EMPIRICAL DISTRIBUTIONS IN LEAST SQUARES ESTIMATION FOR DISTRIBUTED PARAMETER SYSTEMS

B.G. Fitzpatrick † and G. Yin*

Department of Mathematics
North Carolina State University, Raleigh, NC 27695-8205

Department of Mathematics
Wayne State University, Detroit, MI 48202

Abstract. We consider the estimation of error distributions in least squares identification of distributed parameter systems. Asymptotic properties of approximate error sequences are developed. In particular, we examine consistency and asymptotic normality of empirical estimates of the error distribution. The consistency obtained is analogous to the Glivenko-Cantelli theorem. For asymptotic normality, we establish that the normalized sequence of empirical distributions converges to a Gaussian random element, which is the sum of a stretched Brownian bridge plus another Gaussian process.

Keywords. distributed parameter systems, empirical distribution, Brownian bridge, weak convergence.

† Research of this author was supported in part by the Air Force Office of Scientific Research under grant AFOSR-91-0021.

* Research of this author was supported in part by the National Science Foundation under grant DMS-9022139.

1. Introduction

In many applications, ranging from population biology to structural vibrations, one models dynamic behavior with a distributed model, such as a partial differential equation (PDE). In quantitative situations one wishes to make predictions concerning the system based on the mathematical model. To do so requires knowledge of system parameters which may not be directly measurable. Hence, the model must be fit to observed data to determine these parameters. In most cases, observations of the system are corrupted by some sort of measurement errors. An important aspect of many applied problems is assessing the fit of a model to observed data. One particular question that arises is the impact of measurement uncertainty on parameter estimates obtained from such a fit. It is our goal in this work to examine some statistical procedures, based on empirical distributions, which will allow us to analyze the effect of measurement error on parameter estimates.

For an example of a distributed parameter estimation problem, we consider an experiment in which a flexible beam, clamped at one end and free at the other, is deformed to an initial deflection and then released to vibrate freely. A commonly-used model of a vibrating beam is the Euler-Bernoulli equation

$$\rho\frac{\partial^2 u}{\partial t^2}(x,t)+\gamma\frac{\partial u}{\partial t}(x,t)+\frac{\partial^2}{\partial x^2}\left[EI\frac{\partial^2 u}{\partial x^2}(x,t)+c_D I\frac{\partial^3 u}{\partial x^2 \partial t}(x,t)\right]=0,$$

for $t>0, \quad 0<x<\ell$, where $u(x,t)$ denotes the displacement of the beam at position x along its axis and at time t, ρ is the linear mass density, γ is the viscous damping coefficient, EI is the stiffness, and $c_D I$ is the Kelvin-Voigt damping coefficient. The initial displacement $u(x,0)$ is known, and the beam is initially at rest, so that $\partial u/\partial t$ is initially 0. The beam is assumed to be clamped at $x=0$ and free at $x=\ell$, conditions which determine the boundary conditions. We set $q=(\gamma, c_D I, EI)$, to denote the vector of parameters for the problem, and we include explicitly the parameter dependence of the solution with the notation $u(x,t;q)$.

In our example, we place an accelerometer at the tip of the beam, so that we obtain a sequence of acceleration measurements over a period of time. We denote by $\{\widehat{u}_{tt}(\ell, t_k)\}_{k=1}^{n}$ the collection of acceleration measurements. If the model accurately represents the experimental system, one might expect that $\widehat{u}_{tt}(\ell, t_k) \approx u_{tt}(\ell, t_k; q^*)$, for some value of q^*. The usual inverse problem is to estimate the parameter vector q^* from these observations. Statistical tasks center on the relationship of the model to the data and the impact of this relationship on the parameter estimates.

In Fitzpatrick [10], and Banks and Fitzpatrick [3], this nonlinear least squares problem was examined within the context of hypothesis testing for parameter estimators. These results are very much in the spirit of analysis of variance (ANOVA) testing, in which one compares the least squares residuals obtained by different estimators. Further asymptotic results in light of functional invariance principles were obtained in [16]. Applied to certain biological problems, such as fluid transport in brain tissue, these tests provided insight into various modeling questions such as the role of convection in grey and white matter (see [3]). However, in [2], when these tests were applied to structural vibration problems, we found them to be "too powerful," due to the point hypothesis nature of the ANOVA test. In many such problems one needs to test an interval hypothesis. Possible alternatives to ANOVA-type tests are Bayesian methods (see for example [11]) and Monte Carlo methods. Both of these techniques require precise knowledge of the measurement error distribution.

Another interesting question which cannot be approached with the framework of [3] involves testing observation errors for spatial uniformity. In [1], a size structured model for larval striped bass populations was considered. Here a distributed model was fit to observations using least squares techniques, and the above-mentioned ANOVA statistics were used to study changes in mortality rates. These tests seemed to work well only when data from fish of smaller sizes was omitted from the least squares fit, suggesting that the errors were not identically distributed over the size structure. Formally applying Kolmogorov-Smirnov tests rejected at all levels the hypothesis of independent, identically distributed errors; however, the Kolmogorov-Smirnov test is not applicable in cases in which parameters are estimated.

Our goal in this work is to derive a number of asymptotic results for a sequence of empirical or sample distributions for measurement errors in a least squares framework. The impact of such a study is that the sample distribution can be used for various statistical inference tasks, such as testing the data for bias, implementing a Monte Carlo, bootstrap, or empirical Bayes inference scheme. The rest of the paper is arranged as follows. In Section 2, we set up the least squares framework and state some previous results of [2, 3, 10] for easy references in the subsequent sections. Section 3 is devoted to the almost sure convergence of the empirical distributions, which is obtained by means of constructing lower and upper bounds of the empirical distribution function. Then, the rate of convergence is dealt with in Section 4. Employing weak convergence methods, we show that a suitably scaled version of the empirical distribution converges weakly to a Gaussian random element. Section 5 treats further extensions to two-sample and many sample problems, and we conclude with some remarks concerning implementation and further work in Section 6.

2. Empirical Distributions in Least Squares Identification

The least squares identification problem that we consider here can be described as follows. We have a sequence of observations $\{Y_k\}$ with

$$Y_k = f(x_k, q^*) + \varepsilon_k,\ 1 \le k \le n, \tag{2.1}$$

where $\{x_k\}$ (with $x_k \in X \subset \mathbb{R}^r$, $1 \le k \le n$) is a collection of settings at which the observations are made, $f(x, q)$ is a parameterized function, called the model function, and $\{\varepsilon_k\}$ denotes the measurement error. The function f typically arises from a differential equation model describing the system. The idea here is that the "true parameter" q^* is unknown: we wish to estimate it by fitting the observations $\{Y_k\}$ to f. In what follows, we assume q^* lies in a known set Q_{ad}, referred to as the admissible parameter set. This set incorporates various constraints on the parameters, which in general depend on the specific application. To estimate the parameter q^*, we minimize the mean square objective function

$$J_n(q) = \frac{1}{n}\sum_{k=1}^{n}(Y_k - f(x_k, q))^2, \tag{2.2}$$

and we denote by q_n a minimizer over Q_{ad}.

One important ingredient for inference is the distribution of the observation error ε_k. Unfortunately, $\{\varepsilon_k\}$ is not observable. To overcome this difficulty, we estimate the errors by:

$$\varepsilon_k^n = Y_k - f(x_k, q_n), \tag{2.3}$$

where q_n minimizes J_n over Q_{ad}. Since Y_k is the actual observation at time k and q_n is the parameter estimator, $\{\varepsilon_k^n\}$ is effectively an approximating sequence of the observation errors $\{\varepsilon_k\}$. More generally, we define

$$\varepsilon_k(\tilde{q}) = Y_k - f(x_k, \tilde{q}),$$

which represents an estimate of the error based on $\tilde{q}$, which may be a fixed value or a random variable. Note, of course, that $\varepsilon_k^n = \varepsilon_k(q_n)$.

In order to study the distribution of ε_k, we define an empirical distribution function

$$\begin{aligned} \hat{F}_n(t, \tilde{q}) &= \frac{1}{n} \sum_{k=1}^{n} I_{\{\varepsilon_k(\tilde{q}) \le t\}} \\ &= \frac{1}{n} \left(\# \text{ of } \varepsilon_k(\tilde{q});\ k \le n,\, \varepsilon_k(\tilde{q}) \le t\right), \end{aligned} \tag{2.4}$$

which is based on the estimated errors, instead of the unobservable "true" errors.

Of course, the behavior of this empirical distribution depends heavily on the behavior of the estimator $\tilde{q}$ on which it is based. Thus, we review some results on least squares estimators. These results will provide the foundation for our study of the empirical distribution. To proceed, we first make the following assumptions:

(A1) The sequence $\{\varepsilon_k\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with $E\varepsilon_k = 0$, $E\varepsilon_k^2 = \sigma^2 < \infty$.

(A2) The set Q_{ad} is a compact subset of a complete, separable metric space Q. The set X is a compact subset of $\mathbb{R}^r$. The function $f : Q \to C(X)$ is continuous, where $C(X)$ denotes the space of continuous functions defined on X.

(A3) The sequence $\{x_k\}$ is a subset of X and there exists a finite measure μ on X, such that for each bounded and continuous function h, $\frac{1}{n}\sum_{k=1}^{n} h(x_k) \xrightarrow{n} \int_X h d\mu$.

(A4) The functional $J^*(q) = \sigma^2 + \int_X (f(x, q^*) - f(x, q))^2 d\mu$ has a unique minimizer $q^* \in \text{Int } Q_{ad} \subset \text{Int } Q$, where Int G denotes the interior of the set G.

(A5) The set Q is a subset of $\mathbb{R}^r$, and that for each x, $f(x, \cdot)$, is twice continuously differentiable. In addition, the matrix

$$\mathcal{T} = \partial^2 J^*(q^*)/\partial q^2 = 2\int_X \frac{\partial f(x, q^*)}{\partial q}\frac{\partial f'(x, q^*)}{\partial q} d\mu(x)$$

is positive definite.

Assumption (A1) gives us a simple statistical model of the error. The restrictions contained in (A2) are typical of those required in general frameworks for analysis of inverse problems (see in particular, Section IV. 8 of [4]). Assumption (A3) controls the manner in which the measurement sample is increased in size. Assumption (A4) is known as identifiability, and the technical condition that the "true" parameter lie in the interior cannot be removed (see [3, 10]). The extra assumptions of (A5) are needed for the asymptotic distribution theory, which is analyzed through linearization. Verifying (A4) and (A5) is a task that depends heavily on the specific problem. For more on the motivation and applicability of these conditions, we refer the readers to the papers [2, 3, 10].

Note that continuity of f and compactness of Q_{ad} guarantee the existence of a minimizer q_n of J_n. The following results, which can be found in [2, 3, 10], will be used heavily in the sequel.

Proposition 2.1. *Under the above conditions (A1)—(A4)*

(1) for each $q \in Q$, $P(\lim_n J_n(q) = J^*(q)) = 1$ *and the convergence is uniform on each compact subset of* Q.

(2) $P(\lim_n q_n = q^*) = 1$. *If, additionally, we assume that (A5) holds, then we have*

(3) $\sqrt{n}(q_n - q^*) \xrightarrow{n} N(0, 2\sigma^2 \mathcal{T}^{-1})$ *in distribution. In the above* $N(0,S)$ *denotes a normal distribution with mean 0 and covariance* S. $\square$

In fact, more far reaching convergence and rate of convergence results have been obtained: interested readers are referred to [16] for the corresponding functional limit results which exploit the 'dynamic' behavior of the estimation sequence. We now proceed to the study of the empirical distributions $\hat{F}_n$, beginning with consistency.

3. Almost sure convergence of $\hat{F}_n$

As was mentioned before, in a wide variety of situations, one may wish to monitor closely how the random noise comes into play. Since the true error sequence $\{\varepsilon_k\}$ is not observable, it is then necessary to analyze an estimate, such as the sequence $\{\varepsilon_k^n\}$. Notice that the sequence defined by (2.3) actually depends on q_n, so that ε_k^n are not even independent.

For the distribution function defined in (2.4), we wish to derive a result analogous to the well-known Glivenko-Cantelli's Theorem. However, from Taylor's theorem and the assumptions and results of the previous section, we have that

$$\varepsilon_k^n = \varepsilon_k + \left(\frac{\partial f(x_k, \bar{q}_n)}{\partial q} \right)' (q^* - q_n) = \varepsilon_k + o_s(1), \tag{3.1}$$

where $\bar{q}_n$ is a vector with all of its components sitting between q^* and q_n, and $o_s(1) \xrightarrow{n} 0$ w.p.1. uniformly in $k < n$. Eq. (3.1) indicates that $\{\varepsilon_k^n\}$ behaves asymptotically like $\{\varepsilon_k\}$. We shall see below exactly how the empirical distributions based on these sequences are related.

Let the common distribution function of ε_k be $F(t)$. Similar to (2.4), we write

$$F_n(t, q^*) = \frac{1}{n} \sum_{k=1}^{n} I_{\{\varepsilon_k \le t\}}. \tag{3.2}$$

Note that $F_n(t, q^*)$ is the empirical distribution for samples obtained from the sequence $\{\varepsilon_k\}$. Since these errors are not observable, this distribution is not computable from the data. However, as we shall see below, it is of great importance in the proofs of our results.

Theorem 3.1. *Assume that (A1) and (A2) hold, that $\tilde{q}_n \to q^*$ w.p.1, and that $F(\cdot, q)$ is continuous for each $q \in Q$. Then,*

$$P\left(\lim_n \hat{F}_n(t, \tilde{q}_n) = F(t, q^*)\right) = 1,$$

and the convergence is uniform in t.

Proof: We prove this theorem by using a "squeezing" argument; i.e., we shall find lower and upper bounds for ε_k^n and then show the bounds have the same limit. Let

$$l_n = \inf_x (f(x, q^*) - f(x, \tilde{q}_n)), \quad u_n = \sup_x (f(x, q^*) - f(x, \tilde{q}_n)) \tag{3.3}$$

Then,

$$\varepsilon_k + l_n \le \varepsilon_k^n \le \varepsilon_k + u_n. \tag{3.4}$$

This implies that if $\varepsilon_k + u_n \le t$, then $\varepsilon_k^n \le t$ and $\varepsilon_k + l_n \le t$. Consequently,

$$\{\varepsilon_k + u_n \le t\} \subset \{\varepsilon_k^n \le t\} \subset \{\varepsilon_k + l_n \le t\}. \tag{3.5}$$

Therefore, for the corresponding indicator functions, we have

$$I_{\{\varepsilon_k + u_n \le t\}} \le I_{\{\varepsilon_k^n \le t\}} \le I_{\{\varepsilon_k + l_n \le t\}}.$$

This then yields

$$F_n(t - u_n, q^*) \le \hat{F}_n(t, \tilde{q}_n) \le F_n(t - l_n, q^*). \tag{3.6}$$

By virtue of the well-known Glivenko-Cantelli theorem (see for example [5] and the references therein), $F_n(\cdot) \to F(\cdot)$ uniformly in t. Owing to the strong consistency of $\{\tilde{q}_n\}$ and the continuity of f, $t - u_n \xrightarrow{n} t$ w.p.1 and $t - l_n \xrightarrow{n} t$ w.p.1. By virtue of the continuity of $F(\cdot)$, both the left-hand side and the right-hand side of (3.6) tend to $F(t, q^*)$ uniformly in t. The theorem is thus proved. □

As seen in Section 2 above, the assumptions (A1)–(A4) guarantee the strong consistency for the least squares estimator q_n; hence, Theorem 3.1 gives us the first result of interest–strong consistency of the empirical distributions using the least squares estimator. The reason for the slightly more general setting is to allow for other possible estimators–particularly regularized least squares estimators (see [11]).

4. Normalized error sequences

In this section, we wish to establish a weak convergence theorem for the sequence

$$\hat{W}_n(t) = \sqrt{n}(\hat{F}_n(t, q_n) - F(t, q^*)). \tag{4.1}$$

This can be considered as a rate of convergence result. In addition, it provides an approach to estimating distributions of statistics which can test for questions such as biases in data and goodness of fit of certain error distributions.

The convergence result of the previous section states merely that the convergence is uniform, with probability one, and no extra function space structure is required. However, to develop a weak convergence theory, we will rely heavily on general results developed for random processes in a particular space of functions. We denote by $D = D(-\infty, \infty)$ the space of real-valued functions, defined on the interval $(-\infty, \infty)$, which are right continuous, and have left limits. This space allows for the type of discontinuities that are inherent in empirical distributions. Moreover, this space is a complete separable metric space when endowed with the Skorohod metric (functions are "close" in this metric if they are "uniformly close" where one is smooth and if their discontinuities are close). The framework of weak convergence in D provides flexible tools for analysis of our empirical process. For a general development of the Skorohod topology and weak convergence in D, see for example [5], [6], [9], [13], or [14].

To derive the desired result, we need one more assumption:

(A6) The sequence $\{\varepsilon_k\}$ has a common density function $\pi_\varepsilon(\cdot)$ which is continuous on $(-\infty, \infty)$.

We thus restrict ourselves to continuous random variables with finite fourth moment. A typical example satisfying (A6) is a Gaussian distribution. In view of this assumption,

$$F(t, q^*) = P(\varepsilon_k \le t) = \int_{-\infty}^{t} \pi_\varepsilon(s)ds \text{ and } \frac{\partial F(t, q^*)}{\partial t} = \pi_\varepsilon(t). \tag{4.2}$$

Next, we set

$$\tau_k^n = (\partial f(x_k, \bar{q}_n)/\partial q)(q_n - q^*), \tag{4.3}$$

where $\bar{q}_n$ is defined as in (3.1); that is,

$$\varepsilon_k^n = \varepsilon_k - \tau_k^n.$$

By virtue of the definition of $\hat{W}_n(\cdot)$,

$$\begin{aligned}\hat{W}_n(t) &= \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\Big([I_{\{\varepsilon_k \le t+\tau_k^n\}} - F(t+\tau_k^n))] - [F(t+\tau_k^n)) - F(t)]\Big) \\ &= \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\big([I_{\{\varepsilon_k \le t\}}] - F(t)\big) + \frac{1}{\sqrt{n}}\sum_{k=1}^{n}(F(t+\tau_k^n)) - F(t)) \\ &\qquad + \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\Big((I_{\{\varepsilon_k \le t+\tau_k^n\}} - F(t+\tau_k^n)) - (I_{\{\varepsilon_k \le t\}} - F(t))\Big).\end{aligned} \tag{4.4}$$

This type of "splitting up" has been used by Durbin (see [7], [8]) in the case of i.i.d. observations when F is evaluated using an estimated parameter. To establish a weak convergence result for $\{\hat{W}_n(\cdot)\}$, we shall study the terms in (4.4), beginning with the last one. The details of the present case are quite different from those of [7] and [8].

Lemma 4.1. *Under the conditions (A1)-(A6),*

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\Big((I_{\{\varepsilon_k \le t+\tau_k^n\}} - F(t+\tau_k^n)) - (I_{\{\varepsilon_k \le t\}} - F(t))\Big) \xrightarrow{n} 0 \textit{ in probability.} \tag{4.5}$$

Proof: Suppose that $\eta > 0$. We may rewrite the left-hand side of (4.5) as

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\Big((I_{\{\varepsilon_k \le t+\tau_k^n\}} - F(t+\tau_k^n)) - (I_{\{\varepsilon_k \le t\}} - F(t))\Big) I_{\{|\tau_k^n| \le \eta\}} \\ &\quad + \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\Big((I_{\{\varepsilon_k \le t+\tau_k^n\}} - F(t+\tau_k^n)) - (I_{\{\varepsilon_k \le t\}} - F(t))\Big) I_{\{|\tau_k^n| > \eta\}}.\end{aligned} \tag{4.6}$$

Since τ_k^n converges in probability to 0 as $n \to \infty$ uniformly in $k \le n$, for any $\eta > 0$, we have that

$$P[|\tau_k^n| > \eta] < \eta.$$

For the second term in (4.6), using the definition of τ_k^n, an application of the Chebyshev inequality gives,

$$
\begin{aligned}
&E\left|\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\Big((I_{\{\varepsilon_k\le t+\tau_k^n\}}-F(t+\tau_k^n))-(I_{\{\varepsilon_k\le t\}}-F(t))\Big)\,I_{\{|\tau_k^n|>\eta\}}\right| \\
&\quad\le 4\frac{1}{\sqrt{n}}\sum_{k=1}^{n}P(|\tau_k^n|>\eta) \\
&\quad\le \frac{4}{\eta^2\sqrt{n}}\sum_{k=1}^{n}\frac{M}{n}E|\sqrt{n}(q_n-q^*)|^2 \\
&\quad\le \frac{4M}{\sqrt{n}\eta^2},
\end{aligned}
$$

where $M=\sup_{x,q}|\partial f/\partial q|$. Having $\eta>0$ fixed, letting $n\to\infty$, the above term tends to 0. Consequently, we need only look at the first term on the right-hand side of (4.6).

For future use, note that when

$$-\eta\le\tau_k^n\le\eta,$$

$$I_{\{\varepsilon_k\le t+\eta\}}-I_{\{\varepsilon_k\le t+\tau_k^n\}}\ge 0,\ \ F(t+\tau_k^n)-F(t-\eta)\ge 0. \tag{4.7}$$

Therefore,

$$[I_{\{\varepsilon_k\le t+\eta\}}-F(t-\eta)]\ge[I_{\{\varepsilon_k\le t+\tau_k^n\}}-F(t+\tau_k^n)]. \tag{4.8}$$

A straight forward expansion of the first term on the right-hand side of (4.6) leads to:

$$
\begin{aligned}
&E\left|\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\Big([I_{\{\varepsilon_k\le t+\tau_k^n\}}-F(t+\tau_k^n)]-[I_{\{\varepsilon_k\le t\}}-F(t)]\Big)\,I_{\{|\tau_k^n|\le\eta\}}\right|^2 \\
&=\frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}E\Big([I_{\{\varepsilon_k\le t+\tau_k^n\}}-F(t+\tau_k^n)]-[I_{\{\varepsilon_k\le t\}}-F(t)]\Big) \\
&\qquad\times\Big([I_{\{\varepsilon_j\le t+\tau_j^n\}}-F(t+\tau_j^n)]-[I_{\{\varepsilon_j\le t\}}-F(t)]\Big)\,I_{\{|\tau_k^n|\le\eta\}}I_{\{|\tau_j^n|\le\eta\}} \\
&=\frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}E\Bigg([I_{\{\varepsilon_k\le t+\tau_k^n\}}-F(t+\tau_k^n)][I_{\{\varepsilon_j\le t+\tau_j^n\}}-F(t+\tau_j^n)] \\
&\qquad-[I_{\{\varepsilon_k\le t+\tau_k^n\}}-F(t+\tau_k^n)][I_{\{\varepsilon_j\le t\}}-F(t)] \\
&\qquad-[I_{\{\varepsilon_k\le t\}}-F(t)][I_{\{\varepsilon_j\le t+\tau_j^n\}}-F(t+\tau_j^n)] \\
&\qquad+[I_{\{\varepsilon_k\le t\}}-F(t)][I_{\{\varepsilon_j\le t\}}-F(t)]\Bigg)I_{\{|\tau_k^n|\le\eta\}}I_{\{|\tau_j^n|\le\eta\}}.
\end{aligned} \tag{4.9}
$$

Now, let the joint distribution of τ_k^n and τ_j^n be denoted by $P_{k,j}^n(\cdot,\cdot)$, and denote the distribution of τ_k^n by $P_k^n(\cdot)$. Using nested expectation, when $j \neq k$ we have for the first term in the last equality of (4.9) that

$$
\begin{aligned}
&E[I_{\{\varepsilon_k \le t+\tau_k^n\}} - F(t+\tau_k^n)][I_{\{\varepsilon_j \le t+\tau_j^n\}} - F(t+\tau_j^n)]I_{\{|\tau_k^n|\le\eta\}}I_{\{|\tau_j^n|\le\eta\}}\\
=&EE\Big([I_{\{\varepsilon_k \le t+\tau_k^n\}} - F(t+\tau_k^n)]I_{\{|\tau_k^n|\le\eta\}}\\
&\qquad \times [I_{\{\varepsilon_j \le t+\tau_j^n\}} - F(t+\tau_j^n)]I_{\{|\tau_j^n|\le\eta\}}\Big|(\tau_k^n,\tau_j^n)\Big)\\
=&\int E([I_{\{\varepsilon_k \le t+a_k^n\}} - F(t+a_k^n)]I_{\{|a_k^n|\le\eta\}}\\
&\qquad \times [I_{\{\varepsilon_j \le t+a_j^n\}} - F(t+a_j^n)]I_{\{|a_j^n|\le\eta\}})dP_{k,j}^n(a_k^n,a_j^n)\\
=&\int (E[I_{\{\varepsilon_k \le t+a_k^n\}} - F(t+a_k^n)]I_{\{|a_k^n|\le\eta\}}\\
&\qquad \times E[I_{\{\varepsilon_j \le t+a_j^n\}} - F(t+a_j^n)]I_{\{|a_k^n|\le\eta\}})dP_{k,j}^n(a_k^n,a_j^n) = 0.
\end{aligned}
$$

The cross term has no contribution to the limit, inequalities (4.7) and (4.8) then lead to

$$
\begin{aligned}
&E\frac{1}{n}\sum_{j=1}^n\sum_{k=1}^n[I_{\{\varepsilon_k \le t+\tau_k^n\}} - F(t+\tau_k^n)][I_{\{\varepsilon_j \le t+\tau_j^n\}} - F(t+\tau_j^n)]I_{[|\tau_k^n|\le\eta]}I_{[|\tau_j^n|\le\eta]}\\
=&\frac{1}{n}\sum_{k=1}^n E[I_{\{\varepsilon_k \le t+\tau_k^n\}} - F(t+\tau_k^n)]^2 I_{[|\tau_k^n|\le\eta]}\\
\le&\frac{1}{n}\sum_{k=1}^n E[I_{\{\varepsilon_k \le t+\eta\}} - F(t-\eta)]^2\\
\le&F(t+\eta)-F(t-\eta)^2. \qquad (4.10.1)
\end{aligned}
$$

Treating the second term of (4.9) in a similar manner, we see that the cross terms of the double sum drop out as well. The second term, then becomes

$$
\begin{aligned}
&-E\frac{1}{n}\sum_{j=1}^n\sum_{k=1}^n[I_{\{\varepsilon_k \le t+\tau_k^n\}} - F(t+\tau_k^n)][I_{\{\varepsilon_j \le t\}} - F(t)]I_{[|\tau_k^n|\le\eta]}I_{[|\tau_j^n|\le\eta]}\\
=&-\frac{1}{n}\sum_{k=1}^n E[I_{\{\varepsilon_k \le t+\tau_k^n\}} - F(t+\tau_k^n)][I_{\{\varepsilon_k \le t\}} - F(t)]I_{[|\tau_k^n|\le\eta]}\\
=&-\frac{1}{n}\sum_{k=1}^n \int_{-\eta}^{\eta} E[I_{\{\varepsilon_k \le t+a\}} - F(t+a)][I_{\{\varepsilon_k \le t\}} - F(t)]\,dP_k^n(a)\\
\le&\frac{1}{n}\sum_{k=1}^n \int_{-\eta}^{\eta} E[-I_{\{\varepsilon_k \le t-\eta\}}I_{\{\varepsilon_k \le t\}}\\
&\qquad + I_{\{\varepsilon_k \le t+\eta\}}F(t) + F(t+\eta)I_{\{\varepsilon_k \le t\}} - F(t-\eta)F(t)]dP_k^n(a)\\
\le&\frac{1}{n}\sum_{k=1}^n(-F(t)+2F(t+\eta)F(t)-F(t-\eta)F(t))\,P[|\tau_k^n|\le\eta]\\
\le&(-F(t)+2F(t+\eta)F(t)-F(t-\eta)F(t))\,P[M|q_n-q^*|\le\eta], \qquad (4.10.2)
\end{aligned}
$$

where $M = \sup_{x,q} |\partial f/\partial q|$ as above.

The third term of (4.9) produces an identical bound. Applying the above technique to the last term of (4.9), we find

$$
\begin{aligned}
&E\frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}[I_{\{\varepsilon_k\le t\}} - F(t)][I_{\{\varepsilon_j\le t\}} - F(t)]I_{[|\tau_k^n|\le\eta]}I_{[|\tau_j^n|\le\eta]}\\
=&\frac{1}{n}\sum_{k=1}^{n}E[I_{\{\varepsilon_k\le t\}} - F(t)]^2 I_{[|\tau_k^n|\le\eta]}\\
\le&F(t) - F^2(t). \qquad (4.10.3)
\end{aligned}
$$

Combining the bounds in (4.10.1), (4.10.2) and (4.10.3), we have for (4.9) that

$$
\begin{aligned}
&E\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\Big((I_{\{\varepsilon_k\le t+\tau_k^n\}} - F(t+\tau_k^n)) - (I_{\{\varepsilon_k\le t\}} - F(t))\Big)\right)^2\\
\le&F(t+\eta) - F^2(t-\eta) + F(t) - F^2(t)\\
&\qquad + 2\left(-F(t) + 2F(t+\eta)F(t) - F(t-\eta)F(t)\right)P[M|q_n - q^*| \le \eta].
\end{aligned}
$$

Since $P[M|q_n - q^*| \le \eta] \to 1$ for any $\eta > 0$, and since F is continuous, the desired result follows. □

In view of Lemma 4.1,

$$
\begin{aligned}
\hat{W}_n(t) &= \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\Big([I_{\{\varepsilon_k\le t+\tau_k^n\}} - F(t+\tau_k^n))] - [F(t+\tau_k^n)) - F(t)]\Big)\\
&= \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left([I_{\{\varepsilon_k\le t\}}] - F(t)\right) + \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left(F(t+\tau_k^n)) - F(t)\right) + o(1)\\
&= W_n(t) + \frac{1}{\sqrt{n}}\sum_{k=1}^{n}[F(t+\tau_k^n) - F(t)] + o(1),
\end{aligned}
\qquad (4.11)
$$

where $o(1)\xrightarrow{n}0$ in probability, and where W_n is the scaled process associated with the "true" errors:

$$W_n(t) = \sqrt{n}\big(F_n(t) - F(t)\big).$$

Examining the second term on the right-hand side of (4.11) closely, by virtue of the strong convergence of q_n and $\bar{q}_n$, we note that

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{k=1}^{n}[F(t+\tau_k^n) - F(t)] &= \frac{1}{n}\sum_{k=1}^{n}\pi_\varepsilon(\tilde{t})\frac{\partial f'(x_k,\bar{q}_n)}{\partial q}\sqrt{n}(q_n - q^*)\\
&= \frac{1}{n}\sum_{k=1}^{n}\pi_\varepsilon(\tilde{t})\frac{\partial f'(x_k,q^*)}{\partial q}\sqrt{n}(q_n - q^*) + o(1)\\
&= \left(\int_X \frac{\partial f'(x,q^*)}{\partial q}d\mu\right)\pi_\varepsilon(t)\sqrt{n}(q_n - q^*) + o(1)\\
&= \nu_n(t) + o(1)
\end{aligned}
$$

where $\tilde{t}$ takes values between t and $t + \tau_k^n$ and $o(1) \xrightarrow{n} 0$ in probability.

As a result, we have that

$$\hat{W}_n(t) = W_n(t) + \nu_n(t) + o(1) \tag{4.12}$$

where $o(1) \xrightarrow{n} 0$ in probability. Our next lemma characterizes the limiting behavior of the terms in the above expression for $\hat{W}_n$.

Lemma 4.2. *Under the conditions (A1)-(A6), the following hold:*

(i) $W_n(\cdot)$ *converges weakly to a stretched Brownian bridge (cf. [5]) process* $W^0(\cdot)$, *with mean* $EW^0(t) = 0$ *and covariance*

$$EW^0(t_1)W^0(t_2) = \min(F(t_1, q^*), F(t_2, q^*)) - F(t_1, q^*)F(t_2, q^*).$$

(ii) $\nu_n(t)$ *converges weakly to a Gaussian process with* $E\nu(t) = 0$ *and covariance*

$$E\nu(t_1)\nu(t_2) = 2\sigma^2 \pi_\varepsilon(t_1)\pi_\varepsilon(t_2) \times \left(\int_X \frac{\partial f'(x, q^*)}{\partial q} d\mu(x) \mathcal{T}^{-1} \int_X \frac{\partial f(x, q^*)}{\partial q} d\mu(x) \right).$$

Proof: The proof of (i) can be found for example in [5], which uses a standard argument of weak convergence. As for (ii), we note that due to the fact only $\pi_\varepsilon(\cdot)$ depends on t, it is easy to obtain the convergence of finite dimensional distributions of $\nu_n(\cdot)$ to that of $\nu(\cdot)$ in view of the asymptotic normality of $\sqrt{n}(q_n - q^*)$, (cf. Proposition 2.1 (3)). Likewise, it is easily seen that $\{\nu_n(\cdot)\}$ is tight. In addition, $\nu(\cdot)$ must have continuous paths w.p.1 by the continuity of $\pi_\varepsilon(\cdot)$. Furthermore, simple calculations yield the mean and covariance function as stated. □

The following lemma [9, Theorem 3.8.6] (see also [13, Theorem 3.3.3]) will be used to prove the tightness of the sequence $\{\hat{W}_n(\cdot)\}$.

Lemma 4.3. *Suppose that* $y_n(\cdot)$ *is a sequence of processes with paths in* $D(-\infty, \infty)$ *satisfying*

$$\lim_{\kappa \to \infty} \overline{\lim_n} P \left(\sup_{-T \le t \le T} |y_n(t)| \ge \kappa \right) = 0 \text{ for each } T < \infty. \tag{4.13}$$

Let $\mathcal{F}_t^n = \sigma\{y_n(s); s \le t\}$ *and* M_T^n *be the collection of* $\mathcal{F}_t^n$ *stopping times* τ *such that* $|\tau| \le T$ *w.p.1. If*

$$\lim_{\delta \to 0} \overline{\lim_{n \to \infty}} \sup_{\tau \in M_T^n} E \min\{1, |y_n(\tau + \delta) - y_n(\tau)|^2\} = 0, \tag{4.14}$$

then $\{y_n(\cdot)\}$ *is tight in* $D(-\infty, \infty)$. □

Denote $Z_n(\cdot) = W_n(\cdot) + \nu_n(\cdot)$. Then $\hat{W}_n(\cdot) = Z_n(\cdot) + o(1)$ with $o(1) \xrightarrow{n} 0$ in probability. To prove the tightness of $\{\hat{W}_n(\cdot)\}$ is equivalent to establishing the tightness of the sequence $\{Z_n(\cdot)\}$.

Lemma 4.4. *Assume that the conditions of Lemma 4.1 are satisfied. Then* $\{Z_n(\cdot)\}$ *is tight in* $D(-\infty,\infty)$ *and hence* $\{\hat{W}_n(\cdot)\}$ *is tight in* $D(-\infty,\infty)$. *Moreover, the limit of any convergent subsequence has continuous paths w.p.1.*

Proof: We shall verify that (4.13) and (4.14) hold for the sequence $\{Z_n(\cdot)\}$. First, notice that

$$|Z_n(t)| \le |W_n(t)| + |\nu_n(t)|.$$

By virtue of Lemma 4.2, both $\{W_n(\cdot)\}$ and $\{\nu_n(\cdot)\}$ are tight. By using the tightness of $\{W_n(\cdot)\}$ and $\{\nu_n(\cdot)\}$, (4.13) is easily verified.

Next, for each $T<\infty$ and $-T \le t \le T$, let $\mathcal{F}_t^n = \sigma\{Z_n(s);\ s \le t\}$ and M_T^n denote the collection of $\mathcal{F}_t^n$ stopping times τ with $|\tau| \le T$ w.p.1. We have that

$$\begin{aligned}
&\lim_{\delta\to 0}\overline{\lim_{n\to\infty}}\sup_{\tau\in M_T^n} E|Z_n(\tau+\delta)-Z_n(\tau)|^2\\
\le& 2\lim_{\delta\to 0}\overline{\lim_{n\to\infty}}\sup_{\tau\in M_T^n} E\left(E_\tau|W_n(\tau+\delta)-W_n(\tau)|^2 + E_\tau|\nu_n(\tau+\delta)-\nu_n(\tau)|^2\right),
\end{aligned} \tag{4.15}$$

where E_τ denotes the conditioning with respect to the stopping time τ.

Examine the second term on the right-hand side of (4.15). Since $\pi_\varepsilon(\cdot)$ is continuous, it is uniformly continuous on $[-T,T]$. It then follows that for some $K>0$,

$$\begin{aligned}
&\lim_{\delta\to 0}\overline{\lim_{n\to\infty}}\sup_{\tau\in M_T^n} EE_\tau|\nu_n(\tau+\delta)-\nu_n(\tau)|^2\\
=&\lim_{\delta\to 0}\overline{\lim_{n\to\infty}}\sup_{\tau\in M_T^n} EE_\tau(\pi_\varepsilon(\tau+\delta)-\pi_\varepsilon(\tau))^2\\
&\qquad\times\int_X \frac{\partial f'(x,q^*)}{\partial q}d\mu[n(q_n-q^*)(q_n-q^*)']\int_X \frac{\partial f(x,q^*)}{\partial q}d\mu\\
\le& K\lim_{\delta\to 0}\sup_{\tau\in M_T^n} E(\pi_\varepsilon(\tau+\delta)-\pi_\varepsilon(\tau))^2\\
=&0.
\end{aligned}$$

The first term on the right-hand side of (4.15) also goes to zero. This can be seen by using similar argument as in the proof of Lemma 4.1 and some detailed estimates.

Now by virtue of Lemma 4.2, the limit of $W_n(\cdot)$ has continuous paths with probability one since it is a Brownian bridge. The continuity of $\pi_\varepsilon(\cdot)$ also implies that the limit of $\nu_n(\cdot)$ has continuous paths with probability one. This completes the proof of this lemma. □

Lemma 4.5. *$Z_n(\cdot) \Rightarrow W(\cdot)$ (the symbol '$\Rightarrow$' denotes that 'converges weakly'), which is a Gaussian element with mean*

$$EW(t) = 0 \tag{4.16}$$

and covariance function

$$\begin{aligned}
EW(t_1)W(t_2) &= \min\Big(F(t_1), F(t_2)\Big) - F(t_1)F(t_2) \\
&+ 2\pi_\varepsilon(t_2)\Bigg(\int_X \frac{\partial f'(x,q^*)}{\partial q} d\mu \mathcal{T}^{-1} \int_X \frac{\partial f(x,q^*)}{\partial q} d\mu \\
&\qquad \times \Big(E\varepsilon_1\big(I_{\{\varepsilon_1 \le t_1\}} - F(t_1,q^*)\big)\Big)\Bigg) \\
&+ 2\pi_\varepsilon(t_1)\Bigg(\int_X \frac{\partial f'(x,q^*)}{\partial q} d\mu \mathcal{T}^{-1} \int_X \frac{\partial f(x,q^*)}{\partial q} d\mu \\
&\qquad \times \Big(E\varepsilon_1\big(I_{\{\varepsilon_1 \le t_2\}} - F(t_2,q^*)\big)\Big)\Bigg) \\
&+ 2\sigma^2\pi_\varepsilon(t_1)\pi_\varepsilon(t_2)\Big(\int_X \frac{\partial f'(x,q^*)}{\partial q} d\mu \mathcal{T}^{-1} \int_X \frac{\partial f(x,q^*)}{\partial q} d\mu\Big).
\end{aligned} \tag{4.17}$$

Proof: First note that one of the crucial observations made in [3, 10] is that

$$\sqrt{n}(q_n - q^*) = \frac{2\mathcal{T}^{-1}}{\sqrt{n}} \sum_{k=1}^{n} \frac{\partial f(x_k, q^*)}{\partial q} \varepsilon_k + o(1),$$

where $o(1) \xrightarrow{n} 0$ in probability. In view of the structure of $W_n(\cdot)$ and $\nu_n(\cdot)$, together with the above observation, for each $T < \infty$ and each $-T \le t \le T$,

$$\begin{aligned}
&Z_n(t) = W_n(t) + \nu_n(t) \\
=&\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \left([I_{\{\varepsilon_k \le t\}} - F(t)] + 2\pi_\varepsilon(t) \int_X \frac{\partial f'(x,q^*)}{\partial q} d\mu \mathcal{T}^{-1} \frac{\partial f(x_k,q^*)}{\partial q} \varepsilon_k \right) \\
&\qquad + o(1) \\
=&\tilde{Z}_n(t) + o(1),
\end{aligned}$$

where $o(1) \xrightarrow{n} 0$ in probability. Thus the limit problem of $\{Z_n(\cdot)\}$ reduces to that of $\{\tilde{Z}_n(\cdot)\}$. By Lemma 4.4, $\{\tilde{Z}_n(\cdot)\}$ is tight. Pick out any convergent subsequence and denote the limit by $W(\cdot)$. $\tilde{Z}_n(t)$ can be thought of as a normalized sum of independent random variables. There are at most a countable number of points at which

$$P\{W(t) \neq W(t^-)\} > 0.$$

Let T_P denote the complement of this set. Using an argument as in Linderberger-Lévy central limit theorem, it can be shown that for every finite set $\{t_1, \ldots, t_k\} \subset T_P$,

$$(\tilde{Z}_n(t_1), \cdots, \tilde{Z}_n(t_k)) \Rightarrow (W(t_1), \cdots, W(t_k)) \text{ in distribution,}$$

i.e., the finite dimensional distributions converge. Thus, $\tilde{Z}_n(\cdot) \Rightarrow W(\cdot)$ and hence $Z_n(\cdot) \Rightarrow W(\cdot)$ (cf. [5, Theorem 15.6], [9, Theorem 3.7.8]). Finally, direct computations yield the mean (4.16) and covariance (4.17). □

Putting things together, Lemma 4.4 and Lemma 4.5 yield:

Theorem 4.6. *$\hat{W}_n(\cdot)$ converges weakly to a Gaussian element $W(\cdot)$ such that the mean and covariance functions of $W(\cdot)$ are given by (4.16) and (4.17), respectively.* □

5. Two sample problems

We consider the two sample problem. Let us take two samples of observations:

$$Y_k = f(x_k, q^*) + \varepsilon_k, \quad 1 \le k \le n \tag{5.1}$$

$$\tilde{Y}_j = f(\tilde{x}_j, q^*) + \tilde{\varepsilon}_j, \quad 1 \le j \le m, \tag{5.2}$$

with sample size n and m, respectively.

For these samples, using the long-run average cost as before, we construct two sequences of least squares estimates, $\{q_n\}$ and $\{\tilde{q}_m\}$, respectively.

Previously, an empirical stochastic process for two sample problem was considered in [6] and [15]. The connection between the two sample problem and Chernoff-Savage theorem were discussed. Their results cannot be applied to our problem due to the fact that in our case the estimated parameters q_n and $\tilde{q}_m$, rather than the true parameter q^*, are used.

(A7) The assumptions (A1)-(A5) hold for both samples. For the second sample, replace x_k, ε_k, etc. by $\tilde{x}_k$, $\tilde{\varepsilon}_k$, respectively.

Proposition 5.1. *Under (A7), the following claims hold:*

(1) $q_n \xrightarrow{n} q^*$ *w.p.1,* $\tilde{q}_m \xrightarrow{m} q^*$ *w.p.1;*

(2) $\sqrt{n}(q_n - q^*)$ *is asymptotically normally distributed as* $n \to \infty$*, and*

(3) $\sqrt{m}(\tilde{q}_m - q^*)$ *is asymptotically normally distributed as* $m \to \infty$*;*

$$q_n - q^* = \frac{2\mathcal{I}^{-1}}{n} \sum_{k=1}^{n} \frac{\partial f(x_k, q^*)}{\partial q} \varepsilon_k + o_n(1),$$
$$\tilde{q}_m - q^* = \frac{2\mathcal{I}^{-1}}{m} \sum_{j=1}^{m} \frac{\partial f(\tilde{x}_j, q^*)}{\partial q} \tilde{\varepsilon}_j + o_m(1),$$

where $\sqrt{n} o_n(1) \xrightarrow{n} 0$ *and* $\sqrt{m} o_m(1) \xrightarrow{m} 0$ *both in probability.* □

Define ε_k^n, $\hat{F}_n(t)$, $F_n(t)$, $F(t)$ and $\hat{W}_n(t)$ etc. as before and define:

$$\begin{aligned}
&\tilde{\varepsilon}_j^m = \tilde{Y}_j - f(\tilde{x}_j, \tilde{q}_m), \\
&\hat{G}_m(t, \tilde{q}_m) = \frac{1}{m} \sum_{j=1}^{m} I_{\{\tilde{\varepsilon}_j^m \le t\}} \\
&G_m(t) = \frac{1}{m} \sum_{j=1}^{m} I_{\{\tilde{\varepsilon}_j \le t\}} \\
&G(t, q^*) = P(\tilde{\varepsilon}_j \le t) \\
&\hat{B}_m(t) = \sqrt{m}(\hat{G}_m(t, \tilde{q}_m) - G(t, q^*)).
\end{aligned} \tag{5.3}$$

(A8) The sequences $\{\varepsilon_k\}$ and $\{\tilde{\varepsilon}_j\}$ are independent and each has a common continuous density function $\pi_\varepsilon(\cdot)$ and $\tilde{\pi}_{\tilde{\varepsilon}}(\cdot)$, respectively, and $E\varepsilon_1^4 < \infty$, $E\tilde{\varepsilon}_1^4 < \infty$. Furthermore, $n/m \xrightarrow{n,m} c_0 > 0$ a fixed constant.

Theorem 5.2. *(Glivenko-Cantelli Analogue) If (A7) and (A8) hold, then*

$$\sup_t |\hat{F}_n(t, q_n) - F(t, q^*)| \xrightarrow{n} 0 \text{ w.p.1},$$
$$\sup_t |\hat{G}_m(t, \tilde{q}_m) - G(t, q^*)| \xrightarrow{m} 0 \text{ w.p.1}.$$

□

Remark: For the convergence alone, the full capacity of (A8) is not needed. All we need is the continuity of $F(\cdot, q^*)$ and $G(\cdot, q^*)$. The proof of this theorem is essentially the same as the proof of Theorem 3.1.

We wish to study the normalized sequence

$$\begin{aligned}
\hat{C}_{n,m}(t) &= \sqrt{\frac{mn}{n+m}} \left([\hat{F}_n(t, q_n) - F(t, q^*)] - [\hat{G}_m(t, \tilde{q}_m) - G(t, q^*)] \right) \\
&= \sqrt{\frac{1}{n+m}} \left(\sqrt{m} \hat{W}_n(t) - \sqrt{n} \hat{B}_m(t) \right)
\end{aligned} \tag{5.4}$$

as $n \to \infty$ and $m \to \infty$.

Rewrite the expression in (5.4) as

$$\begin{aligned}\hat{C}_{n,m}(t) &= \left(\frac{1}{\frac{n}{m}+1}\right)^{1/2} \hat{W}_n(t) - \left(\frac{1}{\frac{m}{n}+1}\right)^{1/2} \hat{B}_m(t) \\ &= \left(\frac{1}{c_0+1}\right)^{1/2} \hat{W}_n(t) - \left(\frac{1}{c_0^{-1}+1}\right)^{1/2} \hat{B}_m(t) + o_{nm}(1),\end{aligned} \tag{5.5}$$

where $o_{nm}(1) \xrightarrow{n,m} 0$ in probability. The last line above follows from the fact $n/m \to c_0$ and the weak convergence of $\hat{W}_n(\cdot)$ and $\hat{B}_m(\cdot)$. Using similar argument as in Section 4, we can then derive the following asymptotic normality result for the two sample problem.

Theorem 5.3. *Assuming (A7) and (A8), as $n \to \infty$ and $m \to \infty$, $\hat{C}_{n,m}(\cdot)$ converges weakly to $C(\cdot)$ in $D(-\infty, \infty)$ where $C(\cdot)$ is a Gaussian element with mean*

$$E\hat{C}(t) = 0 \tag{5.6}$$

and covariance function

$$\begin{aligned}EC(t_1)C(t_2) = \left(\frac{1}{c_0+1}\right) \Bigg\{ &\min\left(F(t_1,q^*), F(t_2,q^*)\right) - F(t_1,q^*)F(t_2,q^*) \\ &+ 2\pi_\varepsilon(t_2) \\ &\left(\int_X \frac{\partial f'(x,q^*)}{\partial q} d\mu \mathcal{T}^{-1} \int_X \frac{\partial f(x,q^*)}{\partial q} d\mu \left(E\varepsilon_1 \left(I_{\{\varepsilon_1 \le t_1\}} - F(t_1,q^*)\right)\right)\right) \\ &+ 2\pi_\varepsilon(t_1) \\ &\left(\int_X \frac{\partial f'(x,q^*)}{\partial q} d\mu \mathcal{T}^{-1} \int_X \frac{\partial f(x,q^*)}{\partial q} d\mu \left(E\varepsilon_1 \left(I_{\{\varepsilon_1 \le t_2\}} - F(t_2,q^*)\right)\right)\right) \\ &+ 2\sigma^2 \pi_\varepsilon(t_1)\pi_\varepsilon(t_2) \left(\int_X \frac{\partial f'(x,q^*)}{\partial q} d\mu \mathcal{T}^{-1} \int_X \frac{\partial f(x,q^*)}{\partial q} d\mu\right) \Bigg\} \\ &+ \left(\frac{1}{c_0^{-1}+1}\right) \Bigg\{ \min\left(G(t_1,q^*), G(t_2,q^*)\right) - G(t_1,q^*)G(t_2,q^*) \\ &+ 2\tilde{\pi}_{\tilde{\varepsilon}}(t_2) \\ &\left(\int_X \frac{\partial f'(\tilde{x},q^*)}{\partial q} d\mu \tilde{\mathcal{T}}^{-1} \int_X \frac{\partial f(\tilde{x},q^*)}{\partial q} d\mu \left(E\tilde{\varepsilon}_1 \left(I_{\{\tilde{\varepsilon}_1 \le t_1\}} - G(t_1,q^*)\right)\right)\right) \\ &+ 2\tilde{\pi}_{\tilde{\varepsilon}}(t_1) \\ &\left(\int_X \frac{\partial f'(\tilde{x},q^*)}{\partial q} d\mu \tilde{\mathcal{T}}^{-1} \int_X \frac{\partial f(\tilde{x},q^*)}{\partial q} d\mu \left(E\tilde{\varepsilon}_1 \left(I_{\{\tilde{\varepsilon}_1 \le t_2\}} - G(t_2,q^*)\right)\right)\right) \\ &+ 2\tilde{\sigma}^2 \tilde{\pi}_{\tilde{\varepsilon}}(t_1)\tilde{\pi}_{\tilde{\varepsilon}}(t_2) \left(\int_X \frac{\partial f'(\tilde{x},q^*)}{\partial q} d\mu\ \tilde{\mathcal{T}}^{-1} \int_X \frac{\partial f(\tilde{x},q^*)}{\partial q} d\mu\right) \Bigg\}.\end{aligned} \tag{5.7}$$

Proof: Consider the pair $(\hat{W}_n(\cdot), \hat{B}_m(\cdot))$. It is clear that $(\hat{W}_n(\cdot), \hat{B}_m(\cdot))$ is in $D^2(-\infty, \infty)$. We have shown that $\hat{W}_n(\cdot) \Rightarrow W(\cdot)$ as $n \to \infty$ and $\hat{B}_m(\cdot) \Rightarrow B(\cdot)$ as $m \to \infty$. Owing to Theorem 3.2 in [5] (cf. also [5, pp. 26]), the independence of $\hat{W}_n(\cdot)$ and $\hat{B}_m(\cdot)$ implies that

$$(\hat{W}_n(\cdot), \hat{B}_m(\cdot)) \Rightarrow (W(\cdot), B(\cdot)) \text{ as } n \to \infty,\ m \to \infty.$$

Now, (5.5) together with the Cramér-Wold device (cf. [5, pp. 48]) yields that

$$\hat{C}_{n,m} \Rightarrow C(\cdot) = \left(\frac{1}{c_0+1}\right)^{1/2} W(\cdot) - \left(\frac{1}{c_0^{-1}+1}\right)^{1/2} B(\cdot).$$

It is easily seen that $EC(t) = 0$. Upon using the independence and the expression (5.5) again, direct computations lead to the representation of the covariance of $C(\cdot)$. □

Remark: The above theorem can be readily generalized to many sample problems. Suppose we take l-samples with independent observations

$$Y_k^i = f(x_k^i, q^*) + \varepsilon_k^i,\ 1 \le k \le n_i,\ i = 1, \cdots, l, \tag{5.8}$$

where the errors are distributed as F^i.

For these samples, construct the estimates $\{q_{n_i}^i\}$, $i = 1, \cdots, l$ and define the approximate errors

$$\varepsilon_k^{n_i} = Y_k^i - f(x_k^i, q_{n_i}^i). \tag{5.9}$$

Then we may examine the empirical distributions from these samples. Let

$$\begin{aligned} \hat{F}_{n_i}^i(t) &= \frac{1}{n_i}\sum_{k=1}^{n_i} I_{\{\varepsilon_k^{n_i} \le t\}} \\ \hat{W}_{n_i}^i(t) &= \sqrt{n_i}\left(\hat{F}_{n_i}^i(t) - F(t)\right) \\ \hat{C}_N(t) &= \sum_{i=1}^{l}(-1)^{i-1}\left(\frac{N_i}{(l-1)N}\right)^{\frac{1}{2}} \hat{W}_{n_i}^i(t) \end{aligned} \tag{5.10}$$

where $N = \sum_{i=1}^l n_i$ and $N_i = \sum_{j \ne i} n_j$. Our one sample result implies that $\hat{W}_{n_i}^i(\cdot) \Rightarrow W^i(\cdot)$ as $n_i \to \infty$. Assuming that $N_i/N \to \lambda_i > 0$, $\sum_{i=1}^l \lambda_i = l-1$. A straightforward extension of Theorem 5.3 yields that $\hat{C}_N(\cdot) \Rightarrow C(\cdot)$ as $n_i \to \infty$, for $i = 1, \cdots, l$, with $EW(t) = 0$ and the covariance function equal to a linear combination of that of $W^i(\cdot)$, $i = 1, \cdots, l$.

6. Concluding remarks

We have concentrated on the convergence and the rate of convergence of empirical processes of residuals from nonlinear least squares estimation problems. These processes can be used to answer many statistical questions of interest in application problems. We are particularly interested in testing for normality of the measurement errors and in determining whether or not collections of errors are identically distributed.

One of the difficulties with the above results is the nature of the limiting Gaussian process. The covariance depends on the unknown parameter, and it also depends on derivatives of the model function with respect to the parameter. In the applications mentioned in the Introduction, the model function is (a function of) the solution of a partial differential equation, so that approximating the covariance of $\hat{W}_n$ is quite involved.

We have begun studying the use of bootstrap methods for these problems. In bootstrapping, one builds "bootstrap" data sets using q_n and errors which are sampled from the empirical distribution. We have performed a wide range of computational experiments, which are reported in [12]. The results contained therein are very promising. We feel that empirical distribution techniques can provide important tools for inference problems in fitting distributed parameter models to observed data.

References

[1] H. T. Banks, L. Botsford, F. Kappel, and C. Wang, Estimation of Growth and Survival in Size Structured Cohort Data: an Application to Larval Striped Bass (*Morone Saxatilis*), *J. Math. Bio.* to appear.

[2] H. T. Banks and B. G. Fitzpatrick, Inverse Problems for Distributed Systems: Statistical Tests and ANOVA, in *Proceedings of the International Symposium on Mathematical Approaches to Environmental and Ecological Problems*, Lecture Notes in Biomath. **81**, Springer-Verlag, New York, 1989.

[3] H.T. Banks and B.G. Fitzpatrick, Statistical tests for model comparison in parameter estimation problems for distributed systems, *J. Math. Bio.* **28** (1990), 501-527.

[4] Banks, H. T., and K. Kunisch. *Estimation Techniques for Distributed Parameter Systems,* Birkhäuser, Boston, 1989.

[5] P. Billingsley, *Convergence of Probability Measure*, Wiley, New York, 1968.

[6] R.M. Dudley, A course on empirical processes, *Lecture Notes Math.* 1097, 1-133, Springer-Verlag, New York, 1984.

[7] J. Durbin, Weak convergence of the sample distribution function when parameters are estimated, *Ann. Statist.* **1** (1973), 279-290.

[8] J. Durbin, Kolmogorov-Smirnov tests when parameters are estimated, *Lecture Notes Math.* 566, 33-44, Springer-Verlag, New York, 1976.

[9] S.N. Ethier and T.G. Kurtz, *Markov Processes, Characterization and Convergence*, Wiley, New York, 1986.

[10] B. G. Fitzpatrick, Statistical methods in parameter identification and model selection, Ph. D. Dissertation, Division of Applied Mathematics, Brown University, Providence, RI., 1988.

[11] B. G. Fitzpatrick, Bayesian Analysis in Inverse Problems, *Inverse Problems*, **7** (1991), 675-702.

[12] B. G. Fitzpatrick and G. Yin, Bootstrap methods for inference in least squares identification problems, to appear in *Identification and Control in Systems Governed by PDEs*, H.T. Banks, R. Fabiano, and K. Ito, Ed., SIAM, 1993.

[13] H. Kushner, *Approximation and Weak Convergence Methods for Random Processes, with applications to Stochastic Systems Theory*, MIT Press, 1984.

[14] T. Lindvall, Weak convergence of probability measures and random functions in the function space $D[0,\infty)$, *J. Appl. Probab.* **10** (1973), 109-121.

[15] R. Pyke and G.R. Shorack, Weak convergence of two-sample empirical process and a new approach to Chernoff-Savage theorems, *Ann. Math. Statist.* **39** (1968), 755-771.

[16] G. Yin and B.G. Fitzpatrick, On invariance principles for distributed parameter identification algorithms, *Informatica* **3** (1992), 98-118.