Monte Carlo tests with nuisance parameters: a general approach to finite-sample inference and nonstandard asymptotics[1]

Jean-Marie Dufour[2]

September 2000

[1] This work was supported by the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and by the Government of Québec (Fonds FCAR). The author thanks Torben Andersen, Russell Davidson, V. P. Godambe, Joanna Jasiak, Linda Khalaf, Jan Kiviet, James MacKinnon, Charles Manski, Eric Renault, Olivier Torrès, Pascale Valéry, Michael Veall for several useful comments.

[2] Centre de recherche et développement en économique (C.R.D.E.) and Département de sciences économiques, Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal (Québec) H3C 3J7, Canada, Tel. : (514) 343-6557, Fax : (514) 343-5831.

ABSTRACT

The technique of Monte Carlo (MC) tests [Dwass (1957), Barnard (1963)] provides an attractive method of building exact tests from statistics whose finite sample distribution is intractable but can be simulated (provided it does not involve nuisance parameters). We extend this method in two ways: first, by allowing for MC tests based on exchangeable possibly discrete test statistics; second, by generalizing the method to statistics whose null distributions involve nuisance parameters (maximized MC tests, MMC). Simplified asymptotically justified versions of the MMC method are also proposed and it is shown that they provide a simple way of improving standard asymptotics and dealing with nonstandard asymptotics (*e.g.*, unit root asymptotics). Parametric bootstrap tests may be interpreted as a degenerate version of the MMC method (without the general validity properties of the latter).

Key words: Monte Carlo test; finite sample test; exact test; bounds; bootstrap.

Contents

List of Definitions, Propositions and Theorems

1 Introduction

During the last twenty years, the development of faster and cheaper computers has made Monte Carlo techniques more affordable and attractive in statistical analysis. In particular, such techniques may now be used routinely in data analysis. Important developments in this area include the use of bootstrap techniques for improving standard asymptotic approximations [for reviews, see Efron (1982), Efron and Tibshirani (1993), Hall (1992), Jeong and Maddala (1993), Vinod (1993), Shao and Tu (1995), Davison and Hinkley (1997), and Horowitz (1997)] and techniques where estimators and forecasts are obtained from criteria evaluated by simulation [see Mariano and Brown (1993), Hajivassiliou (1993), Keane (1993), Gouriéroux and Monfort (1996) and Gallant and Tauchen (1996)].

With respect to tests and confidence sets, these techniques only have asymptotic justifications and do not yield finite sample inferences that are provably valid (in the sense of correct levels) in finite samples. Here it is of interest to note that the use of simulation in the execution of tests was suggested much earlier than recent bootstrap and simulation-based techniques. For example, randomized tests have been proposed long ago as a way obtaining tests with any given level from statistics with discrete distributions (*e.g.*, sign and rank tests); see Lehmann (1986). A second interesting possibility is the technique of Monte Carlo tests originally suggested by Dwass (1957) for implementing permutation tests and later extended by Barnard (1963) and Birnbaum (1974). This technique has the great attraction of providing *exact* (randomized) tests based on any statistic whose finite sample distribution may be intractable but can be simulated. The validity of the tests so obtained does not depend at all on the number of replications made (which can be small). Only the power of the procedure is influenced by the number of replications, but the power gains associated with lengthy simulations are typically rather small. For further discussion of Monte Carlo tests, see Dufour and Kiviet (1996, 1998), Edgington (1980), Foutz (1980), Jöckel (1986), Kiviet and Dufour (1997) and Marriott (1979).

An important limitation of the technique of Monte Carlo tests is the fact that one needs to have a statistic whose distribution does not depend on nuisance parameters. This obviously limits considerably its applicability. The main objective of this paper is to extend the technique of Monte Carlo tests in order to allow for the presence of nuisance parameters in the null distribution of the test statistic.

In Section 2, we summarize and extend results on Monte Carlo (MC) tests when the null distribution of a test statistic does not involve nuisance parameters. In particular, we put them in a form that will make their extension to cases with nuisance parameters easy and intuitive, and we generalize them by allowing for MC tests based on exchangeable (possibly non independent) replications and statistics with discrete distributions. These generalizations allow, in particular, for various nonparametric tests (*e.g.*, permutation tests) as well as test statistics where certain parameters are themselves evaluated by simulation. In Section 3, we study how the power of Monte Carlo tests is related to the number of replications used and the sensitivity of the conclusions to the randomized nature of the procedure. In particular, given the observed (randomized) p-value of the Monte Carlo test, we see that the probability of an eventual reversal of the conclusion of the procedure (rejection or acceptance at a given level, *e.g.* 5%) can easily be computed.

In Section 4, we present the extension to statistics whose null distribution depends on nuisance parameters. This procedure involves one to consider a simulated p-value function as a function of the nuisance parameters (under the null hypothesis), and we show that maximizing the latter with respect to the nuisance parameters yields a test with provably exact level, irrespective of the sample size and the number replications used. For this reason, we call the latter maximized Monte Carlo (MMC) tests. We also discuss how this maximization can be achieved in practice, *e.g.* through *simulated annealing* techniques.

In the two next sections, we discuss simplified (asymptotically justified) approximate versions of the proposed procedures, which involve the use of consistent set or point estimates of model parameters. In Section 5, we suggest a method [the consistent set estimate MMC method (CSEMMC)] which is applicable when a consistent set estimator of the nuisance parameters [*e.g.*, a random subset of the parameter space whose probability of covering the nuisance parameters converges to one as the sample size goes to infinity] is available. The approach proposed involves maximizing the simulated p-value function over the consistent set estimator over the consistent set estimate, as opposed to the full nuisance parameter space. This procedure may thus be computationally much less costly. Using a consistent set estimator (or confidence set), as opposed to a point estimate, to deal with nuisance parameters is especially useful because it allows one to obtain asymptotically valid tests even when the test statistic does not possess an asymptotic distribution or when the asymptotic distribution depends on nuisance parameters possibly in a discontinuous way. Consequently, there is no need to study the asymptotic distribution of the test statistic considered or even to establish its existence. This consistent set estimator MMC method (CSEMMC) may be viewed as an asymptotic Monte Carlo extension of finite-sample two-stage procedures proposed in Dufour (1990), Dufour and Kiviet (1996, 1998), Campbell and Dufour (1997), and Dufour, Hallin, and Mizera (1998).

In Section 6, we consider the simplest form of Monte Carlo test with nuisance parameters, *i.e.* the one where the consistent set estimate has been replaced by a consistent point estimate. In other words, the distribution of the test statistic is simulated after replacing the nuisance parameters by a consistent point estimate. Such a procedure can be interpreted as a parametric bootstrap test based on the percentile method [see Efron and Tibshirani (1993, Chapter 16) and Hall (1992)]. The term "nonparametric" may however be misleading here, because such MC tests can be applied as well to nonparametric (distribution-free) test statistics. For such methods, it is well known from the literature on the bootstrap, that using a consistent point estimate may fail to provide asymptotically valid tests when the test statistic simulated has an asymptotic distribution involving nuisance parameters [see Athreya (1987), Basawa, Mallik, McCormick, Reeves, and Taylor (1991) and Sriram (1994)]. Further, proofs of the validity of the bootstrap typically require the assumption that the test statistic be asymptotically pivotal as the sample size goes to infinity (*i.e.*, its asymptotic distribution does not involve nuisance parameters). In Section 6, we give general conditions under which parametric bootstrap procedures would yield asymptotically valid tests in cases where the asymptotic of the test statistic involves nuisance parameters. In particular, these continuity involve a smooth (continuous) dependence of the asymptotic distribution upon the nuisance parameters. These conditions, however, are more restrictive and more difficult to check than those under which CSEMMC procedures would be applicable.

2 Monte Carlo tests without nuisance parameters

Let $\{S \geq c\}$ be a critical region for testing an hypothesis H_0, where S a real-valued test statistic whose distribution $F(x) = \mathsf{P}[S \leq x]$ is uniquely determined under H_0, either because H_0 is a simple hypothesis (*i.e.*, H_0 defines a unique data generating process) or because the null distribution of S does not depend on (unknown) nuisance parameters. The constant c is chosen so that

$$\mathsf{P}[S \geq c] = 1 - F(c) + \mathsf{P}[S = c] \leq \alpha \tag{2.1}$$

where α is the desired level of the test $(0 < \alpha < 1)$. Note that the critical region $S \geq c$ can also be put in two useful alternative forms, which are equivalent to $S \geq c$ with probability one (*i.e.*, they can differ from the critical region $S \geq c$ only on a set of zero probability):

$$G(S) \leq G(c) \ , \tag{2.2}$$

$$S \geq F^{-1}\left[(F(c) - \mathsf{P}[S = c])^{+}\right] = F^{-1}\left[(1 - G(c))^{+}\right] \tag{2.3}$$

where

$$G(x) = \mathsf{P}[S \geq x] = 1 - F(x) + \mathsf{P}[S = x] \tag{2.4}$$

is the "tail area" or "p-value function" associated with F, and F^{-1} is the quantile function of F, with the conventions

$$F^{-1}\left(q^{+}\right) = \lim_{\varepsilon \downarrow 0} F^{-1}(q + \varepsilon) = \inf\left\{F^{-1}(q_0) : q_0 > q\right\} \ , \ 0 \leq q \leq 1 \, ,$$

$F^{-1}(1^{+}) = \infty$ and $F^{-1}(0^{+}) = F^{-1}(0)$. For any probability distribution function $F(x)$, the quantile function $F^{-1}(q)$ is defined as follows:

$$\begin{aligned} F^{-1}(q) &= \inf\{x : F(x) \geq q\} \ , \quad \text{if } 0 < q < 1 \, , \\ &= \inf\{x : F(x) > 0\} \ , \quad \text{if } q = 0 \, , \\ &= \sup\{x : F(x) < 1\} \ , \quad \text{if } q = 1 \, ; \end{aligned} \tag{2.5}$$

see Reiss (1989, p. 13). In general, $F^{-1}(q)$ takes its values in the extended real numbers $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ and, for coherence, we set $F(-\infty) = 0$ and $F(\infty) = 1$. Using (2.5), it is easy to see that

$$F^{-1}\left[(F(c) - \mathsf{P}[S = c])^{+}\right] = c$$

when: $0 < F(c) < 1$ and either $\mathsf{P}[S = c] > 0$ or $F(x)$ is continuous and strictly monotonic in an open neighborhood containing c. However, the formulation (2.3) remains valid in all cases.

2.1 Monte Carlo tests based on statistics with continuous distributions

Consider now a situation where the distribution of S under H_0 may not be easy to compute analytically but can be simulated. Let $S_1, \ldots, S_N$ be a sample of independent and identically distributed ($i.i.d.$) real random variables with the same distribution as S. Typically, it is assumed that

$S_1, \ldots, S_N$ are also independent. However, we will observe that the exchangeability of $S_1, \ldots, S_N$ is sufficient for most of the results presented below.[1] The technique of Monte Carlo tests provides a simple method allowing one to replace the theoretical distribution $F(x)$ by its sample analogue based on $S_1, \ldots, S_N$:

$$\hat{F}_N[x; S(N)] = \frac{1}{N}\sum_{i=1}^{N} s(x - S_i) = \frac{1}{N}\sum_{i=1}^{N} 1_{[0,\infty)}(x - S_i) \tag{2.6}$$

where $S(N) = (S_1, \ldots, S_N)'$, $s(x) = 1_{[0,\infty)}(x)$ and $1_A(x)$ is the indicator function associated with the set A:

$$\begin{aligned} 1_A(x) &= 1\,, \text{ if } x \in A\,, \\ &= 0\,, \text{ if } x \notin A\,. \end{aligned} \tag{2.7}$$

We also consider the corresponding sample tail area function:

$$\hat{G}_N[x; S(N)] = \frac{1}{N}\sum_{i=1}^{N} s(S_i - x)\ . \tag{2.8}$$

The sample distribution function is related to the ranks $R_1, \ldots, R_N$ of the variables $S_1, \ldots, S_N$ (when put in ascending order) by the expression:

$$R_j = N\hat{F}_N[S_j; S(N)] = \sum_{i=1}^{N} s(S_j - S_i)\ ,\ j = 1, \ldots, N\,. \tag{2.9}$$

The central property we shall exploit here is the following: to obtain critical values or compute p-values, the "theoretical" null distribution $F(x)$ can be replaced by its simulation-based "estimate" $\hat{F}_N(x)$ in a way that will preserve the level of the test in *finite samples, irrespective of the number N of replications used*. For continuous distributions, this property is expressed by Proposition **2.1.2** below, which is easily proved by using the following simple lemma.

2.1.1 Lemma DISTRIBUTION OF RANKS WHEN TIES HAVE ZERO PROBABILITY. *Let $(y_1, \ldots, y_N)'$ be a $N \times 1$ vector of exchangeable real random variables such that*

$$\mathsf{P}[y_i = y_j] = 0 \textit{ for } i \neq j\ ,\ i,\ j = 1, \ldots, N\,, \tag{2.10}$$

and let $R_j = \sum_{j=1}^{N} s(y_j - y_i)$ be the rank of y_j when $y_1, \ldots, y_N$ are ranked in nondecreasing order $(j = 1, \ldots, N)$. Then, for $j = 1, \ldots, N$,

$$\mathsf{P}[R_j/N \leq x] = I[xN]/N\ ,\ \textit{for } 0 \leq x \leq 1\,, \tag{2.11}$$

[1]The elements of a random vector $(S_1, S_2, \ldots, S_N)'$ are exchangeable if $(S_{r_1}, S_{r_2}, \ldots, S_{r_N})' \sim (S_1, S_2, \ldots, S_N)'$ for any permutation $(r_1, r_2, \ldots, r_N)$ of the integers $(1, 2, \ldots, N)$.

and

$$\begin{aligned} \mathsf{P}\left[R_j/N \geq x\right] &= 1\,, && \textit{if } x \leq 0\,, \\ &= \left(I\left[(1-x)\,N\right]+1\right)/N\,, && \textit{if } 0 < x \leq 1\,, \\ &= 0\,, && \textit{if } x > 1\,, \end{aligned} \tag{2.12}$$

where $I\left[x\right]$ *is the largest integer less than or equal to* x.

It is clear that condition (2.10) is satisfied whenever the variables $y_1, \ldots, y_N$ are independent with continuous distribution functions [see Hájek (1969, pp. 20-21)], or when the vector $(y_1, \ldots, y_N)'$ has an absolutely continuous distribution (with respect to the Lebesgue measure on $\mathbb{R}^N$).

2.1.2 Proposition VALIDITY OF MONTE CARLO TESTS WHEN TIES HAVE ZERO PROBABILITY. *Let* $(S_0, S_1, \ldots, S_N)'$ *be a* $(N+1)\times 1$ *vector of exchangeable real random variables such that*

$$\mathsf{P}\left[S_i = S_j\right] = 0 \textit{ for } i \neq j\,,\ i,\ j = 0, 1, \ldots, N\,, \tag{2.13}$$

let $\hat{F}_N(x) \equiv \hat{F}_N\left[x; S(N)\right]$, $\hat{G}_N(x) = \hat{G}_N\left[x; S(N)\right]$ *and* $\hat{F}_N^{-1}(x)$ *be defined as in* (2.5) - (2.8), *and set*

$$\hat{p}_N(x) = \frac{N\hat{G}_N(x)+1}{N+1}\,. \tag{2.14}$$

Then

$$\mathsf{P}\left[\hat{G}_N(S_0) \leq \alpha_1\right] = \mathsf{P}\left[\hat{F}_N(S_0) \geq 1-\alpha_1\right] = \frac{I\left[\alpha_1 N\right]+1}{N+1}\,,\ \textit{for } 0 \leq \alpha_1 \leq 1\,, \tag{2.15}$$

$$\mathsf{P}\left[S_0 \geq \hat{F}_N^{-1}(1-\alpha_1)\right] = \frac{I\left[\alpha_1 N\right]+1}{N+1}\,,\ \textit{for } 0 < \alpha_1 < 1\,, \tag{2.16}$$

and

$$\mathsf{P}\left[\hat{p}_N(S_0) \leq \alpha\right] = \frac{I\left[\alpha(N+1)\right]}{N+1}\,,\ \textit{for } 0 \leq \alpha \leq 1\,. \tag{2.17}$$

The latter proposition can be used as follows: choose α_1 and N so that

$$\alpha = \frac{I\left[\alpha_1 N\right]+1}{N+1} \tag{2.18}$$

is the desired significance level. Provided N is reasonably large, α_1 will be very close to α; in particular, if $\alpha(N+1)$ is an integer, we can take

$$\alpha_1 = \alpha - \frac{(1-\alpha)}{N}\,,$$

in which case we see easily that the critical region $\hat{G}_N(S_0) \leq \alpha_1$ is equivalent to $\hat{G}_N(S_0) < \alpha$. For $0 < \alpha < 1$, the randomized critical region $S_0 \geq \hat{F}_N^{-1}(1-\alpha_1)$ has the same level (α) as the nonrandomized critical region $S_0 \geq F^{-1}(1-\alpha)$, or equivalently the critical regions $\hat{p}_N(S_0) \leq \alpha$ and $\hat{G}_N(S_0) \leq \alpha_1$ have the same level as the critical region $G(S_0) \equiv 1 - F(S_0) \leq \alpha$.

2.2 Monte Carlo tests based on general statistics

The assumption (2.13), which states that ties have zero probability, plays an important role in proving Proposition **2.1.2**. However, it is possible to prove analogous results for general sequences of exchangeable random variables (which may exhibit ties with positive probability), provided we consider a properly randomized empirical distribution function. For this purpose, we introduce randomized ranks which are obtained like ordinary ranks except that ties are "broken" according to a uniform distribution. More precisely, let us associate with each variable S_j, $j = 1, \ldots, N$, a random variable U_j, $j = 1, \ldots, N$ such that

$$U_1, \ldots, U_N \overset{i.i.d.}{\sim} U(0, 1) , \tag{2.19}$$

$U(N) = (U_1, \ldots, U_N)'$ is independent of $S(N) = (S_1, \ldots, S_N)'$ where $U(0,1)$ is the uniform distribution on the interval $(0,1)$. Then, we consider the pairs

$$Z_j = (S_j, U_j) , \quad j = 1, \ldots, N , \tag{2.20}$$

which are ordered according to the lexicographic order:

$$(S_i, U_i) \geq (S_j, U_j) \Longleftrightarrow [S_i > S_j \text{ or } (S_i = S_j \text{ and } U_i \geq U_j)] . \tag{2.21}$$

After defining the indicator function

$$W[(x_1, u_1) - (x_2, u_2)] = [1 - s(x_2 - x_1)] + \delta(x_1 - x_2)\, s(u_1 - u_2) \tag{2.22}$$

where $\delta(x) = 1_{\{0\}}(x)$, it is easy to see that

$$(S_i, U_i) \geq (S_j, U_j) \Longleftrightarrow W[(S_i, U_i) - (S_j, U_j)] = 1 . \tag{2.23}$$

We now order $S_1, \ldots, S_N$ according to the ranking of the pairs $Z_1, \ldots, Z_N$ by (2.21) yielding the following "randomized ranks":

$$\begin{aligned} \tilde{R}_j[S(N), U(N)] &= \sum_{i=1}^{N} W[(S_j, U_j) - (S_i, U_i)] \\ &= N - \sum_{i=1}^{N} s(S_i - S_j) + \sum_{i=1}^{N} \delta(S_j - S_i) s(U_j - U_i) , \end{aligned} \tag{2.24}$$

$j = 1, \ldots, N$. By the continuity of the uniform distribution, the ranks $\tilde{R}_j = \tilde{R}_j[S(N), U(N)]$, $j = 1, \ldots, N$, are all distinct with probability 1, so that the randomized rank vector $(\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_N)'$ is a permutation of $(1, 2, \ldots, N)'$ with probability 1. Furthermore when $S_j \neq S_i$ for all $j \neq i$, we have $\tilde{R}_j = R_j$: provided (2.13) holds, we have $\tilde{R}_j = R_j$, $j = 1, \ldots, N$, with probability 1 [where R_j is defined in (2.9)]. We can now state the following extension of Lemma **2.1.1**.

2.2.1 Lemma DISTRIBUTION OF RANDOMIZED RANKS. *Let* $y(N) = (y_1, \ldots, y_N)'$ *be a* $N \times 1$

vector of exchangeable real random variables and let $\tilde{R}_j = \tilde{R}_j [y(N), U(N)]$ *be defined as in* (2.24) *where* $U(N) = (U_1, \ldots, U_N)'$ *is a vector of* $i.i.d.$ $U(0, 1)$ *variables independent of* $y(N)$ *. Then, for* $j = 1, \ldots, N$,

$$\mathsf{P}[\tilde{R}_j/N \leq x] = I[xN]/N, \text{ for } 0 \leq x \leq 1\,, \tag{2.25}$$

and

$$\begin{aligned} \mathsf{P}[\tilde{R}_j/N \geq x] &= 1\,, && \text{if } x \leq 0\,, \\ &= (I[(1-x)N]+1)/N\,, && \text{if } 0 < x \leq 1\,, \\ &= 0\,, && \text{if } x > 1\,. \end{aligned} \tag{2.26}$$

To the above randomized rankings, it is natural to associate the following randomized empirical (pseudo-)distribution function:

$$\begin{aligned} \tilde{F}_N [x; U_0, S(N), U(N)] &= \frac{1}{N}\sum_{i=1}^{N} W[(x, U_0) - (S_i, U_i)] \\ &= 1 - \frac{1}{N}\sum_{i=1}^{N} s(S_i - x) + \frac{1}{N}\sum_{i=1}^{N} \delta(x - S_i)\, s(U_0 - U_i) \\ &= 1 - \hat{G}_N [x; S(N)] + T_N [x; U_0, S(N), U(N)] \end{aligned} \tag{2.27}$$

where U_0 is a $U(0,1)$ random variable independent of $S(N)$ and $U(N)$,

$$\begin{aligned} T_N [x; U_0, S(N), U(N)] &= \frac{1}{N}\sum_{i=1}^{N} \delta(x - S_i)\, s(U_0 - U_i) \\ &= \frac{1}{N}\sum_{i \in E_N(x)} s(U_0 - U_i) \end{aligned} \tag{2.28}$$

and $E_N(x) = \{i : S_i = x,\ 1 \leq i \leq N\}$. The function $\tilde{F}_N [x; \cdot]$ retains all the properties of a probability distribution function, except for the fact that it may not be right continuous at some of its jump points (where it may take values between its right and left limits). We can also define the corresponding tail area function:

$$\begin{aligned} \tilde{G}_N [x; U_0, S(N), U(N)] &= \frac{1}{N}\sum_{i=1}^{N} W[(S_i, U_i) - (x, U_0)] \\ &= 1 - \frac{1}{N}\sum_{i=1}^{N} s(x - S_i) + \frac{1}{N}\sum_{i=1}^{N} \delta(S_i - x)\, s(U_i - U_0) \\ &= 1 - \hat{F}_N [x; S(N)] + \bar{T}_N [x; U_0, S(N), U(N)] \end{aligned} \tag{2.29}$$

where

$$\begin{aligned} \bar{T}_N [x; U_0, S(N), U(N)] &= \frac{1}{N}\sum_{i=1}^{N} \delta(S_i - x)\, s(U_i - U_0) \\ &= \frac{1}{N}\sum_{i \in E_N(x)} s(U_i - U_0)\ . \end{aligned} \tag{2.30}$$

From (2.27) – (2.30), we see that the following inequalities must hold:

$$1 - \hat{G}_N[x;\, S(N)] \leq \tilde{F}_N[x;\, U_0,\, S(N),\, U(N)] \leq \hat{F}_N[x;\, S(N)] \;, \tag{2.31}$$

$$1 - \hat{F}_N[x;\, S(N)] \leq \tilde{G}_N[x;\, U_0,\, S(N),\, U(N)] \leq \hat{G}_N[x;\, S(N)] \;. \tag{2.32}$$

When no element of $S(N)$ is equal to x [*i.e.*, when $E_N(x)$ is empty], we have:

$$\begin{aligned} \tilde{G}_N[x;\, U_0,\, S(N),\, U(N)] &= \hat{G}_N[x;\, S(N)] = 1 - \hat{F}_N[x;\, S(N)] \\ &= 1 - \tilde{F}_N[x;\, U_0,\, S(N),\, U(N)] \;. \end{aligned} \tag{2.33}$$

Using the above observations, it is then easy to establish the following proposition.

2.2.2 Proposition VALIDITY OF MONTE CARLO TESTS FOR GENERAL STATISTICS. *Let* $(S_0, S_1, \ldots, S_N)'$ *be a* $(N+1) \times 1$ *vector of exchangeable real random variables, let* $(U_0, U_1, \ldots, U_N)'$ *be a* $(N+1) \times 1$ *vector of* $i.i.d.$ $U(0, 1)$ *random variables independent of* $(S_0, S_1, \ldots, S_N)'$*, let* $\hat{F}_N(x) \equiv \hat{F}_N[x;\, S(N)]$*,* $\hat{G}_N(x) \equiv \hat{G}_N[x;\, S(N)]$*,* $\tilde{F}_N(x) \equiv \tilde{F}_N[x;\, U_0,\, S(N), U(N)]$ *and* $\tilde{G}_N(x) \equiv \tilde{G}_N[x;\, U_0,\, S(N),\, U(N)]$ *be defined as in* (2.6) *–* (2.8) *and* (2.27) *–* (2.29)*, with* $S(N) = (S_1, \ldots, S_N)$ *and* $U(N) = (U_1, \ldots, U_N)$ *, and let*

$$\tilde{p}_N(x) = \frac{N\tilde{G}_N(x) + 1}{N+1} \;. \tag{2.34}$$

Then for $0 \leq \alpha_1 \leq 1$,

$$\begin{aligned} \mathsf{P}[\hat{G}_N(S_0) \leq \alpha_1] &\leq \mathsf{P}[\tilde{G}_N(S_0) \leq \alpha_1] = \mathsf{P}[\tilde{F}_N(S_0) \geq 1 - \alpha_1] \\ &= \frac{I[\alpha_1 N] + 1}{N+1} \leq \mathsf{P}[\hat{F}_N(S_0) \geq 1 - \alpha_1] \end{aligned} \tag{2.35}$$

with $\mathsf{P}[\hat{F}_N(S_0) \geq 1 - \alpha_1] = \mathsf{P}[S_0 \geq \hat{F}_N^{-1}(1 - \alpha_1)]$ *for* $0 < \alpha_1 < 1$*, and defining* $\hat{p}_N(x)$ *as in* (2.14),

$$\mathsf{P}[\hat{p}_N(S_0) \leq \alpha] \leq \mathsf{P}[\tilde{p}_N(S_0) \leq \alpha] = \frac{I[\alpha(N+1)]}{N+1} \;, \text{ for } 0 \leq \alpha \leq 1 \;. \tag{2.36}$$

3 Power functions and concordance probabilities

The procedures described above are *randomized* in the sense that the result of the tests depend on auxiliary simulations. This raises the issue of the sensitivity of the results to these simulations. To study this more closely, let us suppose that:

$$\begin{aligned} &S_0, S_1, \ldots, S_N \text{ are independent with} \\ &\mathsf{P}[S_i \leq x] = F(x) \;,\; \mathsf{P}[S_i \geq x] = G(x) \;,\; \mathsf{P}[S_i = x] = g(x) \;,\; i = 1, \ldots, N \;, \\ &\mathsf{P}[S_0 \leq x] = H(x) \,, \mathsf{P}[S_0 \geq x] = K(x) \;. \end{aligned} \tag{3.1}$$

Then we can compute the conditional probability given (S_0, U_0) of the critical region $\tilde{G}(S_0) \leq \alpha_1$:

$$
\begin{aligned}
\mathsf{P}[\tilde{G}_N(S_0) \leq \alpha_1 \mid (S_0, U_0)] &= \mathsf{P}\left[\sum_{i=1}^{N} W[(S_i, U_i) - (S_0, U_0)] \leq I[\alpha_1 N] \mid (S_0, U_0)\right] \\
&= \mathsf{P}[Bi\left(N, \bar{G}(S_0, U_0)\right) \leq I[\alpha_1 N] \mid (S_0, U_0)] \\
&= \sum_{k=0}^{I[\alpha_1 N]} \binom{N}{k} \bar{G}(S_0, U_0)^k [1 - \bar{G}(S_0, U_0)]^{N-k} \qquad (3.2)
\end{aligned}
$$

where $\binom{N}{k} = N!/[k!(N-k)!]$,

$$
\begin{aligned}
\bar{G}(x, u) &= \mathsf{P}(W[(S_i, U_i) - (x, u)] = 1) \\
&= \mathsf{P}(S_i > x) + \mathsf{P}(S_i = x)\,\mathsf{P}(U_i \geq u) \\
&= 1 - F(x) + g(x)(1 - u) \qquad (3.3)
\end{aligned}
$$

and $Bi(N, p)$ represents a binomial random variable with number of successes N and probability of "success" p. Similarly, we can also write:

$$
\begin{aligned}
\mathsf{P}[\hat{G}_N(S_0) \leq \alpha_1 \mid S_0] &= \mathsf{P}[Bi(N, G(S_0)) \leq I[\alpha_1 N] \mid S_0] \\
&= \sum_{k=0}^{I[\alpha_1 N]} \binom{N}{k} G(S_0)^k (1 - G(S_0))^{N-k} \,. \qquad (3.4)
\end{aligned}
$$

When $F(x)$ is continuous, so that $g(x) = 0$, we have:

$$
\begin{aligned}
\mathsf{P}[\tilde{G}(S_0) \leq \alpha_1 \mid (S_0, U_0)] &= \mathsf{P}[\hat{G}_N(S_0) \leq \alpha_1 \mid S_0] \\
&= \sum_{k=0}^{I[\alpha_1 N]} \binom{N}{k} [1 - F(S_0)]^k F(S_0)^{N-k} \,. \qquad (3.5)
\end{aligned}
$$

Using (3.2), we can find a close-form expression for the power of the randomized test $\tilde{G}(S_0) \leq \alpha_1$ for any null hypothesis which entails that S_0 has the distribution $F(\cdot)$ against an alternative under which its distribution is $H(\cdot)$:

$$
\begin{aligned}
\mathsf{P}[\tilde{G}_N(S_0) \leq \alpha_1] &= \underset{(S_0, U_0)}{\mathsf{E}}\{\mathsf{P}[\tilde{G}_N(S_0) \leq \alpha_1 \mid (S_0, U_0)]\} \\
&= \sum_{k=0}^{I[\alpha_1 N]} \binom{N}{k} \int \int_0^1 \bar{G}(x, u)^k [1 - \bar{G}(x, u)]^{N-k} du\, dH(x) \,. \qquad (3.6)
\end{aligned}
$$

Furthermore, when $F(x)$ is continuous everywhere, the latter expression simplifies and we can write:

$$
\begin{aligned}
\mathsf{P}[\tilde{G}_N(S_0) \leq \alpha_1] &= \mathsf{P}[\hat{G}_N(S_0) \leq \alpha_1] \\
&= \sum_{k=0}^{I[\alpha_1 N]} \binom{N}{k} \int [1 - F(x)]^k F(x)^{N-k} \, dH(x) \,. \qquad (3.7)
\end{aligned}
$$

The above formulae will be useful in establishing the validity of simplified asymptotic Monte Carlo tests in the presence of nuisance parameters. They also allow one to compute the probability that the result of the randomized test $\tilde{G}_N(S_0) \leq \alpha_1$ be different of the result of the corresponding nonrandomized test $G(S_0) \leq \alpha \equiv (I[N\alpha_1]+1)/(N+1)$. For example, let $\hat{\alpha}_0 = G(S_0)$ the "p-value" one would obtain if the function $G(x)$ were easy to compute (the p-value of the "fundamental test"). The latter is generally different from the p-value $\tilde{p}_N(S_0)$ or $\hat{p}_N(S_0)$ obtained from a Monte Carlo test based on $S_1, \ldots, S_N$. An interesting question here is the probability that the Monte Carlo test yields a conclusion different from the one based on $\hat{\alpha}_0$. To study this, we shall consider the test which rejects the null hypothesis H_0 when $\hat{p}_N(S_0) \leq \alpha$ under the assumptions (3.1).

If $\hat{\alpha}_0 > \alpha$ (in which case H_0 is not rejected at level α by the fundamental test), the probability that H_0 be rejected at level α_0 is

$$\begin{aligned} \mathsf{P}\left[\hat{p}_N(S_0) \leq \alpha_0 \mid S_0\right] &= \mathsf{P}[N\hat{G}_N(S_0) \leq (N+1)\alpha_0 - 1 \mid S_0] \\ &= \mathsf{P}\left[Bi(N, \hat{\alpha}_0) \leq (N+1)\alpha_0 - 1 \mid S_0\right] \\ &\leq \mathsf{P}\left[Bi(N, \alpha) \leq (N+1)\alpha_0 - 1\right] \\ &= \mathsf{P}\left[\frac{Bi(N, \alpha) - N\alpha}{[N\alpha(1-\alpha)]^{1/2}} \leq \frac{N(\alpha_0 - \alpha) - (1-\alpha_0)}{[N\alpha(1-\alpha)]^{1/2}}\right] \end{aligned} \tag{3.8}$$

where the inequality follows on observing that $\hat{\alpha}_0 > \alpha$. From (3.8), we can bound the probability that a Monte Carlo p-value as low as α_0 be obtained when the fundamental test is not significant at level α. In particular, for $\alpha_0 < \alpha$, this probability decreases as the difference $|\alpha_0 - \alpha|$ and N get larger. It is also interesting to observe that

$$\lim_{N\to\infty} \mathsf{P}\left[\hat{p}_N(S_0) \leq \alpha_0 \mid S_0\right] = \lim_{N\to\infty} \mathsf{P}\left[\frac{Bi(N,\alpha) - N\alpha}{[N\alpha(1-\alpha)]^{1/2}} \leq \frac{N(\alpha_0 - \alpha) - (1-\alpha_0)}{[N\alpha(1-\alpha)]^{1/2}}\right] = 0 \tag{3.9}$$

for $\alpha_0 < \alpha$, so that the probability of a discrepancy between the fundamental test and the Monte Carlo test goes to zero as N increases.

Similarly, for $\hat{\alpha}_0 < \alpha$ (in which case H_0 is rejected at level α by the fundamental test), the probability that H_0 not be rejected at level α_0 is

$$\begin{aligned} \mathsf{P}\left[\hat{p}_N(S_0) > \alpha_0 \mid S_0\right] &= \mathsf{P}\left[Bi(N, \hat{\alpha}_0) > (N+1)\alpha_0 - 1 \mid S_0\right] \\ &\leq \mathsf{P}\left[\frac{Bi(N, \alpha) - N\alpha}{[N\alpha(1-\alpha)]^{1/2}} \leq \frac{N(\alpha_0 - \alpha) - (1-\alpha_0)}{[N\alpha(1-\alpha)]^{1/2}}\right] \end{aligned} \tag{3.10}$$

hence

$$\lim_{N\to\infty} \mathsf{P}\left[\hat{p}_N(S_0) \geq \alpha_0 \mid S_0\right] = 0\,, \text{ for } \alpha_0 > \alpha\,. \tag{3.11}$$

(3.10) gives an upper bound on the probability of observing a p-value as high as α_0 when the fundamental test is significant at a level lower than α. Again, the probability of a discrepancy between the fundamental test and the Monte Carlo test goes to zero as N increases. The only case where the probability of a discrepancy between the two tests does not go to zero as $N \to \infty$ is when

$\hat{\alpha}_0 = \alpha$ (an event with probability zero for statistics with continuous distributions).

The probabilities (3.8) and (3.10) may be computed *a posteriori* to assess the probability of obtaining p-values as low (or as high) as $\hat{p}_N(S_0)$ when the result of the corresponding fundamental test is actually not significant (or significant) at level α. Note also that similar (although somewhat different) calculations may be used to determine the number N of simulations that will ensure a given probability of concordance between the fundamental and the Monte Carlo test [see Marriott (1979)].

4 Monte Carlo tests with nuisance parameters

We will now study the case where the distribution of the test statistic S depends on nuisance parameters. We consider a family of probability spaces $\{(\mathcal{Z}, \mathcal{A}_{\mathcal{Z}}, \mathsf{P}_\theta) : \theta \in \Omega\}$ and suppose that S is a real valued $\mathcal{A}_{\mathcal{Z}}$ -measurable function whose distribution is determined by $\mathsf{P}_{\bar{\theta}}$ [*i.e.*, $\bar{\theta}$ is the "true" parameter vector]. We wish to test the hypothesis

$$H_0 : \bar{\theta} \in \Omega_0 \tag{4.1}$$

where Ω_0 is a nonempty subset of Ω. Again we take a critical region of the form $S \geq c$, where c is a constant which does not depend on θ. The critical region $S \geq c$ has *level* α if and only if

$$\mathsf{P}_\theta [S \geq c] \leq \alpha, \ \forall \theta \in \Omega_0, \tag{4.2}$$

or equivalently,

$$\sup_{\theta \in \Omega_0} \mathsf{P}_\theta [S \geq c] \leq \alpha. \tag{4.3}$$

Furthermore, $S \geq c$ has *size* α when

$$\sup_{\theta \in \Omega_0} \mathsf{P}_\theta [S \geq c] = \alpha. \tag{4.4}$$

If we define the distribution and p-value functions,

$$F[x \mid \theta] = \mathsf{P}_\theta [S \leq x], \ x \in \overline{\mathbb{R}}, \tag{4.5}$$

$$G[x \mid \theta] = \mathsf{P}_\theta [S \geq x], \ x \in \overline{\mathbb{R}}, \tag{4.6}$$

where $\theta \in \Omega$, it is again easy to see that the critical regions

$$\sup_{\theta \in \Omega_0} G[S \mid \theta] \leq \alpha(c), \tag{4.7}$$

where $\alpha(c) \equiv \sup_{\theta \in \Omega_0} G[c \mid \theta]$, and

$$S \geq \sup_{\theta \in \Omega_0} F^{-1}\left[(1 - G[c \mid \theta])^+ \mid \theta\right] \equiv \bar{c} \tag{4.8}$$

are equivalent to $S \geq c$ in the sense that $c \leq \bar{c}$, with equality holding when $F[x \mid \theta]$ is discontinuous at $x = c$ for all $\theta \in \Omega_0$ or both $F[x \mid \theta]$ and $F^{-1}[q \mid \theta]$ are continuous at $x = c$ and $q = F(c)$ respectively for all $\theta \in \Omega_0$, and

$$\sup_{\theta \in \Omega_0} \mathsf{P}_\theta [S \geq \bar{c}] \leq \sup_{\theta \in \Omega_0} \mathsf{P}_\theta [S \geq c] = \sup_{\theta \in \Omega_0} \mathsf{P}_\theta [\sup \{G[S \mid \theta_0] : \theta_0 \in \Omega_0\} \leq \alpha(c)] \; . \tag{4.9}$$

We shall now extend Proposition **2.1.2** in order to allow for the presence of nuisance parameters. For that purpose, we consider a real random variable S_0 and random vectors of the form

$$S(N, \theta) = (S_1(\theta), \ldots, S_N(\theta))' \, , \; \theta \in \Omega \, , \tag{4.10}$$

all defined on a common probability space $(\mathcal{Z}, \mathcal{A}_\mathcal{Z}, \mathsf{P})$, such that

$$\begin{aligned} &\text{the variables } S_0, \, S_1(\bar{\theta}), \ldots, S_N(\bar{\theta}) \text{ are exchangeable for some } \bar{\theta} \in \Omega \, , \\ &\text{each one with distribution function } F[x \mid \bar{\theta}] = \mathsf{P}[S_0 \leq x] \; . \end{aligned} \tag{4.11}$$

Typically, S_0 will refer to a test statistic computed from observed data when the true parameter vector is $\bar{\theta}$ (*i.e.*, $\theta = \bar{\theta}$), while $S_1(\theta), \ldots, S_N(\theta)$ will refer to $i.i.d.$ replications of the test statistic obtained independently (*e.g.*, by simulation) under the assumption that the parameter vector is θ (*i.e.*, $\mathsf{P}[S_i(\theta) \leq x] = F[x \mid \theta]$). The notation $S_i(\theta)$ *does not mean* that the value of θ is required for *computing* the test statistic $S_i(\theta)$: it simply indicates that the distribution function of $S_i(\theta)$ is $F[x \mid \theta]$. In parametric models, the statistic S may be simulated by first generating an "observation" vector y according to an equation of the form

$$y = g(\theta, u) \tag{4.12}$$

where u has a known distribution (which can de simulated) and then computing

$$S(\theta) \equiv S[g(\theta, u)] \equiv g_S(\theta, u) \, . \tag{4.13}$$

In such cases, the above assumptions can be interpreted as follows: $S_0 = S[y(\bar{\theta}, u_0)]$ and $S_i(\theta) = S[y(\theta, u_i)]$, $i = 1, \ldots, N$, where the random vectors $u_0, u_1, \ldots, u_N$ are $i.i.d.$ (or exchangeable). Note θ may include the parameters of a disturbance distribution in a model, such as covariance coefficients (or even its complete distribution function), so that the assumption u has a known distribution is not restrictive. Assumptions on the structure of the parameter space Ω (*e.g.*, whether it is finite-dimensional) will however entail real restrictions on the data-generating process.

A more general setup that allows for nonparametric models would consist in assuming that the test statistic null distribution depends on θ only through some transformation $T(y)$ of the observation vector y, which in turn only depends upon θ through some transformation $\theta_* = h(\theta)$, *e.g.* a subvector of θ :

$$T(y) = g[h(\theta), u] = g[\theta_*, u] \, , \; \theta_* \in \Omega_* \tag{4.14}$$

where $\Omega_* = h(\Omega)$, hence

$$S(\theta) = S\left(T(y)\right) = S\left(g[h(\theta),\, u]\right) \equiv g_S[h(\theta),\, u] = g_S[\theta_*,\, u]\,. \tag{4.15}$$

The setup (4.14) - (4.15) allows for reductions of the nuisance parameter space (*e.g.*, through invariance). In particular, nonparametric models may be considered by taking appropriate distribution-free statistics (e.g., test statistics based on signs, ranks, permutations, etc.). What matters for our purpose is the possibility of simulating the test statistic, not necessarily the data themselves.

Let also

$$\hat{F}_N\left[x \mid \theta\right] \equiv \hat{F}_N\left[x;\, S\left(N,\, \theta\right)\right]\,,\ \hat{G}_N\left[x \mid \theta\right] \equiv \hat{G}_N\left[x;\, S\left(N,\, \theta\right)\right]\,, \tag{4.16}$$

$$\hat{p}_N\left[x \mid \theta\right] = \frac{N\hat{G}_N\left[x \mid \theta\right] + 1}{N+1} \tag{4.17}$$

and $\hat{F}_N^{-1}\left[x \mid \theta\right]$ be defined as in (2.5) – (2.8), and suppose the variables

$$\begin{aligned}&\sup\{\hat{G}_N\left[S_0 \mid \theta\right] : \theta \in \Omega_0\} \text{ and } \inf\{\hat{F}_N\left[S_0 \mid \theta\right] : \theta \in \Omega_0\}\\ &\text{are } \mathcal{A}_{\mathcal{Z}}\text{-measurable (where } \phi \neq \Omega_0 \subseteq \Omega)\,.\end{aligned} \tag{4.18}$$

We then get the following proposition.

4.1 Proposition VALIDITY OF MMC TESTS WHEN TIES HAVE ZERO PROBABILITY. *Under the assumptions and notations* (4.10), (4.11) *and* (4.16) - (4.18), *set* $S_0\left(\bar{\theta}\right) = S_0$ *and suppose that*

$$\mathsf{P}\left[S_i\left(\bar{\theta}\right) = S_j\left(\bar{\theta}\right)\right] = 0 \text{ for } i \neq j\,,\ i, j = 0, 1, \ldots, N\,. \tag{4.19}$$

If $\bar{\theta} \in \Omega_0$*, then for* $0 \leq \alpha_1 \leq 1$,

$$\begin{aligned}\mathsf{P}\left[\sup\{\hat{G}_N\left[S_0 \mid \theta\right] : \theta \in \Omega_0\} \leq \alpha_1\right] &\leq \mathsf{P}\left[\inf\{\hat{F}_N\left[S_0 \mid \theta\right] : \theta \in \Omega_0\} \geq 1 - \alpha_1\right]\\ &\leq \frac{I\left[\alpha_1 N\right] + 1}{N+1}\,.\end{aligned} \tag{4.20}$$

where $\mathsf{P}[\inf\{\hat{F}_N\left[S_0 \mid \theta\right] : \theta \in \Omega_0\}] \geq 1 - \alpha_1 = \mathsf{P}[S_0 \geq \sup\{\hat{F}_N^{-1}\left[1 - \alpha_1 \mid \theta\right] : \theta \in \Omega_0\}]$ *for* $0 < \alpha_1 < 1$*, and*

$$\mathsf{P}\left[\sup\left\{\hat{p}_N\left[S_0 \mid \theta\right] : \theta \in \Omega_0\right\} \leq \alpha\right] \leq \frac{I\left[\alpha\left(N+1\right)\right]}{N+1}\,,\ \text{for } 0 \leq \alpha \leq 1\,. \tag{4.21}$$

Following the latter proposition, if we choose α_1 and N so that (2.18) holds, the critical region $\sup\{\hat{G}_N\left[S_0 \mid \theta\right] : \theta \in \Omega_0\} \leq \alpha_1$ has level α irrespective of the presence of nuisance parameters in the distribution of the test statistic S under the null hypothesis $H_0 : \bar{\theta} \in \Omega_0$. The same also holds if we use the (almost) equivalent critical regions $\inf\{\hat{F}_N\left[S_0 \mid \theta\right] : \theta \in \Omega_0\} \geq 1 - \alpha_1$ or $S_0 \geq \sup\{\hat{F}_N^{-1}\left[1 - \alpha_1 \mid \theta\right] : \theta \in \Omega_0\}$. We shall call such tests maximized Monte Carlo (MMC) tests.

To be more explicit, if $S(\theta)$ is generated according to expressions of the form (4.14) - (4.15), we have

$$\begin{aligned}\hat{G}_N\left[S_0 \mid \theta\right] &= \frac{1}{N}\sum_{i=1}^{N} s\left(S_i\left(\theta\right) - S_0\right) = \frac{1}{N}\sum_{i=1}^{N} s\left(S\left(g[h(\theta),\, u_i]\right) - S_0\right) \\ &= \frac{1}{N}\sum_{i=1}^{N} s\left(g_S[\theta_*,\, u_i] - S_0\right) . \end{aligned} \tag{4.22}$$

The function $\hat{G}_N\left[S_0 \mid \theta\right]$ (or $\hat{p}_N\left[S_0 \mid \theta\right]$), is then maximized with respect to $\theta \in \Omega_0$ [or equivalently, with respect to $\theta_* \in \Omega_{0*} = h(\Omega_0)$] keeping the observed statistic S_0 and the simulated disturbance vectors $u_1, \ldots, u_N$ fixed. The function $\hat{G}_N\left[S_0 \mid \theta\right]$ is a step-type function which typically has zero derivatives almost everywhere, except on isolated points (or manifolds) where it is not differentiable. So it cannot be maximized with usual derivative-based algorithms. However, the required maximizations can be performed by using appropriate optimization algorithms that do not require differentiability, such as *simulated annealing*. For further discussion of such algorithms, the reader may consult Goffe, Ferrier, and Rogers (1994).

It is easy to extend Proposition **4.1** in order to relax the no-tie assumption (4.19). For that purpose, we generate as in Proposition **2.2.2** a vector $(U_0, U_1, \ldots, U_N)$ of $N+1$ $i.i.d.$ $U\left(0,\, 1\right)$ random variables independent of $S_0,\, S_i\left(\bar{\theta}\right), \ldots\, S_N\left(\bar{\theta}\right)$, and we consider the corresponding randomized distribution, tail area and p-value functions:

$$\tilde{F}_N\left[x \mid \theta\right] \equiv \tilde{F}_N\left[x;\, U_0,\, S\left(N,\, \theta\right),\, U\left(N\right)\right] , \tag{4.23}$$

$$\tilde{G}_N\left[x \mid \theta\right] \equiv \tilde{G}_N\left[x;\, U_0,\, S\left(N,\, \theta\right),\, U\left(N\right)\right] ,\ \tilde{p}_N\left[x \mid \theta\right] = \frac{N\tilde{G}_N\left[x \mid \theta\right] + 1}{N+1} \tag{4.24}$$

where

$$U\left(N\right) = \left(U_1,\, \ldots,\, U_N\right) \text{ , and } U_0,\, U_1,\, \ldots,\, U_N \overset{i.i.d.}{\sim} U\left(0, 1\right) . \tag{4.25}$$

Under the corresponding measurability assumption

$$\begin{aligned}&\sup\{\tilde{G}_N\left[S_0 \mid \theta\right] : \theta \in \Omega_0\} \text{ and } \inf\{\tilde{F}_N\left[S_0 \mid \theta\right] : \theta \in \Omega_0\} \\ &\text{are } \mathcal{A}_{\mathcal{Z}} \text{ -measurable (where } \phi \neq \Omega_0 \subseteq \Omega) ,\end{aligned} \tag{4.26}$$

we can then state the following generalization of Proposition **4.2**.

4.2 Proposition VALIDITY OF MMC TESTS FOR GENERAL STATISTICS. *Under the assumptions and notations* (4.10), (4.11), (4.16) - (4.18) *and* (4.23) - (4.26), *suppose* $U_0, U_1, \ldots, U_N$ *are independent of* $S_0,\, S_1\left(\bar{\theta}\right), \ldots, S_N\left(\bar{\theta}\right)$. *If* $\bar{\theta} \in \Omega_0$, *then for* $0 \leq \alpha_1 \leq 1$,

$$\begin{aligned}\mathsf{P}\left[\sup\{\hat{G}_N\left[S_0 \mid \theta\right] : \theta \in \Omega_0\} \leq \alpha_1\right] &\leq \mathsf{P}\left[\sup\{\tilde{G}_N\left[S_0 \mid \theta\right] : \theta \in \Omega_0\} \leq \alpha_1\right] \\ &\leq \frac{I\left[\alpha_1 N\right] + 1}{N+1} ,\end{aligned} \tag{4.27}$$

$$
\begin{aligned}
\mathsf{P}\left[\sup\{\hat{G}_N\left[S_0 \mid \theta\right] : \theta \in \Omega_0\} \leq \alpha_1\right] &\leq \mathsf{P}\left[\sup\{\tilde{F}_N\left[S_0 \mid \theta\right] : \theta \in \Omega_0\} \geq 1 - \alpha_1\right] \\
&\leq \frac{I\left[\alpha_1 N\right] + 1}{N + 1}\,, \qquad (4.28)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathsf{P}\left[\sup\left\{\hat{p}_N\left[S_0 \mid \theta\right] : \theta \in \Omega_0\right\} \leq \alpha\right] &\leq \mathsf{P}\left[\sup\left\{\tilde{p}_N\left[S_0 \mid \theta\right] : \theta \in \Omega_0\right\} \leq \alpha\right] \\
&\leq \frac{I\left[\alpha\left(N+1\right)\right]}{N + 1}\,, \ \textit{for}\ 0 \leq \alpha \leq 1\,. \qquad (4.29)
\end{aligned}
$$

5 Asymptotic Monte Carlo tests based on consistent set estimators

In this section, we propose simplified approximate versions of the procedures proposed in the previous section when a consistent point or set estimate of θ is available. To do this, we shall need to reformulate the setup used previously in order to allow for an increasing sample size.

Consider

$$
S_{T0},\, S_{T1}\left(\theta\right),\, \ldots,\, S_{TN}\left(\theta\right),\, T \geq I_0\,,\, \theta \in \Omega\,, \qquad (5.1)
$$

real random variables all defined on a common probability space $(\mathcal{Z}\,,\, \mathcal{A}_{\mathcal{Z}}\,,\, \mathsf{P})$, and set

$$
S_T\left(N,\, \theta\right) = \left(S_{T1}\left(\theta\right),\, \ldots,\, S_{TN}\left(\theta\right)\right)\ , T \geq I_0\,. \qquad (5.2)
$$

We will be primarily interested by situations where

$$
\begin{aligned}
&\text{the variables } S_{T0},\, S_{T1}\left(\bar{\theta}\right),\, \ldots,\, S_{TN}\left(\bar{\theta}\right) \text{ are exchangeable for some } \bar{\theta} \in \Omega\,, \\
&\text{each one with distribution function } F_T\left[x \mid \bar{\theta}\right] = \mathsf{P}\left[S_{T0} \leq x\right]\,. \qquad (5.3)
\end{aligned}
$$

Here S_{T0} will normally refer to a test statistic with distribution function $F_T\left[\cdot \mid \theta\right]$ based on a sample of size T, while $S_{T1}\left(\theta\right),\, \ldots,\, S_{TN}\left(\theta\right)$ are $i.i.d.$ replications of the same test statistic obtained independently under the assumption that the parameter vector is θ (*i.e.*, $\mathsf{P}\left[S_{Ti}\left(\theta\right) \leq x\right] = F_T\left[x \mid \theta\right],\, i = 1,\, \ldots,\, N$). Let also

$$
\hat{F}_{TN}\left[x \mid \theta\right] = \hat{F}_N\left[x;\, S_T\left(N,\, \theta\right)\right]\,,\ \hat{G}_{TN}\left[x \mid \theta\right] = \hat{G}_N\left[x;\, S_T\left(N,\, \theta\right)\right]\,, \qquad (5.4)
$$

$$
\hat{p}_{TN}\left[x \mid \theta\right] = \frac{N\hat{F}_{TN}\left[x \mid \theta\right] + 1}{N + 1} \qquad (5.5)
$$

and let $\hat{F}_{TN}^{-1}\left[x \mid \theta\right]$ be defined as in (2.5) – (2.8).

We consider first the situation where p-values are maximized over a subset C_T of Ω (*e.g.*, a confidence set for θ) instead of Ω_0. Consequently, we introduce the following assumption:

$$
\begin{aligned}
&C_T, T \geq I_0 \text{ is a sequence of (possibly random) subsets of } \Omega \text{ such that} \\
&\sup\{\hat{G}_{TN}\left[S_{T0} \mid \theta\right] : \theta \in C_T\} \text{ and } \inf\{\hat{F}_{TN}\left[S_{T0} \mid \theta\right] : \theta \in C_T\} \\
&\text{are } \mathcal{A}_{\mathcal{Z}}\text{-measurable, for all } T \geq I_0,\ \text{where } \emptyset \neq \Omega_0 \subseteq \Omega. \qquad (5.6)
\end{aligned}
$$

Then we have the following proposition.

5.1 Proposition ASYMPTOTIC VALIDITY OF CONFIDENCE-SET RESTRICTED MMC TESTS: CONTINUOUS DISTRIBUTIONS. *Under the assumptions and notations* (5.1) *to* (5.6), *set* $S_{T0}\left(\bar{\theta}\right) = S_{T0}$, *suppose*

$$\mathsf{P}\left[S_{Ti}\left(\bar{\theta}\right) = S_{Tj}\left(\bar{\theta}\right)\right] = 0 \text{ for } i \neq j, \text{ and } i, j = 0, 1, \ldots, N\,, \tag{5.7}$$

for all $T \geq I_0$, *and let* $C_T, T \geq I_0$, *be a sequence of (possibly random) subsets of* Ω *such that*

$$\lim_{T \to \infty} \mathsf{P}\left[\bar{\theta} \in C_T\right] = 1\,. \tag{5.8}$$

If $\bar{\theta} \in \Omega_0$, *then*

$$\begin{aligned}
&\lim_{T \to \infty} \mathsf{P}\left[\sup\{\hat{G}_{TN}\left[S_{T0} \mid \theta\right] : \theta \in C_T\} \leq \alpha_1\right] \\
&\quad \leq \lim_{T \to \infty} \mathsf{P}\left[\inf\{\hat{F}_{TN}\left[S_{T0} \mid \theta\right] : \theta \in C_T\} \geq 1 - \alpha_1\right] \\
&\quad = \lim_{T \to \infty} \mathsf{P}\left[S_{T0} \geq \sup\{\hat{F}_{TN}^{-1}\left[1 - \alpha_1 \mid \theta\right] : \theta \in C_T\}\right] \leq \frac{I\left[\alpha_1 N\right] + 1}{N + 1}
\end{aligned} \tag{5.9}$$

and

$$\lim_{T \to \infty} \mathsf{P}\left[\sup\left\{\hat{p}_{TN}\left[S_{T0} \mid \theta\right] : \theta \in C_T\right\} \leq \alpha\right] \leq \frac{I\left[\alpha(N+1)\right]}{N+1}\,, \text{ for } 0 \leq \alpha \leq 1\,. \tag{5.10}$$

It is quite easy to find a consistent set estimate of $\bar{\theta}$ whenever a consistent point estimate $\widehat{\theta}_T$ of $\bar{\theta}$ is available. Suppose $\Omega \subseteq \mathbb{R}^k$ and

$$\lim_{T \to \infty} \mathsf{P}[\parallel \widehat{\theta}_T - \bar{\theta} \parallel < \varepsilon] = 1\,,\ \forall \varepsilon > 0\,, \tag{5.11}$$

where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^k$ [*i.e.*, $\parallel x \parallel = (x'x)^{1/2}\,,\ x \in \mathbb{R}^k$]. Note that condition (5.8) need only hold for the true value $\bar{\theta}$ of the parameter vector θ. Then any set of the form $C_T = \{\theta \in \Omega : \parallel \widehat{\theta}_T - \theta \parallel < c\}$ satisfies (5.8), whenever c is a fixed positive constant that does not depend on T. More generally, if there is a sequence of (possibly random) matrices A_T and a non-negative exponent δ such that

$$\lim_{T \to \infty} \mathsf{P}[T^{\delta} \parallel A_T(\widehat{\theta}_T - \bar{\theta}) \parallel^2 < c] = 1\,,\ \forall c > 0\,, \tag{5.12}$$

then any set of the form

$$\begin{aligned}
C_T &= \{\theta \in \Omega : (\widehat{\theta}_T - \theta)' A_T' A_T (\widehat{\theta}_T - \theta) < c/T^{\delta}\} \\
&= \{\theta \in \Omega : \parallel A_T(\widehat{\theta}_T - \theta) \parallel^2 < c/T^{\delta}\}\,,\ c > 0
\end{aligned} \tag{5.13}$$

satisfies (5.8), since in this case,

$$
\begin{aligned}
\mathsf{P}\left[\bar{\theta} \in C_T\right] &= \mathsf{P}\left[(\widehat{\theta}_T - \bar{\theta})' A_T' A_T (\widehat{\theta}_T - \bar{\theta}) < c/T^{\delta}\right] \\
&= \mathsf{P}\left[T^{\delta}(\widehat{\theta}_T - \bar{\theta})' A_T' A_T (\widehat{\theta}_T - \bar{\theta}) < c\right] \underset{T\to\infty}{\longrightarrow} 1\,.
\end{aligned}
$$

In particular (5.12) will hold whenever we can find $\bar{\delta} > 0$ (*e.g.*, $\bar{\delta} = 1$) such that $T^{\bar{\delta}/2} A_T(\widehat{\theta}_T - \bar{\theta})$ has an asymptotic distribution as $T \to \infty$ and δ is selected so that $0 \leq \delta < \bar{\delta}$. Whenever $\delta > 0$ and $\underset{T\to\infty}{plim}\,(A_T' A_T) = C_0$ with $\det(C_0) \neq 0$, the diameter of the set C_T goes to zero, a fact which can greatly simplifies the evaluation of the variables $\sup\{\hat{G}_{TN}\}$, $\inf\{\hat{F}_{TN}\}$ and $\sup\{\hat{p}_{TN}\}$ in (5.9) - (5.10).

Again, it is possible to extend Proposition **5.1** to statistics with general (possibly discrete) distributions by considering properly randomized distribution, tail area and p-value functions:

$$
\begin{aligned}
\tilde{F}_{TN}\left[x \mid \theta\right] &= \tilde{F}_N\left[x; U_0,\, S_T\left(N;\, \theta\right),\, U\left(N\right)\right], && (5.14)\\
\tilde{G}_{TN}\left[x \mid \theta\right] &= \tilde{G}_N\left[x; U_0,\, S_T\left(N;\, \theta\right),\, U\left(N\right)\right], && (5.15)\\
\tilde{p}_{TN}\left[x \mid \theta\right] &= \frac{N\tilde{G}_{TN}\left[x \mid \theta\right] + 1}{N+1}, && (5.16)
\end{aligned}
$$

where $\tilde{F}_N\left[\cdot\right]$, $\tilde{G}_N\left[\cdot\right]$, U_0 and $U\left(N\right)$ are defined as in (4.23) - (4.25).

5.2 Proposition ASYMPTOTIC VALIDITY OF CONFIDENCE-SET RESTRICTED MMC TESTS: GENERAL DISTRIBUTIONS. *Under the assumptions and notations* (5.1) - (5.6) *and* (5.14) - (5.16), *suppose the sets* $C_T \subseteq \Omega$, $T \geq I_0$, *satisfy* (5.8). *If* $\bar{\theta} \in \Omega_0$, *then for* $0 \leq \alpha_1 \leq 1$ *and* $0 \leq \alpha \leq 1$,

$$
\begin{aligned}
&\lim_{T\to\infty} \mathsf{P}\left[\sup\{\hat{G}_{TN}\left[S_{T0} \mid \theta\right] : \theta \in C_T\} \leq \alpha_1\right] \\
&\quad \leq \lim_{T\to\infty} \mathsf{P}\left[\sup\{\tilde{G}_{TN}\left[S_{T0} \mid \theta\right] : \theta \in C_T\} \leq \alpha_1\right] \leq \frac{I\left[\alpha_1 N\right] + 1}{N+1}, \qquad (5.17)
\end{aligned}
$$

$$
\begin{aligned}
&\lim_{T\to\infty} \mathsf{P}\left[\sup\{\hat{G}_{TN}\left[S_{T0} \mid \theta\right] : \theta \in C_T\} \leq \alpha_1\right] \\
&\quad \leq \lim_{T\to\infty} \mathsf{P}\left[\sup\{\tilde{F}_{TN}\left[S_{T0} \mid \theta\right] : \theta \in C_T\} \geq 1 - \alpha_1\right] \leq \frac{I\left[\alpha_1 N\right] + 1}{N+1}, \qquad (5.18)
\end{aligned}
$$

$$
\begin{aligned}
&\lim_{T\to\infty} \mathsf{P}\left[\sup\{\hat{p}_{TN}\left[S_{T0} \mid \theta\right] : \theta \in C_T\} \leq \alpha\right] \\
&\quad \leq \lim_{T\to\infty} \mathsf{P}\left[\sup\{\tilde{p}_{TN}\left[S_0 \mid \theta\right] : \theta \in C_T\} \leq \alpha\right] \leq \frac{I\left[\alpha(N+1)\right]}{N+1}\,. \qquad (5.19)
\end{aligned}
$$

6 Bootstrap tests based on asymptotically nonpivotal statistics

Parametric bootstrap tests may be interpreted as a degenerate form of the procedures described in Proposition **5.1** and **5.2** where the consistent confidence set C_T has been replaced by a consistent point estimate $\widehat{\theta}_T$. In other words, the distribution of $S_T(\theta)$, $\theta \in \Omega_0$, is simulated at a single point $\widehat{\theta}_T$. It is well known that such bootstrap tests are not generally valid, unless much stronger regularity conditions are imposed. In the following proposition, we extend considerably earlier proofs of the asymptotic validity of such bootstrap tests. In particular, we allow for the presence of nuisance parameters in the asymptotic distribution of the test statistic considered.

To do this, we will need more restrictive assumptions on the distributions of the statistics $S_T(\theta)$ as functions of the parameter vector θ. We shall use four basic assumptions:

$$\begin{aligned}&S_{T1}(\theta), \dots, S_{TN}(\theta) \text{ are } i.i.d. \text{ according to the distribution}\\ &F_T[x \mid \theta] = \mathsf{P}[S_T(\theta) \leq x]\ ,\ \forall \theta \in \Omega\,;\end{aligned} \tag{6.1}$$

$$\Omega \text{ is a nonempty subset of } \mathbb{R}^k\,; \tag{6.2}$$

$$\begin{aligned}&\text{for every } T \geq I_0\,,\ S_{T0} \text{ is a real random variable and } \widehat{\theta}_T \text{ an estimator of } \theta,\\ &\text{both measurable with respect to the probability space } (\mathcal{Z}\,,\,\mathcal{A}_{\mathcal{Z}}\,,\,\mathsf{P})\,,\\ &\text{and } F_T[S_{T0} \mid \widehat{\theta}_T] \text{ is a random variable;}\end{aligned} \tag{6.3}$$

$$\begin{aligned}&\forall \varepsilon_0 > 0\,,\ \forall \varepsilon_1 > 0\,,\ \exists \delta > 0 \text{ and a sequence of open subsets } D_{T0}(\varepsilon_0) \text{ in } \mathbb{R}\\ &\text{such that } \liminf_{T\to\infty} \mathsf{P}[S_{T0} \in D_{T0}(\varepsilon_0)] \geq 1-\varepsilon_0 \text{ and}\\ &\left\|\theta - \bar{\theta}\right\| \leq \delta \Rightarrow \limsup_{T\to\infty}\left[\sup_{x\in D_{T0}(\varepsilon_0)}\left|F_T[x \mid \theta] - F_T\left[x \mid \bar{\theta}\right]\right|\right] \leq \varepsilon_1\,.\end{aligned} \tag{6.4}$$

The first of these four conditions replaces the exchangeability assumption by an assumption of $i.i.d.$ variables. The two next ones simply make appropriate measurability assumptions, while the last one may be interpreted as a local equicontinuity condition $\left(\text{at } \theta = \bar{\theta}\right)$ on the sequence of distribution functions $F_T(x \mid \theta)\,,\,T \geq I_0$. Note that S_{T0} is not assumed to follow the same distribution as the other variables $S_{T1}(\theta), \dots, S_{TN}(\theta)$. Furthermore, S_{T0} and $F_T(x \mid \theta)$ do not necessarily converge to limits as $T \to \infty$. An alternative assumption of interest would consist in assuming that S_{T0} converges in probability ($\xrightarrow{p}$) to a random variable $\bar{S}_0$ as $T \to \infty$, in which case the "global" equicontinuity condition (6.4) can be weakened to a "local" one:

$$S_{T0} \xrightarrow[T\to\infty]{p} \bar{S}_0\,; \tag{6.5}$$

$$D_0 \text{ is subset of } \mathbb{R} \text{ such that } \mathsf{P}\left[\bar{S}_0 \in D_0 \text{ and } S_{T0} \in D_0 \text{ for all } T \geq I_0\right] = 1\,; \tag{6.6}$$

$$\forall x \in D_0\,,\ \forall \varepsilon > 0\,,\ \exists \delta > 0 \text{ and an open neighborhood } B(x, \varepsilon) \text{ of } x \text{ such that}$$

$$\left\|\theta-\bar{\theta}\right\| \leq \delta \Rightarrow \limsup_{T\to\infty}\left[\sup_{y\in B(x,\varepsilon)\cap D_0}\left|F_T\left[y \mid \theta\right]-F_T\left[y \mid \bar{\theta}\right]\right|\right] \leq \varepsilon\,. \tag{6.7}$$

We can now show that Monte Carlo tests obtained by simulating $S_{Ti}(\theta)$, $i = 1, \ldots, N$, with $\theta = \widehat{\theta}_T$ are equivalent for large T to those based on using the true value $\theta = \bar{\theta}$. To do this, it will be convenient to first prove two lemmas.

6.1 Lemma CONTINUITY OF p-VALUE FUNCTION. *Under the assumptions and notations* (5.1), (5.2), (5.4), (5.5), (5.14) - (5.16) *and* (6.1)*, set*

$$Q_{TN}(\theta,\, x,\, u_0,\, \alpha_1) = \mathsf{P}\left[\tilde{G}_{TN}\left[x \mid \theta\right] \leq \alpha_1 \mid U_0 = u_0\right]\,, \tag{6.8}$$

$$\bar{Q}_{TN}(\theta,\, x,\, \alpha_1) = \mathsf{P}\left[\hat{G}_{TN}\left[x \mid \theta\right] \leq \alpha_1\right]\,,\ 0 \leq \alpha_1 \leq 1\,, \tag{6.9}$$

and suppose U_0 is independent of $S_T(N,\theta)$. For any $\theta,\, \bar{\theta} \in \Omega,\, x \in \mathbb{R}$ and $u_0,\, \alpha_1 \in [0,1]$, the inequality

$$\left|F_T\left[y \mid \theta\right] - F_T\left[y \mid \bar{\theta}\right]\right| \leq \varepsilon\,,\ \forall y \in (x-\delta,\ x+\delta)$$

where $\delta > 0$, entails

$$\left|Q_{TN}(\theta,\, x,\, u_0,\, \alpha_1) - Q_{TN}\left(\bar{\theta},\, x,\, u_0,\, \alpha_1\right)\right| \leq 3C(N,\, \alpha_1)\,\varepsilon \tag{6.10}$$

and

$$\left|\bar{Q}_{TN}(\theta,\, x,\, \alpha_1) - \bar{Q}_{TN}\left(\bar{\theta},\, x,\, u_0\right)\right| \leq 3C(N,\, \alpha_1)\,\varepsilon \tag{6.11}$$

where $C(N,\, \alpha_1) = N \sum_{k=0}^{I[\alpha_1 N]} \binom{N}{k}$.

6.2 Lemma CONVERGENCE OF BOOTSTRAP p-VALUES. *Under the assumptions and notations of Lemma* **6.1***, suppose that* (6.2) *and* (6.3) *also hold. If* $\widehat{\theta}_T \underset{T\to\infty}{\longrightarrow} \bar{\theta}$ *in pr. and condition* (6.4) *or* (6.5) - (6.7) *holds, then*

$$\sup_{0\leq u_0\leq 1}\left|Q_{TN}(\widehat{\theta}_T,\, S_{T0},\, u_0,\, \alpha_1) - Q_{TN}\left(\bar{\theta},\, S_{T0},\, u_0,\, \alpha_1\right)\right| \xrightarrow[T\to\infty]{p} 0 \tag{6.12}$$

and

$$\bar{Q}_{TN}(\widehat{\theta}_T,\, S_{T0},\, \alpha_1) - \bar{Q}_{TN}\left(\bar{\theta},\, S_{T0},\, \alpha_1\right) \xrightarrow[T\to\infty]{p} 0\,. \tag{6.13}$$

We can now prove the following proposition.

6.3 Proposition ASYMPTOTIC VALIDITY OF BOOTSTRAP p-VALUES. *Under the assumptions and notations* (5.1), (5.2), (5.4), (5.5), (5.14) - (5.16) *and* (6.1) - (6.3)*, suppose the random variable S_{T0} and the estimator $\widehat{\theta}_T$ are both independent of $S_T(N,\theta)$ and U_0. If* $\underset{T\to\infty}{plim}\,\widehat{\theta}_T = \bar{\theta}$ *and condition*

(6.4) *or* (6.5) - (6.7) *hold, then for* $0 \leq \alpha_1 \leq 1$ *and* $0 \leq \alpha \leq 1$,

$$\lim_{T\to\infty} \left\{ \mathsf{P}\left[\tilde{G}_{TN}[S_{T0} \mid \widehat{\theta}_T] \leq \alpha_1\right] - \mathsf{P}\left[\tilde{G}_{TN}[S_{T0} \mid \bar{\theta}] \leq \alpha_1\right]\right\}$$
$$= \lim_{T\to\infty} \left\{ \mathsf{P}\left[\hat{G}_{TN}[S_{T0} \mid \widehat{\theta}_T] \leq \alpha_1\right] - \mathsf{P}\left[\hat{G}_{TN}[S_{T0} \mid \bar{\theta}] \leq \alpha_1\right]\right\} = 0\,, \tag{6.14}$$

and

$$\lim_{T\to\infty} \left\{ \mathsf{P}\left[\tilde{p}_{TN}[S_{T0} \mid \widehat{\theta}_T] \leq \alpha\right] - \mathsf{P}\left[\tilde{p}_{TN}[S_{T0} \mid \bar{\theta}] \leq \alpha\right]\right\}$$
$$= \lim_{T\to\infty} \left\{ \mathsf{P}[\hat{p}_{TN}[_{T0}| \, \widehat{\theta}_T] \leq \alpha] - \mathsf{P}[\hat{p}_{TN}[S_{T0} \mid \bar{\theta}] \leq \alpha]\right\} = 0\,. \tag{6.15}$$

It is worth noting that condition (6.7) holds whenever $F_T\,[x \mid \theta]$ converges to a distribution function $F_\infty\,[x \mid \theta]$ which is continuous with respect to $(x,\, \theta')'$, for $x \in D_0$, as follows:

$$\begin{aligned}
&\forall x \in D_0\,,\ \forall \varepsilon > 0\,,\ \exists \delta_1 > 0 \text{ and an open neighborhood } B_1\,(x,\, \varepsilon) \text{ of } x \text{ such that}\\
&\left\|\theta - \bar{\theta}\right\| \leq \delta \Rightarrow\ \limsup_{T\to\infty} \left(\sup_{y \in B_1(x,\varepsilon)\cap D_0} \left|F_T\,[y \mid \theta] - F_\infty\,[y \mid \theta]\right| \right) \leq \varepsilon\,;
\end{aligned} \tag{6.16}$$

$$\begin{aligned}
&\forall x \in D_0\,,\ \forall \varepsilon > 0\,,\ \exists \delta_2 > 0 \text{ and an open neighborhood } B_2\,(x, \varepsilon) \text{ of } x \text{ such that}\\
&\left\|\theta - \bar{\theta}\right\| \leq \delta_2 \Rightarrow \sup_{y \in B_2(x,\varepsilon)\cap D_0} \left|F_\infty\,[y \mid \theta] - F_\infty\left[y \mid \bar{\theta}\right]\right| \leq \varepsilon\,.
\end{aligned} \tag{6.17}$$

It is then easy to see that (6.16) - (6.17) entail (6.7) on noting that

$$\begin{aligned}
\left|F_T\,[x \mid \theta] - F_T\left[x \mid \bar{\theta}\right]\right| \ \leq\ & \left|F_T\,[x \mid \theta] - F_\infty\,[x \mid \theta]\right| + \left|F_T\left[x \mid \bar{\theta}\right] - F_\infty\left[x \mid \bar{\theta}\right]\right|\\
& + \left|F_\infty\,[x \mid \theta] - F_\infty\left[x \mid \bar{\theta}\right]\right|\ ,\ \forall x\,.
\end{aligned}$$

Note also that (6.17) holds whenever $F_\infty\,[x \mid \theta]$ is continuous with respect to $(x,\, \theta')'$, although the latter condition is not at all necessary (*e.g.*, in models where D_0 is a discrete set of points). In particular, (6.16) - (6.17) will hold when $F_T\,[x \mid \theta]$ admits an expansion around a pivotal distribution:

$$F_T\,[x \mid \theta] = F_\infty\,(x) + T^{-\gamma} g\,(x,\, \theta) + h_T\,(x,\, \theta) \tag{6.18}$$

where $F_\infty\,(x)$ is a distribution function that does not depend on $\theta, \gamma > 0$, with the following assumptions on $g\,(x,\, \theta)$ and $h_T\,(x, \theta)$:

$$\begin{aligned}
&\forall x \in D_0\,,\ \exists \text{ an open neighborhood } B\left(x,\, \bar{\theta}\right) \text{ of } (x,\, \bar{\theta}')' \text{ such that}\\
&|g\,(y, \theta)| \leq C\,(x,\, \theta)\ ,\ \text{for all } (y, \theta')' \in B\left(x,\, \bar{\theta}\right) \cap D_0\\
&\text{where } C\,(x,\, \theta) \text{ is a positive constant, and}\\
&T^\gamma h_T\,(y, \theta) \underset{T\to\infty}{\longrightarrow} 0 \text{ uniformly on } B\left(x,\, \bar{\theta}\right) \cap D_0\,.
\end{aligned} \tag{6.19}$$

The latter is quite similar (although somewhat weaker) than the assumption considered by Hall and Titterington (1989, eq. (2.5))].

When S_{T0} is distributed like $S_T\left(\bar{\theta}\right)$, *i.e.*, $\mathsf{P}\left[S_{T0} \leq x\right] = F_T\left[x \mid \bar{\theta}\right]$, we can apply Proposition **2.2.2** and see that $\mathsf{P}[\tilde{G}_{TN}[S_{T0} \mid \bar{\theta}] \leq \alpha_1] = \left(I\left[\alpha_1 N\right] + 1\right) / \left(N+1\right)$, hence

$$\lim_{T\to\infty} \mathsf{P}\left[\tilde{G}_{TN}[S_{T0} \mid \hat{\theta}_T] \leq \alpha_1\right] = \frac{I\left[\alpha_1 N\right] + 1}{N+1}\,. \tag{6.20}$$

A Appendix: Proofs

Proof of Lemma 2.1.1 By condition (2.10), the variables $y_{1,}\, y_2,\, \ldots, y_N$ are all distinct with probability 1, and the rank vector $(R_1,\, R_2,\, \ldots,\, R_N)'$ is with probability 1 a permutation of the integers $(1,\, 2,\, \ldots,\, N)$. Furthermore, since the variables $y_1,\, y_2,\, \ldots, y_N$ are exchangeable, the $N!$ distinct permutations of $(y_1,\, y_2,\, \ldots,\, y_N)$ have the same probability $1/N!$. Consequently, we have

$$\mathsf{P}\left[R_j = i\right] = \frac{1}{N}\,,\ i = 1,\, 2,\, \ldots,\, N\ ,$$

and

$$\mathsf{P}\left[R_j/N \le x\right] = I\left[xN\right]/N\,,\ 0 \le x \le 1\,,$$

from which (2.11) follows and

$$\begin{aligned} \mathsf{P}\left[R_j/N < x\right] \quad &= \left(I\left[xN\right] - 1\right)/N\,, \quad \text{if } xN \in \mathbb{Z}_+\,, \\ &= I\left[xN\right]/N\,, \quad \text{otherwise,} \end{aligned}$$

where $\mathbb{Z}_+$ is the set of the positive integers. Since, for any real number z,

$$\begin{aligned} I\left[N - z\right] \quad &= N - z\,, \quad \text{if } z \text{ is an integer,} \\ &= N - I\left[z\right] - 1\,, \quad \text{otherwise,} \end{aligned}$$

we then have, for $0 \le x \le 1$,

$$\begin{aligned} \mathsf{P}\left[R_j/N \ge x\right] \quad &= 1 - \mathsf{P}\left[R_N/N < x\right] = \left(N - I\left[xN\right] + 1\right)/N\,, \quad \text{if } xN \in \mathbb{Z}_+\,, \\ &= \left(N - I\left[xN\right]\right)/N \quad \text{otherwise,} \end{aligned}$$

hence

$$\begin{aligned} \mathsf{P}\left[R_j/N \ge x\right] \quad &= \left(I\left[(1-x)\,N\right] + 1\right)/N\,, \quad \text{if } 0 < x \le 1\,, \\ &= 1\,, \quad \text{if } x = 0\,, \end{aligned}$$

from which we get (2.12). ∎

Proof of Proposition 2.1.2 Assuming there are no ties among $S_0,\, S_1,\, \ldots,\, S_N$ (an event with probability 1), we have

$$\begin{aligned} \hat{G}_N\left(S_0\right) \quad &= \quad \frac{1}{N}\sum_{i=1}^{N} s\left(S_i - S_0\right) = \frac{1}{N}\sum_{i=1}^{N}\left[1 - s\left(S_0 - S_i\right)\right] = 1 - \frac{1}{N}\sum_{i=1}^{N} s\left(S_0 - S_i\right) \\ &= \quad 1 - \frac{1}{N}\left[-1 + \sum_{i=0}^{N} s\left(S_0 - S_i\right)\right] = \left(N + 1 - R_0\right)/N \end{aligned}$$

where $R_0 = \sum_{i=0}^{N} s\left(S_0 - S_i\right)$ is the rank of S_0 when the $N+1$ variables $S_0,\, S_1,\, \ldots,\, S_N$ are ranked in nondecreasing order. Using (2.13) and Lemma **2.1.1**, it then follows that

$$\begin{aligned} \mathsf{P}\left[\hat{G}_N(S_0) \le \alpha_1\right] &= \mathsf{P}\left[\frac{N+1-R_0}{N} \le \alpha_1\right] = \mathsf{P}\left[\frac{R_0}{N+1} \ge \frac{(1-\alpha_1)N+1}{N+1}\right] \\ &= \frac{I[\alpha_1 N]+1}{N+1} \end{aligned}$$

for $0 \le \alpha_1 \le 1$. Furthermore, $\hat{F}_N(S_0) = 1 - \hat{G}_N(S_0)$ with probability 1, and (2.15) follows. We then get (2.16) on observing that

$$\hat{F}_N(y) \ge q \Longleftrightarrow y \ge \hat{F}_N^{-1}(q)$$

for $y \in \mathbb{R}$ and $0 < q < 1$ [see Reiss (1989, Appendix 1)]. Finally, to obtain (2.17), we note that

$$\hat{p}_N(S_0) \equiv \frac{N\hat{G}_N(S_0)+1}{N+1} \le \alpha \Longleftrightarrow \hat{G}_N(S_0) \le \frac{\alpha(N+1)-1}{N}$$

hence, since that $0 \le \hat{G}_N(S_0) \le 1$ and using (2.15),

$$\begin{aligned} \mathsf{P}[\hat{p}_N(S_0) \le \alpha] &= \mathsf{P}\left[\hat{G}_N(S_0) \le \frac{\alpha(N+1)-1}{N}\right] \\ &= \begin{cases} 0\,, & \text{if } \alpha < 1/(N+1)\,, \\ \frac{I[\alpha(N+1)-1]+1}{N+1} = \frac{I[\alpha(N+1)]}{N+1}\,, & \text{if } \frac{1}{N+1} \le \alpha \le 1\,, \\ 1\,, & \text{if } \alpha > 1\,, \end{cases} \end{aligned}$$

from which (2.17) follows on observing that $I[\alpha(N+1)] = 0$ for $0 \le \alpha < 1(N+1)$. ∎

Proof of Lemma 2.2.1 From (2.21) and the continuity of the $U(0, 1)$ distribution, we see easily that

$$\mathsf{P}[(y_i, U_i) = (y_j, U_j)] \le \mathsf{P}[U_i = U_j] = 0\,, \text{ for } i \ne j\,,$$

from which it follows that $\mathsf{P}\left[\tilde{R}_i = \tilde{R}_j\right] = 0$ for $i \ne j$ and the rank vector $(\tilde{R}_1, \tilde{R}_2, \dots, \tilde{R}_N)$ is with probability 1 a random permutation of the integers $(1, 2, \dots, N)$. Set $V_i = (y_i, U_i)$, $i = 1, \dots, N$. By considering all possible permutations $(V_{r_1}, V_{r_2}, \dots, V_{r_N})$ of $(V_1, V_2, \dots, V_N)$, and since $(V_{r_1}, V_{r_2} \dots, V_{r_N}) \sim (V_1, V_2, \dots, V_N)$ for all permutations $(r_1, r_2, \dots, r_N)$ of $(1, 2, \dots, N)$ [by the exchangeability assumption], the elements of $(\tilde{R}_1, \tilde{R}_2, \dots, \tilde{R}_N)$ are also exchangeable. The result then follows from Lemma 2.1. The reader may note that an alternative proof could be obtained by modifying the proof of Theorem 29A of Hájek (1969) to relax the independence assumption for $y_1, \dots, y_N$. ∎

Proof of Proposition 2.2.2 Since the pairs (S_i, U_i) , $i = 0, 1, \dots, N$, are all distinct with probability 1, we have almost surely:

$$\begin{aligned}\tilde{G}_N(S_0) &= \frac{1}{N}\sum_{i=1}^{N} W[(S_i, U_i) - (S_0, U_0)] = 1 - \frac{1}{N}\sum_{i=1}^{N} W[(S_0, U_0) - (S_i, U_i)] \\ &= 1 - \frac{1}{N}\left\{-1 + \sum_{i=0}^{N} W[(S_0, U_0) - (S_i, U_i)]\right\} = (N + 1 - \tilde{R}_0)/N\end{aligned}$$

where $\tilde{R}_0 = \sum_{i=1}^{N} W[(S_0, U_0) - (S_i, U_i)]$ is the randomized rank of S_0 obtained when ranking in ascending order [according to (2.21)] the $N+1$ pairs (S_i, U_i) , $i = 0, 1, \ldots\ N+1$. Using Lemma **2.2.1**, it follows that

$$\begin{aligned}\mathsf{P}[\tilde{G}_N(S_0) \le \alpha_1] &= \mathsf{P}[(N + 1 - \tilde{R}_0)/N \le \alpha_1] = \mathsf{P}\left[\frac{\tilde{R}_0}{N+1} \ge \frac{(1-\alpha_1)N + 1}{N+1}\right] \\ &= \begin{cases} 0\,, & \text{if } \alpha_1 < 0\,, \\ \frac{I[\alpha_1 N]+1}{N+1}\,, & \text{if } 0 \le \alpha_1 \le 1\,, \\ 1\,, & \text{if } \alpha_1 > 1\,. \end{cases}\end{aligned}$$

Since the pairs (S_i, U_i) , $i = 0, 1, \ldots, N$, are all distinct with probability 1, we also have

$$\tilde{F}_N(S_0) = 1 - \tilde{G}_N(S_0) \text{ with probability 1,}$$

hence using the inequalities (2.31) – (2.32),

$$\mathsf{P}[\hat{G}_N(S_0) \le \alpha_1] \le \mathsf{P}[\tilde{G}_N(S_0) \le \alpha_1] = \mathsf{P}[\tilde{F}_N(S_0) \ge 1 - \alpha_1] \le \mathsf{P}[\hat{F}_N(S_0) \ge 1 - \alpha_1]$$

and (2.35) is established. The identity $\mathsf{P}[\hat{F}_N(S_0) \ge 1 - \alpha_1] = \mathsf{P}[S_0 \ge \hat{F}_N^{-1}(1 - \alpha_1)]$ follows from the equivalence

$$\hat{F}_N(y) \ge q \Longleftrightarrow y \ge \hat{F}_N^{-1}(q)\ ,\ \forall y \in \mathbb{R}\,,\ 0 < q < 1\,.$$

Finally, to obtain (2.17), we observe that

$$\tilde{p}_N(S_0) = \frac{N\tilde{G}_N(S_0) + 1}{N+1} \le \alpha \Longleftrightarrow \tilde{G}_N(S_0) \le \frac{\alpha(N+1) - 1}{N+1}$$

hence, using (2.35),

$$\begin{aligned}\mathsf{P}[\tilde{p}_N(S_0) \le \alpha] &= \mathsf{P}\left[\tilde{G}_N(S_0) \le \frac{\alpha(N+1) - 1}{N+1}\right] \\ &= \begin{cases} 0\,, & \text{if } \alpha < 1/(N+1) \\ \frac{I[\alpha(N+1)-1]+1}{N+1} = \frac{I[\alpha(N+1)]}{N+1}\,, & \text{if } \frac{1}{N+1} \le \alpha \le 1 \end{cases}\end{aligned}$$

from which (2.17) follows on observing that $I[\alpha(N+1)] = 0$ for $0 \le \alpha \le 1/(N+1)$. ∎

Proof of Proposition 4.1 Since

$$\hat{G}_N[S_0 \mid \theta] \geq 1 - \hat{F}_N[S_0 \mid \theta] \,, \tag{A.1}$$

$$\left\|\theta - \bar{\theta}\right\| \leq \delta_2 \Rightarrow \sup_{y \in B_2(x,\varepsilon) \cap D_0} \left|F_\infty[y \mid \theta] - F_\infty\left[y \mid \bar{\theta}\right]\right| \leq \varepsilon \,,\ \forall\theta \,, \tag{A.2}$$

we have

$$\mathsf{P}\left[\sup\{\hat{G}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \leq \alpha_1\right] \leq \mathsf{P}_\theta\left[\inf\{\hat{F}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \geq 1 - \alpha_1\right] \,.$$

When $\bar{\theta} \in \Omega_0$, it is also clear that

$$\inf_{\theta \in \Omega_0} \hat{F}_N[S_0 \mid \theta] \geq 1 - \alpha_1 \Rightarrow \hat{F}_N\left[S_0 \mid \bar{\theta}\right] \geq 1 - \alpha_1 \,,$$

hence, using Proposition **2.1.2**,

$$\mathsf{P}\left[\inf\{\hat{F}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \geq 1 - \alpha_1\right] \leq \mathsf{P}[\hat{F}_N\left[S_0 \mid \bar{\theta}\right] \geq 1 - \alpha_1] = \frac{I[\alpha_1 N] + 1}{N + 1} \,.$$

Furthermore

$$\begin{aligned} \inf_{\theta \in \Omega_0} \hat{F}_N[S_0 \mid \theta] \quad & \geq \quad 1 - \alpha_1 \Longleftrightarrow \hat{F}_N[S_0 \mid \theta] \geq 1 - \alpha_1, \forall\theta \in \Omega_0 \\ & \Longleftrightarrow \quad S_0 \geq \hat{F}_N^{-1}[1 - \alpha_1 \mid \theta], \forall\theta \in \Omega_0 \Longleftrightarrow S_0 \geq \sup_{\theta \in \Omega_0} \hat{F}_N^{-1}[1 - \alpha_1 \mid \theta] \\ & \Longleftrightarrow \quad S_0 \geq \sup_{\theta \in \Omega_0} \hat{F}_N^{-1}[1 - \alpha_1 \mid \theta] \end{aligned} \tag{A.3}$$

so that, using Proposition **2.1.2**,

$$\begin{aligned} \mathsf{P}\left[S_0 \geq \sup\{\hat{F}_N^{-1}[1 - \alpha_1 \mid \theta_0] : \theta \in \Omega_0\}\right] \quad & = \quad \mathsf{P}[\inf\{\hat{F}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \geq 1 - \alpha_1] \\ & \leq \quad \frac{I[\alpha_1 N] + 1}{N + 1} \end{aligned}$$

and (4.20) is established. (4.21) follows in the same way on observing that

$$\sup_{\theta \in \Omega_0} \tilde{p}_N[S_0 \mid \theta] \leq \sup_{\theta \in \Omega_0} \hat{p}_N[S_0 \mid \theta]$$

and

$$\sup_{\theta \in \Omega_0} \tilde{p}_N[S_0 \mid \theta] \leq \alpha \Rightarrow \tilde{p}_N\left[S_0 \mid \bar{\theta}\right] \leq \alpha \,,\ \text{when } \bar{\theta} \in \Omega_0 \,.$$

∎

Proof of Proposition 4.2 Using (2.31) - (2.32), we have

$$\begin{aligned} 1-\hat{F}_N\left[S_0 \mid \theta\right] &\leq \tilde{G}_N\left[S_0 \mid \theta\right] \leq \hat{G}_N\left[S_0 \mid \theta\right] \,,\; \forall\theta \,, \\ 1-\hat{G}_N\left[S_0 \mid \theta\right] &\leq \tilde{F}_N\left[S_0 \mid \theta\right] \leq \hat{F}_N\left[S_0 \mid \theta\right] \,, \end{aligned}$$

hence

$$\begin{aligned} \sup_{\theta\in\Omega_0} \tilde{G}_N\left[S_0 \mid \theta\right] &\leq \sup_{\theta\in\Omega_0} \hat{G}_N\left[S_0 \mid \theta\right] \,, \\ 1-\sup_{\theta\in\Omega_0} \hat{G}_N\left[S_0 \mid \theta\right] &\leq \inf_{\theta\in\Omega_0} \{1-\hat{G}_N\left[S_0 \mid \theta\right]\} \leq \inf_{\theta\in\Omega_0} \tilde{F}_N\left[S_0 \mid \theta\right] \,, \\ \sup_{\theta\in\Omega_0} \tilde{p}_N\left[S_0 \mid \theta\right] &\leq \sup_{\theta\in\Omega_0} \hat{p}_N\left[S_0 \mid \theta\right] \,. \end{aligned}$$

Furthermore, when $\bar{\theta}\in\Omega_0$,

$$\begin{aligned} \sup_{\theta\in\Omega_0} \tilde{G}_N\left[S_0 \mid \theta\right] &\leq \alpha_1 \Rightarrow \tilde{G}_N\left[S_0 \mid \bar{\theta}\right] \leq \alpha_1 \,, \\ \inf_{\theta\in\Omega_0} \tilde{F}_N\left[S_0 \mid \theta\right] &\geq 1-\alpha_1 \Rightarrow \tilde{F}_N\left[S_0 \mid \bar{\theta}\right] \geq 1-\alpha_1 \,, \\ \sup_{\theta\in\Omega_0} \tilde{p}_N\left[S_0 \mid \theta\right] &\leq \alpha_1 \Rightarrow \tilde{p}_N\left[S_0 \mid \bar{\theta}\right] \leq \alpha_1 \,, \end{aligned}$$

hence, using Proposition **2.2.2** ,

$$\begin{aligned} \mathsf{P}\left[\sup\{\hat{G}_N\left[S_0 \mid \theta\right] : \theta\in\Omega_0\} \leq \alpha_1\right] &\leq \mathsf{P}\left[\sup\{\tilde{G}_N\left[S_0 \mid \theta\right] : \theta\in\Omega_0\} \leq \alpha_1\right] \\ &\leq \mathsf{P}[\tilde{G}_N\left[S_0 \mid \bar{\theta}\right] \leq \alpha_1] = \frac{I\left[\alpha_1 N\right]+1}{N+1} \,, \\ \mathsf{P}\left[\sup\{\hat{G}_N\left[S_0 \mid \theta\right] : \theta\in\Omega_0\} \leq \alpha_1\right] &\leq \mathsf{P}\left[\inf\{\tilde{F}_N\left[S_0 \mid \theta\right] : \theta\in\Omega_0\} \geq 1-\alpha_1\right] \\ &\leq \mathsf{P}[\tilde{F}_N\left[S_0 \mid \bar{\theta}\right] \geq 1-\alpha_1] = \frac{I\left[\alpha_1 N\right]+1}{N+1} \end{aligned}$$

for $0\leq\alpha_1\leq 1$, and

$$\begin{aligned} \mathsf{P}\left[\sup\left\{\hat{p}_N\left[S_0 \mid \theta\right] : \theta\in\Omega_0\right\} \leq \alpha\right] &\leq \mathsf{P}\left[\sup\left\{\tilde{p}_N\left[S_0 \mid \theta\right] : \theta\in\Omega_0\right\} \leq \alpha\right] \\ &\leq \mathsf{P}\left[\tilde{p}_N\left[S_0 \mid \bar{\theta}\right] \leq \alpha\right] = \frac{I\left[\alpha\left(N+1\right)\right]}{N+1} \end{aligned}$$

for $0\leq\alpha\leq 1$. ∎

Proof of Proposition 5.1 Using arguments similar to the ones in the proof of Proposition **4.2** [see (A.1) - (A.3)], it is easy to see that

$$\mathsf{P}\left[\sup\{\hat{G}_{TN}\left[S_{T0} \mid \theta\right] : \theta\in C_T\} \leq \alpha_1\right] \leq \mathsf{P}\left[\inf\{\hat{F}_{TN}\left[S_{T0} \mid \theta\right] : \theta\in C_T\} \geq 1-\alpha_1\right]$$

$$= \quad \mathsf{P}\left[S_{T0} \geq \sup\{\hat{F}_{TN}^{-1}\left[1-\alpha_1 \mid \theta\right] : \theta \in C_T\}\right] .$$

Further

$$\begin{aligned}
&\mathsf{P}\left[\inf\{\hat{F}_{TN}\left[S_{T0} \mid \theta\right] : \theta \in C_T\} \geq 1-\alpha_1\right] \\
&\quad = \mathsf{P}\left[\inf\{\hat{F}_{TN}\left[S_{T0} \mid \theta\right] : \theta \in C_T\} \geq 1-\alpha_1 \text{ and } \bar{\theta} \in C_T\right] \\
&\qquad + \mathsf{P}\left[\inf\{\hat{F}_{TN}\left[S_{T0} \mid \theta\right] : \theta \in C_T\} \geq 1-\alpha_1 \text{ and } \bar{\theta} \notin C_T\right] \\
&\quad \leq \mathsf{P}[\hat{F}_{TN}\left[S_{T0} \mid \bar{\theta}\right] \geq 1-\alpha_1] + \mathsf{P}[\bar{\theta} \notin C_T] = \frac{I\left[\alpha_1 N\right]+1}{N+1} + \mathsf{P}[\bar{\theta} \notin C_T]
\end{aligned}$$

where the last inequality follows from Proposition **2.1.2**, hence, since $\lim\limits_{T\to\infty} \mathsf{P}\left[\bar{\theta} \notin C_T\right] = 0$,

$$\begin{aligned}
\lim_{T\to\infty} \mathsf{P}\left[\inf\{\hat{F}_{TN}\left[S_{T0} \mid \theta\right] : \theta \in C_T\} \geq 1-\alpha_1\right] \quad &\leq \quad \frac{I\left[\alpha_1 N\right]+1}{N+1} + \lim_{T\to\infty} \mathsf{P}[\bar{\theta} \notin C_T] \\
&= \quad \frac{I\left[\alpha_1 N\right]+1}{N+1} ,
\end{aligned}$$

from which (5.9) and (5.10) follow. ∎

Proof of Proposition 5.2 The result follows from arguments similar to the ones used in the proofs of Propositions **4.1** and **5.2** (with Ω_0 replaced by C_T). ∎

Proof of Lemma 6.1 It is easy to see [as in (3.2)] that

$$\begin{aligned}
Q_{TN}\left(\theta, x, u_0, \alpha_1\right) \quad &= \quad \mathsf{P}[\tilde{G}_{TN}\left(x \mid \theta\right) \leq \alpha_1 \mid U_0 = u_0] \\
&= \quad \sum_{k=0}^{I[\alpha_1 N]} \binom{N}{k} \bar{G}_T\left(x,\, u_0 \mid \theta\right)^k \left[1 - \bar{G}_T\left(x,\, u_0 \mid \theta\right)\right]^{N-k}
\end{aligned}$$

where

$$\begin{aligned}
\bar{G}_T\left(x,\, u_0 \mid \theta\right) \quad &= \quad \mathsf{P}\left(W\left[\left(S_{Ti}\left(\theta\right),\, U_i\right) - \left(x,\, u_0\right)\right] = 1\right) \\
&= \quad \mathsf{P}\left[S_{Ti}\left(\theta\right) > x\right] + \mathsf{P}\left[S_{Ti}\left(\theta\right) = x\right] \mathsf{P}\left[U_i \geq u_0\right] \\
&= \quad 1 - F_T\left[x \mid \theta\right] + g_T\left(x \mid \theta\right)\left(1 - u_0\right) ,\ 1 \leq i \leq T .
\end{aligned}$$

Note also that

$$g_T\left(x \mid \theta\right) = F_T\left[x \mid \theta\right] - \lim_{\delta_0 \to 0^+} F_T\left[y - \delta_0 \mid \theta\right] .$$

Then the inequality

$$\left|F_T\left[y \mid \theta\right] - F_T\left[y \mid \bar{\theta}\right]\right| \leq \varepsilon ,\ \forall y \in (x-\delta\, ,\, x+\delta)$$

entails the following inequalities:

$$\left|1 - F_T\left[x \mid \theta\right] - 1 + F_T\left[x \mid \bar{\theta}\right]\right| \leq \varepsilon \; ,$$

$$\begin{aligned}\left|g_T\left(x \mid \theta\right) - g_T\left(x \mid \bar{\theta}\right)\right| &= \left| \left\{1 - F_T\left[x \mid \theta\right]\right\} - \left\{1 - F_T\left[x \mid \bar{\theta}\right]\right\} \right. \\ &\qquad \left. + \lim_{\delta_0 \to 0^+} \left\{F_T\left[y - \delta_0 \mid \theta\right] - F_T\left[y - \delta_0 \mid \bar{\theta}\right]\right\} \right| \\ &\leq \left|F_T\left[x \mid \theta\right] - F_T\left[x \mid \bar{\theta}\right]\right| + \lim_{\delta_0 \to 0^+}\left|F_T\left[y - \delta_0 \mid \theta\right] - F_T\left[y - \delta_0 \mid \bar{\theta}\right]\right| \leq 2\varepsilon \, ,\end{aligned}$$

hence

$$\begin{aligned}\left|\bar{G}_T\left(x, u_0 \mid \theta\right) - \bar{G}_T\left(x, u_0 \mid \bar{\theta}\right)\right| &\leq \left|F_T\left[x \mid \theta\right] - F_T\left[x \mid \bar{\theta}\right]\right| \\ &\quad + \left|1 - u_0\right| \left|g_T\left(x \mid \theta\right) - g_T\left(x \mid \bar{\theta}\right)\right| \leq 3\varepsilon \, , \; \forall u_0 \in [0, 1] \; ,\end{aligned}$$

and

$$\begin{aligned}&\left|Q_{TN}\left(\theta, \, x, \, u_0, \, \alpha_1\right) - Q_{TN}\left(\bar{\theta}, \, x, \, u_0, \, \alpha_1\right)\right| \\ &\leq \left| \sum_{k=0}^{I[N\alpha_1]} \binom{N}{k} \bar{G}_T\left(x, u_0 | \theta\right)^k \left[1 - \bar{G}_T\left(x, u_0 | \theta\right)\right]^{N-k} - \bar{G}_T\left(x, u_0 | \bar{\theta}\right)^k \left[1 - \bar{G}_T\left(x, u_0 | \bar{\theta}\right)\right]^{N-k} \right| \\ &\leq \sum_{k=0}^{I[N\alpha_1]} \binom{N}{k} \left|\bar{G}_T\left(x, \, u_0 \mid \theta\right)^k - \bar{G}_T\left(x, \, u_0 \mid \bar{\theta}\right)^k\right| \\ &\qquad + \left|\left[1 - \bar{G}_T\left(x, \, u_0 \mid \theta\right)\right]^{N-k} - \left[1 - \bar{G}_T\left(x, \, u_0 \mid \bar{\theta}\right)\right]^{N-k}\right| \\ &\leq \sum_{k=0}^{I[N\alpha_1]} \binom{N}{k} k \left|\bar{G}_T\left(x, \, u_0 \mid \theta\right) - \bar{G}_T\left(x, \, u_0 \mid \bar{\theta}\right)\right| \\ &\qquad + \left(N - k\right) \left|\bar{G}_T\left(x, \, u_0 \mid \theta\right) - \bar{G}_T\left(x, \, u_0 \mid \bar{\theta}\right)\right| \\ &= C\left(N, \, \alpha_1\right) \left|\bar{G}_T\left(x, \, u_0 \mid \theta\right) - \bar{G}_T\left(x, \, u_0 \mid \bar{\theta}\right)\right| \leq 3C\left(N, \, \alpha_1\right) \varepsilon \; , \; \forall u_0 \in [0, \, 1]\end{aligned}$$

where $C\left(N, \, \alpha_1\right) = N \sum_{k=0}^{I[\alpha_1 N]} \binom{N}{k}$, from which (6.10) follows. The inequality (6.11) follows in a similar way on noting that

$$\bar{Q}_{TN}\left(\theta, x, \alpha_1\right) = \sum_{k=0}^{I[\alpha_1 N]} \binom{N}{k} G_T\left(x \mid \theta\right)^k \left[1 - G_T\left(x \mid \theta\right)\right]^{N-k}$$

where $G_T\left(x \mid \theta\right) = \mathsf{P}\left[S_{Ti}\left(\theta\right) \geq x\right] , \; 1 \leq i \leq N$. ∎

Proof of Lemma 6.2 Let $\alpha_1 \in [0, \, 1]$, $\varepsilon > 0$ and $\varepsilon_0 > 0$ and suppose first that (6.4) holds. Then, using Lemma **6.1**, we can find $\delta > 0$ and T_1 such that

$$x \quad \in \quad D_{T0}\left(\varepsilon_0\right) \; , \; \left\|\theta - \bar{\theta}\right\| \leq \delta \text{ and } T > T_1$$

$$\Rightarrow \quad \left|F_T\left[x \mid \theta\right] - F_T\left[x \mid \bar{\theta}\right]\right| \leq \varepsilon_1 \equiv \varepsilon \,/\, [3C\,(N,\,\alpha_1)]$$
$$\Rightarrow \quad \left|Q_{TN}\,(\theta,\,x,\,u_0,\,\alpha_1) - Q_{TN}\left(\bar{\theta},\,x,\,u_0,\,\alpha_1\right)\right| \leq \varepsilon\,,\ \forall u_0 \in [0,\,1]\ .$$

Thus
$$S_{T0} \in D_{T0}\,(\varepsilon_0)\ \text{ and }\ \|\,\widehat{\theta}_T - \bar{\theta}\,\| \leq \delta \Rightarrow \Delta_{TN}(\widehat{\theta}_T,\,\bar{\theta},\,S_{T0},\,\alpha_1) \leq \varepsilon$$

where

$$\Delta_{TN}(\widehat{\theta}_T,\,\bar{\theta},\,S_{T0},\,\alpha_1) \equiv \sup_{0\leq u_0 \leq 1}\left|Q_{TN}(\widehat{\theta}_T,\,S_{T0},\,u_0,\,\alpha_1) - Q_{TN}(\bar{\theta},\,x,\,u_0,\,\alpha_1)\right|\,,$$

hence

$$\begin{aligned}
\mathsf{P}[\Delta_{TN}(\widehat{\theta}_T,\,\bar{\theta},\,S_{T0},\,\alpha_1) &\leq \varepsilon] \geq \mathsf{P}[S_{T0} \in D_{T0}\,(\varepsilon_0)\ \text{ and }\ \|\,\widehat{\theta}_T - \bar{\theta}\,\| \leq \delta] \\
&\geq 1 - \mathsf{P}\,[S_{T0} \notin D_{T0}\,(\varepsilon_0)] - \mathsf{P}[\|\,\widehat{\theta}_T - \bar{\theta}\,\| > \delta] \\
&= \mathsf{P}\,[S_{T0} \in D_{T0}\,(\varepsilon_0)] - \mathsf{P}[\|\,\widehat{\theta}_T - \bar{\theta}\,\| > \delta]\,.
\end{aligned}$$

Since $\widehat{\theta}_T \xrightarrow{p} \bar{\theta}$,it follows that

$$\liminf_{T\to\infty} \mathsf{P}[\Delta_{TN}(\widehat{\theta}_T,\,\bar{\theta},\,S_{T0},\,\alpha_1) \leq \varepsilon] \geq \liminf_{T\to\infty} \mathsf{P}\,[S_{T0} \in D_{T0}\,(\varepsilon_0)] \geq 1 - \varepsilon_0$$

for any $\varepsilon_0 > 0$, hence
$$\lim_{T\to\infty} \mathsf{P}[\Delta_{TN}(\widehat{\theta}_T,\,\bar{\theta},\,S_{T0},\,\alpha_1) \leq \varepsilon] = 1\,.$$
Since the latter identity holds for any $\varepsilon > 0$, (6.12) is established. (6.13) follows in a similar way upon using (6.11).

Suppose now (6.5) - (6.7) hold instead of (6.4). Then,

$$(S_{T0},\,\hat{\theta}_T) \xrightarrow[T\to\infty]{p} (S_0,\,\bar{\theta}) \tag{A.4}$$

and
$$(S_{\bar{T}_k},\,\hat{\theta}_{\bar{T}_k}) \xrightarrow[T\to\infty]{p} (S_0,\,\bar{\theta}) \tag{A.5}$$

for any subsequence $\{(S_{\bar{T}_k},\,\hat{\theta}_{\bar{T}_k}) : k = 1, 2, \ldots\ \}$ of $\{(S_T,\ \hat{\theta}_T) : T \geq I_0\}$. Since S_{T0} and $\widehat{\theta}_T$, $T \geq I_0$, are random variables (or vectors) defined on $\mathcal{Z}$, we can write $S_{T0} = S_{T0}\,(\omega)$, $\widehat{\theta}_T = \widehat{\theta}_T\,(\omega)$ and $S_0 = S_0\,(\omega)$, $\omega \in \mathcal{Z}$. By (6.6), the event

$$A_0 = \{\omega : S_0\,(\omega) \in D_0 \text{ and } S_{T0}\,(\omega) \in D_0\,,\ \text{for } T \geq I_0\}$$

has probability one. Furthermore, by (A.5), the subsequence $(S_{\bar{T}_k},\,\hat{\theta}'_{\bar{T}_k})'$ contains a further subsequence $(S_{T_k0},\,\widehat{\theta}'_{T_k})'$, $k \geq 1$ such that $(S_{T_k0},\,\widehat{\theta}'_{T_k})' \xrightarrow[T\to\infty]{} (S_0,\ \bar{\theta}')'$ $a.s.$ (where $T_1 < T_2 < \ldots$);

see Bierens (1994, pp. 22-23). Consequently, the set

$$C_0 = \{\omega \in \mathcal{Z} : S_0(\omega) \in D_0\,,\ \lim_{k\to\infty} S_{T_k 0}(\omega) = S_0(\omega) \text{ and } \lim_{k\to\infty} \hat{\theta}_{T_k}(\omega) = \bar{\theta}\}$$

has probability one. Now, let $\varepsilon > 0$. By (6.7), for any $x \in D_0$, we can find $\delta(x, \varepsilon) > 0$, $T(x, \varepsilon) > 0$ and an open neighborhood $B(x, \varepsilon)$ of x such that

$$\| \theta - \bar{\theta} \| \leq \delta(x, \varepsilon) \text{ and } T > T(x, \varepsilon) \Rightarrow \left| F_T[y \mid \theta] - F_T\left[y \mid \bar{\theta}\right] \right| \leq \varepsilon\,,\ \forall y \in B(x, \varepsilon) \cap D_0\,.$$

Furthermore, for $\omega \in C_0$, we can find k_0 such that

$$k \geq k_0 \Rightarrow S_{T_k 0}(\omega) \in B(S_0(\omega), \varepsilon) \cap D_0 \text{ and } \| \hat{\theta}_{T_k} - \bar{\theta} \| \leq \delta(S_0(\omega), \varepsilon)\,,$$

so that $T_k > \max\{T(S_0(\omega), \varepsilon), T_{k_0}\}$ entails

$$\left| F_{T_k}[S_{T_k 0}(\omega) \mid \theta_{T_k}(\omega)] - F_{T_k}\left[S_{T_k 0}(\omega) \mid \bar{\theta}\right] \right| \leq \varepsilon\,.$$

Thus, for $\omega \in C_0$,

$$\lim_{k\to\infty} \{F_{T_k}[S_{T_k 0}(\omega) \mid \theta_{T_k}(\omega)] - F_{T_k}\left[S_{T_k 0}(\omega) \mid \bar{\theta}\right]\} = 0$$

which implies, using Lemma **6.1**,

$$\lim_{k\to\infty} \Delta_{T_k N}\left(\hat{\theta}_{T_k}(\omega), \bar{\theta}, S_{T_k 0}(\omega), \alpha_1\right) = 0\,,$$

hence

$$\Delta_{T_k N}\left(\hat{\theta}_{T_k}, \bar{\theta}, S_{T_k 0}, \alpha_1\right) \underset{k\to 0}{\longrightarrow} 0 \ a.s.$$

This shows that any subsequence of the sequence $\Delta_{TN}(\widehat{\theta}_T, \bar{\theta}, S_{T0}, \alpha_1)$, $T \geq I_0$, contains a further subsequence which converge $a.s.$ to zero. It follows that

$$\Delta_{TN}(\hat{\theta}_T, \bar{\theta}, S_{T0}, \alpha_1) \xrightarrow[T\to\infty]{p} 0$$

and (6.12) is established. The proof of (6.13) under the condition (6.5) - (6.7) is similar. ∎

Proof of Proposition 6.3 Using the fact that $\widehat{\theta}_T$, S_{T0} and U_0 are independent of $S_T(N, \theta)$, we can write

$$\begin{aligned}
&\mathsf{P}[\tilde{G}_{TN}[S_{T0} \mid \widehat{\theta}_T] \leq \alpha_1] - \mathsf{P}[\tilde{G}_{TN}[S_{T0} \mid \bar{\theta}] \leq \alpha_1] \\
&\quad = \mathsf{E}\{\mathsf{P}\left[\tilde{G}_{TN}[S_{T0} \mid \widehat{\theta}_T] \leq \alpha_1 \mid (\widehat{\theta}_T, S_{T0}, U_0)\right] \\
&\qquad - \mathsf{P}\left[\tilde{G}_{TN}\left[S_{T0} \mid \bar{\theta}\right] \leq \alpha_1 \mid (\widehat{\theta}_T, S_{T0}, U_0)\right]\} \\
&\quad = \mathsf{E}[Q_{TN}(\widehat{\theta}_T, S_{T0}, U_0, \alpha_1) - Q_{TN}(\bar{\theta}, S_{T0}, U_0, \alpha_1)]\,.
\end{aligned}$$

From Lemma **6.1** and using the Lebesgue dominated convergence theorem, we then get

$$\begin{aligned}
&\left|\mathsf{P}[\tilde{G}_{TN}[S_{T0} \mid \widehat{\theta}_T] \leq \alpha_1] - \mathsf{P}\left[\tilde{G}_{TN}[S_{T0} \mid \bar{\theta}] \leq \alpha_1\right]\right| \\
&\quad = \left|\{\mathsf{E}[Q_{TN}(\widehat{\theta}_T,\, S_{T0},\, U_0,\, \alpha_1) - Q_{TN}(\bar{\theta},\, S_{T0},\, U_0,\, \alpha_1)]\right| \\
&\quad \leq \mathsf{E}\{\left|Q_{TN}\left(\theta_{TN},\, S_{T0},\, U_0,\, \alpha_1\right) - Q_{TN}\left(\bar{\theta},\, S_{T0},\, U_0,\, \alpha_1\right)\right|\} \\
&\quad \leq \mathsf{E}\left[\sup_{0 \leq u_0 \leq 1}\left|Q_{TN}(\widehat{\theta}_T, S_{T0}, u_0, \alpha_1) - Q_{TN}(\bar{\theta},\, S_{T0},\, u_0,\, \alpha_1)\right|\right] \underset{T \to \infty}{\longrightarrow} 0\,.
\end{aligned}$$

We can show in a similar way that

$$\left|\mathsf{P}[\hat{G}[S_{T0} \mid \widehat{\theta}_T] \leq \alpha_1] - \mathsf{P}[G[S_{T0} \mid \bar{\theta}] \leq \alpha_1]\right| \underset{T \to \infty}{\longrightarrow} 0\,,$$

from which we get (6.14). (6.15) then follows from the definitions of $\tilde{p}_{TN}\,(x \mid \theta)$ and $\hat{p}_{TN}\,(x \mid \theta)$. ∎

References

ATHREYA, K. B. (1987): “Bootstrap of the Mean in the Infinite Variance Case,” *The Annals of Statistics*, 15, 724–731.

BARNARD, G. A. (1963): “Comment on ‘The Spectral Analysis of Point Processes’ by M. S. Bartlett,” *Journal of the Royal Statistical Society, Series B*, 25, 294.

BASAWA, I. V., A. K. MALLIK, W. P. MCCORMICK, J. H. REEVES, AND R. L. TAYLOR (1991): “Bootstrapping Unstable First-Order Autoregressive Processes,” *The Annals of Statistics*, 19, 1098–1101.

BIERENS, H. J. (1994): *Topics in Advanced Econometrics*. Cambridge University Press, New York.

BIRNBAUM, Z. W. (1974): “Computers and Unconventional Test-Statistics,” in *Reliability and Biometry*, ed. by F. Proschan, and R. J. Serfling, pp. 441–458. SIAM, Philadelphia, PA.

CAMPBELL, B., AND J.-M. DUFOUR (1997): “Exact Nonparametric Tests of Orthogonality and Random Walk in the Presence of a Drift Parameter,” *International Economic Review*, 38, 151–173.

DAVISON, A., AND D. HINKLEY (1997): *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge (UK).

DUFOUR, J.-M. (1990): “Exact Tests and Confidence Sets in Linear Regressions with Autocorrelated Errors,” *Econometrica*, 58, 475–494.

DUFOUR, J.-M., M. HALLIN, AND I. MIZERA (1998): “Generalized Runs Tests for Heteroskedastic Time Series,” *Journal of Nonparametric Statistics*, 9, 39–86.

DUFOUR, J.-M., AND J. F. KIVIET (1996): “Exact Tests for Structural Change in First-Order Dynamic Models,” *Journal of Econometrics*, 70, 39–68.

——— (1998): “Exact Inference Methods for First-Order Autoregressive Distributed Lag Models,” *Econometrica*, 66, 79–104.

DWASS, M. (1957): “Modified Randomization Tests for Nonparametric Hypotheses,” *Annals of Mathematical Statistics*, 28, 181–187.

EDGINGTON, E. S. (1980): *Randomization Tests*. Marcel Dekker, New York.

EFRON, B. (1982): *The Jacknife, the Bootstrap and Other Resampling Plans*, CBS-NSF Regional Conference Series in Applied Mathematics, Monograph No. 38. Society for Industrial and Applied Mathematics, Philadelphia, PA.

EFRON, B., AND R. J. TIBSHIRANI (1993): *An Introduction to the Bootstrap*, vol. 57 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, New York.

FOUTZ, R. V. (1980): “A Method for Constructing Exact Tests from Test Statistics that have Unknown Null Distributions,” *Journal of Statistical Computation and Simulation*, 10, 187–193.

GALLANT, A. R., AND G. TAUCHEN (1996): “Which Moments to Match?,” *Econometric Theory*, 12, 657 – 681.

GOFFE, W. L., G. D. FERRIER, AND J. ROGERS (1994): “Global Optimization of Statistical Functions with Simulated Annealing,” *Journal of Econometrics*, 60, 65–99.

GOURIÉROUX, C., AND A. MONFORT (1996): *Simulation-Based Econometric Methods*. Oxford University Press, Oxford, U.K.

HÁJEK, J. (1969): *A Course in Nonparametric Statistics*. Holden-Day, San Francisco.

HAJIVASSILIOU, V. A. (1993): “Simulation Estimation Methods for Limited Dependent Variables,” in Maddala, Rao, and Vinod (1993), pp. 519–543.

HALL, P. (1992): *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York.

HALL, P., AND D. M. TITTERINGTON (1989): “The Effect of Simulation Order on Level Accuracy and Power of Monte Carlo Tests,” *Journal of the Royal Statistical Society, Series B*, 51, 459–467.

HOROWITZ, J. L. (1997): “Bootstrap Methods in Econometrics: Theory and Numerical Performance,” in *Advances in Economics and Econometrics*, ed. by D. Kreps, and K. W. Wallis, vol. 3, pp. 188–222. Cambridge University Press, Cambridge, U.K.

JEONG, J., AND G. S. MADDALA (1993): “A Perspective on Application of Bootstrap Methods in Econometrics,” in Maddala, Rao, and Vinod (1993), pp. 573–610.

JÖCKEL, K.-H. (1986): “Finite Sample Properties and Asymptotic Efficiency of Monte Carlo Tests,” *The Annals of Statistics*, 14, 336–347.

KEANE, M. P. (1993): “Simulation Estimation for Panel Data Models with Limited Dependent Variables,” in Maddala, Rao, and Vinod (1993), pp. 545–571.

KIVIET, J. F., AND J.-M. DUFOUR (1997): “Exact Tests in Single Equation Autoregressive Distributed Lag Models,” *Journal of Econometrics*, 80, 325–353.

LEHMANN, E. L. (1986): *Testing Statistical Hypotheses, 2nd edition*. John Wiley & Sons, New York.

MADDALA, G. S., C. R. RAO, AND H. D. VINOD (eds.) (1993): *Handbook of Statistics 11: Econometrics*. North-Holland, Amsterdam.

MARIANO, R. S., AND B. W. BROWN (1993): “Stochastic Simulation for Inference in Nonlinear Errors-in-Variables Models,” in Maddala, Rao, and Vinod (1993), pp. 611–627.

MARRIOTT, F. H. C. (1979): “Barnard’s Monte Carlo Tests: How Many Simulations?,” *Applied Statistics*, 28, 75–77.

REISS, H. D. (1989): *Approximate Distributions of Order Statistics with Applications to Nonparametric Statistics*, Springer Series in Statistics. Springer-Verlag, New York.

SHAO, S., AND D. TU (1995): *The Jackknife and Bootstrap*. Springer-Verlag, New York.

SRIRAM, T. N. (1994): "Invalidity of Bootstrap for Critical Branching Processes with Immigration," *The Annals of Statistics*, 22, 1013–1023.

VINOD, H. D. (1993): "Bootstrap Methods: Applications in Econometrics," in Maddala, Rao, and Vinod (1993), pp. 629–661.