Median of the p Value Under the Alternative Hypothesis

Bhaskar BHATTACHARYA and Desale HABTZGHI

Due to absence of an universally acceptable magnitude of the Type I error in various fields, p values are used as a well-recognized tool in decision making in all areas of statistical practice. The distribution of p values under the null hypothesis is uniform. However, under the alternative hypothesis the distribution of the p values is skewed. The expected p value (EPV) has been proposed by authors to be used as a measure of the performance of the test. This article proposes the median of the p values (MPV) which is more appropriate for this purpose. We work out many examples to calculate the MPV's directly and also compare the MPV with the EPV. We consider testing equality of distributions against stochastic ordering in the multinomial case and compare the EPV's and MPV's by simulation. A second simulation study for general continuous data is also considered for two samples with different test statistics for the same hypotheses. In both cases MPV performs better than EPV.

KEY WORDS: Comparing test statistics; Expected p value; Median p value; Power; Stochastic order.

1. INTRODUCTION

The theory of hypothesis testing depends heavily on the pre-specified value of the significance level. To avoid the nonuniqueness of the decision of testing the same hypotheses using the same test statistic but different significance levels, it is a popular choice to report the p value. The p value is the smallest level of significance at which an experimenter would reject the null hypothesis on the basis of the observed outcome. The user can compare his/her own significance level with the p value and make his/her own decision. The p values are particularly useful in cases when the null hypothesis is well defined but the alternative is not (e.g., composite) so that Type II error considerations are unclear. In this context to quote Fisher, "The actual value of p obtainable from the table by interpolation indicates the strength of evidence against the null hypothesis."

We will consider tests of the form "Reject H_0 when $T \geq c$," where T is a real-valued test statistic computed from data when testing the null hypothesis H_0 against the alternative H_1. The value c is determined from the prespecified size restrictions such that $P_{H_0}(T \geq c) = \alpha$. Of course, when T is a discrete random variable, one needs to adopt randomization so that all sizes are possible. If t is the observed value of T and the distribution of T under H_0 is given by $F_0(\cdot)$, then the p value is given by $1 - F_0(t)$ which is the probability of finding the test statistic as extreme as,

or more extreme than, the value actually observed. Thus, if the p value is less than the preferred significance level, then one rejects H_0. Over the years several authors have attempted to properly explain p values. Gibbons and Pratt (1975) provided interpretation and methodology of the p values. Recently, Schervish (1996) treated the p values as significance probabilities. Discussions on p values can be found in Blyth and Staudte (1997) and in Dollinger, Kulinskaya, and Staudte (1996). However, these articles do not treat the p values as random.

The p values are based on the test statistics used and hence random. The stochastic nature of the p values was investigated by Dempster and Schatzoff (1965) and Schatzoff (1966) who introduced "expected significance value." Recently, Sackrowitz and Samuel-Cahn (1999) investigated this concept further and renamed it as the expected p value (EPV). Under the null hypothesis, the p values have a uniform distribution over (0, 1) for any sample size. Thus, under H_0, EPV is 1/2 *always*, and there is no way to distinguish p values derived from large studies and those from small-scale studies. Also it would be impossible under H_0 to differentiate between studies well powered to detect a posited alternative hypothesis and the underpowered to detect the same posited alternative value.

In contrast, the distribution of the p values under the alternative hypothesis is a function of the sample size and the true parameter value in the alternative hypothesis. As the p values measure evidence against the null hypothesis, it is of interest to investigate the behavior of the p values under the alternative at various sample sizes. We reject H_0 when p value is small which is expected when H_1 is true. As noted by Hung, O'Neill, Bauer, and Kohne (1997), the distribution of the p values under the alternative is highly skewed. The skewness increases with the sample size and the true parameter value under the alternative reflecting the ability to detect the alternative by increasing power under these situations. Hence it is more appropriate to consider the median of the p value (MPV) instead of the EPV under the alternative as a measure of the center of its distribution which is the main focus of this article. Applications of the distribution of the p values under the alternative in the area of meta-analysis of several studies is considered by Hung et al. (1997). Studying the p value under the alternative is also beneficial over the power of a test and is explained in the next paragraph.

When several test procedures are available for the same testing situation one compares them by means of power. However, power calculations depend on the chosen significance levels, and in discrete cases involves randomization. These steps can be avoided by considering MPV's. Also as the power functions depend on the chosen significance levels, it is difficult to compare them when different power functions use different significance levels. On the other hand, MPV's depend only on the alternative and not on the significance level. The smaller the value of MPV, the stronger the test. The value of an MPV can tell us which alternative an attained p value best represents for a given sam-

Bhaskar Bhattacharya is Associate Professor, Department of Mathematics, Southern Illinois University, Carbondale, IL 62901 (E-mail: Bhaskar@math.siu.edu). Desale Habtzghi is a Graduate Student, Department of Statistics, University of Georgia, Athens, GA 30602.

ple size. Also, it helps to know the behavior of the MPV's for varying sample sizes.

In Section 2, we obtain a general expression for the MPV's. We also derive a computationally favorable form to be used later in the article. In Section 3, we consider several examples to compute the MPV's directly. In Section 4, we perform a simulation study to compute the MPV's in the case of testing against stochastic ordering for multinomial distributions. We also perform a second simulation study for general distributions when testing the same hypotheses with two samples using three test statistics of t test, Mann–Whitney–Wilcoxon, and Kolmogorov–Smirnov. In Section 5, we make concluding remarks.

2. p VALUE AS A RANDOM VARIABLE AND ITS MEDIAN

For the test statistic T with distribution $F_0(\cdot)$ under H_0, let $F_\theta(\cdot)$ be its distribution under H_1. Also, let $F_0^{-1}(\cdot)$ be the inverse function of $F_0(\cdot)$, so that, $F_0(F_0^{-1}(\gamma)) = \gamma$, for any $0 < \gamma < 1$. Since the p value is the probability of observing a more extreme value than the observed test statistic value, as a random variable it can be expressed as

$$X = 1 - F_0(T).$$

As $F_0(T) \sim U(0, 1)$ under H_0, so is X. The power of the test is related to the p value as

$$\begin{aligned} \beta &= P_\theta(X \le \alpha) \\ &= 1 - F_\theta(F_0^{-1}(1 - \alpha)). \end{aligned} \tag{1}$$

Note the above is also the distribution function of the p value under the alternative. As T is stochastically larger under the alternative than under the null hypothesis, it follows that the p value under the alternative is stochastically smaller than under the null (Lehmann 1986). This explains why the distribution of the p value is skewed to the right under the alternative. Hence to estimate the center of the distribution of the p value, the median is a better choice than the mean. The median is any value of $\alpha = \alpha^*$ which satisfies

$$P_\theta(X < \alpha^*) \le .5 \quad \text{and} \quad P_\theta(X > \alpha^*) \le .5.$$

For continuous distributions, the median is the value of $\alpha = \alpha^*$ which satisfies

$$P_\theta(X \le \alpha^*) = .5.$$

It follows from (1) by simple manipulations that

$$\alpha^* = 1 - F_0(F_\theta^{-1}(.5)). \tag{2}$$

Table 1. MPV's in the Testing for the Normal Mean as a Function of the Significance Level α and Power β

α/β	0.40	0.50	0.60	0.70	0.80	0.90	0.95
0.01	0.0191	0.0100	0.0049	0.0022	0.0008	0.0002	0.0000
0.05	0.0820	0.0500	0.0288	0.0150	0.0065	0.0017	0.0005
0.10	0.1519	0.1000	0.0624	0.0355	0.0169	0.0052	0.0017
0.15	0.2168	0.1500	0.0986	0.0593	0.0302	0.0102	0.0036

Since $F_\theta(t) \le F_0(t), \forall t$, it is seen that $F_0(F_\theta^{-1}(.5)) \ge .5$ which implies that $\alpha^* \le .5$ and equality holds when H_0 is true. The smaller the MPV the better it is to detect the alternative. Given the stochastic nature of the p value under the alternative, it is also true that the MPV is smaller than the EPV and hence clearly preferable over the EPV. The MPV being smaller than the EPV produces a smaller indifference region in the sense that higher power is generated closer to the null hypothesis region using MPV than with using EPV.

It is also possible to express the MPV in another way. Let $T \sim F_\theta(\cdot)$ and, independently, $T^* \sim F_0(\cdot)$. If the observed value of T is t, then the p value is simply $g(t) = P(T^* \ge t | T = t)$. The MPV is

$$\text{med } g(t) = \text{MPV}(\theta) = P(T^* \ge \text{med } T). \tag{3}$$

If H_0 is true, then T and T^* are identically distributed, and hence the above probability is .5 for any continuous distribution. For discrete distributions, although the above expressions still hold, MPV will be slightly higher as $P(T^* = \text{med } T) > 0$. For an UMP test, the MPV will be uniformly minimal for all θ values in the alternative as compared to the MPV's of any other test of the same H_0 versus H_1. For computational purposes the above form of the MPV in (3) is very useful.

3. EXAMPLES

This section considers several examples to calculate the MPV's directly.

Example 1. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a $N(\mu, \sigma^2)$ distribution, and we like to test $H_0 : \mu \le \mu_0$ versus $H_1 : \mu > \mu_0$ where σ is known. Using the test statistic $T = \overline{X}$ and a particular value μ_1 in H_1 it follows from (2) that

$$\text{MPV} = \Phi\left(\frac{\sqrt{n}(\mu_0 - \mu_1)}{\sigma}\right), \tag{4}$$

where Φ is the cdf of the standard normal distribution. It is well known that for a size α test to achieve power β at a specified alternative value μ_1 the sample size n satisfies

$$n = \frac{(z_{1-\alpha} + z_\beta)^2 \sigma^2}{(\mu_0 - \mu_1)^2}, \tag{5}$$

where z_γ is the γth quantile of the standard normal distribution. Using (4) and (5), it follows that

$$\text{MPV} = \Phi(-z_{1-\alpha} - z_\beta). \tag{6}$$

In Table 1, we have provided values of the MPV's in (6) using some commonly used values of α and β. Each MPV value is smaller than the corresponding EPV value of Table 1 of Sackrowitz and Samuel-Cahn (1999). We have also graphed the EPV and the MPV in Figure 1 for various values of $\mu_1 > 0$ when $\mu_0 = 0$ at sample sizes 10 and 50. The MPV's decrease from the .5 value at much faster rate than the EPV's although for both the rate increases with the sample size. In Figure 2, we have graphed the EPV and the MPV for various sample sizes at $\mu_1 = .3$ and at $\mu_1 = .5$. It is observed that the MPV's decrease at a faster rate than the EPV's at smaller sample sizes and when closer to the null hypothesis.

Figure 1. EPV's and MPV's at various $\mu_1 > 0$ when testing for the normal mean with $\mu_0 = 0$ for sample sizes $n = 10$ and $n = 50$.

If the value of σ is unknown, the sample standard deviation S (or any other consistent estimator of σ) may be used to replace it for moderately large n. The formula in (4) is approximately correct in this case, consequently, Table 1 is approximately correct for the one-sample t test situation. After observing an actual p value for an approximately normally distributed statistic, the expression in (4) can be used to determine the μ_1 for which the given value would be an MPV.

Example 2. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from $N(\mu_1, \sigma^2)$, and independently, let $Y_1, Y_2, \ldots, Y_m$ be another random sample of size m from $N(\mu_2, \sigma^2)$, and we like
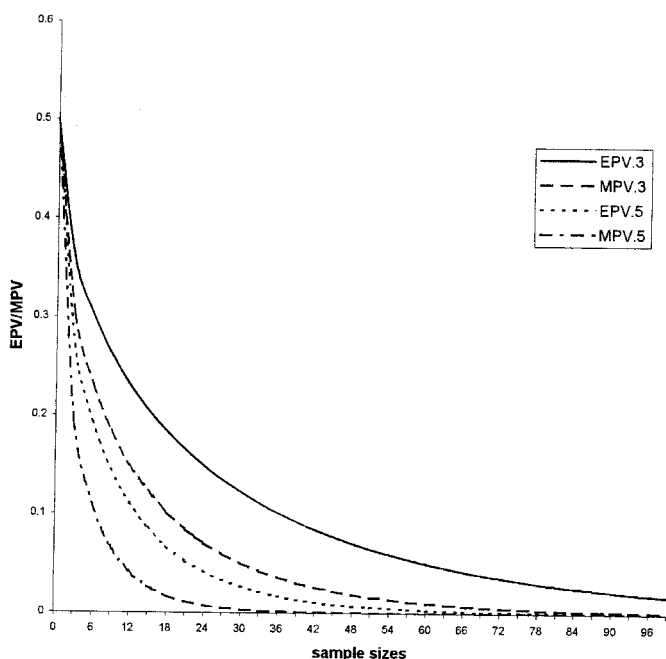
Figure 2. EPV's and MPV's at various sample sizes when testing for the normal mean with $\mu_0 = 0$ and $\mu_1 = .3$ and $\mu_1 = .5$.

to test $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 > \mu_2$ where σ is known. Using the test statistic $T = \overline{X} - \overline{Y}$ and a particular value $\mu_1 - \mu_2$ in H_1 it follows that

$$\text{MPV} = \Phi\left(\sqrt{\frac{mn}{m+n}}\frac{\mu_2 - \mu_1}{\sigma}\right). \qquad (7)$$

When $m = n$, formula (6) is still valid and hence Table 1 is also correct in this case. For unknown σ, a consistent estimator of σ may be used in (7) for moderately large m, n, and the formula in (7) becomes approximately correct in this case.

Example 3. Suppose $X_1, X_2, \ldots, X_n$ is a random sample of size n from a $N(\mu, \sigma^2)$ distribution, and we like to test $H_0 : \sigma \leq \sigma_0$ versus $H_1 : \sigma > \sigma_0$ when μ is unknown. Using the test statistic $(n-1)S^2/\sigma^2$ where S^2 is the sample variance, it follows from (4) that at the alternative point σ_1,

$$\text{MPV} = 1 - G\left(\frac{\sigma_1^2}{\sigma_0^2}G^{-1}(.5)\right), \qquad (8)$$

where G is the cdf of a chi-square distribution with $n-1$ degrees of freedom. Note we need not use the F-distribution as needed for the computation of the EPV in this problem (Sackrowitz and Samuel-Cahn 1999).

The following two examples are concerned with testing the scale and location parameters of the exponential distribution.

Example 4. Suppose T is exponentially distributed with parameter θ (from pdf $f(t) = \theta e^{-t\theta}$ for $t > 0$) denoted by $\exp(\theta)$ and we like to test $H_0 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$. It is seen that med $T = (1/\theta)\ln 2$. For a particular value of $\theta_1 < \theta_0$, if $T^* \sim \exp(\theta_0)$ and $T \sim \exp(\theta_1)$, the MPV is given by

$$P(T^* \geq \text{med } T) = 2^{-\theta_0/\theta_1}.$$

For a size α test with power β, since $\ln \alpha / \ln \beta = \theta_0/\theta_1$, it follows that for $\alpha = .1$ and $\beta = .9$, the MPV is 2.645×10^{-7}. The EPV is .0438 in this case (Sackrowitz and Samuel-Cahn 1999). If a random sample $X_1, X_2, \ldots, X_n$ is available, the test may be based on $T = \sum_{i=1}^n X_i$. Here T has a gamma distribution with shape parameter n and scale parameter θ, we will denote its cdf by $G_{n,\theta}(\cdot)$. Then it follows that the MPV at alternative point θ_1 is given by

$$P(T^* \geq G_{n,\theta_1}^{-1}(.5)) = 1 - G_{n,\theta_0}(G_{n,\theta_1}^{-1}(.5)).$$

Example 5. Suppose $X_1, X_2, \ldots, X_n$ is a random sample of size n from an exponential distribution with parameters μ, θ (with pdf $f(x) = (1/\theta)e^{-(x-\mu)/\theta}$ for $x \geq \mu$) denoted by $\exp(\mu, \theta)$ and we like to test $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$ where θ is known. The test statistic is $T = \min(X_1, X_2, \ldots, X_n)$ whose distribution is $\exp(\mu, \theta/n)$. It is seen that med $T = \mu + (\theta/n)\ln 2$. For a particular alternative $\mu_1 > \mu_0$, the MPV is given by $.5e^{-n(\mu_1 - \mu_0)/\theta}$.

4. SIMULATION STUDIES

We perform two simulation studies to calculate and compare the EPV's and MPV's. For the discrete case we consider the binomial distribution, with m trials and probability of success
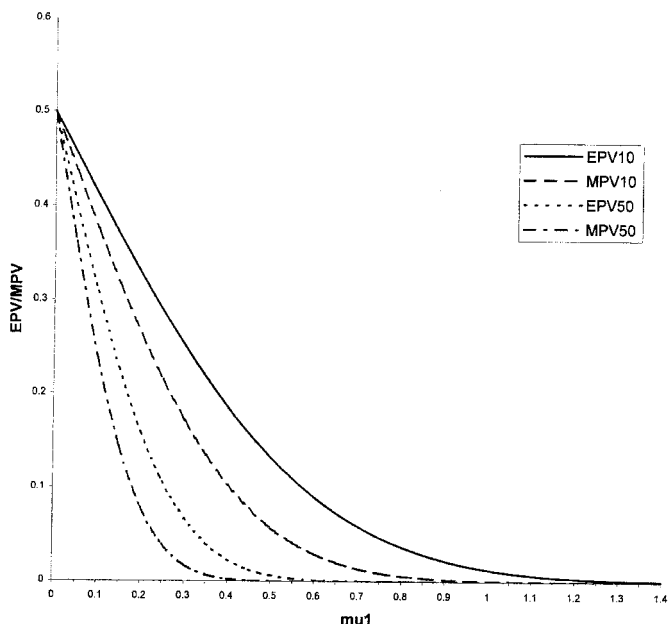
| | m = 3 | | | | m = 6 | | | |
| | EPV | | MPV | | EPV | | MPV | |
p	n = 50	n = 100	n = 50	n = 100	n = 50	n = 100	n = 50	n = 100
0.50	0.509	0.508	0.510	0.506	0.509	0.511	0.505	0.510
0.51	0.454	0.432	0.439	0.407	0.445	0.421	0.421	0.401
0.52	0.404	0.364	0.363	0.309	0.380	0.335	0.332	0.279
0.53	0.350	0.291	0.292	0.222	0.319	0.250	0.248	0.164
0.54	0.300	0.235	0.230	0.143	0.258	0.170	0.168	0.079
0.55	0.252	0.174	0.165	0.081	0.202	0.110	0.103	0.033
0.56	0.208	0.129	0.117	0.045	0.149	0.065	0.055	0.011
0.57	0.169	0.087	0.078	0.020	0.102	0.037	0.027	0.003
0.58	0.135	0.054	0.050	0.007	0.069	0.019	0.012	0.001
0.59	0.102	0.036	0.031	0.003	0.044	0.008	0.006	0.000
0.60	0.075	0.021	0.017	0.001	0.026	0.002	0.002	0.000
0.61	0.054	0.011	0.008	0.000	0.013	0.001	0.001	0.000
0.62	0.037	0.006	0.003	0.000	0.007	0.000	0.000	0.000
0.63	0.026	0.003	0.002	0.000	0.003	0.000	0.000	0.000
0.64	0.017	0.001	0.001	0.000	0.001	0.000	0.000	0.000

p, given by

$$p_j = \binom{m}{j} p^j (1-p)^{m-j}, \quad j = 0, \ldots, m \quad (9)$$

which is symmetric when $p = .5$. When $p > .5$, the binomial distribution is skewed to the left, that is, it becomes stochastically larger than the $p = .5$ case. Let $\mathbf{q} = (q_0, \ldots, q_m)$ and $\mathbf{p} = (p_0, \ldots, p_m)$ be the vectors of binomial probabilities obtained from (9) with $p = .5$ and $p > .5$, respectively. We consider testing $H_0 : \mathbf{p} = \mathbf{q}$ (i.e., $p_i = q_i, \forall i$) against $H_1 : \mathbf{p}$ is stochastically larger than $\mathbf{q}$ (i.e., $\sum_{i=j}^{m} p_i \geq \sum_{i=j}^{m} q_i, \forall j = 1, \ldots, m$, and $\sum_{i=0}^{m} p_i = \sum_{i=0}^{m} q_i = 1$). The

likelihood ratio test statistic is given by

$$T = 2n \sum_{i=0}^{m} \hat{p}_i \ln(\overline{p}_i / q_i)$$

where $\overline{p}_i$ is the ith coordinate of $\overline{\mathbf{p}} = \hat{\mathbf{p}} E_{\hat{\mathbf{p}}}(\mathbf{q}/\hat{\mathbf{p}}|\mathcal{A})$, $E_{\hat{\mathbf{p}}}(\mathbf{q}/\hat{\mathbf{p}}|\mathcal{A})$ is the isotonic regression of $\mathbf{q}/\hat{\mathbf{p}}$ (all multiplications and divisions of vectors are done coordinatewise) onto the nonincreasing cone $\mathcal{A} = \{\mathbf{x} = (x_0, x_1, \ldots, x_m) : x_0 \geq x_1 \geq \cdots \geq x_m\}$ with weights $\hat{\mathbf{p}}$. It is well established (Robertson, Wright, and Dykstra 1988) that under H_0, asymptotically, the statistic T has a chi-bar squared distribution.

Table 3. MPV's for the One-Sided, Two-Sample t, Mann–Whitney–Wilcoxon (MWW), and Kolmogorov–Smirnov (KS) Tests for Various Sample Sizes, Shift Parameters and Underlying Distributions

| | n = 10 | | | n = 20 | | | n = 50 | | |
	t	MWW	KS	t	MWW	KS	t	MWW	KS
					Normal(0,1)				
$\Delta_0 = 0.0000$	0.4943	0.5196	0.6745	0.4964	0.5032	0.6485	0.5066	0.4927	0.6181
$\Delta_1 = 0.3600$	0.2115	0.2497	0.4072	0.1275	0.1368	0.1662	0.0389	0.0413	0.0888
$\Delta_2 = 0.5692$	0.1084	0.1271	0.2039	0.0399	0.0458	0.0886	0.0027	0.0024	0.0100
$\Delta_3 = 0.8050$	0.0446	0.0567	0.0843	0.0085	0.0086	0.0189	0.0001	0.0000	0.0007
					Exponential				
$\Delta_0 = 0.0000$	0.5022	.5125	0.6858	0.4924	0.5078	0.6397	0.4864	0.5094	0.6031
$\Delta_1 = 0.3600$	0.2030	0.1235	0.2157	0.1221	0.0505	0.0913	0.0309	0.0041	0.0017
$\Delta_2 = 0.5692$	0.0905	0.0544	0.0851	0.0336	0.0104	0.0060	0.0013	0.0000	0.0000
$\Delta_3 = 0.8050$	0.0312	0.0168	0.0281	0.0038	0.0012	0.0004	0.0000	0.0000	0.0000
					Chi-square(10)				
$\Delta_0 = 0.0000$	0.4863	0.5202	0.6815	0.5151	0.4878	0.6491	0.4891	0.5050	0.5990
$\Delta_1 = 1.6100$	0.1955	0.2234	0.4172	0.1300	0.1129	0.1682	0.0349	0.0318	0.0533
$\Delta_2 = 2.5456$	0.0947	0.1129	0.2001	0.0376	0.0299	0.0844	0.0023	0.0017	0.0054
$\Delta_3 = 3.6000$	0.0341	0.0376	0.0764	0.0063	0.0049	0.0153	0.0000	0.0000	0.0010
					Uniform(0,1)				
$\Delta_0 = 0.0000$	0.4959	0.5196	0.6745	0.5013	0.5032	0.6485	0.5086	0.4927	0.6181
$\Delta_1 = 0.1039$	0.2154	0.2497	0.4072	0.1305	0.1426	0.2897	0.0365	0.0442	0.0888
$\Delta_2 = 0.1643$	0.1125	0.1271	0.2039	0.0431	0.0506	0.0886	0.0026	0.0036	0.0187
$\Delta_3 = 0.2324$	0.0477	0.0647	0.2039	0.0104	0.0122	0.0416	0.0000	0.0001	0.0022
					Double exponential				
$\Delta_0 = 0.0000$	0.5016	0.5119	0.6830	0.5008	0.5122	0.6507	0.4947	0.5013	0.6074
$\Delta_1 = 0.5091$	0.2119	0.1793	0.2138	0.1260	0.0928	0.1688	0.0329	0.0156	0.0183
$\Delta_2 = 0.8050$	0.0984	0.0841	0.0842	0.0357	0.0197	0.0426	0.0019	0.0003	0.0004
$\Delta_3 = 1.1384$	0.0370	0.0327	0.0286	0.0054	0.0033	0.0068	0.0000	0.0000	0.0000

We create a random sample $T_1, \ldots, T_n$ distributed like T, and independently, another random sample $T_1^*, \ldots, T_n^*$ distributed like T^*. An unbiased estimator of EPV is given by

$$A_E = \frac{1}{n} \sum_{i=1}^{n} I(T_i^* \geq T_i),$$

and an unbiased estimator of MPV is given by

$$A_M = \frac{1}{n} \sum_{i=1}^{n} I(T_i^* \geq \text{med } T_i).$$

The variance of A_E is $\text{EPV}(1 - \text{EPV})/n$ and that of A_M is $\text{MPV}(1 - \text{MPV})/n$

We consider $m = 3$, 6 with sample sizes $n = 50, 100$ and replications 10,000. The results are given in Table 2. When $p = .5$, both of the EPV and MPV start slightly higher than .5 as expected for discrete distributions. For $p > .5$, the MPV's are consistently smaller than the EPV's, both being very close to zero when $p > .65$. The effects are more pronounced for larger n. Note the exact value of the EPV or the MPV is difficult to calculate in this case.

We consider a second simulation study to compare the performance of several tests using the MPV's for a general continuous case. We use the same set up as Sackrowitz and Samuel-Cahn (1999) but calculate the MPV's instead. Thus we consider the two-sample problem of testing $H_0 : F = G$ versus $H_1 : F$ is stochastically larger than G using two independent random samples from F and G respectively. The test statistics considered are the two-sample t test, the Mann–Whitney–Wilcoxon (MWW) test and the Kolmogorov–Smirnov (KS) test. The comparison is made for shift alternatives so that $G(x) = F(x + \Delta)$, where F is chosen as various distributions. We consider F as normal $(0,1)$, exponential, chi-square with degrees of freedom 10, uniform $(0,1)$ and double exponential. We chose sample sizes 10, 20, 50 and $\Delta_0 = 0$, $\Delta_1 = 2.546\sigma/\sqrt{50}$, $\Delta_2 = 2.546\sigma/\sqrt{20}$, $\Delta_3 = 2.546\sigma/\sqrt{10}$, where σ is the actual standard deviation of the underlying distribution F. This choice is made so that when the underlying distribution is normal, a test based on the normal two-sample statistic and n observations would have size $\alpha = .10$ and power $\beta = .7$ (as opposed to $\beta = .9$ of Sackrowitz and Samuel-Cahn 1999, table 4) for $\Delta = (2.546\sigma/\sqrt{n})$ and thus from Table 1 we have MPV = .0355. Note that we have used the same Δ_i values for all n. We consider 10,000 replications.

The MPV values in Table 3 have similar magnitude as the EPV values of Sackrowitz and Samuel-Cahn (1999, table 4) *even* at

$\beta = .7$. They also have similar pattern as the corresponding EPV values. So their conclusions are also valid in our case. However, our KS values of MPV perform worse than the corresponding EPV values of Sackrowitz and Samuel-Cahn (1999) in all cases considered.

5. CONCLUSION

The distribution of the p values under the alternative is a skewed distribution to the right, and hence the median of this distribution is advocated as a more appropriate tool than its mean for determination of the strength of a test for a particular alternative. The alternatives closer to H_0 are detected easily with MPV than with EPV. The MPV is easily computed in most cases and does not depend on the specified significance level of a test. Thus, it may be used as a single number which can help to choose among different test statistics when testing the same hypotheses. For approximately normally distributed statistics, Table 1 can be consulted to relate the MPV value to the usual significance level and power combinations.

[Received January 2001. Revised August 2001.]

REFERENCES

Blyth, C. R., and Staudte, R. G. (1997), "Hypothesis Estimation and Acceptability Profiles for 2×2 Contingency Tables," *Journal of the American Statistical Association*, 92, 694–699.

Dempster, A. P., and Schatzoff, M. (1965), "Expected Significance Level as a Sensibility Index for Test Statistics," *Journal of the American Statistical Association*, 60, 420–436.

Dollinger, M. B., Kulinskaya, E., Staudte, R. G. (1996), "When is a p Value a Good Measure of Evidence," in *Robust Statistics, Data Analysis and Computer Intensive Methods*, ed. H. Reider, New York: Springer-Verlag, pp. 119–134.

Gibbons, J. D., and Pratt, J. W. (1975), "P-values: Interpretation and Methodology," *The American Statistician*, 29, 20–25.

Hung, H. M. J., O'Neill, R. T., Bauer, P., and Kohne, K. (1997), "The Behavior of the p Value When the Alternative Hypothesis is True," *Biometrics*, 53, 11–22.

Lehmann, E. L. (1986), *Testing Statistical Hypotheses*, (2nd ed.), New York: Wiley.

Robertson, T., Wright. F. T., and Dykstra, R. L. (1988), *Order Restricted Statistical Inference*, New York: Wiley.

Sackrowitz, H., and Samuel-Cahn, E. (1999), "P Values as Random Variables-Expected p Values," *The American Statistician*, 53, 326–331.

Schatzoff, M. (1966), "Sensitivity Comparisons Among Tests of the General Linear Hypotheses," *Journal of the American Statistical Association*, 61, 415–435.

Schervish, M. J. (1996), "P Values: What They Are and What They Are Not," *The American Statistician*, 50, 203–206.