ON NONPARAMETRIC TESTING, THE UNIFORM BEHAVIOR OF THE \( t \)-TEST, AND RELATED PROBLEMS

by

Joseph P. Romano

Technical Report No. 2003-10
May 2003

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065
ON NONPARAMETRIC TESTING, THE UNIFORM BEHAVIOR OF THE $t$-TEST, AND RELATED PROBLEMS

by

Joseph P. Romano
Department of Statistics
Stanford University

Technical Report No. 2003-10
May 2003

This research was supported in part by National Science Foundation grant DMS 010392

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065

http://www-stat.stanford.edu
On Nonparametric Testing, the Uniform Behavior of the $t$-test, and Related Problems

Joseph P. Romano  
Department of Statistics  
Stanford University  
Stanford, CA 94305  
romano@stat.stanford.edu

March 20, 2003

Abstract

In this article, we discuss some nonparametric hypothesis testing problems with special emphasis on inference for the mean. First, we recall the classical result of Bahadur and Savage (1956) that delineates the impossibility of constructing useful tests of the value of a mean. Using their ideas, we present some simple results showing how their nonexistence results extend to other testing problems, and we answer a conjecture of theirs. Other examples considered are testing whether or not the mean is rational, testing goodness of fit, and equivalence testing. Next, we discuss the uniform behavior of the classical $t$-test. For most nonparametric models, the size of the $t$-test is one for every sample size. Even if we restrict attention to the family of symmetric distributions supported on a fixed compact set, the $t$-test is not uniformly asymptotically level $\alpha$. However, the convergence of the rejection probability is established uniformly over a large family with a very weak uniform integrability type of condition. Furthermore, under such a restriction, the $t$-test is seen to possess an asymptotic maximin optimality property. We also discuss some other methods that generate hypothesis tests of valid size in finite samples. Some of the results are classical in nature, but they do not appear to be accessible in the literature.

KEY WORDS: Asymptotically Maximin, Confidence Intervals, Hypothesis Tests, Large sample theory.
1 Introduction

Suppose $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) according to a cumulative distribution function (c.d.f) $F$ on the line. Let $\mu(F)$ be the mean of $F$, assuming it exists. The main problems we discuss in this article relate to inference about $\mu(F)$ in a nonparametric setting. In Section 2, we revisit the result of Bahadur and Savage (1956), who proved that it is impossible to construct valid tests of the value of $\mu(F)$ if the model of possible distributions $F$ is suitably large. The nonexistence of reasonable inference procedures in some nonparametric problems (where, roughly speaking, the functional of interest is not continuous in a certain sense) has been studied by several authors. In the context of confidence interval construction, such results are provided by Gleser and Hwang (1987), Donoho (1988) and Pfanzagl (1998). The contribution here is framed in terms of hypothesis tests. By extending the idea used in Bahadur and Savage (1956), we provide a constructive condition that is easy to apply to other hypothesis testing problems (or confidence sets). For example, the problem of testing the existence of the mean is treated, thus solving a conjecture of Bahadur and Savage. Some other examples are considered, such as testing whether or not a mean is rational, as well as testing goodness of fit or equivalence testing.

Next, we discuss methods of constructing tests for the value of $\mu(F)$. In Section 3, the behavior of the classical $t$-test is studied in some detail. With no restrictions, the size of the $t$-test is one (because of the Bahadur Savage result). We show it is one even if we restrict attention to distributions supported on a compact set. The size is bounded away from 1 for the family of symmetric distributions, but still the convergence to the level is not uniform. Next, by imposing a weak condition on the family, we obtain the uniform convergence of the rejection probability under $F$ to the nominal level, where uniform is with respect to distributions $F$ in a large nonparametric family. Our restriction is a uniform integrability condition. We also show that the $t$-test possesses a local asymptotic maximin property over a fairly large collection of distributions. As remarked by Efron (1969) some 34 years ago, “it is a matter of some hubris to claim originality for any topic bearing on the $t$-test.” Indeed, the results are classical in nature, but do not appear to be accessible in the literature.

In Section 4, other procedures are discussed as well, with particular emphasis on those with exact size $\alpha$ in finite samples. In particular, we recall a construction of Anderson-(1967) and one of Romano and Wolf (2000).

It is clear that we must distinguish the exact size of a test from its approximate or asymptotic size. Typically, asymptotic constructions merely assert that the rejection probability of a test tends to the nominal level $\alpha$ under any fixed distribution $P$ satisfying the null hypothesis. Of course, such a result guarantees nothing about the exact finite sample size of the test. A stronger condition requires the convergence to the nominal level is somehow at least uniform in.
To distinguish these concepts, we present some definitions. Suppose that data $X^{(n)}$ comes from a model indexed by a parameter $\theta \in \Omega$. Typically, $X^{(n)}$ refers to an i.i.d. sample of $n$ observations, and an asymptotic approach assumes that $n \to \infty$. Nothing is assumed about the family $\Omega$, so that the problem may be parametric or nonparametric. Consider testing a null hypothesis $H$ that $\theta \in \Omega_H$ versus the alternative hypothesis $K$ that $\theta \in \Omega_K$, where $\Omega_H$ and $\Omega_K$ are two mutually exclusive subsets of $\Omega$. We will be studying sequences of tests $\phi_n(X^{(n)})$.

**Definition 1.1** For a given level $\alpha$, a sequence of tests $\{\phi_n\}$ is pointwise asymptotically level $\alpha$ if, for any $\theta \in \Omega_H$,

$$\limsup_{n \to \infty} E_{\theta}[\phi_n(X^{(n)})] \leq \alpha .$$  \hspace{1cm} (1)

This condition (1) guarantees that for any $\theta \in \Omega_H$ and any $\epsilon > 0$, the level of the test will be less than or equal to $\alpha + \epsilon$ when $n$ is sufficiently large. However, the condition does not guarantee the existence of an $n_0$ (independent of $\theta$) such that

$$E_{\theta}[\phi_n(X^{(n)})] \leq \alpha + \epsilon$$

for all $\theta \in \Omega_H$ and all $n \geq n_0$. We can therefore not guarantee the behavior of the size

$$\sup_{\theta \in \Omega_H} E_{\theta}[\phi_n(X^{(n)})]$$

of the test, no matter how large $n$ is. In order to guarantee the behavior of the limiting size of a test sequence, we require the following stronger condition.

**Definition 1.2** The sequence $\{\phi_n\}$ is uniformly asymptotically level $\alpha$ if

$$\limsup_{n \to \infty} \sup_{\theta \in \Omega_H} E_{\theta}[\phi_n(X^{(n)})] \leq \alpha .$$  \hspace{1cm} (2)

If instead of (2), the sequence $\{\phi_n\}$ satisfies

$$\lim_{n \to \infty} \sup_{\theta \in \Omega_H} E_{\theta}[\phi_n(X^{(n)})] = \alpha ,$$  \hspace{1cm} (3)

then this value of $\alpha$ is called the limiting size of $\{\phi_n\}$.

The distinction between the above two definitions is important and has been largely ignored in the literature; see Romano (1989) Loh and Lehmann (1990). Of course, we also will study the behavior of tests under the alternative hypothesis. We say the sequence $\{\phi_n\}$ is pointwise consistent in power if, for any $\theta$ in $\Omega_K$,

$$E_{\theta}[\phi_n(X^{(n)})] \to 1$$ \hspace{1cm} (4)

as $n \to \infty$. 

3
2 On The Bahadur Savage Theorem

2.1 A Constructive Approach

In 1956, Bahadur and Savage proved that, in a nonparametric setting where not much is assumed about the unknown distribution generating the data, it is impossible to construct a test about the value of the mean of a distribution that has size $\alpha$ and has power greater than $\alpha$ for even one distribution. In this section, we generalize their result slightly, by providing a constructive sufficient condition that applies to other testing problems as well. Although the idea is similar to theirs, it allows one to answer a conjecture of Bahadur and Savage concerning testing the existence of a mean.

Suppose data $X$ is observed on a sample space $S$ with probability law $P$. A model is assumed only in the sense that $P$ is known to belong to $P_1$, some family of distributions on $S$. Consider the problem of testing the null hypothesis $H_0$ that $P$ belongs to $P_0$ versus the alternative hypothesis $H_1$ that $P$ belongs to $P_1$, where $P_0$ and $P_1$ are distinct subsets of $P$ whose union is $P$.

A convenient way to discuss the nonexistence of tests with good power properties is in terms of the total variation metric, defined by

$$\tau(P, Q) = \sup_{\{\phi: |\phi| \leq 1\}} |\int \phi dQ - \int \phi dP|.$$  \hspace{1cm} (5)

The total variation metric has been effectively employed for testing purposes at least as far back as Hoeffding and Wolfowitz (1958). Consider the following condition for the testing problem.

Condition A: For every $Q \in P_1$, there exists a sequence $P_k \in P_0$ such that $\tau(P_k, Q) \to 0$.

Evidently, Condition A asserts that $P_0$ is dense in $P$ with respect to the metric $\tau$. We will also assume $P_0$ and $P_1$ satisfy the following (stronger) condition, which we will refer to as Condition B.

Condition B: For every $Q \in P_1$ and any $\epsilon > 0$, there exists a subset $A = A_\epsilon$ of $S$ satisfying $Q(A_\epsilon) \geq 1 - \epsilon$ and such that, if $X$ has distribution $Q$, then, the conditional distribution of $X$ given $X \in A_\epsilon$ is a distribution in $P_0$.

The set $A_\epsilon$ may depend on $Q$ as well, but we suppress this dependence in the notation. To interpret Condition B, it says that, if $X$ falls in $A$, its conditional distribution is a member of the null hypothesis parameter space, and hence its conditional rejection probability will be bounded by the size of the test. Since such a set $A$ is presumed to exist with probability $Q(A)$ near one, the power of the test under $Q$ will be bounded by the size of the test. Under
Condition A or B, we now prove that no test has power against Q greater than the size of the test.

**Theorem 2.1** Let \( \phi(X) \) be any test of \( P_0 \) versus \( P_1 \).

(i). If Condition A holds, then

\[
\sup_{Q \in P_1} E_Q[\phi(X)] \leq \sup_{P \in P_0} E_P[\phi(X)]. \tag{6}
\]

Hence, if \( \phi \) has size \( \alpha \), then

\[
\sup_{Q \in P_1} E_Q[\phi(X)] \leq \alpha; \tag{7}
\]

that is, the power function is bounded by \( \alpha \).

(ii). Assume Condition B holds. Then, Condition A holds and therefore (6) and (7) hold as well.

**Proof.** To prove (i), fix \( Q \in P_1 \) and let \( \phi \) be any test function. Take any \( \epsilon_k \to 0 \) and let \( P_k \in P_0 \) satisfy \( \tau(Q, P_k) \leq \epsilon_k \). Then,

\[
E_Q(\phi) \leq E_{P_k}(\phi) + \tau(Q, P_k) \leq \sup_{P \in P_0} E_P(\phi) + \epsilon_k.
\]

Let \( \epsilon_k \to 0 \) and the result follows.

To prove (ii), let \( \epsilon_k \to 0 \), let \( A_{\epsilon_k} \) be the subset in Condition B, and let \( P_k \) denote the distribution of \( X \) given \( X \in A_{\epsilon_k} \) when \( X \) has distribution \( Q \). Then, for any \( \phi \),

\[
E_Q[\phi(X)] = E_Q[\phi(X)|X \in A_{\epsilon_k}]Q(A_{\epsilon_k}) + E[\phi(X)|X \in A_{\epsilon_k}^c]Q(A_{\epsilon_k}^c)
\]

\[
\leq E_{P_k}[\phi(X)]Q(A_{\epsilon_k}) + Q(A_{\epsilon_k}^c) \leq E_{P_k}[\phi(X)] + \epsilon_k.
\]

Similarly, for any \( \phi \),

\[
E_Q[\phi(X)] \geq E_{P_k}[\phi(X)](1 - \epsilon_k).
\]

Hence, \( \tau(Q, P_k) \leq \epsilon_k \). Now let \( \epsilon_k \to 0 \).

**Remark 2.1** The total variation metric \( \tau \) is well-known and has many statistical uses; see Le Cam (1986), Chapter 4. Donoho (1988) effectively employed its use in proving the impossibility of constructing useful two-sided confidence intervals for certain functionals of a density. His examples can be also be recast in a hypothesis testing context. In the work here, notice the hypothesis testing framework does not have to be cast in terms of testing a particular parameter as \( P_0 \) and \( P_1 \) are quite general. However, the main point is that Condition B, though stronger than Condition A, is easily verified is some novel examples.
Remark 2.2 When \( X = (X_1, \ldots, X_n) \) is a vector of \( n \) i.i.d. random variables, then it suffices to verify Condition B for \( n = 1 \). To produce the set \( A_\varepsilon \) for \( X \), simply take \( n \)-fold product set \( A_\delta \) obtained from the case \( n = 1 \), where \( \delta \) is taken small enough to guarantee with probability \( 1 - \varepsilon \) all \( n \) observations fall in \( A_\delta \). But, the chance that all observations fall in \( A_\delta \) is at least \( (1 - \delta)^n \). Thus, choose \( \delta \) no bigger than \( 1 - (1 - \varepsilon)^{1/n} \).

Example 2.1 (Finite versus Not Finite Mean) Let \( X \) be \( X_1, \ldots, X_n \), \( n \) i.i.d. observations on the real line. As remarked by Bahadur and Savage (1956), "it would be interesting to know whether, in comparable nonparametric situations, tests of the existence of \( \mu \) are equally unsuccessful"; here, \( \mu \) refers to the mean of an observation. So, let \( P_0 \) be the family of distributions on the real line with a finite mean, and let \( P_1 \) be the distributions without a finite mean. Condition B readily holds. To see why, suppose \( Q \) is a distribution without a mean. Given \( \varepsilon \), let \( A \) be any bounded subset of the real line with probability at least \( (1 - \varepsilon) \) under \( Q \). Moreover, the conditional distribution of an observation given that it falls in \( A \) is some distribution on a bounded set, i.e., a distribution in the null hypothesis parameter space. Hence, the conclusion of Theorem 2.1 holds, and so it is impossible to construct a test with power greater than the size of the test.

As an alternative to Condition B, we now consider the following condition, Condition C.

Condition C: For any \( Q \in P_1 \) and any \( \varepsilon > 0 \), there exists a random variable \( Y \) (on some probability space) with distribution \( P \in P_0 \) and a subset \( A = A_\varepsilon \) with \( P(A) \geq 1 - \varepsilon \) such that the conditional distribution of \( Y \) given \( A_\varepsilon \) is \( Q \).

Condition C also implies a nonexistence result.

Theorem 2.2 Assume Condition C instead of Condition B in Theorem 2.1(ii). Then, Condition A holds and therefore so do (6) and (7).

Proof. Fix \( Q \in P_1 \) and let \( \phi \) be any test function. Let \( \varepsilon_k \to 0 \) and let \( P_k \) be a distribution in \( P_0 \) satisfying Condition C when \( \varepsilon = \varepsilon_k \). Then,

\[
E_{P_k}[\phi(Y)] = E_{P_k}[\phi(Y)|A_{\varepsilon_k}]P_k(A_{\varepsilon_k}) + E_{P_k}[\phi(Y)|A_{\varepsilon_k}^c]P_k(A_{\varepsilon_k}^c) = E_Q[\phi(Y)]P(A_{\varepsilon_k}) + E_{P_k}[\phi(Y)|A_{\varepsilon_k}^c]P_k(A_{\varepsilon_k}^c) \geq E_Q[\phi(X)](1 - \varepsilon_k).
\]

Similarly,

\[
E_{P_k}[\phi(Y)] \leq E_Q[\phi(Y)] + \varepsilon_k.
\]

Therefore, \( \tau(Q, P_k) \leq \varepsilon_k \to 0 \).

Remark 2.3 When \( X_1, \ldots, X_n \) are i.i.d., it suffices to verify Condition C for the case \( n = 1 \).
Example 2.2 (The Bahadur Savage Result) We now show how the classical Bahadur Savage result for testing the value of a mean can be derived by Theorem 2.2. (Note that Bahadur and Savage also consider the sequential case.) It is assumed that $X_1, \ldots, X_n$ are i.i.d. real-valued random variables with unknown c.d.f. $F$. It is now natural to index the underlying model by $F$. It is assumed $F \in \mathbf{F}$, where $\mathbf{F}$ is a large nonparametric class of distributions. Several choices for $\mathbf{F}$ will be considered. Let $\mu(F)$ denote the mean of $F$ and $\sigma^2(F)$ the variance of $F$. The goal is to test the null hypothesis $\mu(F) = 0$ versus $\mu(F) > 0$, or perhaps the two-sided alternative $\mu(F) \neq 0$. Assume that the underlying family of distributions $\mathbf{F}$ is sufficiently large and satisfies properties (1)-(3) below, but otherwise is left unspecified so that the result applies to various families.

Theorem 2.3 (Bahadur and Savage, 1956) Let $\mathbf{F}$ be a family of distributions on $\mathbb{R}$ satisfying:

1. For every $F \in \mathbf{F}$, $\mu(F)$ exists and is finite.

2. For every real $m$, there is an $F \in \mathbf{F}$ with $\mu(F) = m$.

3. The family $\mathbf{F}$ is convex in the sense that, if $F_1 \in \mathbf{F}$ and $\gamma \in [0, 1]$, then $\gamma F_1 + (1 - \gamma) F_2 \in \mathbf{F}$.

Let $X_1, \ldots, X_n$ be i.i.d. $F \in \mathbf{F}$ and let $\phi_n = \phi_n(X_1, \ldots, X_n)$ be any test function. Let $\mathcal{G}_a$ denote the set of distributions $F \in \mathbf{F}$ with $\mu(F) = a$. Then, $\sup_{F \in \mathcal{G}_a} E_F(\phi_n)$ is independent of $a$.

Proof. We will show

$$\sup_{F \in \mathcal{G}_a} E_F(\phi) \leq \sup_{F \in \mathcal{G}_b} E_F(\phi),$$

and the result will follow by interchanging the roles of $a$ and $b$. Apply Theorem 2.2 in the case $n = 1$ with $\mathbf{P}_0$ the set of distributions with mean $b$. Let $U$ denote a random variable with distribution $Q$ having mean $a$, and let $V$ denote a random variable with distribution $R$ having mean $r$ satisfying

$$(1 - \epsilon)a + \epsilon r = b.$$ 

Toss a coin with probability $1 - \epsilon$; let $Y = U$ if it lands heads and let $Y = V$ otherwise. Unconditionally, $Y$ has a distribution with mean $b$, but the conditional distribution of $Y$ given the event the toss is a head is $Q$, as required. ■

In the theorem, $\mathbf{F}$ can be all distributions on the real line with a first moment, all distributions having infinitely many moments, or all distributions with at least two finite moments, for example. It also includes the family of distributions with compact support. However, the family of all distributions on a fixed compact set is excluded because it does not satisfy Property 2. The family of all symmetric distributions as well as the subfamily of symmetric distributions
with infinitely many moments do not satisfy the conditions of the theorem because Property 3 is violated.

**Example 2.3 (Rational versus Irrational Mean)** Again, suppose $X_1, \ldots, X_n$ are i.i.d. real-valued observations with distribution $P$. Assume $P$ is the family of distributions with a finite mean $\mu(P)$. Consider testing $\mu(P)$ is rational versus $\mu(P)$ is irrational. Then, Condition C applies, and it suffices to verify it for the case $n = 1$. Let $Q$ be a distribution with irrational mean $m$. Fix $\epsilon > 0$ and let $r$ be any number such that $(1 - \epsilon)m + \epsilon r$ is rational. Suppose $U$ has distribution $Q$ and $V$ has distribution $R$, where $R$ is any distribution with mean $r$. Toss a coin with probability of heads $1 - \epsilon$, and let $Y = U$ if the toss is a head and $Y = V$ otherwise. Then, unconditionally, $Y$ has a distribution with a rational mean, but conditional on the event $A_\epsilon$ that the toss is a head, $Y$ has distribution $Q$. Thus, the conclusion of Theorem 2.2 holds. A similar argument applies to the reverse hypotheses. In spite of this conclusion, it has been argued (based on different criteria) that there is some basis to distinguish the hypotheses; see Cover (1973), Kulkarni and Zeitouni (1991), and Dembo and Peres (1994).

Note that the above argument follows by the Bahadur and Savage result if we assume $P$ consists of all distributions on the real line with finite mean. However, the above argument using Theorem 2.2 applies if the family is restricted so that the mean lies in a bounded or unbounded interval.

More generally, one can test whether $\mu(P)$ lies in an arbitrary (measurable) set $E$ versus $E^c$. The Bahadur Savage result applies to this situation, which points out that there is nothing particularly special about the case where $E$ is the rational numbers. The conclusion is that it is impossible to find an effective for this situation without a further restriction on $P$.

**Example 2.4 (Multivariate Mean)** The Bahadur and Savage result Theorem 2.3 holds in the multivariate case as well. The theorem reads exactly, except that $F$ refers to a family of distributions on $\mathbb{R}^k$ satisfying 1-3 with $m$ a vector. The proof is identical as well, except $a, b, \text{ and } r$ are vectors.

**Example 2.5 (Not Finite Mean versus Finite Mean)** Suppose $X_1, \ldots, X_n$ are i.i.d. on the real line. Reverse the roles of the hypotheses in Example 2.1 and consider testing the null hypothesis that the mean is infinite or doesn’t exist versus that it is finite. To see that Condition C holds, fix a distribution $Q$ with a finite mean, say $\mu$. Suppose $V$ denotes a random variable with distribution $Q$, and $W$ has distribution $R$, where $R$ is any distribution without a finite mean. Toss a coin with probability of heads $1 - \epsilon$. Let $Y = V$ if the toss lands heads, and let $Y = W$ otherwise. Then, the set $A_\epsilon$ in Condition C is the event that the toss lands heads and so the conditional distribution of $Y$ given $A_\epsilon$ is that of $Z$, which is $Q$, as required.
Example 2.6 (Nonnegative Distributions) Consider the family of all distributions on \([0, \infty)\) having a finite mean (but even this assumed finiteness can be removed). The problem is to test \(\mu(P) \geq \mu_0\) versus \(\mu < \mu_0\), so \(P_1\) is the set of distributions on \([0, \infty)\) with a mean that is less than some fixed \(\mu_0\). Then, Condition C is satisfied. Indeed, fix \(\epsilon > 0\) and suppose \(U\) has distribution \(Q\) with mean \(\mu(Q) = a < \mu_0\). Toss a coin with probability \(1 - \epsilon\) landing heads. Let \(V\) denote a random variable having mean \(r\) satisfying \((1 - \epsilon)a + \epsilon r = \mu_0\). Define \(Y\) to be \(U\) if the coin is heads; otherwise, \(Y = V\). Then, the unconditional distribution of \(Y\) is in \(P_0\), but the conditional distribution of \(Y\) given heads is \(Q\).

If we interchange the roles of \(P_0\) and \(P_1\), then the argument fails. In fact, there are effective tests of \(\mu(P) \leq \mu_0\) versus \(\mu(P) > \mu_0\). Indeed, Markov’s inequality yields

\[
P(\bar{X}_n \geq t) \leq E(\bar{X}_n)/t \leq \mu_0/t.
\]

Hence, the test that rejects when \(\bar{X}_n \geq \mu_0/\alpha\) is level \(\alpha\) and there are distributions for which the power of this test exceeds \(\alpha\). In terms of confidence intervals, one can state there are effective lower bounds for the mean of a nonnegative random variable, but there are not effective upper bounds. For other examples where only one-sided inference is possible, see Donoho (1988).

Example 2.7 (Goodness of Fit Testing) The usual approach to testing goodness of fit runs as follows. Assume \(X_1, \ldots, X_n\) are i.i.d. \(S\)-valued random variables with distribution \(P\). The null hypothesis asserts \(P\) belongs to some class \(\{P_\theta, \theta \in \Omega\}\) and the alternative hypothesis asserts \(P \notin \tilde{P}\), where \(\tilde{P}\) is the family of all other distributions on \(S\). Of course, there are many omnibus type tests such as a Kolmogorov-Smirnov type test that control the size of the test and are consistent in power against any alternative. The point is that one can demonstrate the data are inconsistent with the parametric model \(\{P_\theta, \theta \in \Omega\}\), but acceptance of the null hypothesis does not prove the model holds. Such a test may be more aptly called a lack of fit test (as argued by Wellek (2003)). Instead, consider the problem where the hypotheses are reversed.

For example, consider the quintessential problem of testing uniformity on \(S = (0, 1)\). We show the (fairly obvious) result that it is impossible to test the null hypothesis that \(P\) is not uniform on \((0, 1)\) versus the alternative that \(P\) is uniform on \((0, 1)\), at least not with any degree of power.

More generally, consider the following general condition for testing \(P_0\) versus \(P_1\).

**Condition D.** Assume for any \(Q \in P_1\) and any \(\epsilon > 0\), there exists some distribution \(R\) such that \((1 - \epsilon)Q + \epsilon R\) is not in \(P_1\).

Condition D implies Condition C holds. To see why, flip a coin with probability \(1 - \epsilon\) of heads, and let \(A\) be the event the toss is a head. Conditional on \(A\), let \(Y\) have distribution \(Q\) and conditional on \(\bar{A}\), let \(Y\) have distribution \(R\).
Thus, under Condition D, (6) and (7) hold.

Returning to the example where \( P_1 \) consists of \( P_0 \), the uniform distribution on \((0,1)\), Condition D trivially holds. Indeed, for any other distribution \( R \) and any \( \epsilon > 0 \), \((1-\epsilon)P_0 + \epsilon R\) is not \( P_0 \).

Similar considerations apply when \( P_1 \) is a larger parametric model, such as the family of normal distributions. Of course, the family of distributions other than such a parametric model is much larger, and is infinite dimensional in a certain sense. However, one can also construct examples where the dimension is finite. For example, consider testing a trinomial distribution on three categories, labeled 1, 2 and 3. The Hardy-Weinberg equilibrium model asserts the distribution is \( P_\theta \), which assigns probabilities \((1-\theta)^2, 2\theta(1-\theta), \) and \( \theta^2 \) to the categories 1, 2, and 3, respectively. Suppose we set up the problem where the null hypothesis asserts that the unknown distribution is any other trinomial distribution. Then, Condition D holds. For example, take \( R \) to be the probability distribution which has mass \( 1/2 \) at the points 1 and 2. Then, for any \( \theta \) and any \( \epsilon > 0 \), \((1-\epsilon)P_\theta + \epsilon R\) is not in the functional form of the Hardy-Weinberg model. Thus, condition D holds, and so it is impossible to construct a test of size \( \alpha \) that has power ever bigger than \( \alpha \). In summary, one can apply a lack of fit test (such as a Chi-squared test) to show the data are consistent with the model, but one cannot definitively conclude the model holds. (However, we remark that one can argue that the model holds approximately by changing \( P_1 \) to include all distributions within an appropriate neighborhood of the parametric model; see Chapter 8 of Wellek (2003).) ■

**Example 2.8 (Equivalence Testing)** Suppose \( X_1, \ldots, X_n \) are i.i.d. real-valued observations with c.d.f. \( F \in F \). Let \( \theta(F) \) be some real-valued functional of \( F \). The problem of demonstrating equivalence (or bioequivalence) can be formulated as testing \( H_0 : |\theta(F)| \geq \epsilon \) versus \( H_1 : |\theta(F)| < \epsilon \); see Wellek (2003). In particular, if \( \theta(F) \) is the mean of \( F \), and \( F \) is the set of all distributions with finite mean (or more generally satisfying the conditions of Theorem 2.3), then (6) and (7) hold. Indeed, it is readily checked that Condition D holds.

3 On the Behavior of the t-test

3.1 The size of the t-test.

It follows from the previous section that the goal of constructing an effective test for the mean in a nonparametric setting is unattainable. For testing the mean is zero versus the mean is positive, consider the classical one-sided t-test \( \phi_n \) at level \( \alpha \). Let \( F \) satisfy the conditions of the theorem and contain the family of normal distributions with variance one. In fact, for the sake of argument, take \( F \) to be the family of all distributions having infinitely many moments.
A classical calculation shows the power of the $t$-test against $N(\mu, 1)$ is strictly greater than $\alpha$ and increases to one as $\mu \to \infty$. Hence, it follows that the size of the one-sided $t$-test is one. A similar argument holds for the two-sided $t$-test.

Of course, the poor behavior of the $t$-test applies to any test for this situation: it is impossible to construct a test of the mean that has size $\alpha$, unless the test has power that never exceeds $\alpha$ as well, at least if the model $F$ satisfies the conditions of Theorem 2.3. Evidently, the problem is due to the fact that the mean $\mu(F)$ is quite sensitive to the tails of $F$, and one sample yields little information about the tails. The Bahadur-Savage result does not apply to the family of distributions supported on a compact set. However, if we now restrict attention to distributions supported on a compact set, the size of the $t$-test is still one, as the following calculation demonstrates.

**Example 3.1 (Size of $t$-test on a compact set)** Let $X_1, \ldots, X_n$ be i.i.d. $F$. Consider the one-sided level $\alpha$ $t$-test with test function $\phi_n$, for testing $\mu(F) = 0$ versus $\mu(F) > 0$. Let $F_0$ be the set of distributions supported on $[-1, 1]$, and $G_0$ those distributions on $[-1, 1]$ with mean 0. We will show

$$\sup_{F \in G_0} E_F(\phi_n) = 1$$

for every $n$ greater than or equal to two. It suffices to show that there exists a distribution $F \in G_0$ such that the probability of rejection under $F$ is arbitrarily close to one. Fix $n > 1$ and any $c < 1$. Then, choose $p_n > 0$ so that $(1 - p_n)^n = c$. Let $F = F_{n,c}$ be the distribution that places mass $1 - p_n$ at $p_n$ and mass $p_n$ at $p_n - 1$, so that $\mu(F) = 0$. The idea is that, with probability at least $c$, a random sample of size $n$ from $F$ will be the sample having all observations equal to $p_n > 0$. The $t$-statistic will blow up for such a sample and reject the null hypothesis. In order to make the example more convincing, modify the data distribution as follows to ensure the underlying distribution is continuous and the observations are distinct. Let $X_{n,i}^* = X_{n,i} + U_{n,i}$, where $X_{n,i}$ has the distribution $F_{n,c}$ above and, independently, $U_{n,i}$ is uniform on $[\tau_n, \tau_n]$ (where $\tau_n$ will be determined later). Let $F_n^*$ be the distribution of the $X_{n,i}^*$. Then, with probability at least $c$, all the observations will satisfy $X_{n,i}^* = p_n + U_{n,i}$ and so they will be within $\tau_n$ of $p_n$. Let $S_n^*$ be the sample standard deviation of $X_{n,1}^*, \ldots, X_{n,n}^*$, and let $S_n^U$ denote the sample standard deviation of $U_{n,1}, \ldots, U_{n,n}$. Under the event $X_{n,i} = p_n$ for all $i$, $S_n^* = S_n^U$, Note that

$$(n - 1)[S_n^U]^2 \leq \sum_{i=1}^n U_{n,i}^2 \leq n\tau_n^2. \tag{8}$$

We claim that, for any $b > 0$, the $t$-statistic for the sample of $U_{n,i}$'s satisfies

$$t_n(U_{n,1}, \ldots, U_{n,n}) \equiv \frac{n^{1/2} \bar{U}_n}{S_n^U} \geq b$$

11
with probability at least \(1 - (1/3b^2)\). To see why, by Chebychev’s inequality,

\[
P\{n^{1/2} \bar{U}_n \geq -b\tau_n\} \geq P\{|n^{1/2} \bar{U}_n| \leq b\tau_n\} \geq 1 - \frac{\text{Var}(U_{n,1})}{b^2\tau_n^2} = 1 - \frac{1}{3b^2}.
\]

Thus, using the bound (8), with probability at least \(1 - (1/3b^2)\),

\[
t_n(U_{n,1}, \ldots, U_{n,n}) \geq \frac{-b\tau_n}{(n-1)^{1/2}\tau_n} \geq -b.
\]

Hence, with probability at least \((1 - p_n)(1 - \frac{1}{3b^2})\),

\[
t_n(X_{n,1}^*, \ldots, X_{n,n}^*) = \frac{n^{1/2} \bar{X}_n^*}{S_n^*} = \frac{n^{1/2}(p_n + \bar{U}_n)}{S_n^U}
\]

\[
\geq \frac{n^{1/2}p_n}{(n-1)^{1/2}\tau_n} - b = \frac{(n-1)^{1/2}p_n}{\tau_n} - b.
\]

Now, choose \(\tau_n\) small enough so that

\[
\frac{(n-1)^{1/2}p_n}{\tau_n} - b > t_{1,1-\alpha},
\]

where \(t_{n-1,1-\alpha}\) denotes the \(1 - \alpha\) quantile of the \(t\)-distribution with \(n - 1\) degrees of freedom. Then,

\[
t_n(X_{n,1}^*, \ldots, X_{n,n}^*) > t_{1,1-\alpha} > t_{n-1,1-\alpha}
\]

with probability at least \(c(1 - \frac{1}{3b^2})\). We have shown that the power of the \(t\)-test against \(F_n^*\) satisfies

\[
P_{F_n^*}\{t_n > t_{n-1,1-\alpha}\} \geq c(1 - \frac{1}{3b^2}).
\]

Since \(c\) and \(b\) are arbitrary, this probability can be made near one, and the result follows. (After the preperation of this manuscript, it came to my attention that an alternative construction is presented in Loh and Lehmann (1990).) Also note that using the bootstrap-\(t\) to determine critical values does not salvage the \(t\)-statistic; that is, the size of the bootstrap-\(t\) test is also one; see Romano (1989). ■

As an alternative to the family of distributions supported on a compact set, it is interesting to study the behavior of the \(t\)-test under the assumption of symmetry (so again the Bahadur Savage result does not hold). In this case, there are many reasonable tests of exact size \(\alpha\) in such a semiparametric model. (One can simply take any reasonable test statistic such as \(\bar{X}_n\) or the sample median and consider its randomization distribution determined by randomly changing the signs of the observations; in this way, one can obtain a test of exact size \(\alpha\) by replacing the Student \(t\) or normal critical value by its randomization counterpart.) It is, of course, interesting to study the behavior of the \(t\)-test using the \(t\) critical value and examine its worse case behavior. Some insight is provided in Efron (1969), who considers other notions of
symmetry. Edelman (1990) provides an upper bound to the rejection probability of the $t$-test, which yields the following: if $F$ denotes the family of all distributions symmetric about 0, then for $\alpha < 1/2$,

$$\limsup_n \sup_{F \in \mathcal{F}} P_F \{ t_n \geq t_{n-1,1-\alpha} \} \leq 1 - \Phi(z_{1-\alpha} - 1.5 z_{1-\alpha}^{-1})$$

where $\Phi(\cdot)$ is the standard normal distribution function. Therefore, the size of the $t$-test is bounded away from one for this family. However, we will now demonstrate that the $t$-test is not even uniformly asymptotically level $\alpha$ for the family of symmetric distributions.

**Example 3.2 (Uniform behavior of $t$-test under symmetry)** We will exhibit a sequence of symmetric distributions $F_n$ for which the limiting rejection probability of the $t$-test is greater than $\alpha$. Let $F_n$ be supported on $-1, 0, 1$ with probabilities $p_n/2, 1 - p_n$, and $p_n/2$, respectively. Note that $F_n$ is supported on a compact set as well. (Thus, if had provided an example where $F_n$ is some stable distribution without a variance, we’d more than likely not be satisfied unless we could at least provide an example where $F_n$ had at least a finite variance; in our example, all the moments are uniformly bounded.) We will assume $np_n \to \lambda$, for an appropriate choice of $\lambda$. Suppose you observe a sample of $k$ ones and $n - k$ zeros, where $k > 1$. For such a sample,

$$t_n = \left( \frac{n - 1}{n - k} \right)^{1/2} k^{1/2} \geq k^{1/2}.$$

The probability of such a sample is

$$2^{-k} \binom{n}{k} p_n^k (1 - p_n)^{n-k} \to \exp(-\lambda)(\lambda/2)^k/k!$$

as $n \to \infty$. By summing over $k$, the limiting probability of observing such a sample for any $k \geq 1$ tends to

$$\exp(-\lambda) \sum_{k=1}^{\infty} (\lambda/2)^k/k! = \exp(-\lambda/2)[1 - \exp(-\lambda/2)].$$

By taking $\lambda = 2 \log(2)$, this limiting probability is $1/4$. Therefore, with limiting probability at least $1/4$, the $t_n \geq 1$. So, if $\alpha = 1 - \Phi(1) \approx 0.16$, the limiting rejection probability under $F_n$ is at least $1/4 > \alpha$.

One has to work a little harder to obtain a similar result for smaller values of $\alpha$. To provide another example, consider samples where you observe at least $k = 3$ ones and the remaining $n - k$ zeros (so here the $t$-statistic is at least $3^{1/2} \approx 1.732$). The probability of such a sample under $F_n$ is

$$\sum_{k=3}^{n} 2^{-k} \binom{n}{k} p_n^k (1 - p_n)^{n-k} \to \exp(-2)[1 - 5 \exp(-2)] \approx 0.04375$$

13
if $\lambda = 4$. Next, consider the possibility of observing $k$ nonzeros, and let $d$ denote the difference between the number of positive observations and the number of negative observations. For such a sample,

$$t_n \geq [(n - 1)/n]^{1/2} d/k^{1/2}.$$  

Specifically, consider the case where you observe $k = 6$ and $d = 4$ (so that $[n/(n - 1)]t_n \geq 4/6^{1/2} \approx 1.633$); the limiting probability of such a sample when $\lambda = 4$ is easily calculated to be $8 \exp(-4)/15 \approx 0.00977$. Hence, with limiting probability at least $0.05352$, $t_n \geq 1.63$ (for large enough $n$). But, $1 - \Phi(1.63) = 0.0516$. ■

### 3.2 More on the $t$-test

We have previously shown that the size of the $t$-test is one if the family of distributions $F$ satisfies the Conditions of Theorem 2.3 or if $F$ is the family of distributions supported on a compact set. In such case, the $t$-test is not even uniformly consistent in level over $F$. In particular, it is not uniformly consistent in level over the family of distributions with two finite moments. However, we will now reconsider the $t$-test and show that it is uniformly consistent over certain large subfamilies of distributions with two finite moments. For this purpose, consider a family of distributions $\tilde{F}$ on the real line satisfying

$$\lim_{\lambda \to \infty} \sup_{F \in \tilde{F}} \mathbb{E}_F \left[ \frac{|X - \mu(F)|^2}{\sigma^2(F)} I \left\{ \frac{|X - \mu(F)|}{\sigma(F)} > \lambda \right\} \right] = 0. \tag{9}$$

For example, for any $\epsilon > 0$ and $b > 0$, let $F_b^{2+\epsilon}$ be the set of distributions satisfying

$$\mathbb{E}_F \left[ \frac{|X - \mu(F)|^{2+\epsilon}}{\sigma^{2+\epsilon}(F)} \right] \leq b.$$

Then, $\tilde{F} = F_b^{2+\epsilon}$ satisfies (9). To see why, take expectations of both sides of the inequality

$$\lambda^2 Y^2 I\{|Y| > \lambda\} \leq |Y|^{2+\epsilon}.$$

**Lemma 3.1** Suppose $X_{n,1}, \ldots, X_{n,n}$ are i.i.d. $F_n$ with $F_n \in \tilde{F}$. Assume $\tilde{F}$ satisfies (9). Let $\tilde{X}_n = \sum_{i=1}^{n} X_{n,i}/n$. Then, under $F_n$,

$$n^{1/2} \left[ \frac{\tilde{X}_n - \mu(F_n)}{\sigma(F_n)} \right] \xrightarrow{d} N(0,1).$$

**Proof.** Let $Y_{n,i} = (X_{n,i} - \mu(F_n))/\sigma(F_n)$. We verify the Lindeberg Condition, which in the case of $n$ i.i.d. variables reduces to showing

$$\lim_{n} \sup_{n} \mathbb{E}[Y_{n,i}^2 I\{|Y_{n,i}| > \epsilon n^{1/2}\}] = 0$$  

14
for every $\epsilon > 0$. But, for every $\lambda > 0$,

$$\limsup_{n \to \infty} E_n[|Y_{n,i}|^2 I\{|Y_{n,i}| > \epsilon n^{1/2}\}] \leq E[|Y_{n,i}|^2 I\{|Y_{n,i}| > \lambda\}] .$$

Let $\lambda \to \infty$ and the right side tends to zero. ■

Next, we need an appropriate law of large numbers for triangular arrays. The following is based on Feller (1971), Section VII.7.

**Lemma 3.2** Let $Y_{n,1}, \ldots, Y_{n,n}$ be i.i.d. with c.d.f. $G_n$ and finite mean $\mu(G_n)$ satisfying

$$\lim_{\beta \to \infty} \limsup_{n \to \infty} E[G_n \{ |Y_{n,i} - \mu(G_n)| > \beta \}] = 0 .$$

Let $\bar{Y}_n = \sum_{i=1}^n Y_{n,i}/n$. Then, under $G_n$, $\bar{Y}_n - \mu(G_n) \to 0$ in probability.

**Proof.** Without loss of generality, assume $\mu(G_n) = 0$. Define

$$Z_{n,i} = Y_{n,i} I\{|Y_{n,i}| \leq n\} .$$

Let $m_n = E(Z_{n,i})$ and $\bar{Z}_n = \sum_{i=1}^n Z_{n,i}/n$. Then, the event $\{|\bar{X}_n - m_n| > \epsilon\}$ implies either $\{|\bar{Z}_n - m_n| > \epsilon\}$ occurs or $\{|\bar{Y}_n - \bar{Z}_n| > \epsilon\}$ occurs. Hence, for any $\epsilon > 0$,

$$P\{|\bar{Y}_n - m_n| > \epsilon\} \leq P\{|\bar{Z}_n - m_n| > \epsilon\} + P\{|\bar{Y}_n - \bar{Z}_n| > \epsilon\} .$$

The last term is bounded above by

$$P\bigcup_{i=1}^n \{Y_{n,i} \neq Z_{n,i}\} \leq \sum_{i=1}^n P\{Y_{n,i} \neq Z_{n,i}\} = nP\{|Y_{n,i}| > n\} .$$

The first term on the right side of (11) can be bounded by Markov's inequality, so that

$$P\{|\bar{Y}_n - m_n| > \epsilon\} \leq (n\epsilon^2)^{-1} E[Z_{n,1}^2] + nP\{|Y_{n,1}| > n\} .$$

For $t > 0$, let

$$\tau_n(t) = t[1 - G_n(t) - G_n(-t)]$$

and

$$\kappa_n(t) = \frac{1}{t} \int_{-t}^t x^2 dG_n(t) = -\tau_n(t) + \frac{2}{t} \int_0^t \tau_n(x) dx ;$$

the last equality follows by integration by parts and corrects (7.7), p.235 of Feller (1971). Hence,

$$P\{|\bar{Y}_n - m_n| > \epsilon\} \leq \epsilon^{-2} \kappa_n(n) + \tau_n(n) .$$

But, for any $t > 0$,

$$\tau_n(t) \leq E[|Y_{n,1}| I\{|Y_{n,1}| \geq t\}] ,$$

15
so \( \tau_n(n) \to 0 \) by (10). Fix any \( \epsilon > 0 \) and let \( \beta_0 \) be such that

\[
\limsup_n \text{E} [\{Y_{n,1} \mid |Y_{n,1}| > \beta_0\}] < \frac{\epsilon}{4}.
\]

Then, there is an \( n_0 \) such that, for all \( n \geq n_0 \),

\[
\text{E} [\{Y_{n,1} \mid |Y_{n,1}| > \beta_0\}] < \frac{\epsilon}{2},
\]

and so

\[
\text{E} |Y_{n,1}| \leq \beta_0 + \frac{\epsilon}{2}
\]

for all \( n \geq n_0 \) as well. Then, if \( n \geq n_0 > \beta_0 \),

\[
\frac{1}{n} \int_0^n \tau_n(x) \, dx \leq \frac{1}{n} \int_0^n \text{E} [\{Y_{n,1} \mid |Y_{n,1}| \geq x\}] \, dx
\]

\[
\leq \frac{1}{n} \int_0^{\beta_0} \text{E} |Y_{n,1}| \, dx + \frac{1}{n} \int_{\beta_0}^n \frac{\epsilon}{2} \, dx \leq \frac{\beta_0 (\beta_0 + \frac{\epsilon}{2})}{n} + \frac{\epsilon}{2},
\]

which is less than \( \epsilon \) for all sufficiently large \( n \). Thus, (14) tends to 0 as \( n \to \infty \). Therefore, \( \hat{Y}_n - m_n \to 0 \) in probability. Finally, \( m_n \to 0 \); to see why, observe

\[
0 = \text{E}(Y_{n,1}) = m_n + \text{E} [Y_{n,1} I\{|Y_{n,1}| > n\}],
\]

so that

\[
|m_n| \leq \text{E} [\{Y_{n,1} \mid |Y_{n,1}| > n\}] \to 0,
\]

by assumption (10).

**Corollary 3.1** Let \( \tilde{\mathcal{F}} \) be a family of distributions satisfying (9). Suppose \( X_{n,1}, \ldots, X_{n,n} \) are i.i.d. \( F_n \in \tilde{\mathcal{F}} \) and \( \mu(F_n) = 0 \). Then, under \( F_n \),

\[
\frac{1}{n} \sum_{i=1}^n X_{n,i}^2 \sigma^2(F_n) \to 1 \quad \text{in probability.}
\]

**Proof.** Apply Lemma (3.2) to \( Y_{n,i} = \frac{X_{n,i}^2}{\sigma^2(F_n)} - 1 \). To see that Lemma (3.2) applies, note that if \( \beta > 1 \), then the event \( \{|Y_{n,i}| > \beta\} \) implies \( \frac{X_{n,i}^2}{\sigma^2(F_n)} > \beta + 1 \) (since \( X_{n,i}^2/\sigma^2(F_n) > 0 \)) and also \( |Y_{n,i}| < \frac{X_{n,i}^2}{\sigma^2(F_n)} \). Hence, for \( \beta > 1 \),

\[
\text{E} [\{Y_{n,i} \mid |Y_{n,i}| \geq \beta\}] \leq \text{E} \left[ \frac{X_{n,i}^2}{\sigma^2(F_n)} I\{|X_{n,i}| > \sqrt{\beta + 1}\} \right].
\]

The sup over \( n \) then tends to 0 as \( \beta \to \infty \) by the assumption \( F_n \in \tilde{\mathcal{F}} \). ■

We are now in a position to study the behavior of the \( t \)-test uniformly across a fairly large class of distributions.
Theorem 3.1 Let \( F_n \in \tilde{F} \), where \( \tilde{F} \) satisfies (9). Assume \( n^{1/2} \mu(F_n)/\sigma(F_n) \to \delta \) as \( n \to \infty \) (where \( |\delta| \) is allowed to be \( \infty \)). Let \( X_1, \ldots, X_n \) be i.i.d. with c.d.f. \( F_n \), and consider the t-statistic

\[
t_n = n^{1/2} \frac{\bar{X}_n}{S_n},
\]

where \( \bar{X}_n \) is the sample mean and \( S_n^2 \) is the sample variance. If \( |\delta| < \infty \), then under \( F_n \),

\[
t_n \xrightarrow{L} N(\delta, 1).
\]

If \( \delta \to \infty \) (respectively, \( -\infty \)), then \( t_n \to \infty \) (respectively, \( -\infty \)) in probability under \( F_n \).

Proof. Write

\[
t_n = \frac{n^{1/2} [\bar{X}_n - \mu(F_n)]}{S_n} + \frac{n^{1/2} \mu(F_n)/\sigma(F_n)}{S_n/\sigma(F_n)}.
\]

The proof will follow if we show \( S_n/\sigma(F_n) \to 1 \) in probability under \( F_n \) and if

\[
\frac{n^{1/2} [\bar{X}_n - \mu(F_n)]}{\sigma(F_n)} \xrightarrow{L} N(0, 1). \tag{15}
\]

But the latter follows by Lemma 3.1. To show \( S_n^2/\sigma^2(F_n) \to 1 \) in probability, assume without loss of generality that \( \mu(F_n) = 0 \). Write

\[
\frac{S_n^2}{\sigma^2(F_n)} = \frac{n \left[ \sum_i X_i^2/n - \bar{X}_n^2 \right]}{\sigma^2(F_n)}.
\]

The above asymptotic normality and Slutsky's theorem implies \( \bar{X}_n/\sigma(F_n) \to 0 \) in probability. So, it suffices to show

\[
\frac{\sum_i X_i^2/n}{\sigma^2(F_n)} \to 1
\]

in probability under \( F_n \), but this follows from Corollary 3.1. \( \blacksquare \)

Theorem 3.1 now allows us to deduce that the t-test is uniformly consistent in level, and it also yields a limiting power calculation.

Theorem 3.2 Let \( \tilde{F} \) satisfy (9) and let \( \tilde{F}_0 \) be the set of \( F \) in \( \tilde{F} \) with \( \mu(F') = 0 \) (assumed nonempty).

(i) For testing \( \mu(F') = 0 \) versus \( \mu(F) > 0 \), the t-test that rejects when \( t_n > z_{1-\alpha} \) (or \( t_{n-1,1-\alpha} \)) is uniformly asymptotically level \( \alpha \) over \( \tilde{F}_0 \); that is,

\[
| \sup_{F_n \in \tilde{F}_0} P_F \{ t_n > z_{1-\alpha} \} - \alpha | \to 0 \tag{16}
\]

as \( n \to \infty \).
(ii) Also, the limiting power against $F_n \in \bar{F}$ with $n^{1/2} \mu(F_n)/\sigma(F_n) \to \delta$ is given by

$$\lim_n P_{F_n \{ t_n > z_{1-\alpha} \} = 1 - \Phi(z_{1-\alpha} - \delta). \quad (17)$$

Furthermore,

$$\inf_{\{F \in \bar{F}: n^{1/2} \mu(F)/\sigma(F) \geq \delta \}} P_F \{ t_n > z_{1-\alpha} \} \to 1 - \Phi(z_{1-\alpha} - \delta). \quad (18)$$

**Proof.** To prove (16), if the result failed, one could extract a subsequence $\{F_n\}$ with $F_n \in \bar{F}_0$ such that

$$P_{F_n} \{ t_n > z_{1-\alpha} \} \to \beta \neq \alpha.$$

But this contradicts Theorem 3.1 since $t_n$ is asymptotically standard normal under $F_n$. The proof of (17) follows from Theorem 3.1 as well. To prove (18), again argue by contradiction and assume there exists a subsequence $\{F_n\}$ with $n^{1/2} \mu(F_n)/\sigma(F_n) \geq \delta$ such that

$$P_{F_n} \{ t_n > z_{1-\alpha} \} \to \gamma < 1 - \Phi(z_{1-\alpha} - \delta)$$

The result follows from (17) if $n^{1/2} \mu(F_n)/\sigma(F_n)$ has a limit; otherwise, pass to any convergent subsequence and apply the same argument. □

Part (i) of the Theorem is proved under stronger assumptions in Loh and Lehmann (1990), and they attribute their result to T. Severini. In some sense, the assumptions used here are as weak as possible in the sense that, by a result of Feller (see Billingsley (1986), Theorem 27.4), one cannot remove the Lindeberg Condition in order to obtain asymptotic normality. Thus, part (i) of the above result appears to be the weakest known condition for which the $t$-test is uniformly asymptotic consistent in level. By further assumptions of smoothness and higher moments, Hall and Jing (1995) obtain error bounds for the rejection probability uniformly over large families of distributions.

Note that it is not the case that (18) holds if $\bar{F}$ is replaced by all distributions with finite second moments or finite fourth moments, or even the more restricted family of distributions supported on a compact set. In fact, there exists a sequence of distributions $\{F_n\}$ supported on a fixed compact set and satisfying $n^{1/2} \mu(F_n)/\sigma(F_n) \geq \delta$ such that the limiting power of the $t$-test against this sequence of alternatives is $\alpha$. To construct a sample $X_{n,1}, \ldots, X_{n,n}$ from such an $F_n$, first let $Y_{n,1}, \ldots, Y_{n,n}$ be i.i.d. bernoulli variables with success probability $p_n$, where $np_n = \lambda$ and $\lambda^{1/2} = \delta$. Let $U_{n,1}, \ldots, U_{n,n}$ be i.i.d. uniform variables on $(-\tau_n, \tau_n)$, where $\tau_n^2 = 3p_n^2$. Then, let $X_{n,i} = Y_{n,i} + U_i$, so that $F_n$ is the distribution of $X_{n,i}$. Note that

$$n^{1/2} \mu(F_n)/\sigma(F_n) = (np_n)^{1/2} = \delta$$

so that the sequence $F_n$ qualifies. The behavior of the $t$-test under this $F_n$ is given in the following lemma.
Lemma 3.3 For the above choice of $F_n$,
\[ t_n \xrightarrow{L} V^{1/2}, \]
where $V$ is Poisson with mean $\delta^2$. Thus,
\[ P_{F_n}(t_n > t_{n-1,1-\alpha}) \to P\{V^{1/2} > z_{1-\alpha}\}. \]

Proof. Let $L_n$ be the number of $Y_{n,i}$ equal to 1, so that $L_n$ tends in distribution to the Poisson distribution with mean $\lambda$. Then, the numerator of the $t$-statistic is
\[ n^{1/2} \bar{X}_n = n^{-1/2} L_n + n^{1/2} U_n. \]
Thus, $n \bar{X}_n \xrightarrow{L} V$, since $nU_n$ has mean 0 and variance $nVar(U_{n,1}) = 3n\tau_n^2 \to 0$. Similarly, it can be shown that that the denominator of the $t$-statistic multiplied by $n^{1/2}$ tends in distribution to $V^{1/2}$, and the result follows. 

It follows that, for $\alpha < 1/2$, the limiting power of the $t$-test against $F_n$ satisfies
\[ P\{V^{1/2} > z_{1-\alpha}\} \leq 1 - P\{V = 0\} = 1 - \exp(-\delta^2). \]
This is strictly smaller than $1 - \Phi(z_{1-\alpha} - \delta)$ if and only if
\[ \Phi(z_{1-\alpha} - \delta) > \exp(-\delta^2). \]
Certainly, for small $\delta$, this inequality holds, since the left hand side tends to $1 - \alpha$ as $\delta \to 0$ while the right hand side tends to 1.

In spite of the previous construction, the $t$-test behaves well for typical distributions, as demonstrated in Theorem 3.2. However, it is important to realize the $t$-test does not behave uniformly well across distributions with large skewness, so that the limiting normal theory fails.

3.3 Asymptotically Maximin Tests

We now would like to derive an optimality property of the $t$-test for the mean in a nonparametric setting. Theorem 3.2 implies that the power of the $t$-test is bounded away from $\alpha$ for distributions $F$ whose standardized mean $n^{-1/2}\mu(F)/\sigma(F)$ is bounded away from 0. It is then natural to measure a test sequence by its maximin power over such alternatives, with the goal of finding the test that asymptotically maximizes the minimum power over such alternatives. Consider testing $\mu(F) = 0$ against the alternatives $\mu(F)/\sigma(F) \geq \delta/n^{1/2}$, and assume $F \in \tilde{F}$, where $\tilde{F}$ satisfies (9). By Theorem 3.2, the limiting minimum power of the $t$-test is
$1 - \Phi(z_{1-\alpha} - \delta)$. We now show that this is indeed the optimal limiting maximin power in a nonparametric setting.

First, assume $F$ contains the family $N(\theta, 1)$ for $\theta \geq 0$. Then, an optimality result can be obtained as follows. Suppose $\phi_n = \phi_n(X_1, \ldots, X_n)$ is any sequence of test functions which satisfies $E_F(\phi_n) \to \alpha$ for any $F \in F$ with mean 0. Trivially,

$$\limsup_n \inf_{\{F \in F, \mu(F)/\sigma(F) \geq \delta/n^{1/2}\}} E_F(\phi_n) \leq \limsup_{n} E_{F=N(\delta/n^{1/2}, 1)}(\phi_n). \quad (19)$$

But, the right hand side is the limiting power for testing $\theta = 0$ versus $\theta = \delta/n^{1/2}$ in the normal location model $N(\theta, 1)$ (subject to the constraint $E_F(\phi_n) \to \alpha$ when $F = N(0, 1)$). But, the test that rejects when $n^{1/2} \bar{X}_n \geq z_{1-\alpha}$ is most powerful at level $\alpha_n$. So, the right side is bounded above by the limiting power of this sequence of UMP tests, which is readily seen to be $1 - \Phi(z_{1-\alpha} - \delta)$. The right hand side of (19) is then bounded above by $1 - \Phi(z_{1-\alpha} - \delta)$. Hence, the $t$-test is asymptotically maximin since its limiting minimum power attains this bound (by Theorem 3.2).

On the other hand, if the family of distributions $\tilde{F}$ does not contain the normal distributions, the above argument does not work. However, we can still obtain an optimality result for the $t$-test, as long as $\tilde{F}$ satisfies (9). To see how, we focus on the case where $\tilde{F}$ is supported on $[0, 1]$.

Let $\phi_n$ be any test sequence satisfying $E_F(\phi_n) \to \alpha$ if $F \in \tilde{F}$ and $\mu(F) = 0$. Fix any such $F$ with $\mu(F) = 0$ and $\sigma(F) > 0$. The smallest power over a large class of alternatives can always be bounded above by the smallest power over a smaller class. If the smaller class is chosen appropriately, the testing problem for the smaller model will have relevance for the larger class (the nonparametric model we would like to study). The idea of using a parametric submodel to obtain efficiency results in nonparametric models dates back to Stein (1956). So, introduce the parametric submodel with density

$$p_\theta(x) = \exp(\theta x - C(\theta)) \quad (20)$$

with respect to $F$; evidently, this is a one-parameter exponential family. Let

$$\mu_\theta = \int_{-1}^{1} xp_\theta(x) dF(x)$$

be the mean of $p_\theta$ and let $\sigma_\theta^2$ be its variance. Since $\mu(F) = 0$, $\mu_0 = 0$. In addition, $\mu_\theta = C'(\theta)$ and $\sigma_\theta^2 = C''(\theta)$, so that $C'(0) = 0$ and $C''(0) = \sigma^2(F) > 0$. Then,

$$\frac{\mu_\theta}{\sigma_\theta} = \frac{C'(\theta)}{[C''(\theta)]^{1/2}} = \frac{\theta [C''(0)] + o(\theta)}{[C''(\theta)]^{1/2}} = \theta \sigma(F) + o(\theta)$$

as $\theta \to 0$. Also, for this model, the Fisher Information is given by $I(\theta) = C''(\theta)$, so that $I(0) = \sigma^2(F)$. 

20
With $\delta$ fixed, fix any $\epsilon > 0$ and let $\theta_n$ be any fixed sequence such that $n^{1/2}\theta_n\sigma(F) \to \delta + \epsilon$. Then,
\[ n^{1/2}\mu_{\theta_n}/\sigma_{\theta_n} = n^{1/2}\theta_n/\sigma(F) + o(1) \]
as $\theta_n \to 0$. Thus, $n^{1/2}\mu_{\theta_n}/\sigma_{\theta_n} > \delta$ for all sufficiently large $n$. So, the problem of testing $\theta = 0$ versus $\theta = \theta_n$ is relevant to the nonparametric testing problem because $\theta = 0$ corresponds to a distribution in the null hypothesis parameter space while $\theta = \theta_n$ corresponds to a distribution in the alternative hypothesis parameter space sequence.

Hence, for any test sequence $\phi_n$,
\[ \limsup_n \inf_{F \in \hat{F}, \, n^{1/2}\mu(F)/\sigma(F) \geq \delta} E_F(\phi_n) \leq \limsup_n E_{\theta_n}(\phi_n). \]
The right hand side is bounded above by the optimal limiting power for testing $\theta = 0$ versus $\theta = \theta_n$. But, for the one-parameter submodel, a UMP one-sided test exists, and rejects for large values of $\hat{X}_n$. Thus, a direct calculation (or appeal to Lemma 3.1)) yields the limiting power for this test sequence against $\theta_n$ to be
\[ 1 - \Phi(z_{1-\alpha} - h\sigma(F)) = 1 - \Phi(z_{1-\alpha} - \delta - \epsilon) . \]
Hence, we have shown that
\[ \limsup_n \inf_{F \in \hat{F}, \, n^{1/2}\mu(F)/\sigma(F) \geq \delta} E_F(\phi_n) \leq 1 - \Phi(z_{1-\alpha} - \delta - \epsilon) . \]
Let $\epsilon \to 0$ to obtain
\[ \limsup_n \inf_{F \in \hat{F}, \, n^{1/2}\mu(F)/\sigma(F) \geq \delta} E_F(\phi_n) \leq 1 - \Phi(z_{1-\alpha} - \delta) . \]
But, the $t$-test attains the right hand side, and so is asymptotically maximin.

Close inspection of the above argument shows that we did not really need the assumption that $\hat{F}$ is supported on $[0, 1]$. All that was needed was that, for some $F \in \hat{F}$, the family of distributions having density (20) with respect to $F$ exists for all $\theta$ in some small interval $[0, \theta_1]$. This condition is satisfied if there exists some $F \in \hat{F}$ that has a moment generating function. Thus, we have the following.

**Theorem 3.3** Let $\hat{F}$ satisfy (9). Assume, for some $F \in \hat{F}$, the moment generating function of $F$ exists. Then, the $t$-test is asymptotically maximin in the sense that it maximizes
\[ \limsup_n \inf_{F \in \hat{F}, \, n^{1/2}\mu(F)/\sigma(F) \geq \delta} E_F(\phi_n) \]
among all test sequences $\phi_n$ which satisfy $E_F(\phi_n) \to \alpha$ for all $F \in \hat{F}$ having mean 0.
Note that we are claiming optimality of the t-test among tests that are only pointwise asymptotically level \( \alpha \) (i.e. a large class of test sequences), even though the t-test possesses the stronger property of being uniformly asymptotically level \( \alpha \).

Also, an analogous result holds for the two-sided t-test.

4 Distributions on a Compact Set

For a family \( \mathbf{F} \) satisfying the conditions of Theorem 2.3, it is impossible to construct an effective test of size \( \alpha \). We now consider the family of distributions \( \mathbf{F}_0 \) having support on \([-1, 1]\), which does not satisfy the conditions of Theorem 2.3. For this family, the t-test was shown to have size one for this family. In spite of the poor behavior of the t-test, the following construction gives a test (sequence) that has size \( \alpha \) and is pointwise consistent in power, as long as \( F \) is restricted to \( \mathbf{F}_0 \).

4.1 Anderson's Construction.

Let \( \mathbf{G}_0 \) be the set of distributions on \([-1, 1]\) having mean 0. To exhibit a test that has size \( \alpha \) for any fixed sample size \( n \) and all \( F \in \mathbf{G}_0 \) and is pointwise consistent in power, consider the following construction, due to Anderson (1967). First, recall the Kolmogorov-Smirnov confidence band \( R_{n,1-\alpha} \), constructed as follows. For c.d.f.s \( F \) and \( G \), define the sup distance

\[ d_{KS}(F, G) = \sup_{x \in \mathbb{R}} |F(x) - G(x)|. \]

Let \( \hat{F}_n \) be the empirical c.d.f., that is, the discrete distribution which places mass \( 1/n \) at each of the \( X_i \). The statistic \( d_{KS}(\hat{F}_n, F) \) was introduced in the fundamental paper of Kolmogorov (1933) who also obtained the limiting distribution of \( n^{1/2}d_{KS}(\hat{F}_n, F) \). Note that the sampling distribution of \( n^{1/2}d_{KS}(\hat{F}_n, F) \) and its limiting distribution do not depend on \( F \) as long as \( F \) is continuous. Denote by \( c_n(1-\alpha) \) the \( 1-\alpha \) quantile of the distribution of \( n^{1/2}d_{KS}(\hat{F}_n, F) \) under \( F \) when \( F \) is any continuous distribution. This leads to the following KS uniform confidence bands for \( F \) of nominal level \( 1-\alpha \).

\[ R_{n,1-\alpha} = \{ F \in \mathbf{F} : n^{1/2}d_{KS}(\hat{F}_n, F) \leq c_n(1-\alpha) \}. \]

Note that for any \( F \) (continuous or otherwise)

\[ \text{Prob}_F\{F \in R_{n,1-\alpha}\} \geq 1 - \alpha, \]

where the inequality is an equality if and only if \( F \) is continuous. This leads to a conservative confidence interval \( I_{n,1-\alpha} \) for \( \mu(F) \) as follows. Include the value \( \mu \) in \( I_{n,1-\alpha} \) if and only if there
exists some $G$ in $R_{n,1-\alpha}$ with $\mu(G) = \mu$. Then,

$$\{F \in R_{n,1-\alpha}\} \subset \{\mu(F) \in I_{n,1-\alpha}\}$$

and so

$$P_F\{\mu(F) \in I_{n,1-\alpha}\} \geq P_F\{F \in R_{n,1-\alpha}\} \geq 1 - \alpha,$$

where the last inequality follows by construction of the Kolmogorov-Smirnov confidence bands.

For testing $\mu(F) = 0$ versus $\mu(F) \neq 0$, let $\phi_n$ be the test that accepts the null hypothesis if and only if the value 0 falls in $I_{n,1-\alpha}$. By construction,

$$\sup_{F \in G_0} E_F(\phi_n) \leq \alpha.$$

We claim that

$$I_{n,1-\alpha} \subset \bar{X}_n \pm 2n^{-1/2}c_{n,1-\alpha},$$

where $c_{n,1-\alpha}$ is the $1-\alpha$ quantile of the null distribution of the Kolmogorov-Smirnov test statistic. The result (21) follows from the following lemma.

**Lemma 4.1** Suppose $F$ and $G$ are distributions on $[-1,1]$ with

$$\sup_t |F(t) - G(t)| \leq \epsilon.$$

Then, $|\mu(F) - \mu(G)| \leq 2\epsilon$.

**Proof.** By integration by parts,

$$|\mu(F) - \mu(G)| = \left| \int_{-1}^{1} x d(F - G)(x) \right|$$

$$= \left| \int_{-1}^{1} [F(x) - G(x)] dx \right| \leq 2\epsilon. \quad \blacksquare$$

The result (21) follows by applying the lemma to $F$ and the empirical cdf $\hat{F}_n$.

Let $F$ be a distribution with mean $\mu(F) \neq 0$. Suppose without loss of generality that $\mu(F) > 0$. Also, let $L_{n,1-\alpha}$ be the lower endpoint of the interval $I_{n,1-\alpha}$. Then,

$$E_F(\phi_n) \geq P_F\{L_{n,1-\alpha} > 0\} \geq P_F\{\bar{X}_n > 2n^{-1/2}c_{n,1-\alpha}\} \to 1,$$

by Slutsky’s theorem, since $\bar{X}_n \to \mu(F) > 0$ and $n^{-1/2}c_{n,1-\alpha} \to 0$. Thus, the test is pointwise consistent in power. To obtain a result stronger than pointwise consistent in power, consider any sequence of alternatives $\{F_n\}$ such that $|n^{1/2}\mu(F_n)| \to \infty$. We will now show that the limiting power against such a sequence is one. First, we obtain a crude inequality, valid for finite $n$. 

23
Lemma 4.2 Suppose $F_n \in F_0$ and

$$|n^{1/2} \mu(F_n)| \geq \delta > 2c_{n,1-\alpha}.$$ 

Then,

$$E_{F_n}(\phi_n) \geq 1 - \frac{1}{2(2c_{n,1-\alpha} - \delta)^2}.$$ 

Proof. As in the case of fixed $F$ above, and assuming $\mu(F_n) > 0$,

$$E_{F_n}(\phi_n) \geq P_{F_n}(\bar{X}_n > 2n^{-1/2}c_{n,1-\alpha})$$

$$= P_{F_n}(n^{1/2}(\bar{X}_n - \mu(F_n)) \geq 2c_{n,1-\alpha} - n^{1/2}\mu(F_n)).$$

If $n^{1/2}\mu(F_n) > 2c_{n,1-\alpha}$, then this is bounded below by

$$P_{F_n}(|n^{1/2}(\bar{X}_n - \mu(F_n))| \leq |2c_{n,1-\alpha} - \delta|);$$

now apply Chebychev’s inequality, noting $\sigma(F_n) \leq 1/2$ since $F_n \in F_0$. 

The lemma implies $E_{F_n}(\phi_n) \to 1$ whenever $|n^{1/2}\mu(F_n)| \to \infty$. Moreover, the minimum power of $\phi_n$ over $\{F : n^{1/2}\mu(F) \geq \delta\}$ is at least $1 - [2(2c_{n,1-\alpha} - \delta)^{-2}]$ if $\delta > 2c_{n,1-\alpha}$. Since $c_{n,1-\alpha}$ has a finite limit $c_{1-\alpha}$, the limiting minimum power of $\phi_n$ over $\{F : |n^{1/2}\mu(F_n)| \geq \delta\}$ as at least $1 - [2(2c_{1-\alpha} - \delta)^{-2}]$ if $\delta > 2c_{1-\alpha}$, and this exceeds $\alpha$ for large enough $\delta$.

In summary, Anderson’s procedure has size $\alpha$ for testing $\mu(F) = 0$ against $\mu(F) \neq 0$ as $F$ ranges over all distributions on $[-1,1]$, and its power is bounded away from $\alpha$ against distributions $F$ satisfying $n^{1/2}|\mu(F)| \geq \delta$ (as long as $\delta > 2c_{1-\alpha}$). On the other hand, the procedure is not efficient in the sense that it is not (locally) asymptotically maximin. To see this, let $F$ be uniform on $(-1,1)$ and consider the family of densities (20) with respect to $F$. Let $\theta_n$ satisfy $n^{1/2}\theta_n\sigma(F) \to \delta$. Let $\phi_n$ be the test that rejects if $n^{1/2}|\bar{X}_n| > 2c_{1-\alpha}$. Arguing as in (21), one can show that $\phi_n$ and $\overline{\phi}_n$ are asymptotically equivalent in the sense that

$$E_{\theta_n}[|\phi_n - \overline{\phi}_n|] \to 0$$

for any sequence $\theta_n$. But, the limiting power of $\phi_n$ is easily calculated. By a contiguity argument, under $\theta_n$,

$$n^{1/2} \bar{X}_n \overset{L}{\to} N\left(\delta\sigma(F), \sigma^2(F)\right),$$

where in this case $\sigma^2(F) = 1/3$. So,

$$\lim_{n \to \infty} E_{\theta_n}(\phi_n) = P\{|Z + \delta| > 2c_{1-\alpha}/\sigma(F)\},$$

where $Z$ is standard normal. But, for the $t$-test, the limiting power is $P\{|Z + \delta| > z_{1-\alpha}\}$. Since, $2c_{1-\alpha}/\sigma(F) > z_{1-\alpha}/2$, Anderson’s test is inefficient. (Furthermore, one can argue that, if $\alpha = 0.05$ so that $c_{0.95} = 1.36$, then about $(2 \cdot 1.36 \cdot 3^{1/2}/1.96)^2 \approx 5.78$ times as many observations are needed for $\phi_n$ to obtain the same power against $\theta_n$ as the $t$-test.)
4.2 An Efficient Construction

In fact, there exists a test that is level $\alpha$ for the family of all distributions on $[-1, 1]$, and it possesses an asymptotic maximin property over $\hat{F}_\tau$, the family of all c.d.f.s on $[-1, 1]$ with $\sigma(F) \geq \tau$ (for any small $\tau > 0$). Note that $\hat{F}_\tau$ satisfies the condition (9). Indeed, Romano and Wolf (2000) construct a $1 - \alpha$ level confidence interval, say $I_n$, for $\mu(F)$ that satisfies

$$I_n = \bar{X}_n \pm n^{-1/2} z_{1-\frac{\alpha}{2}} \sigma(F) + O_P \left( \frac{\log(n)}{n} \right),$$

and the error term is uniform over $\hat{F}_\tau$. (Their argument is for fixed $F$, but going over the argument shows the error term is uniformly small if $\sigma(\hat{F}_n)$ can be bounded from below uniformly in $F$; but this follows from Corollary 3.1.) So, let $\phi^*_n$ be the test that rejects $\mu(F) = 0$ if $0 \notin I_n$. Let $F_n \in \hat{F}_\tau$ satisfy $n^{1/2}|\mu(F_n)|/\sigma(F_n) \to \delta$. Then,

$$\lim_n E_{F_n}(\phi^*_n) - \lim_n P\{n^{1/2}|\bar{X}_n| > z_{1-\frac{\alpha}{2}} \sigma(F_n)\} = P\{|Z + \delta| > z_{1-\frac{\alpha}{2}}\},$$

by Theorem 3.1. It follows (by the above argument applied to arbitrary subsequences) that

$$\inf_{F \in \hat{F}_\tau, n^{1/2} |\mu(F)|/\sigma(F) \geq \delta} E_P(\phi^*_n) \to P\{|Z + \delta| > z_{1-\frac{\alpha}{2}}\}.$$ 

This is the optimal maximin power by a two-sided version of Theorem 3.3.

In summary, it is possible to find a test sequence whose power performs asymptotically like that of the $t$-test, whose power is best obtainable in an asymptotic maximin sense. In finite samples, the above construction behaves quite conservatively (see Romano and Wolf (2000) for simulations), and leaves open the possibility that better less conservative methods can be constructed that are still genuinely level $\alpha$. At the very least, but particularly in nonparametric problems, the distinction between pointwise and uniform convergence results point toward the need to find methods that balance finite sample validity and asymptotic efficiency.

Acknowledgment: Special thanks to Erich Lehmann for some helpful suggestions and references.

References


