2002

The Analogy Between Statistical Equivalence and Stochastic Strong Limit Theorems

Lawrence D. Brown
University of Pennsylvania

The Analogy Between Statistical Equivalence and Stochastic Strong Limit Theorems

Abstract
A classical limit theorem of stochastic process theory concerns the sample cumulative distribution function (CDF) from independent random variables. If the variables are uniformly distributed then these centered CDFs converge in a suitable sense to the sample paths of a Brownian Bridge. The so-called Hungarian construction of Komlos, Major and Tusnady provides a strong form of this result. In this construction the CDFs and the Brownian Bridge sample paths are coupled through an appropriate representation of each on the same measurable space, and the convergence is uniform at a suitable rate.

Within the last decade several asymptotic statistical-equivalence theorems for nonparametric problems have been proven, beginning with Brown and Low (1996) and Nussbaum (1996). These theorems consider two sequences of statistical problems. The parameter spaces for these problems are infinite-dimensional; hence the problems are called nonparametric. The approach here to statistical-equivalence is firmly rooted within the asymptotic statistical theory created by L. Le Cam but in some respects goes beyond earlier results.

This paper contains a survey of some of these statistical-equivalence results. It also demonstrates the analogy between these results and those from the coupling method for proving stochastic process limit theorems. These two classes of theorems possess a strong inter-relationship, and technical methods from each domain can profitably be employed in the other. Results in a recent paper by Carter, Low, Zhang and myself will be described from this perspective.

Disciplines
Statistics and Probability

The Analogy Between Statistical Equivalence and Stochastic Strong Limit Theorems. [1]

Lawrence D. Brown[2]

Statistics Department
Wharton School, University of Pennsylvania
Philadelphia, PA 19104-6302
USA
lbrown@wharton.upenn.edu

Abstract

A classical limit theorem of stochastic process theory concerns the sample cumulative distribution function (CDF) from independent random variables. If the variables are uniformly distributed then these centered CDFs converge in a suitable sense to the sample paths of a Brownian Bridge. The so-called Hungarian construction of Komlos, Major and Tusnady provides a strong form of this result. In this construction the CDFs and the Brownian Bridge sample paths are coupled through an appropriate representation of each on the same measurable space, and the convergence is uniform at a suitable rate.

Within the last decade several asymptotic statistical-equivalence theorems for nonparametric problems have been proven, beginning with Brown and Low (1996) and Nussbaum (1996). These theorems consider two sequences of statistical problems. The parameter spaces for these problems are infinite-dimensional; hence the problems are called nonparametric. The approach here to statistical-equivalence is firmly rooted within the asymptotic statistical theory created by L. Le Cam but in some respects goes beyond earlier results.

This paper contains a survey of some of these statistical-equivalence results. It also demonstrates the analogy between these results and those from the coupling method for proving stochastic process limit theorems. These two classes of theorems possess a strong inter-relationship, and technical methods from each domain can profitably be employed in the other. Results in a recent paper by Carter, Low, Zhang and myself will be described from this perspective.

1. Much of this material was contained in an invited address at the International Congress of Mathematicians, Beijing, Aug. 2002. Revised Sept 4, 2002.
2. Research supported in part by NSF-Division of Mathematical Sciences

1. Probability Setting

1.1 Background

Let F be the CDF for a probability on [0,1];. F abs. cont., with

$$f(x) \triangleq \frac{\partial F}{\partial x} \ on \ [0,1].$$

Let X_1, …, X_n iid from F. $\hat{F}_n$ denotes the sample CDF,

$$\hat{F}_n(x) \triangleq \frac{1}{n}\sum_{j=1}^{n} \mathbf{I}_{[0,x]}\left(X_j\right).$$

Let $\hat{Z}_n$ denote the corresponding sample "bridge",

(1) $$\hat{Z}_n(x) \triangleq \hat{F}_n(x) - F(x).$$

Let W(t) denote the standard Wiener process on [0,1] and let $\hat{W}_n$ denote the white noise process with drift f and local variance ${f(t)}/{n}$. Thus $\hat{W}_n$ solves

$$d\hat{W}_n(t) = f(t)dt + \sqrt{\frac{f(t)}{n}}dW\left(t\right).$$

An alternate description of $\hat{W}_n$ is that it is the Gaussian process with mean F(t) and independent increments having

$$\operatorname{var}\left(\hat{W}_n(t) - \hat{W}_n(s)\right) = \frac{1}{n}\left(F(t) - F(s)\right), \ for \ 0 \le s < t \le 1.$$

The analog of $\hat{Z}_n$ is the Gaussian Bridge, defined by

$$\hat{B}_n(t) = \frac{\hat{W}_n(t)}{\hat{W}_n(1)} - F(t).$$

$\hat{B}_n$ can alternately be described as the centered Gaussian process with covariance

$$\operatorname{cov}\left(\hat{B}_n(s), \hat{B}_n(t)\right) = \min\left\{F(s), F(t)\right\} - F(s)F(t).$$

There are various ways of describing the stochastic similarity between $\hat{Z}_n$ and $\hat{B}_n$. For example Komlos, Major, and Tusnady (1975, 1976) proved a result of the form

Theorem (KMT): Given any absolutely continuous F $\{X_1,\ldots,X_n\}$ can be defined on a probability space on which $\hat{B}_n$ can also be defined as a (randomized) function of $\{X_1,\ldots,X_n\}$. This can be done in such a way that $\hat{B}_n$ has the Gaussian Bridge distribution, above, and

(2) $$P_F\left(\sup_{t\in[0,1]} \sqrt{n}\left|\hat{Z}_n(t) - \hat{B}_n(t)\right| > a_n\right) \le c.$$

Here c>0 and $\{a_n\}$ are suitable positive constants with $a_n \sim \log n \big/ \sqrt{n}$. The process $\hat{B}_n$ can be constructed as a (randomized) function of $\hat{Z}_n$, that is, $\hat{B}_n(t) = Q_n\left(\hat{Z}_n(t)\right)$. It should be noted that the construction depends on knowledge of F.

NOTE: If the probability space is equipped with an auxiliary random variable, U, that is uniformly distributed on [0,1] then a randomized function such as Q_n, above, can be written in a measurable way as a non-randomized function of U and the quantity of interest; here $\hat{Z}_n$. (See Wald and Wolfowitz (1951) for the basic theorem from which this can be deduced.) This can be done in all the contexts below that involve randomized functions, and we will generally suppress mention of U in the statement of such results, as we have done above.

[Various authors, such as Csörgö and Revesz (1981) and Bretagnolle and Massart (1989) have given increasingly detailed and precise values for $\{a_n\}$ and $c=c(\{a_n\})$, and also uniform (in n) versions of (2).). These are not our focus.]

1.2 Extensions:

1. Results like the above also extend to functional versions of the process $\hat{Z}_n$. Various authors including Dudley (1978), Massart (1989) and Koltchinskii (1994) have established results of the following form.

Let $q:[0,1]\to\Re$ be of bounded variation. One can define

$$\hat{Z}_n(q) \triangleq \int q d\{\hat{F}_n - F\} = \int \left(F - \hat{F}_n\right) dq .$$

(Thus, $\hat{Z}_n(x) = \hat{Z}_n\left(\mathbf{I}_{[0,x]}\right)$). There is a similar definition for $\hat{B}_n(q)$ as a stochastic integral. (See, for example, Steele (2000).) Then the KMT theorem extends to a fairly broad, but not universal, class of functions, $\mathbb{Q}$. That is, for each F, $\hat{B}_n$ can be defined to satisfy

(3) $$P_F\left(\sup_{q\in\mathbb{Q}} \sqrt{n}\left|\hat{Z}_n(q) - \tilde{B}_n(q)\right| > a'_n\right) \le c \text{ where } \{a'_n\} \text{ depends on } \mathbb{Q}.$$

(For most classes $\mathbb{Q}$, $a'_n\sqrt{n} \big/ \log n \to \infty$ so that $a'_n >> a_n$.)

2. Bretagnolle and Massart (1989) proved a result for inhomogeneous Poisson processes that is similar to (2). Let $\{T_1,\ldots,T_N\}$ be (ordered) observations from an inhomogeneous Poisson process with cumulative intensity function nF and, correpondingly, (local) intensity nf. Note that N~Poisson(n) and conditionally given N the values of $\{T_1,\ldots,T_N\}$ are the order statistics corresponding to an iid sample from the distribution F. In this context we continue to define $\hat{F}_n(t) = n^{-1}\left\{\sum_{j=1}^{N} I_{[0,t]}(T_j)\right\}$ where the term in braces now has a Poisson distribution with mean nF(t). Also, continue to define

$\hat{Z}_n(t) \triangleq \hat{F}_n(t) - F(t)$ as in (1). (But, note that now $\hat{Z}_n(1) \sim n^{-1}\{Poiss(n) - n\}$ rather than $\hat{Z}_n(1) = 0$, w.p.1, as was the case in (1).)

Then versions of the conclusions (2) and (3) remain valid. We give an explicit statement since this result will provide a model for our later development.

Theorem (BM): Given any n and any absolutely continuous F the observations $\{T_1,\ldots,T_N\}$ of the inhomogeneous Poisson process can be defined on a probability space on which $\hat{B}_n$ can also be defined as a (randomized) function of $\{T_1,\ldots,T_N\}$. This can be done in such a way that $\hat{B}_n$ has the Gaussian Bridge distribution, above, and

(4) $$P_F\left(\sup_{t\in[0,1]} \sqrt{n}\left|\hat{Z}_n(t) - \hat{B}_n(t)\right| > a_n\right) \le c.$$

Here c>0 and $\{a_n\}$ are suitable constants with $a_n \sim {\log n}/{\sqrt{n}}$.

Remark: Clearly there must be extensions of (4) that are valid for the Poisson case also, although we are not aware of an explicit treatment in the literature. Such a statement would conclude in this setting that

(5) $$P_F\left(\sup_{q\in\mathbb{Q}} \sqrt{n}\left|\hat{Z}_n(q) - \hat{B}_n(q)\right| > a'_n\right) \le c \text{ where } \{a'_n\} \text{ depends on } \mathbb{Q}.$$

2. Main Results

The objective is a considerably modified version of (3) and (5) that is stronger in several respects and (necessarily) different in others. We will concentrate for most of the following on the statement (5) since our results are slightly stronger and more natural in this setting. The extension of (3) appears in the statement of Theorem 2.

Expression (5) involves the target function $\hat{B}_n$. In the modified version the role of target function is instead played by $\tilde{W}_n$ which is the solution to the stochastic differential equation

(6) $$d\tilde{W}_n(t) = g(t)dt + \frac{1}{2\sqrt{n}}dW(t)$$

where $g(t) = \sqrt{f(t)}$. An alternate description of $\tilde{W}_n$ is thus

(6′) $$\tilde{W}_n = G(t) + {W(t)}/{2\sqrt{n}} \text{ where } G(t) = \int_0^t \sqrt{f(\tau)}d\tau.$$

(In the special case where f is the uniform density, f=1, then $\tilde{W}_n = W_{4n}$.)

The role of the constructed random process $\hat{Z}_n$ is now played by a differently constructed process $\tilde{Z}_n$. As before $\tilde{Z}_n$ depends only on $\{T_1,\ldots,T_N\}$, and not otherwise on their CDF,

F. This version also involves a large set, $\mathcal{F}$, of absolutely continuous CDFs. Both $\tilde{Z}_n$ and $\mathcal{F}$ will be described later in more detail. Here is a statement of the main result.

Theorem 1: Let $\mathcal{F}$ be a set of densities satisfying Assumption A or A′, below. Let $\mathbb{Q}$ be the set of all functions of bounded variation. Let $\{T_1,\ldots,T_N\}$ be an inhomogeneous Poisson process with local intensity nf. The process $\tilde{Z}_n$ can be constructed as a (randomized) function of $\{T_1,\ldots,T_N\}$, with the construction not depending on f. The Gaussian process $\tilde{W}_n$ having the distribution (6′) can also be defined (on this same space) as a (randomized) function of $\{T_1,\ldots,T_N\}$. [This construction depends on f on a set of probability at most c_n. See Remark 9), below.] This can be done in such a way that

$$(7)\qquad \sup_{f\in\mathcal{F}} P_f\left(\sup_{q\in\mathbb{Q}}\sqrt{n}\left|\tilde{Z}_n(q)-\tilde{W}_n(q)\right|>0\right)\le c_n\to 0.$$

Remarks: The conclusion (7) of this theorem should be compared to (5). In many ways (7) is stronger than (5):

1) The set $\mathbb{Q}$ has been enlarged to be the set of all functions of bounded variation.
2) In (7) the bound applies uniformly to $f\in\mathcal{F}$ where $\mathcal{F}$ is a large (but proper) subset of all densities.
3) The constant $a'_n>0$ in (2) has now been set to 0. (The multiplier, $\sqrt{n}$, that appears in (5) – and for analogy in (7) – is then no longer relevant.)
4) At the same time c = c_n is now allowed to depend on n and required to (at least) satisfy $c_n\to 0$. Faster rates of convergence to 0 can also be desirable and are obtainable from more restricted choices of $\mathcal{F}$.
5) The distribution of the target, $\tilde{W}_n$, still has a known, and easily described dependence on f, although this distribution is not that of a Brownian Bridge, as in (5). Indeed, the distribution of $\tilde{W}_n$ turns out to be at least as convenient for statistical applications as that of $\hat{B}_n$.

The tradeoffs are:

6) The construction of $\tilde{Z}_n$ based on $\{T_1,\ldots,T_N\}$ is considerably less convenient and natural than that of the empirical bridge, $\hat{Z}_n$, in the KMT theorem.
7) The construction of $\tilde{W}_n$ depends on f. This is true also in (5), but it needs to be kept in mind by contrast with the construction of $\tilde{Z}_n$ since that depends only on $\{T_1,\ldots,T_N\}$ and not otherwise on f.
8) The construction of $\tilde{W}_n$ is also not quite explicit but turns out to suffice for the statistical application of this result. This should again be noted in order to more fully explain the content of (7).
9) To be more precise, the construction of $\tilde{W}_n$ begins in an explicit fashion via the construction of $\tilde{Z}_n$. This preliminary construction then needs to be altered on an

exceptional set of probability (under f) at most c_n. This set is only implicitly defined. The only part of the construction that depends on f is the determination of the exceptional set, and the modification to be carried out on this set.

For the situation of iid variables, as in (1), a similar result holds. In this case the matching Gaussian process is again $\tilde{W}_n$, rather than the Brownian bridge of the KMT theorem.

Theorem 2: Let $\mathcal{F}$ be a set of densities satisfying Assumption B, below. Let $\mathbb{Q}$ be the set of all functions of bounded variation. Given any n and $f\in\mathcal{F}$, iid variables $\{X_1,\ldots,X_n\}$ with density f can be defined on a probability space. A process $\tilde{\tilde{Z}}_n$ can be constructed as a (randomized) function of $\{X_1,\ldots,X_n\}$, with the construction not depending on f. The Gaussian process $\tilde{W}_n$ having the distribution (6′) can also be defined (on this same space) as a (randomized) function of $\{X_1,\ldots,X_n\}$. [This construction depends on f, but only on a set of probability at most c_n.] This can be done in such a way that

(8) $$\sup_{f\in\mathcal{F}} P_f\left(\sup_{q\in\mathbb{Q}}\sqrt{n}\left|\tilde{\tilde{Z}}_n(q)-\tilde{W}_n(q)\right|>0\right)\le c_n\to 0 .$$

3. Statistical Background

3.1 Settings:

The first purpose of the discussion here is to motivate the probabilistic results described above. A second purpose is to state the result on which to base the proof of Theorem 1. The setting involves two statistical formulations:

Formulation 1 (nonparametric inhomogeneous Poisson process): The observations are **T** = $\{T_1,\ldots,T_N\}$ from the Poisson process with local intensity nf, $f\in\mathcal{F}$. The problem is "nonparametric" because the "parameter space", $\mathcal{F}$, is a very large set – too large to be smoothly parametrized by a mapping from a (subset of) a finite dimensional Euclidean space. Some conventional forms for $\mathcal{F}$ are discussed below. The statistician desires to make some sort of inference, δ, (possibly randomized) based on the observation of **T**.

Formulation 1′ (nonparametric density with random sample size): The relation between Poisson processes and density problems has been mentioned above. As a consequence, Problem 1 is equivalent to a situation where the observations are $\{X_1,...,X_N\}$ with N~Poisson(n) and $\{X_1,...,X_N\}$ the order statistics from a sample of size N from the distribution with density f.

Formulation 1″ (nonparametric density with fixed sample size): This formulation refers to the more conventional density setting in which the observations are $\{X_1,...,X_n\}$ iid with density f and n specified in advance. Clearly, Formulations 1′ and 1″ are closely related, and 1″ has been often referred to as the Poissonization of 1′. See for example van der Vaart and Wellner (1996).

Formulation 2 (white noise with drift): The statistician observes a White noise process $d\tilde{W}_n(t)$, t∈[0,1], with drift g∈$\mathcal{G}$ and local variance $1/4n$. Thus

$$d\tilde{W}_n(t) = g(t)dt + \frac{1}{2\sqrt{n}} dW(t) ,$$

and $\tilde{W}_n(t) - G(t) = W(t)/\sqrt{n}$ where $G(t) = \int_0^t g(\tau)d\tau$. Again $\mathcal{G}$ is a very large – hence "nonparametric" – parameter space. Throughout, $\mathcal{G} \subset \mathcal{L}_2 = \{g: \int g^2 < \infty\}$. In general, there need be no relation between f in Formulation 1 and g in Formulation 2, but such a relation will shortly be assumed in connection with Theorem 1, where

(9) $$g = \sqrt{f} \text{ and } \mathcal{G} = \left\{ \sqrt{f} : f \in \mathcal{F} \right\}.$$

This can alternatively be considered as a statistical formulation having parameter space $\mathcal{F}$ under the identification (9). We take this point of view in the BCLZ theorem, below.

Formulation 2′: (infinite series problem): The Poisson formulation described above is a statistically natural one. (The density problem with fixed n in Problem 1′ is even more natural.) The white noise with drift setting is statistically less familiar and less natural. But it is mathematically very convenient. One of the features that makes it so convenient is that it is equivalent to an infinite series formulation. This equivalence can be derived via any convenient orthonormal basis $\{\varphi_j\}$ of $\mathcal{L}_2$. Let $(\bullet,\bullet)$ denote the inner product in $\mathcal{L}_2$. Let $\theta_j = (\varphi_j, g)$ and $Y_j = (\varphi_j, d\tilde{W}_n)$. Then Y_j, j = 1,… are independent normal variables with mean θ_j and variance 1/n. The parameter space for this form of the problem is $\mathcal{G}^* = \left\{ \{\theta_j\} : \exists g \in \mathcal{G}, \theta_j = (\varphi_j, g), j = 1,... \right\}$.

3.2 Constructive asymptotic statistical equivalence:

Of the various statistical formulations, above, Formulation 2′ is generally the easiest to work with mathematically in order to construct asymptotically desirable statistical procedures. On the other hand, the various versions of Formulations 1 are the more frequently encountered in statistical practice. Hence it is desirable to establish a strong form of asymptotic equivalence between these two different formulations. This enables one to proceed to "solve" the statistical problem in the easier setting of Formulation 2′ and then transfer that "solution" to Formulation 1, which is (usually) the setting of greater practical importance.

Here is one definition of the strongest form of such an equivalence.

Definition (asymptotic equivalence): Let $\mathcal{P}_j^{(n)} = (\mathcal{X}_j^{(n)}, \mathcal{B}_j^{(n)}, \mathcal{F}_j^{(n)})$ j = 1,2, n = 1,2,... be two sequences of statistical problems on the same sequence of parameter spaces, $\Theta^{(n)}$. Hence, $\mathcal{F}_j^{(n)} = \left\{F_{j,\theta}^{(n)} : \theta \in \Theta^{(n)}\right\}$. Then $\mathcal{P}_1$ and $\mathcal{P}_2$ are asymptotically equivalent if there exist (randomized) mappings $Q_j^{(n)} : \mathcal{X}_j^{(n)} \to \mathcal{X}_k^{(n)}$, j,k = 1,2, k≠j, such that

$$\sup_{\theta \in \Theta^{(n)}} \left\| F_{j,\theta}^{(n)}(\bullet) - \int Q_k^{(n)}\left(\bullet \middle| x_k\right) F\left(dx_k\right) \right\|_{TV} = c_n \to 0, \text{ j,k = 1,2, k≠j,} \tag{10}$$

where $\|\bullet\|_{TV}$ denotes the total variation norm.

This definition involves a reformulation of the general theory originated by LeCam (1953, 1964). See also Le Cam (1986), Le Cam and Yang (2000), van der Vaart (2002) and Brown and Low (1996) for background on this theory including several alternate versions of the definition and related concepts, a number of conditions that imply asymptotic equivalence, and many applications to a variety of statistical settings. Note that both Formulations 1 and 2 involve an index, n, and can thus be considered as sequences of statistical problems in the sense of the definition.

3.3 Spaces of densities (or intensities):

Suitable families of densities, $\mathcal{F}$, can be defined via Besov norms with respect to the Haar basis. The Besov norm with index α and shape parameters p = q can most conveniently be defined via the stepwise approximants to f at resolution level k. These approximants are defined as

$$\bar{f}_k(t) = \sum_{\ell=0}^{2^k-1} I_{[\ell/2^k,(\ell+1)/2^k)}(t)\bar{f}_{k,\ell} \text{ where } \bar{f}_{k,\ell} = \int_{\ell/2^k}^{(\ell+1)/2^k} 2^k f,$$

and the Besov(α,p) norm is defined as

$$\|f\|_{\alpha,p} = \left\{ \left|\bar{f}_0\right|^p + \sum_{k=0}^{\infty} 2^{pk\alpha} \left\|\bar{f}_k - \bar{f}_{k+1}\right\|_p^p \right\}^{1/p}.$$

The statement of Theorem 1 can now be completed by stating the assumption on $\mathcal{F}$ needed for its validity.

Assumption A: $\mathcal{F}$ satisfies

$$\mathcal{F} \subset \left\{ f : \inf_{0 \le x \le 1} f(x) \ge \varepsilon_0 \right\} \text{ for some } \varepsilon_0 > 0 \tag{11}$$

and $\mathcal{F}$ is compact in both Besov(1/2,2) and Besov(1/2,4).

Other function spaces are also conventional for nonparametric statistical applications of this type. The most common of these are based on either the Lipshitz norm $\|f\|_\beta^{(L)}$ or the Sobolev norm $\|f\|_\beta^{(S)}$. These are defined for β≤1 by

$$\|f\|_{\beta}^{(L)}=\sup_{0\le x<y\le 1}\frac{|f(y)-f(x)|}{|y-x|^{\beta}},\quad \|f\|_{\beta}^{(S)}=\sum_{-\infty}^{\infty}k^{2\beta}\vartheta_k^2$$

where $\vartheta_k=\int_0^1 f(x)e^{ik2\pi x}dx$ denote the Fourier coefficients of f. (Both spaces have natural definitions for β>1 as well, but we need consider here only the case β≤1.)

The following implies Assumption A and hence also suffices for validity of Theorem 1.

Assumption A′: $\mathcal{F}$ satisfies (11), and is bounded in the Lipshitz norm with index β, and is compact in the Sobolev norm with index α, where α ≥ β and either β > 1/2 or α ≥ 3/4 and α + β ≥ 1.

The following assumption is noticeably stronger than either A′ or A, and is used in Theorem 2.

Assumption B: $\mathcal{F}$ satisfies (11) and is bounded in the Lipshitz norm with index β, where β > 1/2.

For more information about the relation of these spaces in this context see Brown, Cai, Low and Zhang (2002) and Brown, Carter, Low and Zhang (2002) (referred to as BCLZ below).

3.4 Statistical Equivalence Theorems

BCLZ then extended earlier results of Nussbaum (1996) and Klemela and Nussbaum (1998) to prove the following basic result:

Theorem a (BCLZ): Consider the statistical Formulations 1 and 2 with the parameter space $\mathcal{F}$ and the relation (9). Assume $\mathcal{F}$ satisfies Assumption A (or A′). Then the sequences of statistical problems defined in these two formulations are asymptotically statistically equivalent.

NOTE: Assumption A seems to be about the weakest possible assumption under which a conclusion like the above is valid. Brown and Zhang (1998) shows that such a conclusion is not valid under the general assumption that $\mathcal{F}$ is bounded in both the Besov(1/2,2) and Lipshitz(β=1/2) norm . The appearance of the norm Besov(1/2,4) in Assumption A may possibly be a technical artifact of our proof of Theorem a, though perhaps something slightly stronger than compactness in Besov(1/2,2) is required for such a conclusion.

BCLZ describes in detail a construction of $\tilde{Z}_n$ as a (randomized) funcion of $\{T_1,\ldots,T_n\}$. (More precisely, BCLZ describes the construction of the Haar basis representation of $\tilde{Z}_n$, from which $\tilde{Z}_n$ can directly be recovered.) This construction is invertible, in that $\{T_1,\ldots,T_n\}$ can be recovered as a function of $\tilde{Z}_n$. Further, BCLZ shows that both $\tilde{Z}_n$ and $\tilde{W}_n$ can be represented on the same probability space so that their distributions, $P_{\tilde{Z}_n}$ and $P_{\tilde{W}_n}$, say, satisfy

$$\left\| P_{\tilde{Z}_n} - P_{\tilde{W}_n} \right\|_{TV} \to 0 .$$

The mappings $\{Q_j^{(n)}$: j=1,2, n = 1,2,...} that yield the equivalence of the above theorem can then be directly inferred from this construction. We sketch that construction in Section4, and refer the reader to BCLZ for details of the construction and proof. It can be remarked that these bear considerable similarity to parts of the construction and proof in Bretagnolle and Massart (1989) and Koltchinskii (1994) and other proofs of KMT type theorems. But there are also some basic differences, especially those related to the appearance of the square-root in the fundamental relation (9) and the use of the very strong Total Variation metric involved in the statement of the theorem. In addition, the fact that (7) is uniform in $\mathbb{Q}$ and $\mathcal{F}$ entails the need for various refinements in the proof. See Section 4 for a related remark.

Theorem 1 is now an immediate logical consequence of this result from BCLZ and the following lemma.

Lemma: Suppose $\mathcal{P}_j^{(n)}=(\mathcal{X}_j^{(n)}, \mathcal{B}_j^{(n)}, \mathcal{F}_j^{(n)})$ j = 1,2, n = 1,2,... are asymptotically equivalent sequences of statistical problems on the same sequence of parameter spaces, $\Theta^{(n)}$. Let $\{Q_j^{(n)}$: j=1,2, n=1,2,...} denote a sequence of mappings that define this equivalence, as in (10). Then there are non- randomized mappings $\{\tilde{Q}_j^{(n)}$: j=1,2, n = 1,2,...} such that

(12) $$P_f(\tilde{Q}_j^{(n)} = Q_j^{(n)}) \ge 1 - c_n \quad \text{for every } f \in \mathcal{F}_j^{(n)}, \text{ j=1,2, n=1,2,...}$$

and for every $\theta \in \Theta^{(n)}$

(13) $$P_{f_{j,\theta}}\left(\tilde{Q}_j^{(n)}\left(X_j^{(n)}\right) \in A\right) = P_{f_{k,\theta}}\left(X_k^{(n)} \in A\right),\ \theta \in \Theta^{(n)}$$

for every measurable $A \subset \mathcal{X}_k^{(n)}$, j,k=1,2, j≠k, n=1,2,... .

NOTE: The expression (12) refers to the form of the functions $\tilde{Q}_j^{(n)}$ and $Q_j^{(n)}$ in which these are written as a nonrandomized function involving an auxiliary variable U, as described earlier. If these functions are represented in the ordinary form of a randomized mapping and (as is the case) $\mathcal{F}_j^{(n)}$ is a dominated family for each n then the condition (12) is equivalent to

(12′) $$\left\| \tilde{Q}_j^{(n)}\left(\bullet | x_j^{(n)}\right) - Q_j^{(n)}\left(\bullet | x_j^{(n)}\right) \right\|_{TV} \ge 1 - c_n \quad a.e.(f)\ \forall f \in \mathcal{F}_j^{(n)} .$$

Proof of Lemma: Fix n, j, k≠j, $\theta\in\Theta^{(n)}$. Let F_k denote the distribution under θ of $X_k^{(n)}$ and let F_k' denote the distribution under θ of $Q_j^{(n)}\left(X_j^{(n)}\right)$. Let H = min($F_k$, F_k'). Let $\infty \ge f_k' = \frac{dF_k'}{dH} \ge 1$. Then define $\tilde{Q}_j^{(n)}$ as a version of the randomized map satisfying

$$\tilde{Q}_j(B\,|\,x) = \frac{1}{f_k'(x)} Q_j(B\,|\,x) + \frac{f_k' - 1}{f_k'}\left(F_k'(B) - H(B)\right).$$

This completes the proof of the lemma, and consequently also that of Theorem 1. ||

Remark 1: Although not simple, the construction of $Q_j^{(n)}$ is somewhat more natural than that of $\tilde{Q}_j^{(n)}$. In particular, the construction of $Q_j^{(n)}$ does not depend on f∈$\mathcal{F}$. We have included the construction of $\tilde{Q}_j^{(n)}$ in order to complete the statement (7) of Theorem 1 in a form parallel to (4) in Theorem BM. However, it should be noted that $Q_j^{(n)}$ itself satisfies a pleasant property as given in the following corollary.

Corollary 1: Under the setting of Theorem 1 the maps $Q_1^{(n)}$ satisfy

(13) $$\sup_{f\in\mathcal{F}} P_f\left(\sup_{q\in\mathbb{Q}} |\tilde{Z}_n(q) - \left\{Q_1^{(n)}\left(\tilde{Z}_n\right)\right\}(q)| > 0\right) \le c_n \to 0$$

and there is a version of $\tilde{W}_n$ such that

(14) $$\sup_{f\in\mathcal{F}} P_f\left(\sup_{q\in\mathbb{Q}} \left|\tilde{W}_n(q) - \left\{Q_1^{(n)}\left(\tilde{Z}_n\right)\right\}(q)\right| > c_n \|q\|_\infty\right) = 0.$$

Theorem 2 requires a slightly different fundamental construction, and some additional argument for its proof. The proof for this result is based heavily on results in Carter (2001). See the remarks following Theorem 2 of BCLZ.

Theorem b (BCLZ): Consider the statistical Formulations 1″ and 2 with the parameter space $\mathcal{F}$ and the relation (9). Assume $\mathcal{F}$ satisfies Assumption B. Then the sequences of statistical problems defined in these two formulations are asymptotically statistically equivalent.

4. Construction of the Mapping

4.1 A "Natural" Mapping

As already remarked the object, $\hat{Z}_n$, that appears in the KMT theory is a probabilistically and statistically natural quantity. The object $\tilde{Z}_n$ that appears in our main theorem, while explicitly constructable, is far less natural. However, there is a reasonably natural object, call it $\tilde{Z}_n^*$, that resembles $\tilde{Z}_n$. We conjecture below that this object could play a role like that of $\hat{Z}_n$.

By analogy with the KMT theorem one could view $\tilde{Z}_n^*$ as the natural analog of $\hat{Z}_n$ and $\tilde{Z}_n$ as the "coupling" produced to match $\tilde{Z}_n^*$ to the continuous process $\tilde{W}_n$. We conclude this section by sketching the construction of $\tilde{Z}_n$ as an extension of the construction of $\tilde{Z}_n^*$.

The construction of $\tilde{Z}_n$ in BCLZ involves two separate stages. Stripped of several detailed features some of which are needed to obtain the full strength of the theorem there (and others that are present only for technical convenience in the argument there) the first stage of that construction resembles the following definition.

Let $k_0 = k_0(n) = \left[\dfrac{\log_2 n}{2}\right]$. Let $N_{k_0,\ell} = \#\left\{T_i : \ell/2^{k_0} < T_i \le (\ell+1)/2^{k_0}\right\}$. Then let

(15) $$\tilde{Z}_n^*\left(\ell/2^{k_0}\right) = 2^{-k_0}\sum_{j=0}^{\ell-1}\sqrt{\frac{N_{k_0,j}}{n/2^{k_0}}},\ \ \ell = 0,\ldots,2^{k_0}.$$

For $\ell/2^{k_0} < t < (\ell+1)/2^{k_0}$ let

(16) $$\tilde{Z}_n^*(t) = \tilde{Z}_n^*\left(\ell/2^{k_0}\right) + \left(\tilde{Z}_n^*\left((\ell+1)/2^{k_0}\right) - \tilde{Z}_n^*\left(\ell/2^{k_0}\right)\right)\frac{\#\left\{T_j : \ell/2^{k_0} < T_j \le t\right\}}{N_{k_0,\ell}}.$$

The object $\tilde{Z}_n^*$ does not satisfy a conclusion like that of (7) in Theorem 1. However it does seem reasonable to conjecture that $\tilde{Z}_n^*$ can be made uniformly close to $\tilde{W}_n$ with respect to the supremum (Kolmogorov-Smirnov) distance that appears in (2).

Thus, we conjecture that for a suitable sequence of constants a_n (probably with $a_n \sim \log n/\sqrt{n}$), and a suitable c>0, and for $\mathcal{F}$ satisfying Assumption B (and perhaps only the weaker Assumption A) the random quantities $\tilde{Z}_n^*$ and $\tilde{W}_n$ can be represented on the same probability space so that

(17) $$\sup_{f\in\mathcal{F}} P_f\left(\sup_{t\in[0,1]}\sqrt{n}\left|\tilde{Z}_n^*(t) - \tilde{W}_n(t)\right| > a_n\right) \le c.$$

4.2 Definition of $\tilde{Z}_n$

As already mentioned, the definition of $\tilde{Z}_n$ resembles the above definition in several respects, but there are also several differences. This construction begins with a choice of level similar to the choice k_0, above, but here one must choose $k_0(n)$ so that

$$\left(2^{k_0}\big/\sqrt{n}\right) \to \infty \text{ (slowly),}$$

with the allowable speed of convergence depending on the properties of $\mathcal{F}$. (The larger $\mathcal{F}$ is, the slower must be this convergence.) Then let $U_{k,j}$ be independent random variables with $U_{k,j} \sim Uniform(-1/2, 1/2)$,
and let

(18) $$\tilde{Z}_n\left(\ell\big/2^{k_0}\right) = 2^{-k_0}\sum_{j=0}^{\ell-1}\sqrt{\frac{\left|N_{k_0,j}+U_{k_0,j}\right|}{n/2^{k_0}}},\ \ell = 0,\ldots,2^{k_0}.$$

This definition is the same as that in (15) except for the slightly different choice of k_0 and the presence of the uniform random variables whose main purpose is to smooth the distributions of $\tilde{Z}_n\left(\ell\big/2^{k_0}\right)$.

It is helpful to represent the values of $\tilde{Z}_n\left(\ell\big/2^{k_0}\right)$ as the integral of a function defined through a set of Haar coefficients. Thus, let

(19) $$\bar{h}_{k_0,\ell} = 2^{-k_0}\sqrt{\frac{\left|N_{k_0,j}+U_{k_0,j}\right|}{n/2^{k_0}}}, \text{ and } h_{k_0}(t) = \sum_{\ell=0}^{2^{k_0}-1} I_{[\ell/2^{k_0},(\ell+1)/2^{k_0})}(t)\bar{h}_{k_0,\ell}.$$

Then

$$\tilde{Z}_n\left(\ell\big/2^{k_0}\right) = \int_0^{\ell/2^{k_0}} h_{k_0}(\tau)d\tau,\ \ell = 0,\ldots,2^{k_0}.$$

To define the Haar coefficients at finer levels, let $V_m \sim Binomial(p = 1/2)$ and let $F_m^*(x) \triangleq P\left\{V_m + U_{m,2\ell} \le x\right\}$. Then, for $k>k_0$ let

(20) $$-\bar{h}_{k,2\ell+1} \triangleq \bar{h}_{k,2\ell} \triangleq \sigma_{k-1}\Phi^{-1}\left(F^*_{N_{k-1,\ell}}\left(N_{k,2\ell}+U_{k,2\ell}\right)\right),\ \ell = 0,\ldots,2^{k-1},$$

where Φ denotes the standard normal CDF and $\sigma_k^2 = 2^k\big/4n$. Then let

(21) $$h_k(t) = \sum_{\kappa=k_0}^{k}\sum_{\ell=0}^{2^\kappa-1} I_{[\ell/2^\kappa,(\ell+1)/2^\kappa)}(t)\bar{h}_{\kappa,\ell}$$

$$\tilde{Z}_n\left(\ell\big/2^{k}\right) = \int_0^{\ell/2^k} h_k(\tau)d\tau,\ \ell = 0,\ldots,2^k.$$

For the purposes of establishing statistical equivalence it suffices to carry this construction to $k^* = [\log_2 n]$. (See Brown, Cai, Low and Zhang (2002).) If a complete definition of $\tilde{Z}_n$ is desired one can carry the construction in (21) to its limit in $\mathcal{L}_2$ as $k\to\infty$. Alternatively, one can carry the construction down to the value k^* and then define $\tilde{Z}_n(t)$ at values between points of the form $\ell\big/2^{k^*}$ by interpolating suitably scaled independent Brownian Bridges. (This is the continuous analog of the procedure in (16).)

References

Bretagnolle, J and Massart, P. (1989) Hungarian constructions from the nonasymptotic viewpoint, *Ann. Probab., 17*, 239-256.

Brown, L. D., Cai, T. T., Low, M. G. and Zhang, C-H. (2002a) Asymptotic equivalence theory for nonparametric regression with random design, *Ann. Statist. 30*, 688-707.

Brown, L. D., Carter, A., Low, M. G. and Zhang, C-H. (2002b) (BCLZ) Asymptotic equivalence theory for a Poisson process with variable intensity, Preprint, available via http://ljsavage.wharton.upenn.edu/~lbrown/ .

Brown, L. D. and Zhang, C-H (1998) Asymptotic non-equivalence of nonparametric experiments when the smoothness index is ½, *Ann. Statist. 26*, 279-287.

Carter, A. (2001) Deficiency distance between multinomial and multivariate normal experiments under smoothness constraints on the parameter set, Preprint available via http://www.pstat.ucsb.edu/faculty/carter/research.html .

Csörgö, M and Revesz, P. (1981) *Strong Approximations in Probability and Statistics*, Academic Press, NY.

Dudley, R. M. (1978) Central limit theorems for empirical measures, *Ann. Probab. 6*, 899-929.

Klemela, J. and Nussbaum, M. (1998) Constructive asymptotic equivalence of density estimation and Gaussian white noise, Preprint available at http//www.math.cornell.edu/~nussbaum/ .

Koltchinskii, V. I. (1994) Komlos-Major-Tusnady approximation for the general empirical process and Haar expansions of classes of functions, *Jrnl of Theoretical Probab., 7*, 73-118.

Komlos, J., Major, P. and Tusnady, G. (1975) An approximation of partial sums of independent rv's and the sample df. I, *Wahrsch verw Gebiete, 32*, 111-131.

Komlos, J., Major, P. and Tusnady, G. (1976) An approximation of partial sums of independent rv's and the sample df. II, *Wahrsch verw Gebiete, 34*, 33-58.

Le Cam, L and Yang, G. L. (2000) *Asymptotics in Statistics*, Springer, NY.

Le Cam, L. (1953) On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates, *University of California Publications in Statistics, vol 1, 11*, 277-330.

Le Cam, L. (1964) Sufficiency and approximate sufficiency, *Ann. Math. Statist. 35*, 1419-1455.

Le Cam, L. (1986) *Asymptotic Methods in Statistical Decision Theory*, Springer, NY.

Massart, P. (1989) Strong approximations for multivariate empirical and related processes, via KMT constructions, *Ann. Probab*. *17*, 266-291.

Nussbaum, M. (1996) Asymptotic equivalence of density estimation and Gaussian white noise, *Ann. Statist., 24*, 2399-2430.

Steele, J. M. (2000) *Stochastic Calculus and Financial Applications*. Springer, NY.

Van der Vaart, A. W. (2002) The statistical work of Lucien Le Cam, *Ann. Statist.30*, 631-682.

Van der Vaart, A. W. and Wellner, J. A. (1996) *Weak Convergence and Empirical Processes with Applications to Statistics, Springer*, NY.

Wald, A. and Wolfowitz, J. (1951) Two methods of randomization in statistics and the theory of games, *Ann. Math., 53*, 581-586.