The Annals of Statistics
1986, Vol. 14, No. 1, 336–347

FINITE SAMPLE PROPERTIES AND ASYMPTOTIC EFFICIENCY OF MONTE CARLO TESTS

BY KARL-HEINZ JÖCKEL

Bremen Institute for Prevention Research and Social Medicine

Since their introduction by Dwass (1957) and Barnard (1963), Monte Carlo tests have attracted considerable attention. The aim of this paper is to give a unified approach that covers the case of an arbitrary null distribution in order to study the statistical properties of Monte Carlo tests under the null hypothesis and under the alternative. For finite samples we obtain bounds for the power of the Monte Carlo test wrt the original test that allow determination of the required simulation effort. Furthermore the concept of asymptotic (resp. local asymptotic) relative Pitman efficiency (ARPE, resp. LARPE) is adapted to Monte Carlo tests for the study of their asymptotic behaviour. The normal limit case is investigated in more detail, leading to explicit formulas for ARPE and LARPE.

1. Introduction. In many hypothesis testing problems, where the null distribution of the desired test statistic is either unknown or too complicated to evaluate, Monte Carlo techniques are now widely used. Since simulated critical values are subject to sampling error, the use of Monte Carlo tables, such as the famous Lilliefors tables [Lilliefors (1967), (1969)], leads to tests that will not have the *exact* level of significance. The excess, which is unknown to the consumer of such tables, may be considerable, especially for Monte Carlo studies of only some hundred replicates. Furthermore for certain applications, for example permutation tests and intuitive test statistics for the analysis of spatial patterns, a tabulation of Monte Carlo critical values is not feasible. These considerations give rise to the so-called Monte Carlo tests [independently proposed by Dwass (1957) and Barnard (1963)]: For each application of the desired test a simulation experiment of moderate size is carried out, delivering realizations of the test statistic under the null hypothesis. The Monte Carlo test then decides by comparing the simulated values with the observed value of the test statistic.

For a continuous null distribution it has been shown by Hope (1968) and Birnbaum (1974) that an appropriate version of the Monte Carlo test procedure is of exact size α. Dwass (1957) obtained the corresponding result for the case of the Pitman two sample permutation test (which has a discrete null distribution conditional on the order statistics) by letting the sample size of the permutation test tend to infinity. Besides the papers mentioned above, there are others on this subject by Foutz [(1980), (1981)], Jöckel (1981), and Marriott (1979). But there remain open problems:

1. The case of an arbitrary null distribution has not previously been treated.
2. Power considerations are based on rather strict assumptions.

Received December 1983; revised May 1985.
AMS 1980 *subject classifications*. Primary 62C99; secondary 68J99.
Key words and phrases. Monte Carlo tests, testing hypotheses, Pitman efficiency.

3. A general asymptotic efficiency concept does not exist.
4. The simulation sample size is still open to question.

In Section 2 we give a unified approach for arbitrary null distribution of the test statistic. In Sections 3 and 4 we discuss the performance of the Monte Carlo test with respect to power and asymptotic efficiency. The results may be used as recommendations for the simulated sample size.

It should be noticed that Monte Carlo tests have already become a part of applied statistics; for examples see Besag and Diggle (1977), Green (1977), Hollander (1971), Ripley (1977), and Tsutokawa and Yang (1974).

2. Definition of the Monte Carlo test procedure, performance under the null hypothesis. Let $\{P_\theta: \theta \in \Theta\}$ be a family of probability measures on some sample space (χ, γ) (normally a Euclidean space with the Borel sets). We want to test

$$\Theta_0 \text{ versus } \Theta_1 = \Theta - \Theta_0,$$

having observed $x \in \chi$. Consider an appropriate test statistic T with values in the reals. The following assumption often will be fulfilled:

$$\text{There exists } \theta^* \in \Theta_0, \text{ such that}$$

$$P_{\theta_0}(T \le t) \le P_{\theta^*}(T \le t) \le P_{\theta_1}(T \le t) \tag{2.1}$$

$$\text{for all } (\theta_0, \theta_1) \in \Theta_0 \times \Theta_1.$$

Furthermore we adopt

$$F_{\theta^*}(t) = P_{\theta^*}(T \le t) \text{ is a continuous distribution function.} \tag{2.2}$$

Then

$$\phi_\alpha(x) = 1_{(-\infty,\, {}^{-1}F_{\theta^*}(\alpha)]}(T(x)),$$

where

$${}^{-1}F_{\theta^*}(\alpha) = \sup\{\xi: F_{\theta^*}(\xi) \le \alpha\}$$

is of size α and according to (2.1) unbiased. Although in many practical situations neither $F_{\theta^*}(T(x))$ nor ${}^{-1}F_{\theta^*}(\alpha)$ are manageable, it is often possible to simulate random elements $x_1, \ldots, x_m$ distributed according to P_{θ^*}. If the testing level α is an integer multiplier k of $1/(m+1)$ the Monte Carlo test (of simulation size m), ϕ_m rejects Θ_0 if the observed value of the test statistic $t = T(x)$ is less than or equal to the kth order statistic $t_{k:m}$ of the simulated values $t_i = T(x_i)$.

For convenience ϕ_m will be called the *MC test* (wrt T) or the MC test corresponding to ϕ_α. It should be observed that ϕ_m is a randomized test depending on an independent simulation experiment. Because of (2.2), ϕ_m may be equivalently rewritten as

$$\phi_m(t) = \begin{cases} 1 & \text{if } \sum_{i=1}^{m} 1_{(-\infty, t]}(t_i) \le k - 1, \\ 0 & \text{if } \sum_{i=1}^{m} 1_{(-\infty, t]}(t_i) > k - 1. \end{cases}$$

Denote

$$F_{\theta^*}^{-1}(\xi) = \inf\{x \colon F_{\theta^*}(x) \geq \xi\}$$

and let for $\alpha \in (0; 1)$, $b(\alpha, m, \xi)$ be the density of a beta distribution on the unit interval with parameters $p = (m + 1)\alpha$ and $q = (m + 1)(1 - \alpha)$; then by some calculus we have the following identities:

$$E_\theta \phi_m = \int_{-\infty}^{\infty} \sum_{\nu=0}^{k-1} \binom{m}{\nu} [F_{\theta^*}(t)]^\nu [1 - F_{\theta^*}(t)]^{m-\nu} \, dF_\theta(t)$$

$$= \int_{-\infty}^{\infty} \int_{F_{\theta^*}(t)}^{1} b(\alpha, m, \xi)\, d\xi \, dF_\theta(t) = \int_0^1 F_\theta(F_{\theta^*}^{-1}(\xi)) b(\alpha, m, \xi)\, d\xi.$$

Similar results have been obtained (under more restrictive conditions on the testing problem) by Hope (1968) and Birnbaum (1974). Obviously

$$\int_0^1 F_\theta(F_{\theta^*}^{-1}(\xi)) b(\alpha, m, \xi)\, d\xi = \int_0^1 F_\theta({}^{-1}F_{\theta^*}(\xi)) b(\alpha, m, \xi)\, d\xi.$$

For a fixed alternative parameter, $\theta_1 \in \Theta_1$, say,

$$\beta_{\theta_1}(\alpha) = F_{\theta_1}({}^{-1}F_{\theta^*}(\alpha))$$

is the power of the tests ϕ_α for $\{\theta^*\}$ versus $\{\theta_1\}$, considered as a function of α.

The corresponding quantity for the MC test may thus be calculated as

$$E_{\theta_1}\phi_m = \int_0^1 \beta_{\theta_1}(\xi) b(\alpha, m, \xi)\, d\xi$$

and furthermore we have

$$E_{\theta_0}\phi_m \leq E_{\theta^*}\phi_m = \alpha \leq E_{\theta_1}\phi_m \quad \text{for all } (\theta_0, \theta_1) \in \Theta_0 \times \Theta_1.$$

If the alternative parameter is fixed as we shall assume in the sequel, the subscript θ_1 will be suppressed. Although in our context the function

$$\alpha \mapsto \int_0^1 \beta(\xi) b(\alpha, m, \xi)\, d\xi =: \beta_m(\alpha)$$

makes sense only for values $\alpha \in \{1/(m+1), \dots, m/(m+1)\}$, it is convenient to extend the range and regard $\beta_m(\alpha)$ as the power of the MC test as a function of the level $\alpha \in (0; 1)$.

The remainder of this section is devoted to the case of an arbitrary null distribution F_{θ^*}, apparently an open problem. So *for the rest of this section we drop* (2.2). In this case the appropriate test ϕ_α of size α is

$$\phi_\alpha(x) = \begin{cases} 1 & \text{if } T(x) < {}^{-1}F_{\theta^*}(\alpha), \\ \gamma(\alpha) & \text{if } T(x) = {}^{-1}F_{\theta^*}(\alpha), \\ 0 & \text{if } T(x) > {}^{-1}F_{\theta^*}(\alpha), \end{cases} \tag{2.3}$$

where the randomizing constant $\gamma(\alpha)$ is chosen such that $E_{\theta^*}\phi_\alpha = \alpha$. Let us consider the "test statistic"

$$\tilde{T}(t, U) = F_{\theta^*}(t) - U \cdot P_{\theta^*}(T = t),$$

where U denotes an independent uniformly (on(0; 1)) distributed random variable. $\tilde{T}$ is not a test statistic in the strict sense, but may be regarded as some randomized kind of test statistic and the test $\tilde{\phi}_\alpha$ based on $\tilde{T}$ is completely equivalent to (2.3), viz.

$$E_\theta \phi_\alpha = E_\theta \tilde{\phi}_\alpha \quad \text{for all } \theta \in \Theta,$$

as may easily be seen.

Consequently the distribution of $\tilde{T}$ under θ^* is uniform on the unit interval. Unfortunately the calculation of the corresponding MC test

$$\tilde{\phi}_m(t) = \begin{cases} 1 & \text{if } \sum_{i=1}^{m} 1_{(-\infty, \tilde{T}(t,u))}(\tilde{T}(t_i, u_i)) \leq k - 1 \\ 0 & \text{if } \sum_{i=1}^{m} 1_{(-\infty, \tilde{T}(t,u))}(\tilde{T}(t_i, u_i)) > k - 1 \end{cases} \tag{2.4}$$

still involves the unknown quantities $F_{\theta^*}(t)$, $F_{\theta^*}(t_i)$. (The u and u_i's, respectively, are realizations of independent uniformly distributed random variables.)

This problem is solved by

PROPOSITION 2.1. *The MC test corresponding to* ϕ_α *may be written as* $(k/(m+1) = \alpha)$

$$\psi_m = \begin{cases} 1 & \text{if } \sum_{i=1}^{m} 1_{(-\infty, t]}(t_i) \leq k - 1, \\ 0 & \text{if } \sum_{i=1}^{m} 1_{(-\infty, t)}(t_i) \geq k, \\ \gamma & \textit{otherwise}, \end{cases} \tag{2.5}$$

where

$$\gamma = \left(k - \sum_{i=1}^{m} 1_{(-\infty, t)}(t_i)\right)\left(\left(\sum_{i=1}^{m} 1_{\{t\}}(t_i)\right) + 1\right)^{-1}.$$

This test is of exact size α. *The power of* ψ_m *may be calculated as*

$$\beta_m(\alpha) = \int_0^1 \beta(\xi) b(\alpha, m, \xi)\, d\xi, \tag{2.6}$$

where $\beta(\alpha)$ *denotes the power of* ϕ_α.

PROOF. One first observes that the tests ψ_m and $\tilde{\phi}_m$ ((2.4) and (2.5), respectively) are equivalent: Given $t, t_1, t_2, \ldots, t_m$ both tests reject the hypothesis with the same probability as may be easily verified. Since $\tilde{\phi}_m$ is the MC test wrt $\tilde{T}$ (possessing a continuous null distribution), an appeal to the results obtained for the continuous case completes the proof. □

3. Power considerations for MC tests. This section is devoted to the study of the power $\beta_m(\alpha)$ of the MC test under a fixed alternative parameter θ_1

(if not stated otherwise). We make use of the terminology of the preceding section and additionally assume $\lim_{\alpha \to 0} \beta(\alpha) = 0$ and $\lim_{\alpha \to 1} \beta(\alpha) = 1$. An essential assumption for $\beta(\cdot)$ is its concavity. This assumption is satisfied if the ϕ_α's are most powerful size-α tests based on T, a condition that will often be met. For a more detailed discussion on this point we refer to Jöckel (1982). The next proposition shows that the concavity property carries over to the corresponding MC test.

PROPOSITION 3.1. *If $\beta(\alpha)$ is a concave function of α then so is $\beta_m(\alpha)$.*

PROOF. Since the densities $b(\alpha, m, \xi)$ constitute a one-parameter exponential family in α, they are by a result of Karlin (1968, page 18) strictly totally positive of order ∞. Since

$$\int_0^1 b(\alpha, m, \xi)\, d\xi = 1, \qquad \int_0^1 \xi b(\alpha, m, \xi)\, d\xi = \alpha,$$

we may apply Proposition 3 in Karlin (1968, page 23) and conclude that the mapping

$$\alpha \mapsto \int \beta(\xi) b(\alpha, m, \xi)\, d\xi$$

is concave. □

REMARK. It may be shown that $\beta_m(\cdot)$ is strictly concave unless $\beta(\cdot)$ is linear.

From a practical point of view the question arises whether an increased simulation sample size m yields an increased power. This question has already been treated by Hope (1968) for a special case, but the general case has remained unsolved. We give

THEOREM 3.2. *If $\beta(\cdot)$ is concave, then the power of the corresponding MC test $\beta_m(\alpha)$ is a monotone increasing function of simulated sample size for all α. Furthermore $\beta_m(\alpha) \uparrow \beta(\alpha)$ uniformly in $\alpha \in [0,1]$ as $m \to \infty$.*

PROOF. Denoting

$$B_m = \frac{\Gamma((m+1)\alpha)\Gamma((m+1)(1-\alpha))}{\Gamma(m+1)}$$

we have by (2.6)

$$\beta_{m+1}(\alpha) - \beta_m(\alpha) = \int_0^1 \frac{\xi^\alpha (1-\xi)^{1-\alpha} B_m - B_{m+1}}{B_{m+1}} b(\alpha, m, \xi) \beta(\xi)\, d\xi,$$

the integrand being

$$\begin{aligned} &> 0 \quad \text{iff } \xi \in (z_1(\alpha), z_2(\alpha)), \\ &= 0 \quad \text{iff } \xi = z_1(\alpha) \text{ or } z_2(\alpha), \\ &< 0 \quad \text{iff } \xi \notin [z_1(\alpha), z_2(\alpha)], \end{aligned}$$

with some suitable functions $z_i(\cdot)$, determined such that

$$(z_i(\alpha))^{\alpha}(1 - z_i(\alpha))^{1-\alpha} = B_{m+1}/B_m,\ i = 1,2 \text{ and } 0 \le z_1(\alpha) \le z_2(\alpha) \le 1.$$

Considering

$$L_\alpha(\xi) = \beta(z_1(\alpha)) + \frac{\beta(z_2(\alpha)) - \beta(z_1(\alpha))}{z_2(\alpha) - z_1(\alpha)}(\xi - z_1(\alpha))$$

with

$$L_\alpha(\xi) \le \beta(\xi) \quad \text{for} \quad \xi \in [z_1(\alpha), z_2(\alpha)],$$
$$L_\alpha(\xi) \ge \beta(\xi) \quad \text{for} \quad \xi \notin [z_1(\alpha), z_2(\alpha)],$$

we obtain

$$\int_0^1 \beta(\xi)(b(\alpha, m+1, \xi) - b(\alpha, m, \xi))\, d\xi$$
$$\ge \int_0^1 L_\alpha(\xi)(b(\alpha, m+1, \xi) - b(\alpha, m, \xi))\, d\xi = 0,$$

which had to be shown. The uniformity is, as in the following corollary, a simple consequence of Dini's theorem. □

REMARK. If $\beta(\cdot)$ is continuous at α, then by L^2-convergence of $b(\alpha, m, \cdot)$ to α we have $\beta_m(\alpha) \to \beta(\alpha)$ as $m \to \infty$. This has already been shown by Birnbaum (1974) and Hope (1968) under more restrictive conditions.

COROLLARY 3.3. *Let $\Theta = \Theta_0 \cup \Theta_1$ be a topological space. Consider for the testing problem Θ_0 versus Θ_1 the level-α tests ϕ_α and assume that the power $\beta_{\theta_1}(\alpha) = \int_0^1 \phi_\alpha\, dF_{\theta_1}$ is a concave function in α, and that for fixed $\alpha \in (0; 1)$*

$$\theta_1 \mapsto \beta_{\theta_1}(\alpha) \text{ is continuous.}$$

Then

$$\int \beta_{\theta_1}(\xi) b(\alpha, m, \xi)\, d\xi \to \beta_{\theta_1}(\alpha) \quad \text{as } m \to \infty$$

uniformly on every compact subset of Θ_1.

From the results obtained so far it should be clear that, with respect to power, an increase in simulation sample size is always desirable. Since, however, the increase of simulation effort considerably increases computer costs, there remain two important questions:

(3.1) How much power is lost by using MC tests; more precisely what may we learn about the minimal value of $\beta_m(\alpha)/\beta(\alpha)$?

(3.2) How many simulations are needed so that the ratio $\beta_m(\alpha)/\beta(\alpha)$ exceeds a given constant?

The first problem has already been tackled by Dwass (1957) in the special case of the Pitman two sample permutation test. His result is obtained under more restrictive conditions and only in the limit as the number of observations of the permutation test tends to infinity.

THEOREM 3.4. *Let* $\alpha \in (0; 1)$ *be fixed, let* $\gamma \in [\beta(\alpha), 1]$ *and* $\tilde{\beta}(\xi)$ *be the power of the test defined by*

$$\tilde{\phi}_\xi = \begin{cases} \dfrac{\xi}{\alpha}\phi_\alpha & \text{if } \xi \le \alpha, \\ \dfrac{1-\xi}{1-\alpha}\phi_\alpha + \dfrac{\xi-\alpha}{1-\alpha}\gamma & \text{if } \xi > \alpha. \end{cases} \tag{3.3}$$

If for all $\xi \in [0; 1]$

$$\tilde{\beta}(\xi) \le \beta(\xi) \tag{3.4}$$

holds, then we have

$$1 - \frac{\beta_m(\alpha)}{\beta(\alpha)} \le \frac{1 - (\alpha\gamma/\beta(\alpha))}{2\alpha(1-\alpha)} E|Z_{m,\alpha} - \alpha|, \tag{3.5}$$

where $Z_{m,\alpha}$ *has a beta distribution with parameters* $p = \alpha(m+1)$ *and* $q = (1-\alpha)(m+1)$. *Furthermore we note that*

$$\begin{aligned} E|Z_{m,\alpha} - \alpha| &= \frac{2}{m+1}\left(\alpha^\alpha(1-\alpha)^{1-\alpha}\right)^{m+1} \\ &\quad\times \frac{\Gamma(m+1)}{\Gamma((m+1)\alpha)\Gamma((m+1)(1-\alpha))} \end{aligned} \tag{3.6}$$

and (for integer values of $\alpha(m+1)$*)*

$$c_m \le \left(\frac{\pi(m+1)}{2\alpha(1-\alpha)}\right)^{1/2} E|Z_{m,\alpha} - \alpha| \le d_m \le 1, \tag{3.7}$$

where

$$\begin{aligned} c_m &= \exp\left(\frac{1}{12(m+1)+1} - \frac{1}{12\alpha(m+1)} - \frac{1}{12(1-\alpha)(m+1)}\right) \to 1, \\ d_m &= \exp\left(\frac{1}{12(m+1)} - \frac{1}{12\alpha(m+1)+1} - \frac{1}{12(1-\alpha)(m+1)+1}\right) \to 1 \quad \text{as } m \to \infty. \end{aligned} \tag{3.8}$$

REMARK. Note that $\tilde{\beta}(0) = 0$, $\tilde{\beta}(\alpha) = \beta(\alpha)$, and $\tilde{\beta}(1) = \gamma$ hold and that $\tilde{\beta}$ is linear on $[0, \alpha]$ and $[\alpha, 1]$. Consequently (3.4) is satisfied if β is concave.

PROOF. By (3.4) and (2.6) we have

$$\begin{aligned}\beta(\alpha) - \beta_m(\alpha) &= E(\beta(\alpha) - \beta(Z_{m,\alpha}))\\ &\le E(\beta(\alpha) - \tilde{\beta}(Z_{m,\alpha}))\\ &= E\left\{\frac{\beta(\alpha) - \alpha\gamma}{2\alpha(1-\alpha)}|Z_{m,\alpha} - \alpha| + \frac{2\alpha\beta(\alpha) - \beta(\alpha) - \alpha\gamma}{2\alpha(1-\alpha)}(Z_{m,\alpha} - \alpha)\right\},\end{aligned}$$

which implies (3.5) in view of $EZ_{m,\alpha} = \alpha$. Furthermore, (3.6) is easily proven by partial integration. Finally we have, by Stirling's formula,

$$\begin{aligned}\Gamma(m+1) &= (m+1)!/(m+1)\\ &\le \sqrt{2\pi}(m+1)^{m+1/2}e^{-(m+1)}e^{1/(12(m+1))}\end{aligned} \tag{3.9}$$

and

$$\begin{aligned}\Gamma((m+1)\alpha) &= (\alpha(m+1))!/(\alpha(m+1))\\ &\ge \sqrt{2\pi}((m+1)\alpha)^{(m+1)\alpha - 1/2}e^{-(m+1)\alpha}\, e^{1/(12(m+1)+1)}.\end{aligned} \tag{3.10}$$

Since, obviously,

$$\frac{1}{12(m+1)} - \frac{1}{12(m+1)\alpha + 1} - \frac{1}{12(m+1)(1-\alpha)+1} \le 0$$

holds, (3.6), (3.9), and (3.10) together yield the right-hand side of (3.7). Another application of Stirling's formula gives us the other part of this inequality. □

Note that (3.5) with $\gamma = \beta(\alpha)$ and (3.6) yield

$$1 - \frac{\beta_m(\alpha)}{\beta(\alpha)} \le \frac{E|Z_{m,\alpha} - \alpha|}{2\alpha} = \frac{\left(\alpha^\alpha(1-\alpha)^{1-\alpha}\right)^{m+1}}{(m+1)\alpha B((m+1)\alpha, (m+1)(1-\alpha))}$$

and that

$$1 - \frac{E|Z_{m,\alpha} - \alpha|}{2\alpha} = \frac{1}{\alpha}\int_0^\alpha \int_u^1 b(\alpha, m, \xi)\, d\xi\, du. \tag{3.11}$$

The quantity given by (3.11) will be called $e^D_{m,\alpha}$, the *Dwass efficiency* (for simulation sample size m), which gives us a lower bound for $\beta_m(\alpha)/\beta(\alpha)$ under the conditions mentioned above. From the asymptotic approximation (3.7) together with (3.8) we are able to tabulate the Dwass efficiency for different values of α and m, presented in Table 1. They are in good accordance with values reported by Dwass (1957).

The asymptotic approximation may be used for a quick determination of the simulation sample size m to achieve Dwass efficiency of at least e^D, viz.

$$m \doteq \frac{1}{2\pi}\frac{1-\alpha}{\alpha}\frac{1}{(1-e^D)^2}.$$

Practical consequences of these considerations and some more applied aspects are discussed in Jöckel (1984).

TABLE 1

Tabulation of Dwass efficiency $e^D_{m,\alpha}$ based on the approximation (3.7). *For the digits displayed the upper and lower bounds coincide.*

α	0.01	0.02	0.025	0.05	0.10
$m = 19$				0.64	0.743
$m = 39$			0.64	0.736	0.815
$m = 59$				0.782	0.848
$m = 99$	0.63	0.732		0.829	0.881
$m = 199$	0.73	0.807	0.827	0.878	0.916
$m = 299$	0.777	0.841		0.900	0.931
$m = 399$	0.806	0.862	0.876	0.913	0.940
$m = 499$	0.825	0.876		0.922	0.947
$m = 599$	0.840	0.887	0.899	0.929	0.951
$m = 699$	0.852	0.895		0.934	0.955
$m = 799$	0.861	0.902	0.912	0.939	0.958
$m = 899$	0.869	0.907		0.942	0.960
$m = 999$	0.876	0.912	0.921	0.945	0.962

4. Pitman efficiency of MC tests. The last section has confirmed that the power properties of MC tests can be satisfactory. In situations where the use of MC tests is indicated, power calculations for the original tests are not normally available. In these cases a widely used and accepted concept for the comparison of tests is that of asymptotic relative Pitman efficiency (ARPE). The aim of this section is to study the ARPE of MC tests.

Throughout this section we shall assume that we wish to test

$$\theta = \theta_0 \text{ versus } \theta \in \Theta - \{\theta_0\},$$

where the parameter space Θ is a subset of the reals, such that the connected component of θ_0, $C(\theta_0) \neq \{\theta_0\}$. Furthermore let (ϕ_n^α) be a sequence of consistent, unbiased level $-\ \alpha$ tests, which are assumed to be asymptotically normal, viz.

$$E_{\theta_n}\phi_n^\alpha \to \Phi\big(\Phi^{-1}(\alpha) + \delta\eta\big), \qquad \delta > 0 \text{ as } n \to \infty, \tag{4.1}$$

where Φ denotes the Gaussian distribution function and θ_n is any sequence of the form

$$\theta_n = \theta_0 + \eta/\sqrt{n} + o\big(1/\sqrt{n}\big), \qquad \eta > 0. \tag{4.2}$$

The parameter δ is called the slope.

The concept of ARPE of MC tests may be developed in a more general framework, Θ being an arbitrary topological space and (ϕ_n^α) possessing different limiting distributions. For a more detailed discussion on this point the reader is referred to Jöckel (1982). Here, however, we will restrict ourselves to the most important case (4.1), which we shall call the *normal limit* case. The derivation in the general case follows essentially the same lines as indicated here.

If $\tilde{\phi}_n^\alpha$ denotes the MC test corresponding to ϕ_n^α (for simulation sample size m) we have for any sequence fulfilling (4.2)

$$E_{\theta_n}\tilde{\phi}_n^\alpha \to G_{m,\alpha}(\delta \cdot \eta), \tag{4.3}$$

where

$$G_{m,\alpha}(x) = \int_0^1 \Phi(\Phi^{-1}(u) + x) b(\alpha, m, u)\, du.$$

To calculate the ARPE of the MC tests wrt the original tests we may apply a result due to Rothe (1981).

We first observe that the tests $(\tilde{\phi}_n^\alpha)$ satisfy the conditions A, B, and C of Rothe's paper with functions $g(\theta) = (\theta - \theta_0)^2$ and $H(\eta) = G_{m,\alpha}(\delta \cdot \sqrt{\eta})$. Hence for fixed α and β, $0 < \alpha < \beta < 1$, an appeal to Theorem 3 in Rothe (1981, page 666) yields that

$$e_m(\alpha, \beta) = \left(\frac{\Phi^{-1}(\beta) - \Phi^{-1}(\alpha)}{G_{m,\alpha}^{-1}(\beta)} \right)^2$$

is the ARPE of the MC tests wrt the original sequence of tests. This quantity depends on α and β, but turns out to be independent of the slope. To study the limiting behaviour of $e_m(\alpha, \beta)$ $(m \to \infty)$ one first observes that the limiting function in (4.1), considered as a function in α, fulfills the conditions of the power function $\beta(\alpha)$ in Section 3. Furthermore it is concave. Thus by virtue of Corollary 3.3. and Theorem 3.2. it is easy to show that

$$e_m(\alpha, \beta) \uparrow 1 \quad \text{as } m \to \infty.$$

This convergence is uniform in β on every compact subset of $(\alpha, 1)$. If simulation sample size m is a nondecreasing function of n and we let $m = m(n)$ and n simultaneously tend to infinity then the MC test and the exact test are asymptotically equivalent.

If we are considering two sequences of tests with slopes δ_1 and δ_2, respectively, the ARPE of the corresponding MC test is (in an obvious terminology)

$$\begin{aligned} e_{\mathrm{MC1,MC2}}(\alpha, \beta) &= e_{\mathrm{MC1},1}(\alpha, \beta) \cdot e_{1,2} \cdot e_{2,\mathrm{MC2}}(\alpha, \beta) \\ &= e_{1,2} = \delta_1^2/\delta_2^2. \end{aligned}$$

Thus in the normal limit case the ARPE of two sequences of MC tests is the same as that of the corresponding original tests.

The problem of ARPE of MC tests in the normal limit case has already been tackled in the literature. Hope (1968) conjectured that for fixed $\alpha > 0$ there exists a constant $\gamma < 1$, with

$$G_{m,\alpha}(\eta) = \Phi(\Phi^{-1}(\alpha) + \gamma \cdot \sqrt{\eta}).$$

This would imply that $e_m(\alpha, \beta) \equiv \gamma^2$, which is obviously not the case, and thus demonstrates why she did not succeed in developing a satisfactory efficiency concept for MC tests.

In order to achieve an asymptotic efficiency measure independent of β we use the concept of local asymptotic Pitman efficiency, slightly modifying a proposal by Hájek and Šidák (1977).

DEFINITION 4.1. If $e_{1,2}(\alpha, \beta)$ denotes the ARPE of two sequences of tests and the limit

$$e_{1,2}(\alpha) = \lim_{\beta \downarrow \alpha} e_{1,2}(\alpha, \beta)$$

TABLE 2
Local asymptotic relative Pitman efficiency of MC tests for simulation sample size m in the normal limit case (in percent).

	α			
m	0.01	0.025	0.05	0.1
19			81.0	86.8
39			89.8	93.0
99			95.6	97.1
499	97.4	98.6	99.1	99.1
999	98.7	99.4	99.5	99.7

exists, then $e_{1,2}(\alpha)$ is called the local asymptotic relative Pitman efficiency (LARPE).

By using l'Hospital's rule and interchanging differentiation and integration it is easy to show that in the normal limit case the LARPE of the MC tests wrt the original tests is

$$e_m(\alpha) = \left[\frac{\int_0^1 f(\Phi^{-1}(u))b(\alpha, m, u)\,du}{f(\Phi^{-1}(\alpha))}\right]^2,$$

where f denotes the normal density. Furthermore $\lim_{m \to \infty} e_m(\alpha) = 1$. By means of numerical integration and some integral transformations $e_m(\alpha)$ is calculated in Table 2.

Acknowledgments. I want to thank B. D. Ripley and a referee for their very helpful comments on an earlier draft of this paper. Another referee helped to considerably improve the presentation and the proof of Theorem 3.4.

REFERENCES

BARNARD, G. A. (1963). Discussion of Professor Bartlett's paper. *J. Roy. Statist. Soc. Ser. B* **25** 294.
BESAG, J. and DIGGLE, D. J. (1977). Simple Monte Carlo tests for spatial pattern. *Appl. Statist.* **26** 327–333.
BIRNBAUM, Z. W. (1974). Computers and unconventional test statistics. In *Reliability and Biometry. Statistical Analysis of Lifelength* 441–458. SIAM, Philadelphia.
DWASS, M. (1957). Modified randomization tests for nonparametric hypotheses. *Ann. Math. Statist.* **28** 181–187.
FOUTZ, R. V. (1980). A method for constructing exact tests from test statistics that have unknown null distributions. *J. Statist. Comput. Simulation* **10** 187–193.
FOUTZ, R. V. (1981). On the superiority of Monte Carlo tests. *J. Statist. Comput. Simulation* **12** 135–137.
GREEN, B. F. (1977). A practical interactive program for randomization tests of location. *Amer. Statist.* **31** 37–39.
HÁJEK, J. and ŠIDÁK, Z. (1977). *Theory of Rank Tests.* Academic, New York.
HOLLANDER, M. (1971). A nonparametric test for bivariate symmetry. *Biometrika* **58** 203–212.

HOPE, A. C. A. (1968). A simplified Monte Carlo significance test procedure. *J. Roy. Statist. Soc. Ser. B* **30** 582–598.

JÖCKEL, K.-H. (1981). A comment on the construction of exact tests from test statistics that have unknown null distributions. *J. Statist. Comput. Simulation* **12** 133–134.

JÖCKEL, K.-H. (1982). Eigenschaften und effektive Anwendung von Monte-Carlo-Tests. Dissertation, Abteilung Statistik der Universität Dortmund.

JÖCKEL, K.-H. (1984). Computational aspects of Monte Carlo tests. *Proceedings of COMPSTAT* **84** 183–188.

KARLIN, S. (1968). *Total Positivity* **1.** Stanford Univ. Press.

LILLIEFORS, H. W. (1967). On the Kolmogorov–Smirnov-tests for normality with mean and variance unknown. *J. Amer. Statist. Assoc.* **62** 399–402.

LILLIEFORS, H. W. (1969). On the Kolmogorov–Smirnov-tests for the exponential distribution with mean unknown. *J. Amer. Statist. Assoc.* **64** 387–389.

MARRIOTT, F. H. C. (1979). Barnard's Monte Carlo tests: How many simulations. *Appl. Statist.* **28** 75–77.

PEIZER, D. B. and PRATT, H. W. (1968). A normal approximation for binomial, F, beta and other common, related tail probabilities, I and II. *J. Amer. Statist. Assoc.* **63** 1416–1483.

RIPLEY, B. D. (1977). Modelling spatial patterns. *J. Roy. Statist. Soc. Ser. B* **39** 172–212.

ROTHE, G. (1981). Some properties of the asymptotic relative Pitman efficiency. *Ann. Statist.* **9** 663–669.

TSUTAKAWA, R. K. and YANG, S. L. (1974). Permutation tests applied to antibiotic drug resistance. *J. Amer. Statist. Assoc.* **69** 87–92.

BREMEN INSTITUTE FOR PREVENTION RESEARCH AND SOCIAL MEDICINE (BIPS)
DIVISION OF BIOMETRY AND DATA PROCESSING
PRÄSIDENT-KENNEDY-PLATZ 1
D-2800 BREMEN 1
WEST GERMANY