Statistica Sinica 4(1994), 493-504

A CONCEPT OF TYPE-2 p-VALUE

Kesar Singh and Robert H. Berk

Rutgers University

Abstract: A concept of p-value is introduced and developed in this article which is more suitable for the bootstrap. Unlike the classical p-value, which goes to zero under H_1 like type-1 error, the proposed p-value goes to zero under H_1 like type-2 error. The name for this concept is derived from this fact. Corrections are obtained which make the type-2 p-value go to $U[0,1]$ faster under H_0 and yet do not make its behavior under H_1 worse, in terms of slope. In view of the limiting $U[0,1]$ distribution under H_0, this new p-value is to be used in testing just like the classical p-value.

Key words and phrases: p-value, bootstrap, jackknife.

1. Introduction

This study was motivated by the need for a concept like the classical p-value, commonly used for reporting the result of a statistical testing of hypothesis, based on modern resampling techniques like the bootstrap and the jackknife. Under H_0, the classical p-value related to a test, denoted by p_n here, has a uniform or subuniform distribution on $[0,1]$. Under H_1, it goes to zero, as $n \to \infty$, like the type-1 error α of the test when the type-2 error β is held fixed at a nonzero level. To be precise, if β is held fixed at some nonzero level, α_n denotes its type-1 error while p_n is the classical p-value, then, typically

$$\frac{1}{n}\log p_n \sim \frac{1}{n}\log \alpha_n \quad \text{a.s.}$$

under H_1. See Bahadur (1972). The type-2 p-value (denoted by p_n^*) introduced here has an asymptotic $U[0,1]$ distribution under H_0, and under H_1, it goes to zero like the type-2 error β of the test when the type-1 error α is fixed at a nonzero level. Thus if α is fixed at a nonzero level, and β_n denotes the corresponding type-2 error, then, under H_1,

$$\frac{1}{n}\log p_n^* \sim \frac{1}{n}\log \beta_n \quad \text{a.s.}$$

The name for this new concept is derived from this fact.

Suppose we have data $X_1, \ldots, X_n$ from a population F, univariate or multivariate. Let θ be a univariate functional of F, such as the mean, median, variance, correlation, etc. Let T_n be an estimator of θ. Consider the hypotheses: $H_0 : \theta = \theta_0$ and $H_1 : \theta > \theta_0$. Even though H_0 appears to be a simple hypothesis it is, in most problems that statisticians encounter, a composite hypothesis involving finitely many or infinitely many additional parameters needed to specify F completely. The classical p-value is defined as

$$p_n = \sup_{F \in H_0} P_F(T_n \geq T_{nc}),$$

where T_{nc} denotes the computed value of T_n.

In practice though, p_n is almost never evaluated exactly, following the above definition, unless of course when there is a single null distribution. Usually, an approximation to p_n is obtained using asymptotics, with estimated functionals appearing in the limiting distribution. The limit of $-\frac{1}{n} \log p_n$, under H_1, is known as Bahadur's slope, which is used for determining how fast p_n goes to zero under H_1. Indeed, the faster the better. The null distribution of a typical approximated p-value is asymptotically uniform and Bahadur's slope is not retained by the approximations. If a user is testing at a level α, then he would reject H_0 if and only if $p_n < \alpha$. Notice that when one does not know whether X-data are from H_0 or H_1, resampling can not be used directly to approximate p_n. The concept of type-2 p-value proposed here is defined as follows:

$$p_n^* = P_{\hat{F}}(T_n^* \leq \theta_0),$$

where $\hat{F}$ is the estimated underlying population. Here, T_n^* denotes the statistics T_n computed on $Y_1, \ldots, Y_n$, a random sample from $\hat{F}$. In most applications $\hat{F} = F_n$, the empirical distribution of the X-data. In case one has enough faith in a parametric setup one should use a parametric estimator of F to sample from, instead of F_n. It makes some of the asymptotics easier but then p_n^* loses its nonparametric character.

Heuristically speaking, the motivation for this definition comes from the fact that, when $\theta = \theta_0$, p_n^* is expected to swing to both sides of $\frac{1}{2}$ equally. On the other hand when $\theta > \theta_0$, p_n^* is expected to dwindle towards zero. Incidentally, p_n^* is also the aimed coverage probability of the smallest one sided confidence interval based on Efron's percentile method which includes θ_0. When H_1 is $\theta \neq \theta_0$, i.e. H_1 is two sided, the definition of p_n^* modifies to

$$p_n^* = 2 \min\{P_{\hat{F}}(T_n^* \leq \theta_0),\ 1 - P_{\hat{F}}(T_n^* \leq \theta_0)\}.$$

In the testing situation $H_0 : \theta \in [a, b]$ vs. $H_1 \in [a, b]^c$, where $a < b$, one could define the type-2 p-value as

$$p_n^* = \min\{P_{\hat{F}}(T_n^* \geq a),\ P_{\hat{F}}(T_n^* \leq b)\}.$$

For further development along this line in the multiparameter setting with a general H_0, see Liu and Singh (1993).

For an illustration, consider the binomial problem with $n = 20$. Let $H_0 : p = \frac{1}{2}$ vs. $H_1 : p > \frac{1}{2}$. Let us say $\hat{p} = 15/20$. In this problem, the usual p-value $= 2.1\%$ and the type-2 p-value $= 1.4\%$.

In the next section, we study this type-2 p-value under the null hypothesis. We establish a result to the effect that its asymptotic distribution is $U[0,1]$. Thus in the terminology of Beran (1987), p_n^* is a prepivoted statistic under the null. The typical rate of convergence to uniformity is $O(1/\sqrt{n})$. We also correct p_n^* to make the convergence rate $O(1/n)$. Our correction is analogous to Efron's abc (accelerated bias correction); see Efron (1987) and Hall (1988).

There are two notable features of the modified p_n^* (denoted by p_n^{**}): I. It combines two most popular resampling procedures namely the bootstrap and the jackknife, and it seems that any one of the two alone is inadequate for the purpose. II. The correction leaves the index of exponential rate of decay under H_1 unchanged or makes it larger. More precisely,

$$\liminf_{n\to\infty} \left[\frac{1}{n}\log p_n^* - \frac{1}{n}\log p_n^{**}\right] \geq 0$$

as $n \to \infty$, under H_1. This section hinges on some known results on Edgeworth expansion. The technical discussion is limited to the so-called "functions of multivariate mean". It remains to be explored if the same formula for p_n^{**} will carry to other general classes of statistics. In view of the limiting $U[0,1]$ distribution of p_n^* or p_n^{**}, a user testing at α level will reject H_0 if and only if p_n^* (or p_n^{**}) is less than α.

In Section 3, we study the type-2 p-value under H_1. The discussion of p_n^* under H_1, takes us to the large deviation probabilities for bootstrap distributions, a topic of interest on its own. We establish a result to the effect that

$$\frac{1}{n}\log P_{\hat{F}}(T_n^* \leq \theta_0) \sim \frac{1}{n}\log P_F(T_n \leq \theta_0)$$

under H_1. The limit of the latter is the same as the index of exponential decay of β when α is held fixed. This index is known as the Hodges-Lehman efficacy in the literature (see Hodges and Lehman (1956)). Here is an interesting property of p_n^* under H_1. Suppose F_θ is a parametric family and the test "Reject H_0 if $T_n \geq c$" enjoys optimality in the Hodges-Lehman sense, i.e. it maximizes $-\lim \frac{1}{n}\log\beta_n$ when α is held fixed. Then, among all the p_n^* based on different statistics, the p_n^* based on T_n maximizes the limit of $-\frac{1}{n}\log p_n^*$, even if one ignores the parametric family in defining p_n^*, i.e. one simply takes $\hat{F} = F_n$. On the other hand, if the model F_θ is wrong but T_n is a legitimate estimator of θ, then p_n^* is still a perfectly legitimate p-value. For example, in the normal model $N(\theta, \sigma^2)$ case, let $T_n = \bar{X}$. Here, $\lim \frac{1}{n} \log P_{F_n}(T_n \leq \theta_0) = -\frac{(\theta-\theta_0)^2}{2\sigma^2}$. This is same as the Hodges-Lehman efficiency of the t-test or the z-test with estimated σ^2. It should be remarked here that, under the normal model, Hodges-Lehman efficiency does not discriminate between the t-test and the z-test, unlike the Bahadur slope.

2. p_n^* under H_0

We begin this section with an elementary result which proves that the limiting distribution of p_n^* is uniform on $[0,1]$. The result assumes that $\sqrt{n}(T_n - \theta)$ has a limiting normal distribution, though it will extend to the cases where $n^\alpha(T_n - \theta)$ has a symmetric and continuous limiting distribution, provided there is an analogous valid bootstrap approximation.

Theorem 2.1. *Assume that*

$$\sqrt{n}(T_n - \theta) \xrightarrow{\mathcal{L}} N(0, V_F^2)$$

and

$$\sqrt{n}(T_n^* - T_n) \xrightarrow{\mathcal{L}} N(0, V_F^2)$$

under $\hat{F}$, a.s. *Then, under $\theta = \theta_0$, as $n \to \infty$,*

$$P(p_n^* \leq t) \to t,$$

uniformly in $t \in [0,1]$.

Proof. By definition:

$$\begin{aligned} p_n^* &= P_{\hat{F}}(T_n^* \leq \theta_0) \\ &= P_{\hat{F}}(\sqrt{n}(T_n^* - T_n) \leq \sqrt{n}(\theta_0 - T_n)). \end{aligned}$$

Polya's theorem (see Rao (1973), Ch. 2c) states that if the limiting distribution in a weak convergence is continuous, then the distributional convergence is uniform. Thus, we can write the above as

$$= \Phi(\sqrt{n}(\theta_0 - T_n)/V_F) + o(1) \text{ a.s.}$$

using Polya's theorem. But, the limiting distribution of $\Phi(\sqrt{n}(T_n - \theta_0)/V_F)$ is uniform on $[0,1]$, under $\theta = \theta_0$. The result follows, using the symmetry of the normal distribution.

Next, we prepare the framework for a valid one term expansion for the distribution of p_n^*, assuming the required conditions on T_n and T_n^*. Let us suppose that

$$P\Big(\sqrt{n}\frac{T_n - \theta}{V_F} \leq x\Big) = \Phi(x) + \frac{p_F(x)}{\sqrt{n}}\phi(x) + r_n(x)$$

and

$$P\Big(\sqrt{n}\frac{T_n-\theta}{V_{\hat{F}}}\le x\Big)=\Phi(x)+\frac{q_F(x)}{\sqrt{n}}\phi(x)+w_n(x),$$

where both $\sup_x|r_n(x)|$ and $\sup_x|w_n(x)|$ are $o(n^{-1/2})$, and $p_F(\cdot)$, $q_F(\cdot)$ are even second degree polynomials with coefficients depending upon F. Further, suppose that there are corresponding bootstrap expansions, i.e.

$$P_{\hat{F}}\Big(\sqrt{n}\frac{T_n^*-T_n}{V_{\hat{F}}}\le x\Big)=\Phi(x)+\frac{p_F(x)}{\sqrt{n}}\phi(x)+r_n^*(x)$$

and

$$P_{\hat{F}}\Big(\sqrt{n}\frac{T_n^*-T_n}{V_{\hat{F}^*}}\le x\Big)=\Phi(x)+\frac{q_F(x)}{\sqrt{n}}\phi(x)+w_n^*(x),$$

where $\hat{F}^*$ is the estimate of $\hat{F}$ based on the bootstrap data, and both $\sup_x|r_n^*(x)|$ and $\sup_x|w_n^*(x)|$ are $o(n^{-1/2})$ a.s. (See Singh (1981) and Babu and Singh (1983, 1984) for the validity of such expansions.) For our purpose here, though, $o(n^{-1/2})$ a.s. is inadequate. We require a different version of $o_p(n^{-1/2})$ which we denote by $\bar{o}_p(n^{-1/2})$.

Definition. We say $R_n=\bar{o}_p(n^{-1/2})$ if for every $\epsilon>0$,

$$P(|R_n|>\epsilon/\sqrt{n})=o(n^{-1/2}).$$

We assume here that both $\sup_x|r_n^*(x)|$ and $\sup_x|w_n^*(x)|$ are $\bar{o}_p(n^{-1/2})$. (See Hall (1986) which has a stronger bound $\bar{o}_p(n^{-1})$ for the class of statistics known as functions of multivariate means, under stronger moment conditions than would be needed for $\bar{o}_p(n^{-1/2})$.)

Theorem 2.2. *Let $\Phi(z_t)=t$. Under the assumed conditions,*

$$P_{\hat{F}}(p_n^*\le t)=t-\frac{1}{\sqrt{n}}[p_F(z_t)+q_F(z_t)]+u_n(t),$$

where $u_n(t)\to 0$, uniformly in $t\in[0,1]$.

This establishes the fact that the rate of convergence of the distribution of p_n^* to uniformity is $O(n^{-1/2})$, like the rate in the classical Berry-Esseen bound (see Feller (1971), Ch. 16).

Example. In the most basic case of one-dimensional mean $\bar{X}$, $p_F(x)=-\frac{1}{6}\frac{\mu_3}{\sigma^3}(x^2-1)$ and $q_F(x)=\frac{1}{6}\frac{\mu_3}{\sigma^3}(2x^2+1)$. In this case, the existence of six moments of the population and its continuity, will suffice for $\bar{o}_p(n^{-1/2})$ rate. Thus, in this case, one has

$$P(p_n^*\le t)=t-\frac{1}{6\sqrt{n}}\frac{\mu_3}{\sigma^3}[x^2+2]\phi(z_t)+o(n^{-1/2}).$$

Proof of Theorem 2.2. Let $t_n = (T_n - \theta_0)/(V_{\hat{F}}/\sqrt{n})$. Fix $\epsilon > 0$. Let $B_n = \{|t_n| \le \log n \text{ and } \sup_x |r_n^*(x)| \le \epsilon/\sqrt{n}\}$. Clearly, $P(B_n^c) = o(n^{-1/2})$. We restrict $(X_1, \ldots, X_n)$ to B_n in the proof that follows. By definition,

$$\begin{aligned} p_n^* &= P_{\hat{F}}\Big(\frac{T_n^* - T_n}{V_{\hat{F}}/\sqrt{n}} \le -t_n\Big) \\ &= \Phi(-t_n) + \frac{p_F(-t_n)}{\sqrt{n}}\phi(-t_n) + r_n(-t_n) \\ &= \Phi\Big(-t_n + \frac{p_F(-t_n)}{\sqrt{n}}\Big) + a_n, \end{aligned}$$

where $|a_n| \le \frac{2\epsilon}{\sqrt{n}}$, after certain n onward. For all $u \in [\frac{2\epsilon}{\sqrt{n}}, 1 - \frac{2\epsilon}{\sqrt{n}}]$,

$$P(p_n^* \le u) = P\Big(-t_n + \frac{p_F(-t_n)}{\sqrt{n}} \le z_{u-a_n}\Big). \tag{2.1}$$

At this stage, we use the following expansion which can be established using arguments parallel to Theorem 1 of Abramovitch and Singh (1985).

$$P\Big(-t_n + \frac{p_F(-t_n)}{\sqrt{n}} \le x\Big) = \Phi(x) - \frac{p_F(x) + q_F(x)}{\sqrt{n}}\phi(x) + o(n^{-1/2}). \tag{2.2}$$

It is deduced from (2.1) and (2.2) that uniformly in $u \in [\frac{2\epsilon}{\sqrt{n}}, 1-\frac{2\epsilon}{\sqrt{n}}]$, the expansion stated in Theorem 2 holds.

Now for $u \le \frac{2\epsilon}{\sqrt{n}}$ or $> 1 - \frac{2\epsilon}{\sqrt{n}}$, one argues using the tools developed above that

$$P(p_n^* \le u) \le \text{ const.}\frac{\epsilon}{\sqrt{n}}, \quad \text{if } u \le \frac{2\epsilon}{\sqrt{n}},$$

and

$$P(p_n^* \ge u) \le \text{ const.}\frac{\epsilon}{\sqrt{n}}, \quad \text{if } \ge 1 - \frac{2\epsilon}{\sqrt{n}}.$$

Combining these bounds and the fact that $\epsilon > 0$ is arbitrary, the theorem follows.

Utilizing the expansion provided by Theorem 2, we now develop a correction formula for p_n^*, based entirely on resampling, such that the corrected version converges in law to the uniform distribution more rapidly.

Theorem 2.3. *Assume that*

$$P(|p_{\hat{F}}(z_{p_n^*}) - p_F(z_{p_n^*})| > \epsilon) = o(n^{-1/2})$$

and

$$P(|q_{\hat{F}}(z_{p_n^*}) - q_F(z_{p_n^*})| > \epsilon) = o(n^{-1/2}).$$

Define

$$p_n^{**} = p_n^* - \frac{1}{\sqrt{n}}(p_{\hat{F}}(z_{p_n^*}) + q_{\hat{F}}(z_{p_n^*}))\phi(z_{p_n^*}).$$

Then,

$$\sup_{0\le t\le 1} |P(p_n^{**} \le t) - t| = o(n^{-1/2}).$$

The above theorem resembles Theorem 1 of Abramovitch and Singh (1985) and its proof is very similar; and hence it is deleted. The assumed conditions in the theorem are quite reasonable. It will almost always hold under moderate moment conditions.

Let us write

$$\frac{1}{\sqrt{n}}\{p_F(x) + q_F(x)\} = a_F x^2 + b_F.$$

The dependence of a_F and b_F on n has been suppressed in the notations. In view of the fact that

$$P_{\hat{F}}\Big(\frac{T_n^* - T_n}{V_{\hat{F}}} \le 0\Big) = P_{\hat{F}}\Big(\frac{T_n^* - T_n}{V_{\hat{F}^*}} \le 0\Big) = P_{\hat{F}}(T_n^* \le T_n),$$

it follows that

$$\eta_n = 2\Big[P_{\hat{F}}(T_n^* \le T_n) - \frac{1}{2}\Big] = b_{\hat{F}}\phi(0) = \frac{b_{\hat{F}}}{\sqrt{2\pi}}.$$

Hall (1988) has shown that, when T_n is a function of multivariate means, then

$$a_F = \frac{1}{6}\rho_F,$$

where $\rho_F =$ the third moment of the linear approximation of $\frac{\sqrt{n}(T_n-\theta)}{V_F}$.

At this point, one may be tempted to use the following estimate for a_F :

$$\hat{a}_F = \frac{1}{6}E_{\hat{F}}(\sqrt{n}(T_n^* - T_n))^3/\{E_{\hat{F}}(\sqrt{n}(T_n^* - T_n))^2\}^{3/2}.$$

The denominator above estimates V_F^3 consistently, but the numerator does not estimate the third moment of the linear approximation of $\sqrt{n}(T_n - \theta)$ up to the desired order $o(1/\sqrt{n})$. This happens because of the fact that the terms beyond the linear approximation also contribute typically a term of order $O(1/\sqrt{n})$ to the third moment as does the leading linear approximation. To illustrate this, we consider the example of $t_n = \sqrt{n}\frac{\bar{X}-\mu}{S_n}$, whose linear approximation is $\sqrt{n}\bar{Z}_n = \sqrt{n}\frac{\bar{x}-\mu}{\sigma}$. It is well known that $Et_n^3 = -2\mu_3/\sigma^3\sqrt{n}+o(1/\sqrt{n})$ whereas $E(\sqrt{n}\bar{Z}_n)^3 = \mu_3/\sigma^3\sqrt{n}$.

Fortunately, we may resort to the other popular resampling method known as the jackknife for estimating the third moment of the linear approximation of $\sqrt{n}(T_n - \theta)$. Define the jackknife pseudo values:

$$J_i = nT_n - (n-1)T_{ni},$$

where T_{ni} is computed just like T_n on the data set $\{X_1, \ldots, X_n\} - \{X_i\}$. If one can write

$$T_n - \theta = \frac{1}{n}\sum \xi_F(X_i) + o_p(1/\sqrt{n})$$

with $E\xi_F(X_i) = 0$, then typically

$$J_i - T_n = \xi_F(X_i) + r_i,$$

where $\max_i r_i \to 0$ in probability (see Babu and Singh (1983), Singh and Liu (1990)). As a result of this representation, it follows that the third moment of $\frac{1}{\sqrt{n}}\sum \xi_F(X_i) = \frac{1}{\sqrt{n}}E\xi_F^3(X_i)$ can be estimated by

$$\frac{1}{n^{3/2}}\sum(J_i - \bar{J})^3 \quad \text{or} \quad \frac{1}{n^{3/2}}\sum(J_i - T_n)^3$$

up to an error $o(1/\sqrt{n})$. It may be remarked appropriately here that the failure of the jackknife in estimating the third moment of $\sqrt{n}(T_n - \theta)$ consistently, turns out to be a blessing in this instance. Define

$$\hat{a}_F(BJ) = n^{-3/2}\frac{1}{6}\sum(J_i - \bar{J})^3/\{E_{\hat{F}}[\sqrt{n}(T_n - \theta)]^2\}^{3/2}.$$

We now have a formula for corrected p_n^* based entirely on resampling.

Formula

$$p_n^{**}(BJ) = p_n^* - \{\eta_n\sqrt{2\pi} + \hat{a}_F(BJ)z_{p_n^*}^2\}\phi(z_{p_n^*}) \quad \text{(truncated between 0 and 1)}.$$

Under appropriate regularity conditions, p_n^{**} in Theorem 2.3 can be replaced by $p_n^{**}(BJ)$.

3. P_n^* under H_1

In this section it is assumed that the true value of the parameter θ is greater than θ_0, which is the null value. This section critically depends on the large deviation bounds established in Ellis (1984). We restrict T_n to be a function of a multivariate mean. More precisely, let $T_n = f(\bar{W}_n)$ where $\bar{W}_n = (\bar{W}_{n1}, \ldots, \bar{W}_{nk})$ and

$$\bar{W}_{nj} = \frac{1}{n}\sum_{i=1}^{n} g_j(X_i).$$

Here, $g_1, g_2, \ldots, g_k$ are continuous functions from $\mathbb{R}^k$ to $\mathbb{R}$. The function f is also assumed to be continuous. The parameter θ is assumed to be $f(E\bar{W}_n)$. For technical reasons, we have to restrict our X_i's to be bounded r.v.'s. Although, this setup covers all real life r.v.'s, it leaves out even the most basic models including the normal r.v.'s.

The standard assumption of finite m.g.f. in an interval around zero, does suffice in the most basic case when T_n is a univariate mean. However, we are unable to use Ellis's result in our setup with just a condition of finite m.g.f.

Consider a test based on T_n, using its asymptotic distribution for setting up the critical region, for $H_0 : \theta = 0$ vs. $H_1 : \theta > \theta_0$. Let β_n denote the type-2 error of this test, when its type-1 error is fixed at some $\alpha : 0 < \alpha < 1$. Clearly β_n depends on α.

Theorem 3.1. *If $\theta > \theta_0$ prevails, then under the assumed conditions of this section,* a.s.

$$\lim_{n\to\infty} \frac{1}{n} \log p_n^* = \lim_{n\to\infty} \frac{1}{n} \log \beta_n = -HL(\theta).$$

Here $HL(\theta)$ is the well-known Hodges-Lehman efficacy.

Proof. The claim of the theorem is based on the fact that both $\frac{1}{n} \log P(T_n \leq \theta_n)$, where $\theta_n \to \theta_0$, and $\frac{1}{n} \log P^*(T_n^* \leq \theta_0)$ have the same limit. This follows from the fact that both $\bar{W}_n$ and $\bar{W}_n^*$ (the corresponding bootstrap mean) have the same large deviation index. The latter claim basically reduces, in view of Ellis's work (see Theorem II.2 of Ellis (1984) or Theorem II.6.1 of Ellis (1985)), to the elementary fact that

$$E_{F_n}(e^{t\cdot Y_1}) \to E_F(e^{t\cdot X_1})$$

uniformly on a fixed compact subset of $\mathbb{R}^k$, a.s. The details of the proofs are omitted.

Remark 3.0. The modified type-2 p-values proposed in the previous section tend to zero under H_1 exponentially at least as fast as the unmodified type-2 p-value. This can be concluded from the following facts: $\Phi(z_{p_n^*}) = p_n^*$, $(z_{p_n^*})^{-1}\phi(z_{p_n^*}) \sim -p_n^*$; thus $\phi(z_{p_n^*}) \sim -p_n^* z_{p_n^*}$. Furthermore, $z_{p_n^*} = O(\sqrt{n})$ and the multipliers of $\phi(z_{p_n^*})$ in the corrections are $O(\sqrt{n})$. In fact, $p_n^{**} = O(np_n^*)$ if $a_F \geq 0$. If $a_F < 0$, then $p_n^{**} = 0$ under H_1 for all large n a.s.

If one carries out a test at α level based on p_n^* and the limiting null distribution is used for setting up the critical value, then the test would be

$$\{\text{Reject } H_0 \text{ iff } p_n^* \leq \alpha\}.$$

How does the type-2 error β_n^* of this test go to zero under H_1? It turns out that the index of exponential rate of decay of β_n^* is the same as that of β_n which is equal to the type-2 error of a test based on T_n itself whose type-1 error is fixed at α. This property is established in the next (and the last) theorem.

Theorem 3.2. *Assume that*

$$\text{(C)} \qquad M_n = E_{\hat{F}}(T_n^* - T_n)^2 \leq \epsilon_n,$$

where ϵ_n is a sequence of positive numbers $\to 0$. Then for any $\alpha \in (0, .5]$ and $F \in H_1$,

$$\frac{1}{n} \log P_F(p_n^* \geq \alpha) \sim \frac{1}{n} \log \beta_n.$$

Proof. Let us fix a positive K such that $\frac{1}{K^2} < \alpha$. Clearly $\beta_n^* = P(p_n^* \geq \alpha)$ under $F \in H_1$. If α is fixed at some nonzero level, then the corresponding β_n is of the form $P(T_n \leq \theta_0 + \delta_n)$ for some $\delta_n \to 0$, under $F \in H_1$. Since $\alpha \leq 1 - \alpha$, we have

$$\{p_n^* \geq 1 - \alpha\} \subseteq \{p_n^* \geq \alpha\}.$$

The result is deduced from the above and the following two set inequalities.

$$\{p_n^* \geq \alpha\} \subseteq \{T_n \leq \theta_0 + K\sqrt{\epsilon_n}\} \tag{3.1}$$

and

$$\{T_n \leq \theta_0 - K\sqrt{\epsilon_n}\} \subseteq \{p_n^* \geq 1 - \alpha\}. \tag{3.2}$$

To prove (3.1), we note that if $T_n > \theta_0$,

$$\begin{aligned} p_n^* &= P_{\hat{F}}(T_n^* \leq \theta_0) = P_{\hat{F}}(T_n^* - T_n \leq \theta_0 - T_n) \\ &\leq E_{\hat{F}}(T^* - T_n)^2/(T_n - \theta_0)^2 \leq \epsilon_n (T_n - \theta_0)^{-2} \end{aligned}$$

using Markov inequality. Thus if $T_n - \theta_0 > K\sqrt{\epsilon_n}$, then

$$p_n^* < \epsilon_n / K^2 \epsilon_n = 1/K^2 < \alpha.$$

This proves (3.1). Similar arguments prove (3.2).

The condition C in Theorem 3.2 does seem to be excessive, but it will normally hold when the underlying r.v.'s are bounded. Condition C can be relaxed to the following condition which will hold in many unbounded cases: For every $\lambda > 0$, there exists $\delta_n \to 0$ such that

$$P(E_{\hat{F}}(T_n^* - T_n)^2 \geq \delta_n) \leq \text{ const. } \exp(-\lambda n).$$

The proof given above for Theorem 3.2 does critically use $\alpha \leq 1/2$. Note that the last set inequality holds without condition C. Thus the slope of $P(p_n^* > \alpha)$ is at least as big as that of β_n.

We conclude this article with some remarks.

Remark 3.1. One implication of Theorem 3.2 is that if a test based on T_n is optimal in the Hodges-Lehman sense, then so is the test based on p_n^*. Note that p_n^* does not require any parametric model.

Remark 3.2. In order to achieve the actual slope of p_n^* using a simulation based bootstrap, one would require the number of replications to go to infinity at an exponential rate. This suggests that one should use a method like the empirical saddle point approximation for evaluating p_n^*. From the practical viewpoint, though, a number of replications like 5,000 should be quite adequate. Under H_1, $Np_{n,s}^*$ should behave like a Poisson r.v. (N being large and p_n^* being small) with mean equal to NP_n^* where N equals the number of replications and $p_{n,s}^*$ is the simulation based approximation of p_n^*.

Remark 3.3. Here is an alternative way of defining a bootstrap based p-value which is second order correct under H_0, in the sense that its distribution converges to $U[0,1]$ at the rate $O(1/n)$:

$$p_n^*(A) = P_{\hat{F}}\Big(\frac{T_n^* - T_n}{V_{\hat{F}^*}} \geq \frac{T_n - \theta_0}{V_{\hat{F}}}\Big),$$

where $V_{\hat{F}}$ and $V_{\hat{F}^*}$ are the Studentizing statistics of T_n and T_n^*. One clear disadvantage of this definition is that one needs to know the analytic form of V_F. Also, under H_1, $p_n^*(A)$ does not seem to have any nice interpretation. How do $p_n^*(A)$ and p_n^{**} of Section 2 compare under H_1? We have the answer to this question in a very special case, i.e. $T_n = \bar{X}_n$, F is $N(\mu, \sigma^2)$ and $\theta = \mu$. In this situation, p_n^{**} (or $p_n^{**}(BJ)$) is superior in the sense of bigger slope. The slope of $p_n^{**} = \frac{1}{2}\frac{\mu^2}{\sigma^2}$ (like the z-test) and that of $p_n^*(A) = \frac{1}{2}\log(1 + \frac{\mu^2}{\sigma^2})$ (like the t-test).

Acknowledgement

Referees' suggestions helped improve the presentation. This research was supported by NSF Grant DMS-90-04658.

References

Abramovitch, L. and Singh, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *Ann. Statist.* **13**, 116-132.

Babu, G. J. and Singh, K. (1983). Inference on means using the bootstrap. *Ann. Statist.* **11**, 999-1003.

Babu, G. J. and Singh, K. (1984). On one term Edgeworth correction by Efron's bootstrap. *Sankhyā Ser. A* **46**, 219-232.

Bahadur, R. R. (1972). *Limit Theorems in Statistics.* SIAM, Philadelphia.

Beran, R. (1987). Prepivoting to reduce error of confidence sets. *Biometrika* **74**, 457-468.

Efron, B. (1987). Better bootstrap confidence intervals (with Discussion). *J. Amer. Statist. Assoc.* **82**, 171-200.

Ellis, R. S. (1984). Large deviations for a general class of random vectors. *Ann. Probab.* **12**, 1-12.

Ellis, R. S. (1985). *Entropy, Large Deviations, and Statistical Mechanics.* Springer-Verlag, New York.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications.* John Wiley, New York.

Hall, P. (1986). On the bootstrap and confidence intervals. *Ann. Statist.* **14**, 1431-1452.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.* **16**, 927-953.

Hodges, J. L. and Lehmann, E. L. (1956). The efficiency of some nonparametric competitors of the t-test. *Ann. Math. Statist.* **27**, 324-335.

Liu, R. Y. and Singh, K. (1993). Generalized p-values: Using data depth and the bootstrap to define concepts of p-values. Submitted for publication.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications.* John Wiley, New York.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9**, 1187-1195.

Singh, K. and Liu, R. Y. (1990). On the validity of the jack-knife procedure. *Scand. J. Statist.* **17**, 11-21.

Department of Statistics, Rutgers University, Busch Campus, New Brunswick, NJ 08903, U.S.A.

(Received July 1992; accepted February 1994)