CHAPTER I

ASYMPTOTIC THEORY FOR THE GENERALIZED BOOTSTRAP OF STATISTICAL DIFFERENTIABLE FUNCTIONALS

1) Introduction

Let \( T \) be a statistical functional defined on a space \( \mathcal{P} \) of probability measures (p.m.'s) on a locally compact Banach space \( B \). Let \( X, X_1, \ldots, X_n \) be a sequence of independent and identically distributed (i.i.d.) random variables (r.v.) with common probability \( P \in \mathcal{P} \), and let us define

\[
P_n := n^{-1} \sum_{i=1}^{n} \delta_{X_i},
\]

the empirical measure, where \( \delta_{X_i} \) denotes the Dirac measure at \( X_i \). When \( T \) is smooth in a neighborhood of \( P \), a natural estimator of \( T(P) \) is its empirical counterpart \( T(P_n) \) (see Von Mises (1947), Huber (1981), Manski (1988)).

To define the generalized bootstrapped empirical measure of Mason and Newton (1993), consider \( W_n := \{ W_{i,n} : 1 \leq i \leq n, n \geq 1 \} \) a triangular array of row exchangeable r.v. with joint distribution \( W_n \) on the \( n \)-th row, and introduce the random measure

\[
P_{W_n} := \sum_{i=1}^{n} W_{i,n} \delta_{X_i}.
\]

The idea of the (generalized) bootstrap is to approximate the distribution of \( n^{1/2}(T(P_n) - T(P)) \) by that of \( n^{1/2}(T(P_{W_n}) - T(P_n)) \) conditionally on \( P_n \). Efron's bootstrap as defined in Efron (1979, 1982) consists in taking \( W_n = n^{-1} \text{Multinomial}(n, 1/n, \ldots, 1/n) \). In this case \( W_{i,n} \) is the frequency of \( X_i \) in the bootstrapped sample. This approach with random weights (called the resampling plan in Efron (1982) or Wu (1986)) emphasizes the double random aspect of the bootstrap. Other choices for the weights may be found in Praestgaard and Wellner (1993).

For the generalized bootstrap, Mason and Newton (1992) studied the sample mean and the empirical and quantile processes when \( B = \mathbb{R} \). In particular,
they proved the consistency of the bootstrap for the mean assuming the following conditions on the weights,

\[ \sum_{i=1}^{\infty} W_{i,n} = 1, \]

\[ \max_{1 \leq k \leq n} \left( W_{i,n}^{-1} \right)^2 / \sum_{i=1}^{n} (W_{i,n}^{-1})^2 = o_p(1) \quad \text{as } n \to \infty. \]

In Haeusler, Mason and Newton (1992), the distribution \( W_n \) of the weights is data dependent. However, in this chapter, we will assume that

\[ W_n \text{ and the } W_{i,n} \text{'s are independent of the sample,} \]

which holds for Efron's (1979), Rubin's (1981),... bootstraps.

Our purpose in this chapter is to give some asymptotic results for the generalized bootstrap of a large class of statistics in the spirit of Dudley (1990) and Giné and Zinn (1990), without restricting the \( X_i \)'s to be real valued as generally required to apply the quantile-transform-empirical-process approach as in Csörgö and Mason (1989) or Einmahl and Mason (1992) (see also the volume edited by Hahn, Mason and Weiner (1991) for this approach).

Now, let us explain what kind of results we obtain. Among other things, we prove that the convergence of the generalized bootstrapped distribution to the true distribution holds in probability along the sample for first order Fréchet differentiable functionals (Theorem 3.1). Related work on Efron's bootstrap of differentiable functionals may be found in Bickel and Freedman (1981), Lohse (1984,1987), Parr (1985), Gill (1989), Pons and Turkheim (1989). The originality of our approach is that we do not make use of continuous differentiability assumption at \( P \). This is generally made possible by an adequate choice of the distance (maybe depending on the unknown \( P \) ) which metrizes the topology on \( \mathcal{P} \).

Another topic in this chapter is to investigate the convergence of the generalized bootstrap variance estimator

\[ S_n^2 := E_W(n^{1/2}(T(P_{W,n})-T(P_n))^2 / \text{Var}(nW_{i,n})), \]

where \( E_W \) is the expectation under the distribution \( W_n \) of the resampling plan, conditioned on the sample (Theorem 3.2). We show that this generalized bootstrap variance estimator is generally efficient (Theorem 3.2). Our results