IMS Lecture Notes–Monograph Series
Optimality: The Third Erich L. Lehmann Symposium
Vol. 57 (2009) 93–116

DOI: 10.1214/09-LNMS5708

On Bootstrap Tests of Hypotheses

Wei-Yin Loh[1,*] **and Wei Zheng**[2,*]

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]Department of Statistics, University of Wisconsin–Madison, 1300 University Avenue, WI 53706, e-mail: loh@stat.wisc.edu
[2]Department of Statistics, University of Wisconsin–Madison, 1300 University Avenue, WI 53706, e-mail: zheng@stat.wisc.edu
*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.
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. 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 $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ be a vector of n independent observations from F_μ. In the bootstrap method, we first find an estimate $\hat{\mu}_0$ of μ under H_0 and estimate F_μ 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_μ denote the distribution function of S. Let $\mathbf{X}_n^* = (X_1^*, X_2^*, \ldots, X_n^*)$ 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 α $(0 < \alpha < 1)$, let $c_\alpha(\hat{\mu}_0)$ be the upper-α 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 $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$, a normal distribution with mean μ and variance σ^2. Let $\phi(x)$ and $\Phi(x)$ denote the density and distribution functions of the $N(0, 1)$ distribution and let z_α be its upper-α critical value, that is, $1 - \Phi(z_\alpha) = \alpha$. Consider testing

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

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 μ is $\hat{\mu} = \bar{X}_n$. Let $\hat{\mu}_i$ be the MLE of μ 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 $X_1^*, X_2^*, \ldots, X_n^*$ 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 $0 < \alpha \le 1/2$, 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 $\bar{X}_n \ge z_\alpha n^{-1/2}$. 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 $\bar{X}_n I(\bar{X}_n > 0) \ge z_\alpha n^{-1/2}$. If $0 < \alpha \le 1/2$, then $z_\alpha \ge 0$ 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 μ_0 and μ_1, let

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

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

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

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 $\theta \ge 0$. For fixed $0 < \alpha < 1$, the function*

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

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

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

which is greater than α 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 $S = \bar{X}_n I(\bar{X}_n > 0) \ge c_\alpha(\hat{\mu}_0)$, 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

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

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. □

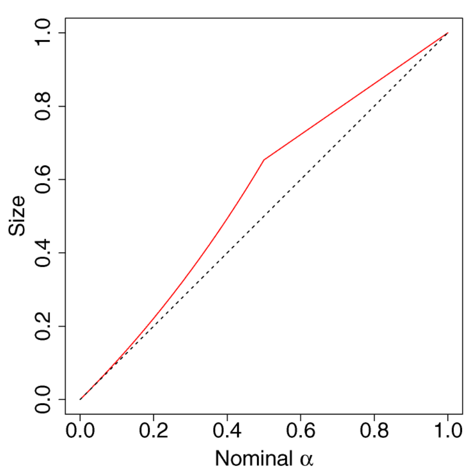

Fig 1. *Size (2.3) of bootstrap test for the normal mean based on the standard likelihood ratio, for known σ. The dashed line is the identity function.*

2.1.3. Cox Likelihood Ratio Statistic

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

$$L(\hat\mu_0, \hat\mu_1, \mathbf{X}_n) = \log\left\{\sup_{H_1}\prod_{i=1}^{n}\phi(x_i-\mu) \Big/ \sup_{H_0}\prod_{i=1}^{n}\phi(x_i-\mu)\right\}.$$

For the current problem,

$$\begin{aligned} L(\hat\mu_0, \hat\mu_1, \mathbf{X}_n) &= n\{\bar X_n(\hat\mu_1-\hat\mu_0) - (\hat\mu_1^2-\hat\mu_0^2)/2\} \\ &= n(\bar X_n|\bar X_n| - \bar X_n|\bar X_n|/2) \\ &= n\bar X_n^2 \operatorname{sgn}(\bar X_n)/2. \end{aligned}$$

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

2.2. Unknown Variance

Now suppose we test the hypotheses (2.1) without assuming that σ 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 μ and σ^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

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

The standard log-likelihood ratio statistic is

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

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

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

Proof. Let Z denote a standard normal variable independent of χ^2_ν. Since

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

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

Theorem 2.4. *If σ is unknown, the size of the nominal level-α 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 $n\to\infty$, the bound tends to (2.3), the size for the case where σ 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

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

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 $t_{n-1,\delta,\alpha} \le 0$ occurs if and only if $\alpha \ge P(T_n^* > 0 \,|\, \hat{\mu}_0, \hat{\sigma}_0)$. 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 $t_{n-1,\delta,\alpha} \le 0$ if and only if $\delta \le -z_\alpha$.

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

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

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

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

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 σ^2 is unknown, the size of the bootstrap test of (2.1) based on the Cox likelihood ratio has lower bound*

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

Proof. The Cox log-likelihood ratio statistic is

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

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

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

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

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

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

$$\sqrt{(n-1)/2}\left\{\frac{n^2}{(n-1)^2\chi^2_{n-1,\alpha}} - 1\right\} \to -z_\alpha$$

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 μ and σ unknown, and let $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ denote the unrestricted MLE of σ^2.

3.1. $H_0 : \sigma^2 \le 1$ *vs.* $H_1 : \sigma^2 > 1$

Let $\hat{\sigma}_i^2$ be the MLE of σ^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* $H_0 : \sigma^2 \le 1$ *vs.* $H_1 : \sigma^2 > 1$*, with* μ *unknown, is bounded below by*

$$\begin{aligned}&\min\{\alpha, 1 - \Psi_{n-1}(n)\} \\ &\quad + E\left[\frac{\alpha - 1 + \Psi_{n-1}(n^2/\chi^2_{n-1})}{\Psi_{n-1}(n^2/\chi^2_{n-1})} I\left\{\chi^2_{n-1} \le \min\left(n, \frac{n^2}{\chi^2_{n-1,\alpha}}\right)\right\}\right].\end{aligned} \tag{3.1}$$

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

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

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 χ^2_{n-1} 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 $\chi^2_{n-1,\alpha}$. 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 $\hat{\sigma}^2\chi^2_{n-1}$ left truncated at n. Thus the test does not reject H_0 if $\hat{\sigma}^2\chi^2_{n-1,\alpha} > n$. On the other hand, if $\hat{\sigma}^2\chi^2_{n-1,\alpha} \leq n$, 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 $n < \chi^2_{n-1,\alpha}$, we have

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

and

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

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

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

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

3.1.2. Cox Likelihood Ratio Statistic

Theorem 3.2. *If μ 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 α and is UMP for $\chi^2_{n-1,\alpha} > n$. It rejects H_0 w.p.1 for other values of α.*

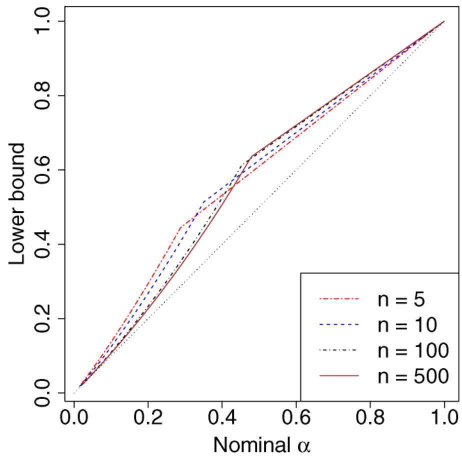

FIG 2. *Lower bounds (3.1) on the size of the bootstrap test of $H_0 : \sigma^2 \le 1$ vs. $H_1 : \sigma^2 > 1$ based on the standard likelihood ratio, for $n = 5, 10, 100$, and 500. The 45-degree line is the identity function.*

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

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

and

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

we have

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

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 $\hat{\sigma}_0^2\chi^2_{n-1}$, the bootstrap critical point of $\hat{\sigma}^2$ is $n^{-1}\hat{\sigma}_0^2\chi^2_{n-1,\alpha}$. Thus the bootstrap test rejects H_0 if $\hat{\sigma}^2\,\hat{\sigma}_0^{-2} > n^{-1}\chi^2_{n-1,\alpha}$, or equivalently, $\max(\hat{\sigma}^2, 1) > n^{-1}\chi^2_{n-1,\alpha}$. If α is so large that $n^{-1}\chi^2_{n-1,\alpha} \le 1$, the bootstrap test rejects H_0 regardless of the data. On the other hand, if $n^{-1}\chi^2_{n-1,\alpha} > 1$, the test rejects H_0 if $\hat{\sigma}^2 > n^{-1}\chi^2_{n-1,\alpha}$, which coincides with the UMP test [10, p. 88]. □

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

Next suppose we reverse the hypotheses and test $H_0 : \sigma^2 \ge 1$ 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 μ unknown, the size of the bootstrap test for $H_0 : \sigma^2 \ge 1$ vs. $H_1 : \sigma^2 < 1$, based on the standard likelihood ratio, is bounded* *below by*

(3.2)

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

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 χ^2_{n-1} 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 $\chi^2_{n-1,1-\alpha}$. 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 $\hat{\sigma}^2\chi^2_{n-1}$ right-truncated at n. The test does not reject H_0 if $\hat{\sigma}^2\chi^2_{n-1,1-\alpha} < n$. Otherwise, if $\hat{\sigma}^2\chi^2_{n-1,1-\alpha} \geq n$, 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 $\chi^2_{n-1,1-\alpha} < n$,

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

and

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

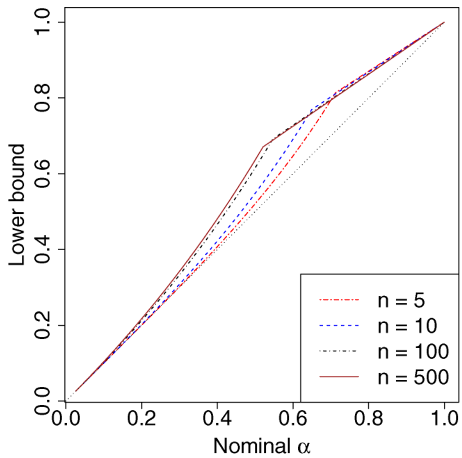

FIG 3. *Lower bounds (3.2) on the size of the bootstrap test of $H_0 : \sigma^2 \geq 1$ vs. $H_1 : \sigma^2 < 1$ based on the standard likelihood ratio $M_n^{(1)}$. The 45-degree line is the identity function.*

Therefore

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

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

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 α and is UMP if $\chi^2_{n-1,\alpha} > n$. 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

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

we have

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

a strictly decreasing function of $\hat\sigma^2$. Thus the test statistic rejects H_0 for small values of $\hat\sigma^2$. The bootstrap null distribution of $\hat\sigma^2$ is $n^{-1}\hat\sigma_0^2\chi^2_{n-1}$, with critical value $n^{-1}\hat\sigma_0^2\chi^2_{n-1,1-\alpha}$. Hence the bootstrap test rejects H_0 if $\hat\sigma^2\,\hat\sigma_0^{-2} < n^{-1}\chi^2_{n-1,1-\alpha}$. But the left side of the inequality is never greater than 1, because

$$\hat\sigma^2\,\hat\sigma_0^{-2} = \begin{cases} \hat\sigma^2, \text{if } \hat\sigma^2 < 1, \\ 1, \quad \text{otherwise.} \end{cases}$$

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

4. Testing Difference of Two Normal Means

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

$$H_0 : \eta \le \mu \quad \text{vs.} \quad H_1 : \eta > \mu. \tag{4.1}$$

The likelihood function for this case is

$$\begin{aligned} &L(\mu,\tau) \\ &= (2\pi)^{-(m+n)/2}\sigma^{-m}\tau^{-n}\exp\Big\{-(2\sigma^2)^{-1}\sum(X_i-\mu)^2 - (2\tau^2)^{-1}\sum(Y_j-\eta)^2\Big\} \\ &= (2\pi)^{-(m+n)/2}\sigma^{-m}\tau^{-n}\exp\Big\{-(2\sigma^2)^{-1}\sum(X_i-\bar X_m)^2 - (2\tau^2)^{-1}\sum(Y_j-\bar Y_n)^2 \\ &\qquad - m(2\sigma^2)^{-1}(\mu-\bar X_m)^2 - n(2\tau^2)^{-1}(\eta-\bar Y_n)^2\Big\} \end{aligned}$$

and the unrestricted MLE of (μ,η) is $(\hat\mu,\hat\eta) = (\bar X_m,\bar Y_n)$.

4.1. Known Variances

Let $V = (m\tau^2\bar X_m + n\sigma^2\bar Y_n)/(m\tau^2+n\sigma^2)$. The MLE of (μ,η) under H_0 is

$$(\hat\mu_0,\hat\eta_0) = \begin{cases} (\bar X_m,\bar Y_n), & \bar Y_n \le \bar X_m, \\ (V,V), & \bar Y_n > \bar X_m, \end{cases}$$

and that under H_1 is

$$(\hat\mu_1,\hat\eta_1) = \begin{cases} (V,V), & \bar Y_n > \bar X_m, \\ (\bar X_m,\bar Y_n), & \bar Y_n \le \bar X_m. \end{cases}$$

4.1.1. Difference of Means Statistic

Theorem 4.1. *The size of the bootstrap test of (4.1) based on $\bar Y_n - \bar X_m$ is α if $\alpha < 1/2$ and is 1 if $\alpha \ge 1/2$.*

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 $\hat\eta_0 - \hat\mu_0 = S\,I(S<0)$ and variance $m^{-1}\sigma^2 + n^{-1}\tau^2$. Thus its nominal level-α bootstrap critical value is $S\,I(S<0) + z_\alpha\{m^{-1}\sigma^2+n^{-1}\tau^2\}^{1/2}$ 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 α. On the other hand, if $\alpha \ge 1/2$, then $z_\alpha \le 0$ 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. $\bar Y_n \le \bar X_m$. Then $S = 0$, $\hat\eta_0 - \hat\mu_0 = \bar Y_n - \bar X_m$, and $\Phi(W) \ge 1/2$. If $1-\Phi(W) < \alpha$, the test is randomized and rejects H_0 with probability $\{\alpha - 1 + \Phi(W)\}/\Phi(W)$. Otherwise, if $1-\Phi(W) \ge \alpha$, 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 $\alpha \ge 1/2$, then the critical value is 0 and the test rejects w.p.1.

Therefore

$$\begin{aligned}&P(\text{Reject } H_0)\\&= P(\text{Reject } H_0,\, \bar Y_n \le \bar X_m,\, 1-\Phi(W) < \alpha)\\&\quad + P(\text{Reject } H_0,\, \bar Y_n > \bar X_m)\, I(\alpha < 1/2) + P(\text{Reject } H_0,\, \bar Y_n > \bar X_m)\, I(\alpha \ge 1/2)\\&= E[\Phi(W)^{-1}\{\alpha - 1 + \Phi(W)\}\, I(W > z_\alpha^+)]\\&\quad + P(W < -z_\alpha)\, I(\alpha < 1/2) + P(W < 0)\, I(\alpha > 1/2)\\&= E[\Phi(W)^{-1}\{\alpha - 1 + \Phi(W)\}\, I(W > z_\alpha^+)] + P(W < -z_\alpha^+)\end{aligned}$$

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* α *if* $\alpha < 1/2$ *and is 1 if* $\alpha \ge 1/2$.

Proof. The Cox log-likelihood ratio statistic is

$$\log\left\{\frac{L(\hat\mu_1,\hat\tau_1)}{L(\hat\mu_0,\hat\tau_0)}\right\} = \frac{mn(\bar Y_n - \bar X_m)^2}{2(m\tau^2+n\sigma^2)}\{I(\bar Y_n > \bar X_m) - I(\bar Y_n \le \bar X_m)\}.$$

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

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

The corresponding MLEs under H_0 and H_1 are, respectively,

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

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

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 Υ_ν denote the t distribution function with ν degrees of freedom and let $s^2 = \hat{\sigma}^2 N(N-2)^{-1}$ be the usual unbiased estimate of σ^2.

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

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

as $N \to \infty$.

Proof. The hypothesis H_0 is rejected if

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

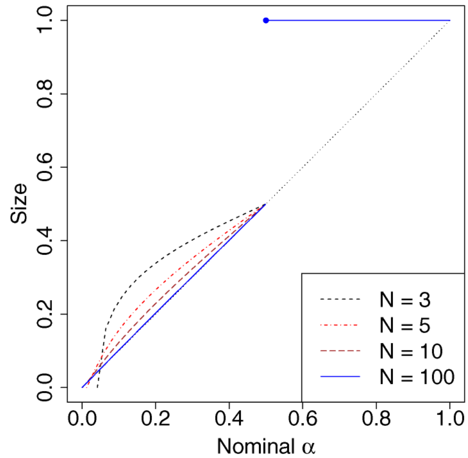

FIG 4. *Size of bootstrap test for a difference of normal means based on the difference of sample means, for equal but unknown variances (4.2). The 45-degree line is the identity function.*

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

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

Therefore if $\alpha \le 1 - \Phi(\sqrt{N})$, 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. $\alpha \ge 1/2$. The test rejects H_0 w.p.1 because $z_\alpha < 0$.

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

4.2.2. Standard Likelihood Ratio Statistic

The log-likelihood ratio statistic is

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

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*

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

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

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

as $N \to \infty$ 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 $S = \sqrt{mn/N}s^{-1}(\bar{Y}_n - \bar{X}_m)$. 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 Exp(θ, τ) denote the distribution with density $\tau^{-1}\exp\{-\tau^{-1}(x-\theta)\}$, $x \geq \theta$. We consider testing hypotheses about θ with $\tau = 1$. The likelihood for a sample $X_1, \ldots, X_n$ from an Exp(θ,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 θ 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 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 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 Exp$(0,1)$ with upper-α critical point $\log(1/\alpha)$, and the test rejects H_0 if $nX_{(1)} > -\log\alpha$.

2. $X_{(1)} \le 0$. Then $S = 0$, $\hat\theta_0 = X_{(1)}$, and the distribution of S^* is $\mathrm{Exp}(nX_{(1)}, 1)$, 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 $\mathrm{Exp}(n\theta, 1)$ distribution,

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

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

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

Therefore

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

where

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

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

5.1.2. Cox Likelihood Ratio Statistic

The Cox log-likelihood ratio statistic is

$$\begin{aligned}
S &= \sum_{i=1}^{n} \log\left[\frac{\exp\{-(X_i - \hat\theta_1)\} I(X_{(1)} \ge \hat\theta_1)}{\exp\{-(X_i - \hat\theta_0)\} I(X_{(1)} \ge \hat\theta_0)}\right]\\
&= \begin{cases} -\infty, & X_{(1)} < 0,\\ nX_{(1)}, & X_{(1)} \ge 0.\end{cases}
\end{aligned}$$

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 $H_0 : \theta \le 0$ vs. $H_1 : \theta > 0$ for a sample from an Exp(θ, 1) distribution, the bootstrap tests based on the standard and Cox likelihood ratios have size α.*

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

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

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

$$\phi(x) - \phi(x-\theta) \begin{cases} > 0, \text{ if } x < \theta/2, \\ < 0, \text{ if } x > \theta/2. \end{cases} \tag{A.1}$$

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 (A.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, $J_1(0) = \lim_{\theta\to\infty} J_1(\theta) = \alpha(1-\alpha)^{-1}$. 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 $g(x) \to -\alpha^2/(1-\alpha)$ as $x \to \infty$. Hence $g(x) < 0$ for $x \geq z_\alpha$.

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

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

(b) $\theta > 2z_\alpha$. From (A.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 $K_2(\theta) \to 0$ as $\theta \to \infty$. Thus $J_2(0) - J_2(\theta) \ge 0$.

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

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

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

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

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

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

References

[1] Ané, C., Burleigh, J. G., McMahon, M. M. and Sanderson, M. J. (2005). Covarion structure in plastid genome evolution: A new statistical test. *Mol. Biol. Evol.* **22** 914–924.

[2] Carstens, B. C., Bankhead, III, A., Joyce, P. and Sullivan, J. (2005). Testing population genetic structure using parametric bootstrapping and MIGRATE-N. *Genetica* **124** 71–75.

[3] Chung, P. J. and Moura, J. (2004). A GLRT and bootstrap approach to detection in magnetic resonance force microscopy. In *IEEE International Conference on Acoustics, Speech, and Signal Processing.*

[4] Cox, D. R. (1961). Tests of separate families of hypotheses. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* **1** 105–123. Univ. California Press, Berkeley, CA.

[5] Dalla, V. and Hidalgo, J. (2004). A parametric bootstrap test for cycles. *J. Econometrics* **129** 219–261.

[6] Dameusa, A., Richter, F. G. C., Brorsen, B. W. and Sukhdial, K. P. (2002). AIDS versus the Rotterdam demand system: A Cox test with parametric bootstrap. *Journal of Agricultural and Resource Economics* **27** 335–347.

[7] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.

[8] Hunsberger, S., Albert, P. S., Follmann, D. A. and Suh, E. (2002). Parametric and semiparametric approaches to testing for seasonal trend in serial count data. *Biostatistics* **3** 289–298.

[9] Kaitibie, S., Nganje, W. E., Wade Brorsen, B. and Epplin, F. M. (2007). A Cox parametric bootstrap test of the von Liebig hypotheses. *Canadian Journal of Agricultural Economics* **55** 15–25.

[10] Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. Springer, New York.

[11] Solow, A. R., Costello, C. J. and Ward, M. (2003). Testing the power law model for discrete size data. *The American Naturalist* **162** 685–689.

[12] Walters, S. J. and Campbell, M. J. (2004). The use of bootstrap methods for analysing health-related quality of life outcomes (particularly the SF-36). *Health Qual Life Outcomes* **2** 70.

[13] Wilson, E. B. and Hilferty, M. M. (1931). The distribution of chi-square. *Proc. Nat. Acad. Sci.* **17** 684–688.