

# On Bootstrap Tests of Hypotheses\*

Wei-Yin Loh<sup>1</sup>, and Wei Zheng<sup>2</sup>

*University of Wisconsin–Madison*

**Abstract:** The size of the bootstrap test of hypotheses is studied for the normal and exponential one and two-sample problems. It is found that the size depends not only on the problem, but on the choice of test statistic and the nominal level. In some special cases, the bootstrap test is UMP, but in other cases, it can be totally useless, such as being completely randomized or rejecting the null hypothesis with probability one. More importantly, the size is usually greater than the nominal level, even in the limit as the sample size goes to infinity.

## Contents

1	Introduction . . . . .	93
2	Testing a Normal Mean . . . . .	93
3	2.1 Known Variance . . . . .	94
4	2.1.1 Sample Mean Statistic . . . . .	94
5	2.1.2 Standard Likelihood Ratio Statistic . . . . .	94
6	2.1.3 Cox Likelihood Ratio Statistic . . . . .	95
7	2.2 Unknown Variance . . . . .	96
8	2.2.1 Standard Likelihood Ratio Statistic . . . . .	97
9	2.2.2 Cox Likelihood Ratio Statistic . . . . .	99
10	3 Testing a Normal Variance, Mean Unknown . . . . .	100
11	3.1 $H_0 : \sigma^2 \leq 1$ vs. $H_1 : \sigma^2 > 1$ . . . . .	100
12	3.1.1 Standard Likelihood Ratio Statistic . . . . .	100
13	3.1.2 Cox Likelihood Ratio Statistic . . . . .	102
14	3.2 $H_0 : \sigma^2 \geq 1$ vs. $H_1 : \sigma^2 < 1$ . . . . .	103
15	3.2.1 Standard Likelihood Ratio Statistic . . . . .	103
16	3.2.2 Cox Likelihood Ratio Statistic . . . . .	104
17	4 Testing Difference of Two Normal Means . . . . .	105
18	4.1 Known Variances . . . . .	105
19	4.1.1 Difference of Means Statistic . . . . .	105
20	4.1.2 Standard Likelihood Ratio Statistic . . . . .	106
21	4.1.3 Cox Likelihood Ratio Statistic . . . . .	106
22	4.2 Unknown but Equal Variances . . . . .	107
23	4.2.1 Difference of Means Statistic . . . . .	107
24	4.2.2 Standard Likelihood Ratio Statistic . . . . .	109

\*This material is based upon work partially supported by the National Science Foundation under grant DMS-0402470 and the U.S. Army Research Laboratory and the U.S. Army Research Office under grant W911NF-05-1-0047.

<sup>1</sup>Department of Statistics, University of Wisconsin–Madison, 1300 University Avenue, WI 53706, email: [loh@stat.wisc.edu](mailto:loh@stat.wisc.edu)

<sup>2</sup>Department of Statistics, University of Wisconsin–Madison, 1300 University Avenue, WI 53706, email: [zheng@stat.wisc.edu](mailto:zheng@stat.wisc.edu)

AMS 2000 subject classifications: Primary 62F03, 62F40; secondary 62F05

Keywords and phrases: likelihood ratio, maximum likelihood, normal distribution, exponential distribution, resampling, size, significance level, uniformly most powerful

1	4.2.3 Cox Likelihood Ratio Statistic . . . . .	110	1
2	5 Testing an Exponential Location Parameter . . . . .	110	2
3	5.1 $H_0 : \theta \leq 0$ vs. $H_1 : \theta > 0$ . . . . .	110	3
4	5.1.1 Standard Likelihood Ratio Statistic . . . . .	111	4
5	5.1.2 Cox Likelihood Ratio Statistic . . . . .	112	5
6	5.2 $H_0 : \theta \geq 0$ vs. $H_1 : \theta < 0$ . . . . .	112	6
7	5.2.1 Standard Likelihood Ratio Statistic . . . . .	112	7
8	5.2.2 Cox Likelihood Ratio Statistic . . . . .	112	8
9			9
10	6 Conclusion . . . . .	113	10
11	Appendix . . . . .	113	11
12	References . . . . .	115	12
13			13
14			14
15			15
16			16

## 1. Introduction

Owing to its practical convenience and wide applicability, the bootstrap method [7] is used to test statistical hypotheses in many research studies. A sample of recent applications includes evolutionary molecular biology [1], genetic structure [2], gene frequency [11], cancer epidemiology [8], microscopy [3], quality of life [12], economic cycles [5], livestock management [9], and meat demand [6]. Despite its popularity, however, there have been few detailed studies of the theoretical validity of the bootstrap for hypothesis testing. This article addresses this issue for some simple parametric problems where the bootstrap null distributions can be studied analytically. Specifically, we consider one and two-sample problems involving normally and exponentially distributed observations. Our goal is to determine the finite-sample or limiting sizes of the bootstrap tests and compare them with those of the traditional tests.

First, we recall some definitions. Let  $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$  be a vector of  $n$  independent observations from  $F_\mu$ . In the bootstrap method, we first find an estimate  $\hat{\mu}_0$  of  $\mu$  under  $H_0$  and estimate  $F_\mu$  with  $\hat{F} = F_{\hat{\mu}_0}$ . Given a test statistic  $S = S(\mathbf{X}_n)$  for which large values lead to rejection of  $H_0$ , let  $G_\mu$  denote the distribution function of  $S$ . Let  $\mathbf{X}_n^* = (X_1^*, X_2^*, \dots, X_n^*)$  be a vector of  $n$  independent observations from  $\hat{F}$  and define  $S^* = S(\mathbf{X}_n^*)$ . The distribution function  $\hat{G} = G_{\hat{\mu}_0}$  of  $S^*$  is the bootstrap distribution function of  $S$ , i.e.,  $\hat{G}$  is the distribution of  $S$  under  $\hat{F}$ .

For any nominal level of significance  $\alpha$  ( $0 < \alpha < 1$ ), let  $c_\alpha(\hat{\mu}_0)$  be the upper- $\alpha$  quantile of  $\hat{G}$ . Thus  $c_\alpha(\hat{\mu}_0)$  is the smallest value such that  $\hat{G}(c_\alpha(\hat{\mu}_0)) \geq 1 - \alpha$ . The nominal level- $\alpha$  bootstrap test rejects  $H_0$  with probability 1 if  $S > c_\alpha(\hat{\mu}_0)$ , and with probability  $[\alpha - 1 + \hat{G}(c_\alpha(\hat{\mu}_0))]/[\hat{G}(c_\alpha(\hat{\mu}_0)) - \hat{G}(c_\alpha(\hat{\mu}_0)-)]$  if  $S = c_\alpha(\hat{\mu}_0)$  and  $\hat{G}(c_\alpha(\hat{\mu}_0)) > \hat{G}(c_\alpha(\hat{\mu}_0)-)$ .

## 2. Testing a Normal Mean

Let  $X_1, X_2, \dots, X_n$  be a random sample from  $N(\mu, \sigma^2)$ , a normal distribution with mean  $\mu$  and variance  $\sigma^2$ . Let  $\phi(x)$  and  $\Phi(x)$  denote the density and distribution functions of the  $N(0, 1)$  distribution and let  $z_\alpha$  be its upper- $\alpha$  critical value, that is,  $1 - \Phi(z_\alpha) = \alpha$ . Consider testing

$$(2.1) \quad H_0 : \mu \leq 0 \quad \text{vs.} \quad H_1 : \mu > 0.$$

## 2.1. Known Variance

We assume without loss of generality that  $\sigma^2 = 1$ . Let  $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ . The unrestricted MLE of  $\mu$  is  $\hat{\mu} = \bar{X}_n$ . Let  $\hat{\mu}_i$  be the MLE of  $\mu$  under  $H_i$  ( $i = 0, 1$ ). Then  $\hat{\mu}_0 = \bar{X}_n I(\bar{X}_n < 0)$ ,  $\hat{\mu}_1 = \bar{X}_n I(\bar{X}_n > 0)$ , and  $X_1^*, X_2^*, \dots, X_n^*$  is a bootstrap random sample drawn from  $N(\hat{\mu}_0, 1)$ . Let  $\bar{X}_n^* = n^{-1} \sum_{i=1}^n X_i^*$ .

### 2.1.1. Sample Mean Statistic

**Theorem 2.1.** *If  $0 < \alpha \leq 1/2$ , the bootstrap test based on  $\bar{X}_n$  is uniformly most powerful (UMP), but if  $1/2 < \alpha < 1$ , the test rejects  $H_0$  with probability 1.*

*Proof.* Recall that the UMP test rejects  $H_0$  if  $\bar{X}_n \geq z_\alpha n^{-1/2}$ . Since  $\bar{X}_n$  is normal with mean  $\hat{\mu}_0$  and variance  $n^{-1}$ , its critical value is  $c_\alpha(\hat{\mu}_0) = \hat{\mu}_0 + z_\alpha n^{-1/2} = \bar{X}_n I(\bar{X}_n < 0) + z_\alpha n^{-1/2}$ . Therefore the bootstrap test rejects  $H_0$  if  $\bar{X}_n^* I(\bar{X}_n^* > 0) \geq z_\alpha n^{-1/2}$ . If  $0 < \alpha \leq 1/2$ , then  $z_\alpha \geq 0$  and the bootstrap test is the UMP test. If  $1/2 < \alpha < 1$ , then  $z_\alpha < 0$  and the test rejects  $H_0$  w.p.1.

### 2.1.2. Standard Likelihood Ratio Statistic

Given the data and values  $\mu_0$  and  $\mu_1$ , let

$$L(\mu_0, \mu_1, \mathbf{X}_n) = \log \left\{ \prod_{i=1}^n \phi(x_i - \mu_1) \middle/ \prod_{i=1}^n \phi(x_i - \mu_0) \right\}.$$

A general statistic for testing  $H_0$  is the log-likelihood ratio

$$L(\hat{\mu}_0, \hat{\mu}, \mathbf{X}_n) = \log \left\{ \sup_{\mu} \prod_{i=1}^n \phi(x_i - \mu) \middle/ \sup_{\mu \in H_0} \prod_{i=1}^n \phi(x_i - \mu) \right\}.$$

Throughout this article, we let  $Z$  denote the standard normal variable and  $z_\alpha^+ = \max(z_\alpha, 0)$ . We need the following lemma whose proof is given in the Appendix.

**Lemma 2.1.** *Let  $\theta \geq 0$ . For fixed  $0 < \alpha < 1$ , the function*

$$(2.2) \quad P(|Z + \theta| > z_\alpha^+) - (1 - \alpha) E\{\Phi(Z + \theta)^{-1} I(Z + \theta > z_\alpha^+)\}$$

*is maximized at  $\theta = 0$  with maximum value*

$$(2.3) \quad \min(2\alpha, 1) + (1 - \alpha) \log\{1 - \min(\alpha, 1/2)\}$$

*which is greater than  $\alpha$  for all  $0 < \alpha < 1$ .*

**Theorem 2.2.** *The size of the bootstrap test based on the standard likelihood ratio is  $\min(2\alpha, 1) + (1 - \alpha) \log\{1 - \min(\alpha, 1/2)\}$ .*

*Proof.* Since

$$\begin{aligned} n^{-1} L(\hat{\mu}_0, \hat{\mu}, \mathbf{X}_n) &= \bar{X}_n(\hat{\mu} - \hat{\mu}_0) - (\hat{\mu}^2 - \hat{\mu}_0^2)/2 \\ &= \bar{X}_n(\bar{X}_n - \hat{\mu}_0) - (\bar{X}_n^2 - \hat{\mu}_0^2)/2 \\ &= (\bar{X}_n - \hat{\mu}_0)^2/2 \\ &= \bar{X}_n^2 I(\bar{X}_n > 0)/2 \end{aligned}$$

the test rejects  $H_0$  if  $S = \bar{X}_n^* I(\bar{X}_n^* > 0) \geq c_\alpha(\hat{\mu}_0)$ , where the critical value is to be determined. Let  $S^* = \bar{X}_n^* I(\bar{X}_n^* > 0)$  and consider two cases.

1.  $\bar{X}_n > 0$ . Then  $S > 0$ ,  $\hat{\mu}_0 = 0$ , and  $\bar{X}_n^*$  has a  $N(0, n^{-1})$  distribution. For any  $x \geq 0$ ,  $P(S^* \leq x) = P(\bar{X}_n^* \leq x) = \Phi(xn^{1/2})$ . Therefore if  $0 < \alpha < 1/2$ ,  $c_\alpha(\hat{\mu}_0) = z_\alpha n^{-1/2}$ . Otherwise, if  $\alpha \geq 1/2$ , then  $c_\alpha(\hat{\mu}_0) = 0$  and the bootstrap test rejects  $H_0$  w.p.1. Thus for all  $0 < \alpha < 1$ ,  $c_\alpha(\hat{\mu}_0) = z_\alpha^+ n^{-1/2}$ .
2.  $\bar{X}_n \leq 0$ . Then  $S = 0$ ,  $\hat{\mu}_0 = \bar{X}_n \leq 0$ , and  $S^*$  has a  $N(\bar{X}_n, n^{-1})$  distribution left-truncated at 0 with  $P(S^* = 0) = P(\bar{X}_n^* \leq 0) = \Phi(-n^{1/2}\bar{X}_n)$ . Thus

$$c_\alpha(\hat{\mu}_0) = \begin{cases} \bar{X}_n + n^{-1/2}z_\alpha, & \text{if } \bar{X}_n + n^{-1/2}z_\alpha > 0 \\ 0, & \text{if } \bar{X}_n + n^{-1/2}z_\alpha \leq 0. \end{cases}$$

$$= (\bar{X}_n + n^{-1/2}z_\alpha)^+.$$

Since  $S = 0$ , the bootstrap test never rejects  $H_0$  if  $\bar{X}_n + n^{-1/2}z_\alpha > 0$ . Otherwise, the test is randomized and rejects  $H_0$  with probability  $\{\alpha - 1 + \Phi(-n^{1/2}\bar{X}_n)\}/\Phi(-n^{1/2}\bar{X}_n)$ .

Thus for  $0 < \alpha < 1$ ,

$$\begin{aligned} P\{\text{Reject } H_0\} &= P\{\text{Reject } H_0, \bar{X}_n > 0\} + P\{\text{Reject } H_0, \bar{X}_n < 0\} \\ &= P(S > z_\alpha^+, \bar{X}_n > 0) \\ &\quad + P\{\text{Reject } H_0, \bar{X}_n + n^{-1/2}z_\alpha \leq 0, \bar{X}_n < 0\} \\ &= P(\bar{X}_n > z_\alpha^+ n^{-1/2}) \\ &\quad + E[\{\alpha - 1 + \Phi(-n^{1/2}\bar{X}_n)\}/\Phi(-n^{1/2}\bar{X}_n)] I(-n^{1/2}\bar{X}_n \geq z_\alpha^+) \\ &= P(|W| > z_\alpha^+) - (1 - \alpha)E\{\Phi(W)^{-1} I(W > z_\alpha^+)\} \end{aligned}$$

where  $W$  is normally distributed with mean  $-n^{1/2}\mu$  and variance 1. By Lemma 2.1, the supremum of the rejection probability under  $H_0$  is attained when  $\mu = 0$  and is given by (2.3). Figure 1 shows a plot of this function.

### 2.1.3. Cox Likelihood Ratio Statistic

Cox (1961) proposed the following alternative likelihood ratio statistic for testing separate families of hypotheses:

$$L(\hat{\mu}_0, \hat{\mu}_1, \mathbf{X}_n) = \log \left\{ \sup_{H_1} \prod_{i=1}^n \phi(x_i - \mu) \middle/ \sup_{H_0} \prod_{i=1}^n \phi(x_i - \mu) \right\}.$$

For the current problem,

$$\begin{aligned} L(\hat{\mu}_0, \hat{\mu}_1, \mathbf{X}_n) &= n\{\bar{X}_n(\hat{\mu}_1 - \hat{\mu}_0) - (\hat{\mu}_1^2 - \hat{\mu}_0^2)/2\} \\ &= n(\bar{X}_n|\bar{X}_n| - \bar{X}_n|\bar{X}_n|/2) \\ &= n\bar{X}_n^2 \text{sgn}(\bar{X}_n)/2. \end{aligned}$$

Therefore rejecting  $H_0$  for large values of  $L(\hat{\mu}_0, \hat{\mu}_1, \mathbf{X}_n)$  is equivalent to rejecting for large values of  $\bar{X}_n$ , and the next theorem follows directly from Theorem 2.1.

**Theorem 2.3.** *If  $0 < \alpha \leq 1/2$ , the bootstrap test based on the Cox likelihood ratio has size  $\alpha$  and is UMP. If  $1/2 < \alpha < 1$ , it rejects  $H_0$  with probability 1.*

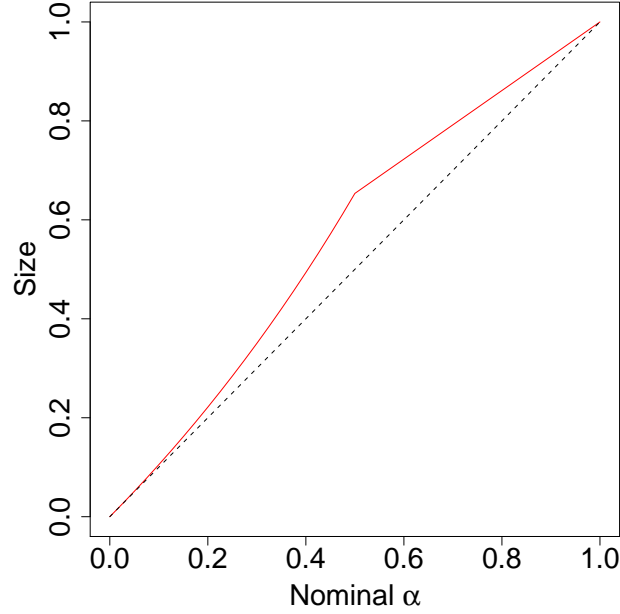


FIG 1. Size (2.3) of bootstrap test for the normal mean based on the standard likelihood ratio, for known  $\sigma$ . The dashed line is the identity function.

## 2.2. Unknown Variance

Now suppose we test the hypotheses (2.1) without assuming that  $\sigma$  is known. The log-likelihood function is

$$l(\mu, \sigma) = -n \log \sigma - \sum_i (X_i - \mu)^2 / (2\sigma^2) - (n/2) \log(2\pi)$$

and its derivatives are  $\partial l / \partial \mu = -\sigma^{-2} \sum (X_i - \mu)$  and  $\partial l / \partial \sigma = -n\sigma^{-1} + \sigma^{-3} \sum (X_i - \mu)^2$ . Hence the unrestricted and restricted (under  $H_0$  and  $H_1$ ) maximum likelihood estimates (MLEs) of  $\mu$  and  $\sigma^2$  are, respectively,

$$\begin{aligned} \hat{\mu} &= \bar{X}_n & \hat{\sigma}^2 &= n^{-1} \sum (X_i - \bar{X}_n)^2 \\ \hat{\mu}_0 &= \bar{X}_n I(\bar{X}_n < 0) & \hat{\sigma}_0^2 &= n^{-1} \sum (X_i - \hat{\mu}_0)^2 \\ \hat{\mu}_1 &= \bar{X}_n I(\bar{X}_n > 0) & \hat{\sigma}_1^2 &= n^{-1} \sum (X_i - \hat{\mu}_1)^2 \end{aligned}$$

giving the log-likelihood ratio statistics:

**Standard:**  $n \log(\hat{\sigma}_0 / \hat{\sigma}) = (n/2) \log\{\sum (X_i - \hat{\mu}_0)^2 / \sum (X_i - \bar{X}_n)^2\}$

**Cox:**  $n \log(\hat{\sigma}_0 / \hat{\sigma}_1) = (n/2) \log\{\sum (X_i - \hat{\mu}_0)^2 / \sum (X_i - \hat{\mu}_1)^2\}$ .

The corresponding bootstrap tests reject  $H_0$  for large values of  $\sum (X_i - \hat{\mu}_0)^2 / \sum (X_i - \bar{X}_n)^2$  and  $\sum (X_i - \hat{\mu}_0)^2 / \sum (X_i - \hat{\mu}_1)^2$ , respectively.

2.2.1. Standard Likelihood Ratio Statistic

Let

$$(2.4) \quad T_n = n^{1/2} \bar{X}_n \left\{ \sum (X_i - \bar{X}_n)^2 / (n-1) \right\}^{-1/2}.$$

The standard log-likelihood ratio statistic is

$$\begin{aligned} \frac{\sum (X_i - \hat{\mu}_0)^2}{\sum (X_i - \bar{X}_n)^2} &= \frac{\sum \{X_i - \bar{X}_n I(\bar{X}_n < 0)\}^2}{\sum (X_i - \bar{X}_n)^2} \\ &= \begin{cases} 1, & \text{if } \bar{X}_n < 0 \\ \sum X_i^2 / \sum (X_i - \bar{X}_n)^2, & \text{if } \bar{X}_n \geq 0 \end{cases} \\ &= \begin{cases} 1, & \text{if } \bar{X}_n < 0 \\ 1 + n\bar{X}_n^2 / \sum (X_i - \bar{X}_n)^2, & \text{if } \bar{X}_n \geq 0 \end{cases} \\ &= \begin{cases} 1, & \text{if } \bar{X}_n < 0 \\ 1 + (n-1)^{-1} T_n^2, & \text{if } \bar{X}_n \geq 0. \end{cases} \end{aligned}$$

Thus  $H_0$  is rejected for large values of  $S = T_n I(T_n > 0)$ . Let  $t_{\nu, \delta}$  denote the noncentral  $t$ -distribution with  $\nu$  degrees of freedom and noncentrality parameter  $\delta$  and let  $t_{\nu, \delta, \alpha}$  denote its upper- $\alpha$  critical point.

**Lemma 2.2.** For any  $\nu$  and  $\alpha$ ,  $t_{\nu, \delta, \alpha}$  is an increasing function of  $\delta$ .

*Proof.* Let  $Z$  denote a standard normal variable independent of  $\chi_\nu^2$ . Since

$$\begin{aligned} P(t_{\nu, \delta} \leq x) &= P\left(\frac{Z + \delta}{\sqrt{\chi_\nu^2/\nu}} \leq x\right) \\ &= P\left(\frac{Z}{\sqrt{\chi_\nu^2/\nu}} \leq x - \frac{\delta}{\sqrt{\chi_\nu^2/\nu}}\right) \end{aligned}$$

we see that  $P(t_{\nu, \delta} \leq x)$  is a decreasing function of  $\delta$ . Therefore  $t_{\nu, \delta, \alpha}$  is an increasing function of  $\delta$ .

**Theorem 2.4.** If  $\sigma$  is unknown, the size of the nominal level- $\alpha$  test of  $H_0 : \mu \leq 0$  vs.  $H_1 : \mu > 0$  based on the standard likelihood ratio has lower bound

$$\min(\alpha, 1/2) + E \left\{ \frac{\alpha - 1 + \Phi\left(-t_{n-1} \sqrt{n/(n-1)}\right)}{\Phi\left(-t_{n-1} \sqrt{n/(n-1)}\right)} I\left(t_{n-1} \sqrt{\frac{n}{n-1}} < -z_\alpha^+\right) \right\}$$

where  $t_{n-1}$  has a (central)  $t$ -distribution with  $n-1$  degrees of freedom. As  $n \rightarrow \infty$ , the bound tends to (2.3), the size for the case where  $\sigma$  is known and  $n$  is finite.

*Proof.* Again, consider two cases.

1.  $\bar{X}_n > 0$ . Then  $S > 0$  and  $\hat{\mu}_0 = 0$ . The bootstrap distribution of  $T_n^*$  is a central  $t_{n-1}$ -distribution and that of  $S^*$  is a central  $t_{n-1}$ -distribution left-truncated at 0. If  $0 < \alpha < 1/2$ , the test rejects  $H_0$  whenever  $T_n > t_{n-1, 0, \alpha}$ . Otherwise, if  $\alpha \geq 1/2$ , the test rejects  $H_0$  with probability 1.
2.  $\bar{X}_n < 0$ . Then  $S = 0$ ,  $\hat{\mu}_0 < 0$ , and  $S^*$  has a left-truncated noncentral  $t_{n-1, \delta}$ -distribution with  $n-1$  degrees of freedom and noncentrality parameter

$$(2.5) \quad \delta = n^{1/2} \hat{\mu}_0 / \hat{\sigma}_0 = n \bar{X}_n / \sqrt{\sum (X_i - \bar{X}_n)^2} = T_n \sqrt{n/(n-1)}$$

and probability  $P(\bar{X}_n^* \leq 0) = P\{n^{1/2}(\bar{X}_n^* - \hat{\mu}_0)/\hat{\sigma}_0 \leq -n^{1/2}\hat{\mu}_0/\hat{\sigma}_0\} = \Phi(-\delta)$  at 0.

If  $t_{n-1,\delta,\alpha} > 0$ , the bootstrap test does not reject  $H_0$  because  $S = 0$ . Otherwise, if  $t_{n-1,\delta,\alpha} \leq 0$ , the test is randomized and rejects  $H_0$  with probability  $\{\alpha - 1 + \Phi(-\delta)\}/\Phi(-\delta)$ . Note that the event  $t_{n-1,\delta,\alpha} \leq 0$  occurs if and only if  $\alpha \geq P(T_n^* > 0 | \hat{\mu}_0, \hat{\sigma}_0)$ . But

$$P(T_n^* > 0 | \hat{\mu}_0, \hat{\sigma}_0) = P(\bar{X}_n^* > 0 | \hat{\mu}_0, \hat{\sigma}_0) = 1 - \Phi(-\delta).$$

Therefore  $t_{n-1,\delta,\alpha} \leq 0$  if and only if  $\delta \leq -z_\alpha$ .

Let  $P_{\eta,\tau}$  denote probabilities when  $\mu = \eta$  and  $\sigma = \tau$ . The size of the test for  $0 < \alpha < 1$  is

$$\begin{aligned} & \sup_{H_0} P_{\mu,\sigma}\{\text{Reject } H_0\} \\ &= \sup_{H_0} [P_{\mu,\sigma}\{\text{Reject } H_0, \bar{X}_n > 0\} + P_{\mu,\sigma}\{\text{Reject } H_0, \bar{X}_n < 0\}] \\ &= \sup_{H_0} P_{\mu,\sigma}\{[T_n I(T_n > 0) > t_{n-1,0,\alpha}, \bar{X}_n > 0] \\ &\quad + P_{\mu,\sigma}\{\text{Reject } H_0, t_{n-1,\delta,\alpha} \leq 0, \bar{X}_n < 0\}\} \\ &= \sup_{H_0} [P_{\mu,\sigma}\{T_n > \max(t_{n-1,0,\alpha}, 0)\} \\ &\quad + E_{\mu,\sigma}\{[\alpha - 1 + \Phi(-\delta)]/\Phi(-\delta)\} I\{\delta < \min(-z_\alpha, 0)\}] \\ &\geq P_{0,1}\{T_n > \max(t_{n-1,0,\alpha}, 0)\} + E_{0,1}\{[\alpha - 1 + \Phi(-\delta)]/\Phi(-\delta)\} I\{\delta < -z_\alpha^+\} \\ &= \min(\alpha, 1/2) + E \left\{ \frac{\alpha - 1 + \Phi\left(-t_{n-1}\sqrt{\frac{n}{n-1}}\right)}{\Phi\left(-t_{n-1}\sqrt{\frac{n}{n-1}}\right)} I\left(t_{n-1}\sqrt{\frac{n}{n-1}} < -z_\alpha^+\right) \right\} \end{aligned}$$

by equation (2.5). Since  $t_{n-1} \rightarrow Z$  in distribution as  $n \rightarrow \infty$ , where  $Z$  is a standard normal variable,

$$\begin{aligned} & \lim_{n \rightarrow \infty} E \left\{ \frac{\alpha - 1 + \Phi\left(-t_{n-1}\sqrt{\frac{n}{n-1}}\right)}{\Phi\left(-t_{n-1}\sqrt{\frac{n}{n-1}}\right)} I\left(t_{n-1}\sqrt{\frac{n}{n-1}} < -z_\alpha^+\right) \right\} \\ & \rightarrow E \left\{ \frac{\alpha - 1 + \Phi(-Z)}{\Phi(-Z)} I(Z < -z_\alpha^+) \right\} \\ &= (\alpha - 1) \int_{-\infty}^{-z_\alpha^+} \phi(z)/\Phi(-z) dz + \Phi(-z_\alpha^+) \\ &= (\alpha - 1) \int_{z_\alpha^+}^{\infty} \phi(z)/\Phi(z) dz + \min(\alpha, 1/2) \\ &= (1 - \alpha) \log \Phi(z_\alpha^+) + \min(\alpha, 1/2) \\ &= (1 - \alpha) \log\{\max(1 - \alpha, 1/2)\} + \min(\alpha, 1/2) \\ &= (1 - \alpha) \log\{1 - \min(\alpha, 1/2)\} + \min(\alpha, 1/2) \end{aligned}$$

Thus the limiting size is  $2 \min(\alpha, 1/2) + (1 - \alpha) \log\{1 - \min(\alpha, 1/2)\} > \alpha$ .

2.2.2. Cox Likelihood Ratio Statistic

**Theorem 2.5.** *If  $\sigma^2$  is unknown, the size of the bootstrap test of (2.1) based on the Cox likelihood ratio has lower bound*

$$\min(\alpha, 1/2) + P(t_{n-1, t_{n-1}\sqrt{n/(n-1)}, \alpha} < t_{n-1} < 0) \geq \alpha.$$

*Proof.* The Cox log-likelihood ratio statistic is

$$\begin{aligned} \frac{\sum(X_i - \hat{\mu}_0)^2}{\sum(X_i - \hat{\mu}_1)^2} &= \frac{\sum\{X_i - \bar{X}_n I(\bar{X}_n < 0)\}^2}{\sum\{X_i - \bar{X}_n I(\bar{X}_n > 0)\}^2} \\ &= \begin{cases} \{1 + (n-1)T_n^2\}^{-1}, & \text{if } \bar{X}_n < 0 \\ 1, & \text{if } \bar{X}_n = 0 \\ 1 + (n-1)T_n^2, & \text{if } \bar{X}_n > 0 \end{cases} \end{aligned}$$

where  $T_n$  is defined in (2.4). Thus rejecting for large values of the statistic is equivalent to rejecting for large values of  $S = T_n$ .

1.  $\bar{X}_n > 0$ . Then  $\hat{\mu}_0 = 0$ ,  $T_n > 0$ , and  $T_n^*$  has a central  $t$ -distribution with  $n-1$  degrees of freedom. Thus the test rejects  $H_0$  if  $T_n > t_{n-1, 0, \alpha}$ . If  $1/2 \leq \alpha < 1$ , then  $t_{n-1, 0, \alpha} \leq 0$  and the test rejects w.p.1.
2.  $\bar{X}_n < 0$ . Then  $\hat{\mu}_0 < 0$ ,  $T_n < 0$ , and  $T_n^*$  has a noncentral  $t$ -distribution with  $n-1$  degrees of freedom and noncentrality parameter  $\delta$  given in (2.5). Hence  $H_0$  is rejected if  $T_n > t_{n-1, \delta, \alpha}$ . Since  $T_n < 0$ , rejection occurs only if  $t_{n-1, \delta, \alpha} < 0$ .

If  $0 < \alpha < 1/2$ ,

$$\begin{aligned} \sup_{H_0} P_{\mu, \sigma}(\text{Reject } H_0) &= \sup_{H_0} [P_{\mu, \sigma}\{\text{Reject } H_0, \bar{X}_n > 0\} \\ &\quad + P_{\mu, \sigma}\{\text{Reject } H_0, \bar{X}_n < 0\}] \\ &= \sup_{H_0} [P_{\mu, \sigma}(T_n > t_{n-1, 0, \alpha}) + P_{\mu, \sigma}(t_{n-1, \delta, \alpha} < T_n < 0)] \\ &\geq P_{0,1}(T_n > t_{n-1, 0, \alpha}) + P_{0,1}(t_{n-1, \delta, \alpha} < T_n < 0) \\ &= \alpha + P(t_{n-1, \delta, \alpha} < t_{n-1} < 0) \\ &\geq \alpha. \end{aligned}$$

If  $1/2 \leq \alpha < 1$ ,

$$\begin{aligned} \sup_{H_0} P_{\mu, \sigma}(\text{Reject } H_0) &= \sup_{H_0} [P_{\mu, \sigma}\{\text{Reject } H_0, \bar{X}_n > 0\} \\ &\quad + P_{\mu, \sigma}\{\text{Reject } H_0, \bar{X}_n < 0\}] \\ &= \sup_{H_0} [P_{\mu, \sigma}(\bar{X}_n > 0) + P_{\mu, \sigma}(t_{n-1, \delta, \alpha} < T_n < 0)] \\ &\geq P_{0,1}(\bar{X}_n > 0) + P_{0,1}(t_{n-1, \delta, \alpha} < T_n < 0) \\ &= 1/2 + P(t_{n-1, \delta, \alpha} < t_{n-1} < 0) \\ &> 1/2 + P(t_{n-1, 0, \alpha} < t_{n-1} < 0) \\ &= \alpha \end{aligned}$$

by Lemma 2.2.

### 3. Testing a Normal Variance, Mean Unknown

Let  $\chi_\nu^2$  denote a chi-squared random variable with  $\nu$  degrees of freedom,  $\chi_{\nu,\alpha}^2$  its upper- $\alpha$  point, and  $\Psi_\nu(\cdot)$  its cumulative distribution function.

**Lemma 3.1.**  $\Psi_{n-1}(n^2/\chi_{n-1,\alpha}^2) \rightarrow \alpha$  and  $\Psi_{n-1}(n) \rightarrow 1/2$  as  $n \rightarrow \infty$ .

*Proof.* Let  $Z_1, Z_2, \dots$  be independent  $N(0, 1)$  variables. Then

$$\begin{aligned} \Psi_{n-1}(n^2/\chi_{n-1,\alpha}^2) &= P\left(\sum_{i=1}^{n-1} Z_i^2 \leq n^2/\chi_{n-1,\alpha}^2\right) \\ &= P\left(\frac{\sum_{i=1}^{n-1} (Z_i^2 - 1)}{\sqrt{2(n-1)}} \leq \sqrt{\frac{n-1}{2}} \left\{ \frac{n^2}{(n-1)^2 \chi_{n-1,\alpha}^2} - 1 \right\}\right) \\ &\approx \Phi\left(\sqrt{(n-1)/2} \left\{ \frac{n^2}{(n-1)^2 \chi_{n-1,\alpha}^2} - 1 \right\}\right) \text{ as } n \rightarrow \infty. \end{aligned}$$

By the Wilson-Hilferty (1931) approximation,  $\nu/\chi_{\nu,\alpha}^2 = 1 - z_\alpha(2/\nu)^{1/2} + o(\nu^{-1})$ . Therefore

$$\sqrt{(n-1)/2} \left\{ \frac{n^2}{(n-1)^2 \chi_{n-1,\alpha}^2} - 1 \right\} \rightarrow -z_\alpha$$

which yields the first result. The second result is similarly proved.

Let  $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$  be a vector of  $n$  independent observations from  $N(\mu, \sigma^2)$ , with  $\mu$  and  $\sigma$  unknown, and let  $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$  denote the unrestricted MLE of  $\sigma^2$ .

#### 3.1. $H_0 : \sigma^2 \leq 1$ vs. $H_1 : \sigma^2 > 1$

Let  $\hat{\sigma}_i^2$  be the MLE of  $\sigma^2$  under  $H_i$  ( $i = 0, 1$ ). Then  $\hat{\sigma}_0^2 = \min(\hat{\sigma}^2, 1)$  and  $\hat{\sigma}_1^2 = \max(\hat{\sigma}^2, 1)$ . Define the log-likelihood ratio

$$M(\mu_0, \mu_1, \sigma_0, \sigma_1, \mathbf{X}_n) = \log\left(\frac{\prod_i \sigma_1^{-1} \phi\{\sigma_1^{-1}(x_i - \mu_1)\}}{\prod_i \sigma_0^{-1} \phi\{\sigma_0^{-1}(x_i - \mu_0)\}}\right).$$

##### 3.1.1. Standard Likelihood Ratio Statistic

**Theorem 3.1.** *The size of the bootstrap test based on the standard likelihood ratio for testing  $H_0 : \sigma^2 \leq 1$  vs.  $H_1 : \sigma^2 > 1$ , with  $\mu$  unknown, is bounded below by*

$$(3.1) \quad \min\{\alpha, 1 - \Psi_{n-1}(n)\} + E\left[\frac{\alpha - 1 + \Psi_{n-1}(n^2/\chi_{n-1}^2)}{\Psi_{n-1}(n^2/\chi_{n-1}^2)} I\left\{\chi_{n-1}^2 \leq \min\left(n, \frac{n^2}{\chi_{n-1,\alpha}^2}\right)\right\}\right].$$

*Proof.* The standard log-likelihood ratio statistic is  $M(\hat{\mu}, \hat{\mu}, \hat{\sigma}_0, \hat{\sigma}, \mathbf{X}_n)$  and

$$\begin{aligned} 2n^{-1} M(\hat{\mu}, \hat{\mu}, \hat{\sigma}_0, \hat{\sigma}, \mathbf{X}_n) &= \log(\hat{\sigma}_0^2 \hat{\sigma}^{-2}) + \hat{\sigma}^2 (\hat{\sigma}_0^{-2} - \hat{\sigma}^{-2}) \\ &= \hat{\sigma}^2 \hat{\sigma}_0^{-2} - \log(\hat{\sigma}^2 \hat{\sigma}_0^{-2}) - 1 \\ &= \begin{cases} 0, & \text{if } \hat{\sigma}^2 \leq 1 \\ \hat{\sigma}^2 - \log(\hat{\sigma}^2) - 1, & \text{otherwise.} \end{cases} \end{aligned}$$

Since the function  $x - \log(x) - 1$  increases monotonically from 0 for  $x > 1$ , rejecting for large values of the statistic is equivalent to rejecting for large values of

$$S = n \max(\hat{\sigma}^2, 1) = \max \left\{ \sum (X_i - \bar{X}_n)^2, n \right\}.$$

Let  $S^*$  denote the bootstrap version of  $S$  under resampling from  $N(\bar{X}_n, \hat{\sigma}_0^2)$ . To find the critical point of the distribution of  $S^*$ , consider two cases.

1.  $\hat{\sigma}^2 > 1$ . Then  $S > n$ ,  $\hat{\sigma}_0^2 = 1$ , and the distribution of  $S^*$  is  $\chi_{n-1}^2$  left-truncated at  $n$ , i.e., it has probability mass  $\Psi_{n-1}(n)$  at  $n$ . If  $0 < \alpha < 1 - \Psi_{n-1}(n)$ , the critical point of the bootstrap distribution is  $\chi_{n-1, \alpha}^2$ . Otherwise, the critical point is  $n$  and the test rejects  $H_0$  w.p.1.
2.  $\hat{\sigma}^2 \leq 1$ . Then  $S = n$ ,  $\hat{\sigma}_0^2 = \hat{\sigma}^2$ , and the distribution of  $S^*$  is  $\hat{\sigma}^2 \chi_{n-1}^2$  left truncated at  $n$ . Thus the test does not reject  $H_0$  if  $\hat{\sigma}^2 \chi_{n-1, \alpha}^2 > n$ . On the other hand, if  $\hat{\sigma}^2 \chi_{n-1, \alpha}^2 \leq n$ , then the critical point is  $n$  and the test rejects  $H_0$  randomly with probability  $\{\alpha - 1 + \Psi_{n-1}(n\hat{\sigma}_0^{-2})\} / \Psi_{n-1}(n\hat{\sigma}_0^{-2})$ .

Since  $\alpha < 1 - \Psi_{n-1}(n)$  if and only if  $n < \chi_{n-1, \alpha}^2$ , we have

$$\begin{aligned} P_{\mu, \sigma}(\text{Reject } H_0, \hat{\sigma}^2 > 1) &= \begin{cases} P_{\mu, \sigma}(S > \chi_{n-1, \alpha}^2, n\hat{\sigma}^2 > n), & \text{if } n < \chi_{n-1, \alpha}^2 \\ P_{\mu, \sigma}(n\hat{\sigma}^2 > n), & \text{otherwise} \end{cases} \\ &= \begin{cases} P_{\mu, \sigma}(n\hat{\sigma}^2 > \chi_{n-1, \alpha}^2), & \text{if } n < \chi_{n-1, \alpha}^2 \\ P_{\mu, \sigma}(n\hat{\sigma}^2 > n), & \text{otherwise} \end{cases} \\ &= \begin{cases} 1 - \Psi_{n-1}(\sigma^{-2} \chi_{n-1, \alpha}^2), & \text{if } n < \chi_{n-1, \alpha}^2 \\ 1 - \Psi_{n-1}(n\sigma^{-2}), & \text{otherwise} \end{cases} \\ &= 1 - \Psi_{n-1}(\sigma^{-2} \max\{n, \chi_{n-1, \alpha}^2\}). \end{aligned}$$

and

$$\begin{aligned} P_{\mu, \sigma}(\text{Reject } H_0, \hat{\sigma}^2 \leq 1) &= P_{\mu, \sigma}(\text{Reject } H_0, \hat{\sigma}^2 \chi_{n-1, \alpha}^2 \leq n, n\hat{\sigma}^2 \leq n) \\ &= E_{\mu, \sigma} \left[ \frac{\alpha - 1 + \Psi_{n-1}(n\hat{\sigma}_0^{-2})}{\Psi_{n-1}(n\hat{\sigma}_0^{-2})} I(\hat{\sigma}_0^2 \chi_{n-1, \alpha}^2 \leq n, n\hat{\sigma}^2 \leq n) \right] \\ &= E \left[ \frac{\alpha - 1 + \Psi_{n-1}(n^2 \sigma^{-2} / \chi_{n-1}^2)}{\Psi_{n-1}(n^2 \sigma^{-2} / \chi_{n-1}^2)} I(\chi_{n-1}^2 \leq \sigma^{-2} \min\{n, n^2 / \chi_{n-1, \alpha}^2\}) \right]. \end{aligned}$$

The choice  $\sigma^2 = 1$  yields the lower bound

$$\begin{aligned} &\sup_{H_0} P_{\mu, \sigma}(\text{Reject } H_0) \\ &\geq P_{\mu, 1}(\text{Reject } H_0, \hat{\sigma}^2 > 1) + P_{\mu, 1}(\text{Reject } H_0, \hat{\sigma}^2 \leq 1) \\ &= 1 - \Psi_{n-1}(\max\{n, \chi_{n-1, \alpha}^2\}) \\ &\quad + E \left[ \frac{\alpha - 1 + \Psi_{n-1}(n^2 / \chi_{n-1}^2)}{\Psi_{n-1}(n^2 / \chi_{n-1}^2)} I \left\{ \chi_{n-1}^2 \leq \min \left( n, \frac{n^2}{\chi_{n-1, \alpha}^2} \right) \right\} \right] \\ &= \min\{\alpha, 1 - \Psi_{n-1}(n)\} \\ &\quad + E \left[ \frac{\alpha - 1 + \Psi_{n-1}(n^2 / \chi_{n-1}^2)}{\Psi_{n-1}(n^2 / \chi_{n-1}^2)} I \left\{ \chi_{n-1}^2 \leq \min \left( n, \frac{n^2}{\chi_{n-1, \alpha}^2} \right) \right\} \right]. \end{aligned}$$

Figure 2 shows graphs of the lower bound (3.1) for  $n = 5, 10, 100$ , and 500.

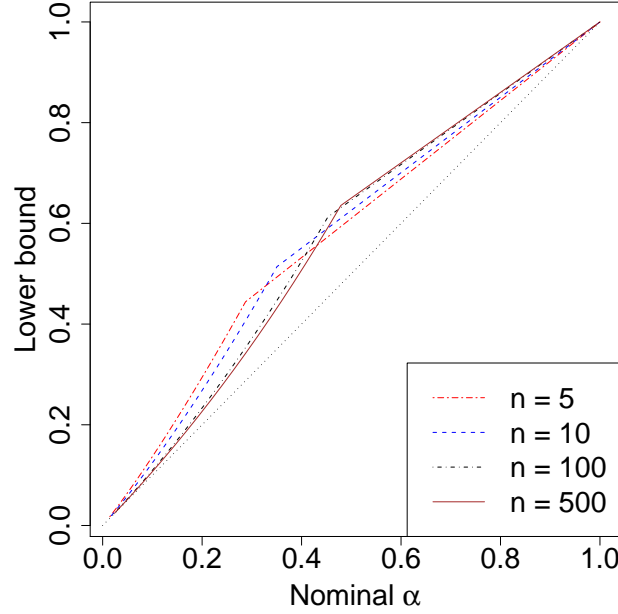


FIG 2. Lower bounds (3.1) on the size of the bootstrap test of  $H_0 : \sigma^2 \leq 1$  vs.  $H_1 : \sigma^2 > 1$  based on the standard likelihood ratio, for  $n = 5, 10, 100$ , and  $500$ . The 45-degree line is the identity function.

### 3.1.2. Cox Likelihood Ratio Statistic

**Theorem 3.2.** *If  $\mu$  is unknown, the bootstrap test of  $H_0 : \sigma^2 \leq 1$  vs.  $H_1 : \sigma^2 > 1$  based on the Cox likelihood ratio has size  $\alpha$  and is UMP for  $\chi_{n-1, \alpha}^2 > n$ . It rejects  $H_0$  w.p.1 for other values of  $\alpha$ .*

*Proof.* The Cox log-likelihood ratio statistic is  $M(\hat{\mu}, \hat{\mu}, \hat{\sigma}_0, \hat{\sigma}_1, \mathbf{X}_n)$ . Since

$$\hat{\sigma}_0^2 \hat{\sigma}_1^{-2} = \begin{cases} \hat{\sigma}^2, & \text{if } \hat{\sigma}^2 \leq 1 \\ \hat{\sigma}^{-2}, & \text{if } \hat{\sigma}^2 > 1 \end{cases}$$

and

$$\hat{\sigma}_0^{-2} - \hat{\sigma}_1^{-2} = \begin{cases} \hat{\sigma}^{-2} - 1, & \text{if } \hat{\sigma}^2 \leq 1 \\ 1 - \hat{\sigma}^{-2}, & \text{if } \hat{\sigma}^2 > 1 \end{cases}$$

we have

$$\begin{aligned} 2n^{-1}M(\hat{\mu}, \hat{\mu}, \hat{\sigma}_0, \hat{\sigma}_1, \mathbf{X}_n) &= \log(\hat{\sigma}_0^2 \hat{\sigma}_1^{-2}) + \hat{\sigma}^2(\hat{\sigma}_0^{-2} - \hat{\sigma}_1^{-2}) \\ &= \begin{cases} \log(\hat{\sigma}^2) - \hat{\sigma}^2 + 1, & \text{if } \hat{\sigma}^2 \leq 1 \\ -\log(\hat{\sigma}^2) + \hat{\sigma}^2 - 1, & \text{if } \hat{\sigma}^2 > 1 \end{cases} \end{aligned}$$

which is strictly increasing in  $\hat{\sigma}^2$ . Therefore rejecting for large values of the statistic is equivalent to rejecting for large values of  $\hat{\sigma}^2$ . Since the bootstrap null distribution of  $n\hat{\sigma}^2$  is  $\hat{\sigma}_0^2 \chi_{n-1}^2$ , the bootstrap critical point of  $\hat{\sigma}^2$  is  $n^{-1}\hat{\sigma}_0^2 \chi_{n-1, \alpha}^2$ . Thus the bootstrap test rejects  $H_0$  if  $\hat{\sigma}^2 \hat{\sigma}_0^{-2} > n^{-1}\chi_{n-1, \alpha}^2$ , or equivalently,  $\max(\hat{\sigma}^2, 1) > n^{-1}\chi_{n-1, \alpha}^2$ . If  $\alpha$  is so large that  $n^{-1}\chi_{n-1, \alpha}^2 \leq 1$ , the bootstrap test rejects  $H_0$  regardless of the data. On the other hand, if  $n^{-1}\chi_{n-1, \alpha}^2 > 1$ , the test rejects  $H_0$  if  $\hat{\sigma}^2 > n^{-1}\chi_{n-1, \alpha}^2$ , which coincides with the UMP test [10, p. 88].

3.2.  $H_0 : \sigma^2 \geq 1$  vs.  $H_1 : \sigma^2 < 1$

Next suppose we reverse the hypotheses and test  $H_0 : \sigma^2 \geq 1$  versus  $H_1 : \sigma^2 < 1$ . Then  $\hat{\sigma}_0^2 = \max(\hat{\sigma}^2, 1)$  and  $\hat{\sigma}_1^2 = \min(\hat{\sigma}^2, 1)$ .

3.2.1. Standard Likelihood Ratio Statistic

**Theorem 3.3.** For any  $0 < \alpha < 1$  and  $\mu$  unknown, the size of the bootstrap test for  $H_0 : \sigma^2 \geq 1$  vs.  $H_1 : \sigma^2 < 1$ , based on the standard likelihood ratio, is bounded below by

$$(3.2) \quad \min\{\alpha, \Psi_{n-1}(n)\} + E \left[ \frac{\alpha - \Psi_{n-1}(n^2/\chi_{n-1}^2)}{1 - \Psi_{n-1}(n^2/\chi_{n-1}^2)} I(\chi_{n-1}^2 \geq \max\{n, n^2/\chi_{n-1,1-\alpha}^2\}) \right].$$

*Proof.* Direct computation yields

$$2n^{-1}M(\hat{\mu}, \hat{\mu}, \hat{\sigma}_0, \hat{\sigma}, \mathbf{X}_n) = \begin{cases} \hat{\sigma}^2 - \log(\hat{\sigma}^2) - 1, & \text{if } \hat{\sigma}^2 \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Since the function  $x - \log(x) - 1$  decreases monotonically for  $0 < x \leq 1$ , the test rejects  $H_0$  for small values of  $S = n \min(\hat{\sigma}^2, 1)$ . Let  $S^*$  denote the bootstrap version of  $S$  under resampling from  $N(\bar{X}_n, \hat{\sigma}_0^2)$ .

1.  $\hat{\sigma}^2 < 1$ . Then  $\hat{\sigma}_0 = 1$  and the distribution of  $S^*$  is  $\chi_{n-1}^2$  right-truncated at  $n$ , with probability mass  $1 - \Psi_{n-1}(n)$  there. If  $0 < \alpha < \Psi_{n-1}(n)$ , the critical point of the bootstrap distribution is  $\chi_{n-1,1-\alpha}^2$ . Otherwise, the critical point is  $n$  and the test rejects w.p.1, because  $S < n$ .
2.  $\hat{\sigma}^2 \geq 1$ . Then  $\hat{\sigma}_0^2 = \hat{\sigma}^2$ ,  $S = n$  and the distribution of  $S^*$  is  $\hat{\sigma}^2 \chi_{n-1}^2$  right-truncated at  $n$ . The test does not reject  $H_0$  if  $\hat{\sigma}^2 \chi_{n-1,1-\alpha}^2 < n$ . Otherwise, if  $\hat{\sigma}^2 \chi_{n-1,1-\alpha}^2 \geq n$ , the test rejects  $H_0$  with probability  $\{\alpha - \Psi_{n-1}(n\hat{\sigma}^{-2})\}/\{1 - \Psi_{n-1}(n\hat{\sigma}^{-2})\}$ .

Since  $\alpha < \Psi_{n-1}(n)$  if and only if  $\chi_{n-1,1-\alpha}^2 < n$ ,

$$\begin{aligned} & P_{\mu,\sigma}(\text{Reject } H_0, \hat{\sigma}^2 < 1) \\ &= \begin{cases} P_{\mu,\sigma}(S < \chi_{n-1,1-\alpha}^2, n\hat{\sigma}^2 < n), & \text{if } 0 < \alpha < \Psi_{n-1}(n) \\ P_{\mu,\sigma}(n\hat{\sigma}^2 < n), & \text{otherwise} \end{cases} \\ &= \begin{cases} P_{\mu,\sigma}(n\hat{\sigma}^2 < \chi_{n-1,1-\alpha}^2, n\hat{\sigma}^2 < n), & \text{if } 0 < \alpha < \Psi_{n-1}(n) \\ P_{\mu,\sigma}(n\hat{\sigma}^2 < n), & \text{otherwise} \end{cases} \\ &= \begin{cases} P_{\mu,\sigma}(n\hat{\sigma}^2 < \chi_{n-1,1-\alpha}^2), & \text{if } 0 < \alpha < \Psi_{n-1}(n) \\ P_{\mu,\sigma}(n\hat{\sigma}^2 < n), & \text{otherwise} \end{cases} \\ &= P_{\mu,\sigma}(n\hat{\sigma}^2 < \min\{\chi_{n-1,1-\alpha}^2, n\}) \\ &= \Psi_{n-1}(\sigma^{-2} \min\{\chi_{n-1,1-\alpha}^2, n\}) \end{aligned}$$

and

$$\begin{aligned} & P_{\mu,\sigma}(\text{Reject } H_0, \hat{\sigma}^2 \geq 1) \\ &= P_{\mu,\sigma}(\text{Reject } H_0, \hat{\sigma}^2 \chi_{n-1,1-\alpha}^2 \geq n, \hat{\sigma}^2 \geq 1) \\ &= E_{\mu,\sigma} \left[ \frac{\alpha - \Psi_{n-1}(n\hat{\sigma}^{-2})}{1 - \Psi_{n-1}(n\hat{\sigma}^{-2})} I(\hat{\sigma}^2 \chi_{n-1,1-\alpha}^2 \geq n, n\hat{\sigma}^2 \geq n) \right] \\ &= E \left[ \frac{\alpha - \Psi_{n-1}(n^2\sigma^{-2}/\chi_{n-1}^2)}{1 - \Psi_{n-1}(n^2\sigma^{-2}/\chi_{n-1}^2)} I(\sigma^2 \chi_{n-1}^2 \geq \max\{n, n^2/\chi_{n-1,1-\alpha}^2\}) \right]. \end{aligned}$$

Therefore

$$\begin{aligned}
 & \sup_{H_0} P_{\mu, \sigma}(\text{Reject } H_0) \\
 & \geq \Psi_{n-1}(\min\{\chi_{n-1, 1-\alpha}^2, n\}) \\
 & \quad + E \left[ \frac{\alpha - \Psi_{n-1}(n^2/\chi_{n-1}^2)}{1 - \Psi_{n-1}(n^2/\chi_{n-1}^2)} I(\chi_{n-1}^2 \geq \max\{n, n^2/\chi_{n-1, 1-\alpha}^2\}) \right] \\
 & = \min\{\alpha, \Psi_{n-1}(n)\} \\
 & \quad + E \left[ \frac{\alpha - \Psi_{n-1}(n^2/\chi_{n-1}^2)}{1 - \Psi_{n-1}(n^2/\chi_{n-1}^2)} I(\chi_{n-1}^2 \geq \max\{n, n^2/\chi_{n-1, 1-\alpha}^2\}) \right].
 \end{aligned}$$

Figure 3 shows graphs of the lower bound (3.2) for  $n = 5, 10, 100,$  and  $500$ .

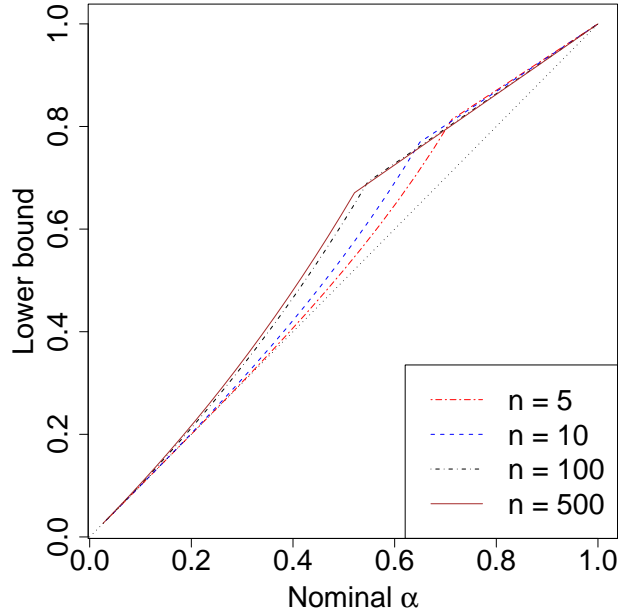


FIG 3. Lower bounds (3.2) on the size of the bootstrap test of  $H_0 : \sigma^2 \geq 1$  vs.  $H_1 : \sigma^2 < 1$  based on the standard likelihood ratio  $M_n^{(1)}$ . The 45-degree line is the identity function.

### 3.2.2. Cox Likelihood Ratio Statistic

**Theorem 3.4.** *The bootstrap test of  $H_0 : \sigma^2 \geq 1$  vs.  $H_1 : \sigma^2 < 1$  based on the Cox likelihood ratio has size  $\alpha$  and is UMP if  $\chi_{n-1, \alpha}^2 > n$ . Otherwise, it rejects  $H_0$  w.p.1.*

*Proof.* Since

$$\hat{\sigma}_0^2 \hat{\sigma}_1^{-2} = \begin{cases} \hat{\sigma}^{-2}, & \text{if } \hat{\sigma}^2 < 1 \\ \hat{\sigma}^2, & \text{if } \hat{\sigma}^2 \geq 1 \end{cases}$$

and

$$\hat{\sigma}_0^{-2} - \hat{\sigma}_1^{-2} = \begin{cases} 1 - \hat{\sigma}^{-2}, & \text{if } \hat{\sigma}^2 < 1 \\ \hat{\sigma}^{-2} - 1, & \text{if } \hat{\sigma}^2 \geq 1 \end{cases}$$

1 we have

$$\begin{aligned}
 2n^{-1}M(\hat{\mu}, \hat{\mu}, \hat{\sigma}_0, \hat{\sigma}_1, \mathbf{X}_n) &= \log(\hat{\sigma}_0^2 \hat{\sigma}_1^{-2}) + \hat{\sigma}^2(\hat{\sigma}_0^{-2} - \hat{\sigma}_1^{-2}) \\
 &= \begin{cases} -\log(\hat{\sigma}^2) + \hat{\sigma}^2 - 1, & \text{if } \hat{\sigma}^2 < 1 \\ \log(\hat{\sigma}^2) - \hat{\sigma}^2 + 1, & \text{if } \hat{\sigma}^2 \geq 1 \end{cases}
 \end{aligned}$$

2 a strictly decreasing function of  $\hat{\sigma}^2$ . Thus the test statistic rejects  $H_0$  for small  
 3 values of  $\hat{\sigma}^2$ . The bootstrap null distribution of  $\hat{\sigma}^2$  is  $n^{-1}\hat{\sigma}_0^2\chi_{n-1}^2$ , with critical value  
 4  $n^{-1}\hat{\sigma}_0^2\chi_{n-1,1-\alpha}^2$ . Hence the bootstrap test rejects  $H_0$  if  $\hat{\sigma}^2 \hat{\sigma}_0^{-2} < n^{-1}\chi_{n-1,1-\alpha}^2$ . But  
 5 the left side of the inequality is never greater than 1, because

$$\hat{\sigma}^2 \hat{\sigma}_0^{-2} = \begin{cases} \hat{\sigma}^2, & \text{if } \hat{\sigma}^2 < 1 \\ 1, & \text{otherwise.} \end{cases}$$

6 Therefore, if  $\alpha$  is so large that  $n^{-1}\chi_{n-1,1-\alpha}^2 \geq 1$ , the bootstrap test rejects  $H_0$   
 7 w.p.1. Otherwise, if  $n^{-1}\chi_{n-1,1-\alpha}^2 < 1$ , the test rejects  $H_0$  if  $\hat{\sigma}^2 < n^{-1}\chi_{n-1,1-\alpha}^2$ ,  
 8 which coincides with the classical UMP unbiased test [10, pp. 154].

9 **4. Testing Difference of Two Normal Means**

10 Let  $X_1, \dots, X_m$  and  $Y_1, \dots, Y_n$  be independent random samples from  $N(\mu, \sigma^2)$  and  
 11  $N(\eta, \tau^2)$ , respectively, and  $N = m + n > 2$ . We want to test

$$(4.1) \quad H_0 : \eta \leq \mu \quad \text{vs.} \quad H_1 : \eta > \mu.$$

12 The likelihood function for this case is

$$\begin{aligned}
 L(\mu, \tau) &= (2\pi)^{-(m+n)/2} \sigma^{-m} \tau^{-n} \exp \left\{ -(2\sigma^2)^{-1} \sum (X_i - \mu)^2 - (2\tau^2)^{-1} \sum (Y_j - \eta)^2 \right\} \\
 &= (2\pi)^{-(m+n)/2} \sigma^{-m} \tau^{-n} \exp \left\{ -(2\sigma^2)^{-1} \sum (X_i - \bar{X}_m)^2 - (2\tau^2)^{-1} \sum (Y_j - \bar{Y}_n)^2 \right. \\
 &\quad \left. - m(2\sigma^2)^{-1} (\mu - \bar{X}_m)^2 - n(2\tau^2)^{-1} (\eta - \bar{Y}_n)^2 \right\}
 \end{aligned}$$

13 and the unrestricted MLE of  $(\mu, \eta)$  is  $(\hat{\mu}, \hat{\eta}) = (\bar{X}_m, \bar{Y}_n)$ .

14 **4.1. Known Variances**

15 Let  $V = (m\tau^2\bar{X}_m + n\sigma^2\bar{Y}_n)/(m\tau^2 + n\sigma^2)$ . The MLE of  $(\mu, \eta)$  under  $H_0$  is

$$(\hat{\mu}_0, \hat{\eta}_0) = \begin{cases} (\bar{X}_m, \bar{Y}_n), & \bar{Y}_n \leq \bar{X}_m \\ (V, V), & \bar{Y}_n > \bar{X}_m \end{cases}$$

16 and that under  $H_1$  is

$$(\hat{\mu}_1, \hat{\eta}_1) = \begin{cases} (V, V), & \bar{Y}_n > \bar{X}_m \\ (\bar{X}_m, \bar{Y}_n), & \bar{Y}_n \leq \bar{X}_m. \end{cases}$$

17 **4.1.1. Difference of Means Statistic**

18 **Theorem 4.1.** *The size of the bootstrap test of (4.1) based on  $\bar{Y}_n - \bar{X}_m$  is  $\alpha$  if*  
 19  $\alpha < 1/2$  *and is 1 if  $\alpha \geq 1/2$ .*

*Proof.* Let  $S = \bar{Y}_n - \bar{X}_m$ . The bootstrap test statistic  $S^* = \bar{Y}_n^* - \bar{X}_m^*$  has a normal distribution with mean  $\hat{\eta}_0 - \hat{\mu}_0 = SI(S < 0)$  and variance  $m^{-1}\sigma^2 + n^{-1}\tau^2$ . Thus its nominal level- $\alpha$  bootstrap critical value is  $SI(S < 0) + z_\alpha\{m^{-1}\sigma^2 + n^{-1}\tau^2\}^{1/2}$  and the rejection region is  $\max(S, 0) > z_\alpha\{m^{-1}\sigma^2 + n^{-1}\tau^2\}^{1/2}$ . Clearly, the size of the test is attained at the boundary  $\mu = \eta$ . If  $\alpha < 1/2$ , the probability of rejecting  $H_0$  when  $\mu = \eta$  is exactly  $\alpha$ . On the other hand, if  $\alpha \geq 1/2$ , then  $z_\alpha \leq 0$  and the test rejects  $H_0$  w.p.1.

#### 4.1.2. Standard Likelihood Ratio Statistic

**Theorem 4.2.** *The size of the bootstrap test of (4.1) based on the standard likelihood ratio is  $\min(2\alpha, 1) + (1 - \alpha) \log\{1 - \min(\alpha, 1/2)\} > \alpha$ .*

*Proof.* The log-likelihood ratio statistic is

$$\begin{aligned} & \log\{L(\hat{\mu}, \hat{\tau})/L(\hat{\mu}_0, \hat{\tau}_0)\} \\ &= \{m(2\sigma^2)^{-1}(V - \bar{X}_m)^2 + n(2\tau^2)^{-1}(V - \bar{Y}_n)^2\}I(\bar{Y}_n > \bar{X}_m) \\ &= mn(m\tau^2 + n\sigma^2)^{-2}(\bar{Y}_n - \bar{X}_m)^2 I(\bar{Y}_n > \bar{X}_m). \end{aligned}$$

Thus the test statistic is equivalent to  $S = (\bar{Y}_n - \bar{X}_m)I(\bar{Y}_n > \bar{X}_m)$ . The bootstrap distribution of  $S^*$  is normal with mean  $\hat{\eta}_0 - \hat{\mu}_0$  and variance  $n^{-1}\tau^2 + m^{-1}\sigma^2$ , left-truncated at 0 with  $P(S^* = 0) = \Phi\{(\hat{\mu}_0 - \hat{\eta}_0)(n^{-1}\tau^2 + m^{-1}\sigma^2)^{-1/2}\}$ . Let  $\delta = (\mu - \eta)(n^{-1}\tau^2 + m^{-1}\sigma^2)^{-1/2}$  and  $W = (\bar{X}_m - \bar{Y}_n)(n^{-1}\tau^2 + m^{-1}\sigma^2)^{-1/2} \sim N(\delta, 1)$ . We consider two cases.

1.  $\bar{Y}_n \leq \bar{X}_m$ . Then  $S = 0$ ,  $\hat{\eta}_0 - \hat{\mu}_0 = \bar{Y}_n - \bar{X}_m$ , and  $\Phi(W) \geq 1/2$ . If  $1 - \Phi(W) < \alpha$ , the test is randomized and rejects  $H_0$  with probability  $\{\alpha - 1 + \Phi(W)\}/\Phi(W)$ . Otherwise, if  $1 - \Phi(W) \geq \alpha$ , the test does not reject  $H_0$ .
2.  $\bar{Y}_n > \bar{X}_m$ . Then  $S = \bar{Y}_n - \bar{X}_m > 0$ ,  $\hat{\eta}_0 - \hat{\mu}_0 = 0$ , and  $P(S^* = 0) = 1/2$ . If  $\alpha < 1/2$ , then the test rejects  $H_0$  if  $\bar{Y}_n - \bar{X}_m > z_\alpha(n^{-1}\tau^2 + m^{-1}\sigma^2)^{-1/2}$ , i.e.,  $W < -z_\alpha$ . Otherwise, if  $\alpha \geq 1/2$ , then the critical value is 0 and the test rejects w.p.1.

Therefore

$$\begin{aligned} & P(\text{Reject } H_0) \\ &= P(\text{Reject } H_0, \bar{Y}_n \leq \bar{X}_m, 1 - \Phi(W) < \alpha) \\ &\quad + P(\text{Reject } H_0, \bar{Y}_n > \bar{X}_m)I(\alpha < 1/2) + P(\text{Reject } H_0, \bar{Y}_n > \bar{X}_m)I(\alpha \geq 1/2) \\ &= E[\Phi(W)^{-1}\{\alpha - 1 + \Phi(W)\}I(W > z_\alpha^+)] \\ &\quad + P(W < -z_\alpha)I(\alpha < 1/2) + P(W < 0)I(\alpha > 1/2) \\ &= E[\Phi(W)^{-1}\{\alpha - 1 + \Phi(W)\}I(W > z_\alpha^+)] + P(W < -z_\alpha^+) \end{aligned}$$

and the result follows from Lemma 2.1.

#### 4.1.3. Cox Likelihood Ratio Statistic

**Theorem 4.3.** *The bootstrap test of (4.1) based on the Cox likelihood ratio statistic is the same as that based on the difference of sample means; its size is  $\alpha$  if  $\alpha < 1/2$  and is 1 if  $\alpha \geq 1/2$ .*

*Proof.* The Cox log-likelihood ratio statistic is

$$\log \left\{ \frac{L(\hat{\mu}_1, \hat{\tau}_1)}{L(\hat{\mu}_0, \hat{\tau}_0)} \right\} = \frac{mn(\bar{Y}_n - \bar{X}_m)^2}{2(m\tau^2 + n\sigma^2)} \{I(\bar{Y}_n > \bar{X}_m) - I(\bar{Y}_n \leq \bar{X}_m)\}.$$

Thus the test statistic is equivalent to  $S = \bar{Y}_n - \bar{X}_m$  and the result follows from Theorem 4.1.

#### 4.2. Unknown but Equal Variances

Suppose that  $\tau^2 = \sigma^2$  but their value is unknown. Then the likelihood function is

$$L(\mu, \tau, \sigma) = (2\pi\sigma^2)^{-N/2} \exp \left[ -(2\sigma^2)^{-1} \left\{ \sum_i (X_i - \mu)^2 + \sum_j (Y_j - \eta)^2 \right\} \right]$$

giving the unrestricted MLE

$$(\hat{\mu}, \hat{\eta}, \hat{\sigma}^2) = \left( \bar{X}_m, \bar{Y}_n, N^{-1} \left\{ \sum_i (X_i - \bar{X}_m)^2 + \sum_j (Y_j - \bar{Y}_n)^2 \right\} \right).$$

Let  $V = N^{-1}(m\bar{X}_m + n\bar{Y}_n)$  and

$$\begin{aligned} \tilde{\sigma}^2 &= N^{-1} \left\{ \sum_i (X_i - V)^2 + \sum_j (Y_j - V)^2 \right\} \\ &= \hat{\sigma}^2 + mnN^{-2}(\bar{Y}_n - \bar{X}_m)^2. \end{aligned}$$

The corresponding MLEs under  $H_0$  and  $H_1$  are, respectively,

$$\begin{aligned} (\hat{\mu}_0, \hat{\eta}_0, \hat{\sigma}_0^2) &= \begin{cases} (\bar{X}_m, \bar{Y}_n, \hat{\sigma}^2), & \text{if } \bar{Y}_n < \bar{X}_m \\ (V, V, \tilde{\sigma}^2), & \text{if } \bar{Y}_n \geq \bar{X}_m \end{cases} \\ (\hat{\mu}_1, \hat{\eta}_1, \hat{\sigma}_1^2) &= \begin{cases} (V, V, \tilde{\sigma}^2), & \text{if } \bar{Y}_n < \bar{X}_m \\ (\bar{X}_m, \bar{Y}_n, \hat{\sigma}^2), & \text{if } \bar{Y}_n \geq \bar{X}_m. \end{cases} \end{aligned}$$

##### 4.2.1. Difference of Means Statistic

Suppose  $S = \bar{Y}_n - \bar{X}_m$ . Then  $S^* = \bar{Y}_n^* - \bar{X}_m^*$  has a normal distribution with mean  $\hat{\eta}_0 - \hat{\mu}_0$  and variance  $N(mn)^{-1}\hat{\sigma}_0^2$ . Let  $\Upsilon_\nu$  denote the  $t$  distribution function with  $\nu$  degrees of freedom and let  $s^2 = \hat{\sigma}^2 N(N-2)^{-1}$  be the usual unbiased estimate of  $\sigma^2$ .

**Theorem 4.4.** *The size of the bootstrap test of (4.1) based on  $\bar{Y}_n - \bar{X}_m$  is*

$$\begin{aligned} &\sup_{H_0} P(\text{Reject } H_0) \\ (4.2) \quad &= \begin{cases} 0, & \text{if } \alpha \leq 1 - \Phi(\sqrt{N}) \\ 1 - \Upsilon_{N-2} \left( z_\alpha \sqrt{\frac{N-2}{N-z_\alpha^2}} \right), & \text{if } 1 - \Phi(\sqrt{N}) < \alpha < 1/2 \\ 1, & \text{if } \alpha \geq 1/2 \end{cases} \\ &\rightarrow \begin{cases} \alpha, & \text{if } \alpha < 1/2 \\ 1, & \text{if } \alpha \geq 1/2 \end{cases} \end{aligned}$$

as  $N \rightarrow \infty$ .

*Proof.* The hypothesis  $H_0$  is rejected if

$$\begin{aligned} \bar{Y}_n - \bar{X}_m &> \hat{\eta}_0 - \hat{\mu}_0 + z_\alpha \hat{\sigma}_0 \sqrt{N/(mn)} \\ &= \begin{cases} \bar{Y}_n - \bar{X}_m + z_\alpha \hat{\sigma} \sqrt{N/(mn)}, & \text{if } \bar{Y}_n < \bar{X}_m \\ z_\alpha \tilde{\sigma} \sqrt{N/(mn)}, & \text{if } \bar{Y}_n \geq \bar{X}_m. \end{cases} \end{aligned}$$

1.  $\alpha < 1/2$ . If  $\bar{Y}_n < \bar{X}_m$ , the test does not reject  $H_0$ . Otherwise, if  $\bar{Y}_n \geq \bar{X}_m$ , the test rejects  $H_0$  if

$$\begin{aligned} (\bar{Y}_n - \bar{X}_m)^2 &> z_\alpha^2 \tilde{\sigma}^2 N/(mn) \\ \iff (\bar{Y}_n - \bar{X}_m)^2 (1 - N^{-1} z_\alpha^2) &> N z_\alpha^2 \hat{\sigma}^2 / (mn) \\ \iff \alpha > 1 - \Phi(\sqrt{N}) \quad \text{and} \quad \sqrt{\frac{mn}{N}} \frac{\bar{Y}_n - \bar{X}_m}{s} &> \sqrt{\frac{N-2}{N-z_\alpha^2}}. \end{aligned}$$

Therefore if  $\alpha \leq 1 - \Phi(\sqrt{N})$ , the test does not reject  $H_0$ . Otherwise, if  $1 - \Phi(\sqrt{N}) < \alpha < 1/2$ , the rejection probability is maximized when  $\eta = \mu$  at  $1 - \Upsilon_{N-2} \left( z_\alpha \sqrt{(N-2)/(N-z_\alpha^2)} \right)$ .

2.  $\alpha \geq 1/2$ . The test rejects  $H_0$  w.p.1 because  $z_\alpha < 0$ .

Hence the result (4.2). The limit is due to  $\Upsilon_\nu(x) \rightarrow \Phi(x)$  as  $\nu \rightarrow \infty$ , for every  $x$ . Figure 4 plots the size function (4.2) for  $N = 3, 5, 10$ , and 100.

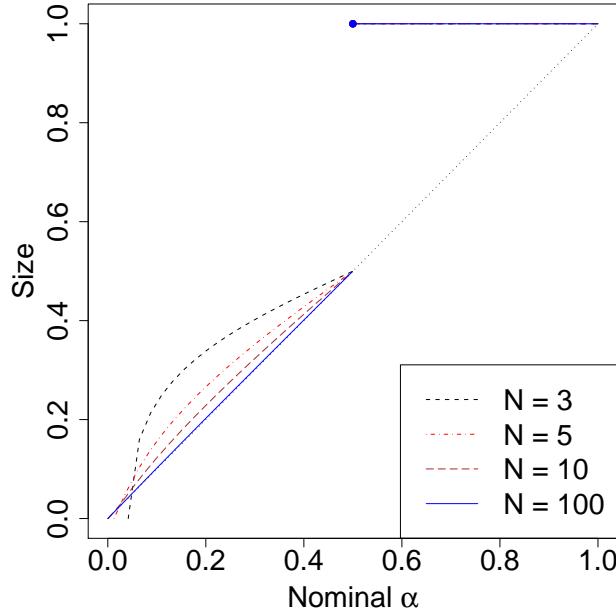


FIG 4. Size of bootstrap test for a difference of normal means based on the difference of sample means, for equal but unknown variances (4.2). The 45-degree line is the identity function.

4.2.2. Standard Likelihood Ratio Statistic

The log-likelihood ratio statistic is

$$(N/2) \log(\hat{\sigma}^2/\hat{\sigma}^2) I(\bar{Y}_n > \bar{X}_m) = (N/2) \log\{1+mnN^{-2}(\bar{Y}_n-\bar{X}_m)^2\hat{\sigma}^{-2}\} I(\bar{Y}_n > \bar{X}_m)$$

which is equivalent to the positive part of the  $t$ -statistic:

$$S = \sqrt{mn/N}s^{-1}(\bar{Y}_n - \bar{X}_m)^+.$$

**Theorem 4.5.** *The size of the bootstrap test of (4.1) based on the standard likelihood ratio is bounded below by*

$$\begin{aligned} \min(\alpha, 1/2) + E \left[ \frac{\alpha - 1 + \Phi\left(\sqrt{N/(N-2)}t_{N-2}\right)}{\Phi\left(\sqrt{N/(N-2)}t_{N-2}\right)} I\left(\sqrt{N/(N-2)}t_{N-2} \geq z_\alpha^+\right) \right] \\ \rightarrow \min(2\alpha, 1) + (1 - \alpha) \log\{1 - \min(\alpha, 1/2)\}, \quad N \rightarrow \infty. \end{aligned}$$

*Proof.* We again consider two situations.

1.  $\bar{Y}_n > \bar{X}_m$ . The bootstrap distribution of  $S^*$  is a  $t_{N-2}$  distribution left-truncated at 0 with probability 1/2. If  $\alpha < 1/2$ , then the test rejects  $H_0$  if  $S > t_{N-2,\alpha}$ . Otherwise, if  $\alpha \geq 1/2$ , the test rejects  $H_0$  w.p.1.
2.  $\bar{Y}_n \leq \bar{X}_m$ . The bootstrap distribution of  $S^*$  consists of the positive part of a noncentral  $t$  with  $N - 2$  degrees of freedom and noncentrality parameter  $\delta = \sqrt{mn/N}\hat{\sigma}^{-1}(\bar{Y}_n - \bar{X}_m) = S\sqrt{N/(N-2)}$  and probability at 0 equal to  $\Phi(-\delta)$ . Since  $S = 0$ , the test does not reject  $H_0$  if  $\alpha < 1 - \Phi(-\delta) \iff -\delta < z_\alpha$ . Otherwise, if  $-\delta \geq z_\alpha$ , then the test is randomized, rejecting  $H_0$  with probability  $\{\alpha - 1 + \Phi(-\delta)\}/\Phi(-\delta)$ .

Thus

$$\begin{aligned} P(\text{Reject } H_0) &= P(S > t_{N-2,\alpha}, \bar{Y}_n > \bar{X}_m) I(\alpha < 1/2) + P(\bar{Y}_n > \bar{X}_m) I(\alpha \geq 1/2) \\ &\quad + P(\text{Reject } H_0, -\delta \geq z_\alpha, \bar{Y}_n \leq \bar{X}_m) \\ &= P(S > t_{N-2,\alpha}) I(\alpha < 1/2) + (1/2)I(\alpha \geq 1/2) \\ &\quad + E \left\{ \frac{\alpha - 1 + \Phi(-\delta)}{\Phi(-\delta)} I(-\delta \geq z_\alpha, \bar{Y}_n \leq \bar{X}_m) \right\} \\ &= \min(\alpha, 1/2) + E \left\{ \frac{\alpha - 1 + \Phi(-\delta)}{\Phi(-\delta)} I(-\delta \geq z_\alpha, \bar{Y}_n \leq \bar{X}_m) \right\} \\ &= \min(\alpha, 1/2) + E \left\{ \frac{\alpha - 1 + \Phi(-\delta)}{\Phi(-\delta)} I(-\delta \geq z_\alpha^+) \right\}. \end{aligned}$$

Evaluating this probability at  $\mu = \eta$  yields

$$\begin{aligned} \sup_{H_0} P(\text{Reject } H_0) &\geq \min(\alpha, 1/2) \\ &\quad + E \left[ \frac{\alpha - 1 + \Phi\left(\sqrt{N/(N-2)}t_{N-2}\right)}{\Phi\left(\sqrt{N/(N-2)}t_{N-2}\right)} I\left(\sqrt{N/(N-2)}t_{N-2} \geq z_\alpha^+\right) \right] \\ &\rightarrow \min(2\alpha, 1) + (1 - \alpha) \log\{1 - \min(\alpha, 1/2)\} \end{aligned}$$

as  $N \rightarrow \infty$  by Lemma 2.1.

### 4.2.3. Cox Likelihood Ratio Statistic

**Theorem 4.6.** *The size of the bootstrap test of (4.1) based on the Cox likelihood ratio or the ordinary  $t$ -statistic is*

$$P(t_{N-2, t_{N-2}\sqrt{N/(N-2)}, \alpha} < t_{N-2} < 0) + P(t_{N-2} > t_{N-2, \alpha}^+) \geq \alpha.$$

*Proof.* The Cox log-likelihood ratio simplifies to

$$(N/2) \log\{1 + mn(\bar{Y}_n - \bar{X}_m)^2 N^{-2} \hat{\sigma}^{-2}\} \{I(\bar{Y}_n \geq \bar{X}_m) - I(\bar{Y}_n < \bar{X}_m)\}$$

which is an increasing function of the Student  $t$  statistic  $S = \sqrt{mn/N} \hat{\sigma}^{-1} (\bar{Y}_n - \bar{X}_m)$ . The bootstrap distribution of  $S$  is a noncentral  $t_{N-2, \delta}$  with  $N-2$  degrees of freedom and noncentrality parameter

$$\begin{aligned} \delta &= \sqrt{mn/N} \hat{\sigma}_0^{-1} (\hat{\eta}_0 - \hat{\mu}_0) \\ &= \begin{cases} \sqrt{mn/N} \hat{\sigma}^{-1} (\bar{Y}_n - \bar{X}_m), & \text{if } \bar{Y}_n < \bar{X}_m \\ 0, & \text{if } \bar{Y}_n \geq \bar{X}_m \end{cases} \\ &= \begin{cases} \sqrt{N/(N-2)} S, & \text{if } \bar{Y}_n < \bar{X}_m \\ 0, & \text{if } \bar{Y}_n \geq \bar{X}_m. \end{cases} \end{aligned}$$

Therefore

$$\begin{aligned} P(\text{Reject } H_0) &= P(S > t_{N-2, \delta, \alpha}, \bar{Y}_n < \bar{X}_m) + P(S > t_{N-2, \alpha}, \bar{Y}_n \geq \bar{X}_m) \\ &= P(t_{N-2, \delta, \alpha} < S < 0) + P(S > t_{N-2, \alpha}^+). \end{aligned}$$

Evaluating the probabilities at  $\mu = \eta$  yields

$$\begin{aligned} \sup_{H_0} P(\text{Reject } H_0) &\geq P(t_{N-2, t_{N-2}\sqrt{N/(N-2)}, \alpha} < t_{N-2} < 0) + P(t_{N-2} > t_{N-2, \alpha}^+) \\ &\geq P(t_{N-2, 0, \alpha} < t_{N-2} < 0) + \min(\alpha, 1/2) \\ &= (\alpha - 1/2) I(\alpha > 1/2) + \min(\alpha, 1/2) \\ &= \alpha. \end{aligned}$$

## 5. Testing an Exponential Location Parameter

Let  $\text{Exp}(\theta, \tau)$  denote the distribution with density  $\tau^{-1} \exp\{-\tau^{-1}(x - \theta)\}$ ,  $x \geq \theta$ . We consider testing hypotheses about  $\theta$  with  $\tau = 1$ . The likelihood for a sample  $X_1, \dots, X_n$  from an  $\text{Exp}(\theta, 1)$  distribution is  $\prod \exp\{-(x_i - \theta)\} I(x_{(1)} \geq \theta)$ , where  $x_{(1)}$  is the smallest order statistic. The unconstrained MLE is  $\hat{\theta} = X_{(1)}$ .

### 5.1. $H_0 : \theta \leq 0$ vs. $H_1 : \theta > 0$

The MLE of  $\theta$  is  $\hat{\theta}_0 = \min(X_{(1)}, 0)$  and  $\hat{\theta}_1 = \max(X_{(1)}, 0)$  under  $H_0$  and  $H_1$ , respectively. Given  $X_{(1)}$ , the bootstrap data are independent observations from an  $\text{Exp}(\hat{\theta}_0, 1)$  distribution.

5.1.1. Standard Likelihood Ratio Statistic

The standard log-likelihood ratio statistic is

$$\begin{aligned}
 S &= \sum_{i=1}^n \log \left[ \frac{\exp\{-(X_i - \hat{\theta})\} I(X_{(1)} \geq \hat{\theta})}{\exp\{-(X_i - \hat{\theta}_0)\} I(X_{(1)} \geq \hat{\theta}_0)} \right] \\
 &= n(\hat{\theta} - \hat{\theta}_0) \\
 &= \begin{cases} 0, & X_{(1)} \leq 0 \\ nX_{(1)}, & X_{(1)} \geq 0. \end{cases}
 \end{aligned}$$

Given  $\hat{\theta}_0$ , the bootstrap distribution of  $S$  is  $\text{Exp}(n\hat{\theta}_0, 1)$ , left-truncated at 0 with probability mass  $1 - \exp(n\hat{\theta}_0)$  there.

1.  $X_{(1)} \geq 0$ . Then  $S = nX_{(1)}$ ,  $\hat{\theta}_0 = 0$ , the distribution of  $S^*$  is  $\text{Exp}(0, 1)$  with upper- $\alpha$  critical point  $\log(1/\alpha)$ , and the test rejects  $H_0$  if  $nX_{(1)} > -\log \alpha$ .
2.  $X_{(1)} \leq 0$ . Then  $S = 0$ ,  $\hat{\theta}_0 = X_{(1)}$ , and the distribution of  $S^*$  is  $\text{Exp}(nX_{(1)}, 1)$ , left-truncated at 0 with probability  $1 - \exp(nX_{(1)})$  there. If  $\alpha < \exp(nX_{(1)})$ , the test never rejects  $H_0$ . Otherwise, the test rejects  $H_0$  with probability  $\{\alpha - \exp(nX_{(1)})\} / \{1 - \exp(nX_{(1)})\}$ .

Since  $nX_{(1)}$  has an  $\text{Exp}(n\theta, 1)$  distribution,

$$\begin{aligned}
 P_\theta\{\text{Reject } H_0\} &= P_\theta\{\text{Reject } H_0, X_{(1)} \geq 0\} + P_\theta\{\text{Reject } H_0, X_{(1)} < 0\} \\
 &= P_\theta(nX_{(1)} > -\log \alpha, X_{(1)} \geq 0) \\
 &\quad + P_\theta\{\text{Reject } H_0, X_{(1)} < 0, \exp(n\theta) < \exp(nX_{(1)}) \leq \alpha\} \\
 &= P_\theta(nX_{(1)} > -\log \alpha) + E_\theta \left[ \frac{\alpha - \exp(nX_{(1)})}{1 - \exp(nX_{(1)})} I(n\theta < nX_{(1)} \leq \log \alpha) \right] \\
 &= \begin{cases} \alpha \exp(n\theta), & n\theta \geq \log \alpha \\ \alpha \exp(n\theta) + \int_{n\theta}^{\log \alpha} \{\alpha - \exp(y)\} \exp(n\theta - y) / \{1 - \exp(y)\} dy, & n\theta \leq \log \alpha. \end{cases}
 \end{aligned}$$

Now for  $n\theta < \log \alpha$ ,

$$\begin{aligned}
 &\int_{n\theta}^{\log \alpha} \frac{\alpha - \exp(y)}{1 - \exp(y)} \exp\{-(y - n\theta)\} dy \\
 &= \exp(n\theta) \int_{\exp(n\theta)}^{\alpha} \frac{\alpha - z}{z^2(1 - z)} dz \\
 &= \exp(n\theta) \int_{\exp(n\theta)}^{\alpha} [\alpha z^{-2} - (1 - \alpha)\{z^{-1} + (1 - z)^{-1}\}] dz \\
 &= \exp(n\theta) [-\alpha z^{-1} + (1 - \alpha)\{\log(1 - z) - \log z\}]_{\exp(n\theta)}^{\alpha} \\
 &= \exp(n\theta)[(1 - \alpha) \log(\alpha^{-1} - 1) - 1 + \alpha \exp(-n\theta) - (1 - \alpha) \log\{\exp(-n\theta) - 1\}].
 \end{aligned}$$

Therefore

$$P_\theta\{\text{Reject } H_0\} = \begin{cases} \alpha \exp(n\theta), & n\theta \geq \log \alpha \\ g_\alpha(\exp(n\theta)), & n\theta \leq \log \alpha \end{cases}$$

where

$$g_\alpha(z) = \alpha + z(1 - \alpha)[\log(\alpha^{-1} - 1) - \log(z^{-1} - 1) - 1], \quad 0 < z < \alpha.$$

Since  $\lim_{z \rightarrow 0} g_\alpha(z) = \alpha$ ,  $\lim_{z \rightarrow \alpha} g_\alpha(z) = \alpha^2$ , and  $g''_\alpha(z) > 0$  for  $0 < z < \alpha$ , we conclude that  $\sup_{H_0} P_\theta\{\text{Reject } H_0\} = \lim_{\theta \rightarrow -\infty} g_\alpha(\exp(n\theta)) = \alpha$ .

### 5.1.2. Cox Likelihood Ratio Statistic

The Cox log-likelihood ratio statistic is

$$\begin{aligned} S &= \sum_{i=1}^n \log \left[ \frac{\exp\{-(X_i - \hat{\theta}_1)\} I(X_{(1)} \geq \hat{\theta}_1)}{\exp\{-(X_i - \hat{\theta}_0)\} I(X_{(1)} \geq \hat{\theta}_0)} \right] \\ &= \begin{cases} -\infty, & X_{(1)} < 0 \\ nX_{(1)}, & X_{(1)} \geq 0. \end{cases} \end{aligned}$$

It follows that the bootstrap test behaves the same as that based on the standard likelihood ratio. We therefore have the following theorem.

**Theorem 5.1.** *For testing  $H_0 : \theta \leq 0$  vs.  $H_1 : \theta > 0$  for a sample from an  $\text{Exp}(\theta, 1)$  distribution, the bootstrap tests based on the standard and Cox likelihood ratios have size  $\alpha$ .*

### 5.2. $H_0 : \theta \geq 0$ vs. $H_1 : \theta < 0$

The MLEs under  $H_0$  and  $H_1$  are  $\hat{\theta}_0 = \max(X_{(1)}, 0)$  and  $\hat{\theta}_1 = \min(X_{(1)}, 0)$ , respectively.

#### 5.2.1. Standard Likelihood Ratio Statistic

**Theorem 5.2.** *The bootstrap test of  $H_0 : \theta \geq 0$  vs.  $H_1 : \theta < 0$  based on the standard likelihood ratio test is completely randomized.*

*Proof.* The standard log-likelihood ratio statistic is

$$\begin{aligned} S &= \sum_{i=1}^n \log \left[ \frac{\exp\{-(X_i - \hat{\theta})\} I(X_{(1)} \geq \hat{\theta})}{\exp\{-(X_i - \hat{\theta}_0)\} I(X_{(1)} \geq \hat{\theta}_0)} \right] \\ &= \begin{cases} \infty, & X_{(1)} < 0 \\ 0, & X_{(1)} \geq 0. \end{cases} \end{aligned}$$

Since  $\hat{\theta}_0 \geq 0$ , the distribution of  $S^*$  is degenerate at 0. On the other hand,  $S = 0$  w.p.1 under  $H_0$ . Therefore the bootstrap test based on  $S$  is completely randomized.

#### 5.2.2. Cox Likelihood Ratio Statistic

**Theorem 5.3.** *The bootstrap test of  $H_0 : \theta \geq 0$  vs.  $H_1 : \theta < 0$  based on the Cox likelihood ratio test rejects  $H_0$  w.p.1 for any  $0 < \alpha < 1$ .*

*Proof.* The Cox log-likelihood ratio statistic is

$$\begin{aligned} S &= \sum_{i=1}^n \log \left[ \frac{\exp\{-(X_i - \hat{\theta}_1)\} I(X_{(1)} \geq \hat{\theta}_1)}{\exp\{-(X_i - \hat{\theta}_0)\} I(X_{(1)} \geq \hat{\theta}_0)} \right] \\ &= \begin{cases} \infty, & X_{(1)} < 0 \\ n(\hat{\theta}_1 - \hat{\theta}_0), & X_{(1)} \geq 0 \end{cases} \\ &= \begin{cases} \infty, & X_{(1)} < 0 \\ -nX_{(1)}, & X_{(1)} \geq 0. \end{cases} \end{aligned}$$

1.  $X_{(1)} < 0$ . Then  $\hat{\theta}_0 = 0$ , the bootstrap data have an  $\text{Exp}(0, 1)$  distribution, and the distribution of  $S^*$  is the negative of an  $\text{Exp}(0, 1)$  distribution. Since  $S = \infty$ , the test rejects  $H_0$  w.p.1 for any  $0 < \alpha < 1$ .
2.  $X_{(1)} \geq 0$ . Then  $\hat{\theta}_0 > 0$ , and the bootstrap data have an  $\text{Exp}(X_{(1)}, 1)$  distribution. The distribution of  $S^*$  is the negative of an  $\text{Exp}(nX_{(1)}, 1)$  distribution, with support  $(-\infty, -nX_{(1)})$ . Since  $S = -nX_{(1)}$ , the test rejects  $H_0$  w.p.1 for any  $0 < \alpha < 1$ .

## 6. Conclusion

The results show that the size of the bootstrap test of hypotheses is unpredictable. It depends on the problem as well as the choice of test statistic. For example, in the case of testing a normal mean with known variance, the test based on the sample mean or the Cox likelihood ratio is UMP for  $0 < \alpha \leq 1/2$ , but it is sub-optimal when it is based on the standard likelihood ratio. On the other hand, if  $\alpha > 1/2$ , the test often rejects  $H_0$  w.p.1. The overall conclusion is that the size of the test is typically larger than its nominal level. This may explain the high power that the test is found to possess in simulation experiments.

## Appendix

### *Proof of Lemma 2.1.*

First note that

$$(6.1) \quad \phi(x) - \phi(x - \theta) \begin{cases} > 0, & \text{if } x < \theta/2 \\ < 0, & \text{if } x > \theta/2. \end{cases}$$

Let  $f(\theta)$  denote the function (2.2). We consider two cases.

1.  $\alpha \geq 1/2$ . Since  $z_\alpha^+ = 0$ , we have  $f(\theta) = 1 - (1 - \alpha) \int_0^\infty \phi(x - \theta)/\Phi(x) dx$  and

$$\begin{aligned} \frac{f(0) - f(\theta)}{1 - \alpha} &= \int_0^\infty \frac{\phi(x - \theta) - \phi(x)}{\Phi(x)} dx \\ &= \int_0^{\theta/2} \frac{\phi(x - \theta) - \phi(x)}{\Phi(x)} dx + \int_{\theta/2}^\infty \frac{\phi(x - \theta) - \phi(x)}{\Phi(x)} dx \\ &\geq 2 \int_0^{\theta/2} \{\phi(x - \theta) - \phi(x)\} dx + \int_{\theta/2}^\infty \{\phi(x - \theta) - \phi(x)\} dx \\ &= 2\{\Phi(-\theta/2) - \Phi(-\theta) - \Phi(\theta/2) + 1/2\} - \Phi(-\theta/2) + \Phi(\theta/2) \\ &= 2\{\Phi(\theta) - \Phi(\theta/2)\} \\ &\geq 0 \end{aligned}$$

where we use (6.1) in the first inequality. Hence

$$\begin{aligned} f(\theta) &\leq f(0) \\ &= 1 - (1 - \alpha) \int_0^\infty \phi(x)/\Phi(x) dx \\ &= 1 + (1 - \alpha) \log(1/2). \end{aligned}$$

2.  $\alpha < 1/2$ . Write  $f(\theta) = J_1(\theta) + J_2(\theta)$ , where

$$\begin{aligned} J_1(\theta) &= \alpha(1-\alpha)^{-1}\Phi(\theta - z_\alpha) + \Phi(-z_\alpha - \theta) \\ J_2(\theta) &= E \left\{ \frac{(1-2\alpha)\Phi(Z+\theta) - (1-\alpha)^2}{(1-\alpha)\Phi(Z+\theta)} I(Z+\theta \geq z_\alpha) \right\}. \end{aligned}$$

Since

$$\begin{aligned} \partial J_1(\theta)/\partial \theta &= \alpha(1-\alpha)^{-1}\phi(\theta - z_\alpha) - \phi(\theta + z_\alpha) \\ &= \phi(\theta + z_\alpha)\{\alpha(1-\alpha)^{-1}\exp(2\theta z_\alpha) - 1\}, \end{aligned}$$

$J_1(\theta)$  is decreasing-increasing, with minimum at  $\theta_0 = (2z_\alpha)^{-1} \log\{(1-\alpha)/\alpha\} > 0$ . Further,  $J_1(0) = \lim_{\theta \rightarrow \infty} J_1(\theta) = \alpha(1-\alpha)^{-1}$ . Therefore  $J_1(\theta) < J_1(0)$  for  $\theta > 0$ .

To obtain a similar result for  $J_2$ , let

$$g(x) = \{(1-2\alpha)\Phi(x) - (1-\alpha)^2\}/\{(1-\alpha)\Phi(x)\}$$

which is increasing in  $x$  with  $g(z_\alpha) = -\alpha/(1-\alpha)$  and  $g(x) \rightarrow -\alpha^2/(1-\alpha)$  as  $x \rightarrow \infty$ . Hence  $g(x) < 0$  for  $x \geq z_\alpha$ .

(a)  $0 < \theta \leq 2z_\alpha$ . Since  $\phi(x) \leq \phi(x-\theta)$  for  $x \geq z_\alpha$ ,

$$J_2(0) - J_2(\theta) = \int_{z_\alpha}^{\infty} g(x)\phi(x) dx - \int_{z_\alpha}^{\infty} g(x)\phi(x-\theta) dx \geq 0.$$

(b)  $\theta > 2z_\alpha$ . From (6.1),

$$\begin{aligned} &J_2(0) - J_2(\theta) \\ &= \int_{z_\alpha}^{\infty} g(x)\phi(x) dx - \int_{z_\alpha}^{\infty} g(x)\phi(x-\theta) dx \\ &= \int_{z_\alpha}^{\theta/2} g(x)[\phi(x) - \phi(x-\theta)] dx + \int_{\theta/2}^{\infty} g(x)[\phi(x) - \phi(x-\theta)] dx \\ &> -\frac{\alpha}{1-\alpha} \int_{z_\alpha}^{\theta/2} [\phi(x) - \phi(x-\theta)] dx - \frac{\alpha^2}{1-\alpha} \int_{\theta/2}^{\infty} [\phi(x) - \phi(x-\theta)] dx \\ &= \alpha(1-\alpha)^{-1}[-\{\Phi(\theta/2) - (1-\alpha) - \Phi(-\theta/2) + \Phi(z_\alpha - \theta)\} \\ &\quad + \alpha\{\Phi(\theta/2) - \Phi(-\theta/2)\}] \\ &= \alpha(1-\alpha)^{-1}\{K_1(\theta) + K_2(\theta)\} \end{aligned}$$

where

$$\begin{aligned} K_1(\theta) &= \Phi(-\theta/2) - \Phi(z_\alpha - \theta) \\ K_2(\theta) &= 1 - \alpha - \Phi(\theta/2) + \alpha\Phi(\theta/2) - \alpha\Phi(-\theta/2). \end{aligned}$$

Now  $K_1(\theta) > 0$  for  $\theta > 2z_\alpha$ ,  $K_2'(\theta) = (\alpha - 1/2)\phi(\theta/2) < 0$ , and  $K_2(\theta) \rightarrow 0$  as  $\theta \rightarrow \infty$ . Thus  $J_2(0) - J_2(\theta) \geq 0$ .

Therefore  $f(\theta) \leq f(0) = 2\Phi(-z_\alpha) - (1-\alpha) \int_{z_\alpha}^{\infty} \phi(x)/\Phi(x) dx = 2\alpha + (1-\alpha) \log(1-\alpha)$ .

It remains to show that  $f(0) > \alpha$  for all  $0 < \alpha < 1$ . Let  $h(\alpha) = f(0) - \alpha$ . Then

$$h(\alpha) = \begin{cases} \alpha + (1-\alpha) \log(1-\alpha), & \text{if } 0 < \alpha \leq 1/2 \\ 1 - \alpha - (1-\alpha) \log 2, & \text{if } 1/2 \leq \alpha < 1 \end{cases}$$

1 and  $h$  is continuous with  $h(0) = h(1) = 0$ ,  $h(1/2) = (1 - \log 2)/2 > 0$ , and 1

$$2 \quad h'(\alpha) = \begin{cases} -\log(1 - \alpha) > 0, & \text{if } 0 < \alpha \leq 1/2 \\ -1 + \log 2 < 0, & \text{if } 1/2 \leq \alpha < 1. \end{cases} \quad 3$$

4  
5 Therefore  $h(\alpha) > 0$  for  $0 < \alpha < 1$ , concluding the proof. 6

7  
8 **References** 8

9  
10 [1] ANÉ, C., BURLEIGH, J. G., MCMAHON, M. M. and SANDERSON, M. J. (2005). Covarion structure in 10  
11 plastid genome evolution: A new statistical test. *Molecular Biology and Evolution*, **22**, 914–924. 11  
12 [2] CARSTENS B. C., BANKHEAD III, A. AND JOYCE P. and SULLIVAN J. (2005). Testing population genetic 12  
13 structure using parametric bootstrapping and MIGRATE-N. *Genetica*, **124**, 71–75. 13  
14 [3] CHUNG P. J. and MOURA J. (2004). A GLRT and bootstrap approach to detection in magnetic 14  
15 resonance force microscopy. In *IEEE International Conference on Acoustics, Speech, and Signal*  
16 *Processing*. 15  
17 [4] COX, D. R. (1961). Tests of separate families of hypotheses. In *Proceedings of the Fourth Berkeley*  
18 *Symposium on Mathematical Statistics and Probability*, **1**. Berkeley, CA, University of California  
19 Press. 105–123. 17  
20 [5] DALLA V. and HIDALGO J. (2004). A parametric bootstrap test for cycles. *Journal of Econometrics*, 18  
21 **129**, 219–261. 18  
22 [6] DAMEUSA A., RICHTER, F. G. C., BRORSEN, B. W. and SUKHDIAL K. P. (2002). AIDS versus the 19  
23 Rotterdam Demand System: a Cox Test with Parametric Bootstrap. *Journal of Agricultural and*  
24 *Resource Economics*, **27**, 335–347. 20  
25 [7] EFRON, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.*, **7**, 1–26. 21  
26 [8] HUNSBERGER S., ALBERT, P. S., FOLLMANN, D. A. and SUH E. (2002). Parametric and semiparametric 22  
27 approaches to testing for seasonal trend in serial count data. *Biostatistics*, **3**, 289–298. 23  
28 [9] KAITIBIE, S., NGANJE, W. E., WADE BRORSEN, B. and EPLIN, F. M. (2007). A Cox Parametric Boot- 24  
25 strap Test of the von Liebig Hypotheses. *Canadian Journal of Agricultural Economics*, **55**, 15–25. 24  
26 [10] LEHMANN E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. Springer, New 25  
27 York. 26  
28 [11] SOLOW A. R., COSTELLO, C. J. and WARD, M. (2003). Testing the Power Law Model for Discrete Size 27  
29 Data. *The American Naturalist*, **162**, 685–689. 28  
30 [12] WALTERS S. J. and CAMPBELL, M. J. (2004). The use of bootstrap methods for analysing health- 29  
31 related quality of life outcomes (particularly the SF-36). *Health Qual Life Outcomes*, **2**, 70. 29  
32 [13] WILSON, E. B. and HILFERTY, M. M. (1931). The distribution of chi-square. *Proceedings of the* 30  
33 *National Academy of Sciences*, **17**, 684–688. 30  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51