathscr{B}_N$, the random vectors $(\tilde{\alpha}_B^{(N)}, \tilde{\beta}_B^{(N)})$ and $\{(\hat{\alpha}_B^{(N,i)}, \hat{\beta}_B^{(N,i)})\colon i \le N\}$ are iid with distribution P_N, the empirical distribution of $\{(\hat{\alpha}^{(i)}, \hat{\beta}^{(i)})\colon i \le N\}$. Finally, write $P_B^{(N)}$ for $P(\cdot|\mathscr{B}_N)$ and $E_B^{(N)}$ for expectations computed with respect to $P_B^{(N)}$. All conditional probabilities are assumed to be regular.

Assume that an angular rotation f is fixed. Let

$$\bar{\hat{\alpha}} = \bar{\hat{\alpha}}(N) = N^{-1} \sum_{j=1}^{N} \hat{\alpha}^{(i)};$$

$\bar{\hat{\beta}} = \bar{\hat{\beta}}(N)$ and $\hat{\sigma}_f = \hat{\sigma}_f(N)$ are defined by analogy. Also, set

$$\bar{\hat{\alpha}}_B = \bar{\hat{\alpha}}_B(N) = N^{-1} \sum_{j=1}^{N} \hat{\alpha}_B^{(N,j)}$$

and define $\bar{\hat{\beta}}_B$ and $\hat{\sigma}_{f,B}$ by analogy. For expository convenience we here use a

conventional 0–2π scale by writing $\psi = 2\pi\theta/100$. The probability of the event (10) (with $M = 1$ and not 6) is thus

$$P\left(\bigcap_{\psi}\left[\left|(\bar{\hat{\alpha}} - \tilde{\alpha})\cos\psi + \left(\bar{\hat{\beta}} - \tilde{\beta}\right)\sin\psi\right| \le \hat{\sigma}_f(\psi)m\right]\right). \tag{14}$$

Introduce the two quantities

$$\mathrm{I} = P\left\{\bigcap_{\psi}\left[\left|(\bar{\hat{\alpha}} - \tilde{\alpha})\cos\psi + \left(\bar{\hat{\beta}} - \tilde{\beta}\right)\sin\psi\right| \le \hat{\sigma}_f(\psi)m\right]\right\}$$

and

$$\mathrm{II} = P_B^{(N)}\left\{\bigcap_{\psi}\left[\left|\left(\bar{\hat{\alpha}}_B - \tilde{\alpha}_B^{(N)}\right)\cos\psi + \left(\bar{\hat{\beta}} - \tilde{\beta}_B^{(N)}\right)\sin\psi\right| \le \hat{\sigma}_{f,B}(\psi)m\right]\right\}.$$

The point of what we show is that I and II are close. In order to complete our program, we assume that

$$\begin{aligned}&\tilde{\alpha} \text{ and } \tilde{\beta} \text{ both have finite } (8+\delta)\text{th moments for some positive}\\ &\delta \text{ and } (\tilde{\alpha}, \tilde{\beta}) \text{ has a bounded density.}\end{aligned} \tag{15}$$

From the results of Rubin and Sethuraman (1965) and the Borel–Cantelli lemma we conclude that almost surely

$$\bar{\hat{\alpha}} = \bar{\hat{\alpha}}_B + O\left(\sqrt{\frac{\log N}{N}}\right), \tag{16}$$

$$\bar{\hat{\beta}} = \bar{\hat{\beta}}_B + O\left(\sqrt{\frac{\log N}{N}}\right) \tag{17}$$

and

$$\hat{\sigma}_f(\psi) = \hat{\sigma}_{f,B}(\psi) + O\left(\sqrt{\frac{\log N}{N}}\right), \tag{18}$$

where the last equality is uniform in ψ. Also,

$$\sup_{N,\psi}\left(\max\left(\left|\bar{\hat{\alpha}} - \bar{\hat{\alpha}}_B\right|, \left|\bar{\hat{\beta}} - \bar{\hat{\beta}}_B\right|, \left|\hat{\sigma}_f(\psi) - \hat{\sigma}_{f,B}(\psi)\right|\right)\right)$$

is integrable. So

$$\begin{aligned}\mathrm{I} = {} & P\left\{\bigcap_{\psi}\left[\left|\left(\bar{\hat{\alpha}}_B - \tilde{\alpha}\right)\cos\psi + \left(\bar{\hat{\beta}}_B - \tilde{\beta}\right)\sin\psi\right| \le \hat{\sigma}_f(\psi)m\right]\right\}\\ & + O\left(\sqrt{\frac{\log N}{N}}\right).\end{aligned} \tag{19}$$

Denote the first term on the right-hand side of (19) by I*. Write $\mathscr{D}_N = \mathscr{B}_N \vee \sigma(\{(\hat{\alpha}_B^{(N,j)}, \hat{\beta}_B^{(N,j)}): j \le N\})$, where $\sigma(\cdot)$ is the σ-field spanned by the parenthetic

random vectors. Then

$$\mathrm{I}^* = E\left\{P\left(\bigcap_{\psi}\left[|(\bar{\hat{\alpha}}_B - \tilde{\alpha})\cos\psi + (\bar{\hat{\beta}}_B - \tilde{\beta})\sin\psi| \le \hat{\sigma}_{f,B}(\psi)m\right]\mathscr{D}_N\right)\right\},$$

$$\text{(20)}\quad \mathrm{II} = \mathrm{II}^* = E_B^{(N)}\left\{P_B^{(N)}\left(\bigcap_{\psi}\left[|(\bar{\hat{\alpha}}_B - \tilde{\alpha}_B^{(N)})\cos\psi + (\bar{\hat{\beta}}_B - \tilde{\beta}_B^{(N)})\sin\psi| \le \hat{\sigma}_{f,B}(\psi)m\right]\mathscr{D}_N\right)\right\}.$$

Of course, part of the conditioning in II* is redundant; it could as well have been restricted to $\sigma(\{(\hat{\alpha}_B^{(N,j)}, \hat{\beta}_B^{(N,j)}): j \le N\})$. Note that after the conditioning, the remaining randomness in the integrand of I* is in $(\tilde{\alpha}, \tilde{\beta})$ and that the remaining randomness in the integrand in II* is in $(\tilde{\alpha}_B^{(N)}, \tilde{\beta}_B^{(N)})$. The latter has conditional distribution P_N.

The notion of Vapnik–Chervonenkis class figures in the arguments that we present now. Our definition is that of Kuelbs and Dudley (1980), who begin with a set $\mathscr{X}$ and a collection $\mathscr{S}$ of subsets of $\mathscr{X}$. Let $V(\mathscr{S})$ be the smallest n such that for every $F \subset \mathscr{X}$ with n elements, not every subset of F is of the form $F \cap C$, $C \in \mathscr{S}$. If $V(\mathscr{S}) < \infty$, then $\mathscr{S}$ is a *Vapnik–Chervonenkis class*.

In the $(\tilde{\alpha}, \tilde{\beta})$ plane, the set whose conditional probability is computed in I* is an ellipse. So from the fact that the set of ellipses is a Vapnik–Chervonenkis class [Steele (1975) and Dudley (1978)] and a law of the iterated logarithm due to Kuelbs and Dudley (1980), we see that

$$\text{(21)}\quad \sup_{a,b,c,m}\left\{\left|P_N\left(\bigcap_{\psi}\left[|(a - \tilde{\alpha}_B^{(N)})\cos\psi + (b - \tilde{\beta}_B^{(N)})\sin\psi| \le cm\right]\right) - P\left(\bigcap_{\psi}\left[|(a - \tilde{\alpha})\cos\psi + (b - \tilde{\beta})\sin\psi| \le cm\right]\right)\right|\right\}$$

is almost surely

$$O\left(\sqrt{\frac{\log\log N}{N}}\right).$$

Since the integrands in I* and II* do not exceed 1, from (21) it follows that

$$\sup_m|\mathrm{I} - \mathrm{II}| = O\left(\sqrt{\frac{\log N}{N}}\right) \quad \text{a.s.}$$

The discussion so far can be summarized thus:

THEOREM 1. *Assume that* $M = 1$ *and* (15). *Then the difference between the probability of the event* (10) *and the* (*theoretical*) *bootstrap probability of* (12) *is almost surely* $O(\sqrt{\log N/N})$, *uniformly in* m.

From Theorem 7.2 and Proposition 7.12 of Dudley (1978), it follows that (21) can be extended immediately to arbitrary $M < \infty$ if the intersection in (10), (11) and (14) and the corresponding maximum in (12) are restricted to (any) finite set. Of course, data are gathered only at a bounded number of discrete points, so the cited extension applies to our practical situation. However, we prefer not to have the restriction in theoretical results; and the extension of Theorem 1 to $M > 1$ does follow from Theorem 2, which is a special case of recent results of Stengle and Yukich (1989). Their result implies that for any fixed $M < \infty$, the sets that arise from (10) and that correspond to the cited ellipses in (20) are a Vapnik–Chervonenkis class. Therefore, the Kuelbs–Dudley law of the iterated logarithm that applied to (21) applies as well to this more general case.

THEOREM 2 [Stengle and Yukich (1989)]. *Assume that p, q, r and s are positive integers and that $\mathcal{G}$ is the set of functions of the form*

$$G(\mathbf{x}) = \max_{\psi} \left\{ \sum_{j=l}^{r} P_j(\mathbf{x}) Q_j(\psi) \right\},$$

where for each j, P_j is a polynomial of degree s in the p-dimensional real variable $\mathbf{x}$ and Q_j is a polynomial of degree q in $\cos\psi$ and $\sin\psi$. Then the class of subsets of Euclidean p-space of the form $\{\mathbf{x}: G(\mathbf{x}) > 0\}$ for some $G \in \mathcal{G}$ is a Vapnik–Chervonenkis class.

Acknowledgments. In the long course of our work in this paper, we have benefited from conversations with various colleagues. Particular thanks are due Peter Bickel, Morris Eaton, Bradley Efron, Soren Johansen, Iain Johnstone, Charles McCulloch, John Rice, Joe Yukich and Joel Zinn. Also, an Associate Editor and referees made helpful suggestions. As usual, none of them bears any responsibility for the deficiencies that remain.

REFERENCES

ABRAMOVITCH, L. and SINGH, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *Ann. Statist.* **13** 116–132.

BABU, G. J. and SINGH, K. (1983). Inference on means using the bootstrap. *Ann. Statist.* **11** 999–1003.

BAI, C. (1988). Asymptotic properties of some sample reuse methods for prediction and classification. Ph.D. dissertation, Dept. Mathematics, Univ. California, San Diego.

BAI, C. and OLSHEN, R. A. (1988). Discussion of "Theoretical comparison of bootstrap confidence intervals" by P. Hall. *Ann. Statist.* **16** 953–956.

BAI, C., BICKEL, P. J. and OLSHEN, R. A. (1990). Hyperaccuracy of bootstrap-based prediction. *Probability in Banach Spaces: Proceedings of the Seventh International Conference* (E. Eberlein, J. Kuelbs and M. B. Marcus, eds.). Birkhäuser, Boston. To appear.

CAPPOZZO, A., LEO, T. and PEDOTTI, A. (1975). A general computing method for the analysis of human locomotion. *J. Biomechanics* **8** 307–320.

CRAMÉR, H. (1936). Über eine Eigenschaft der normalen Verteilungsfunktion. *Math. Z.* **41** 405–414.

DUDLEY, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6** 899–929; correction **7** (1979) 909–911.

EFRON, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans.* SIAM, Philadelphia.

HALL, P. (1988). Theoretical comparison of bootstrap confidence intervals (with discussion). *Ann. Statist.* **16** 927–985.

HARTIGAN, J. A. (1986). Comment on "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy" by B. Efron and R. Tibshirani. *Statist. Sci.* **1** 75–77.

HOTELLING, H. (1939). Tubes and spheres in N-space, and a class of statistical problems. *Amer. J. Math.* **61** 440–460.

JOHANSEN, S. and JOHNSTONE, I. (1990). Hotelling's theorem on the volume of tubes: Some illustrations in simultaneous inference and data analysis. *Ann. Statist.* **18**. To appear.

KUELBS, J. and DUDLEY, R. M. (1980). Log log laws for empirical measures. *Ann. Probab.* **8** 405–418.

OLSHEN, R. A. (1986). Comment on "Jackknife, bootstrap, and other resampling methods in regression analysis" by C. F. J. Wu. *Ann. Statist.* **14** 1318–1320.

POWELL, M. J. D. (1981). *Approximation Theory and Methods*. Cambridge Univ. Press, Cambridge.

RAO, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika* **52** 447–458.

RICE, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12** 1215–1230.

RUBIN, H. and SETHURAMAN, J. (1965). Probabilities of moderate deviations. *Sankhyā Ser. A* **27** 325–346.

SCLOVE, S. L., MORRIS, C. and RADHAKRISHNAN, R. (1972). Non-optimality of preliminary-test estimators for the mean of a multivariate normal distribution. *Ann. Math. Statist.* **43** 1481–1490.

SHIBATA, R. (1981). An optimal selection of regression variables. *Biometrika* **68** 45–54.

STEELE, J. M. (1975). Combinatorial entropy and uniform limit laws. Ph.D. dissertation, Dept. Mathematics, Stanford Univ.

STENGLE, G. and YUKICH, J. (1989). Some new Vapnik–Chervonenkis classes. *Ann. Statist.* **17** 1441–1446.

SUTHERLAND, D. H. (1984). *Gait Disorders in Childhood and Adolescence*. Williams and Wilkins, Baltimore.

SUTHERLAND, D. H., OLSHEN, R. A., BIDEN, E. N. and WYATT, M. P. (1988). *The Development of Mature Walking*. MacKeith Press, London.

SUTHERLAND, D. H., OLSHEN, R. A., COOPER, L. and WOO, S. (1980). The development of mature gait. *J. Bone and Joint Surgery* **62A** 336–353.

SUTHERLAND, D. H., OLSHEN, R. A., COOPER, L., WYATT, M., LEACH, J., MUBARAK, S. and SCHULTZ, P. (1981). The pathomechanics of gait in Duchenne muscular dystrophy. *Develop. Med. Child Neurol.* **23** 3–22.

RICHARD A. OLSHEN
DIVISION OF BIOSTATISTICS
HRP BUILDING, ROOM 114
STANFORD UNIVERSITY SCHOOL OF MEDICINE
STANFORD, CALIFORNIA 94305-5092

EDMUND N. BIDEN
DEPARTMENT OF MECHANICAL ENGINEERING
UNIVERSITY OF NEW BRUNSWICK
P.O. BOX 4400
FREDERICTON, NEW BRUNSWICK
CANADA E3B 5A3

MARILYNN P. WYATT
DAVID H. SUTHERLAND
MOTION ANALYSIS LABORATORY
CHILDREN'S HOSPITAL AND HEALTH CENTER
8001 FROST STREET
SAN DIEGO, CALIFORNIA 92123