Alternatives to asymptotic approximations: data-resampling based techniques

- In point estimation and hypothesis testing, we have resorted to asymptotic theory (which holds when $N \to \infty$) to approximate the *finite-sample* ($n < \infty$) distribution of statistics.
- Here, consider alternatives to this approaches.
- Focus on alternatives based on **data resampling**.

General idea:

- Data are a finite sample $x_1, \ldots, x_N \sim i.i.d.\ F$
- Using data, construct finite-sample statistic $W_N \equiv W(x_1, \ldots, x_N)$ (this could be a point estimator, or a test statistic).
- Now we want to get idea about the sampling variability in W_N (some measure of the variance, or standard error). For concreteness, in this lecture we will focus on alternatives for estimating the *variance* of W_N, equal to $E_F W_N^2 - (E_F W_N)^2$, where expectation is taken over the random variables $x_1, \ldots, x_N$. (Standard error would be the square root of this.)
- The "true" variance is calculated with respect to F:

$$
\begin{aligned}
&E_F W_N^2 - (E_F W_N)^2 \\
&= \int \cdots \int W(x_1, \ldots, x_N)^2 dF(x_1) \ldots dF(x_N) - \left[\int \cdots \int W(x_1, \ldots, x_N) dF(x_1) \ldots dF(x_N) \right]^2 .
\end{aligned} \tag{1}
$$

- Asymptotic approach: we make assumptions such that $\sqrt{N}\,(W_N - W_0) \xrightarrow{d} N(0, V)$. Then approximate the finite-sample distribution of $W_N \overset{A}{\sim} N(W_0, \frac{1}{N} V)$, so that $\frac{1}{N}\hat{V}$ is estimate of W_N's variance.. *Problem:* this approximation can be bad, especially if N is small.

- Resampling approach: we approximate the finite-sample distribution of W_N by the *exact* distribution of $W(x_1^*, \ldots, x_M^*)$, where

$$x_1^*, \ldots, x_M^* \ \sim \ F^* (\cdots ; x_1, \ldots, x_N)$$

where F^* is defined as the resampling distribution. Note that F^* explicitly depends on the original observations $x_1, \ldots, x_N$; for this reasons, $x_1^*, \ldots, x_M^*$ is called a resampled dataset. Furthermore, the size of the resampled dataset M is usually an increasing function of N, but does not have to coincide with N. Examples of resampling distributions F^* are given below.

- Then, we approximate the finite-sample CDF of W_N by the exact CDF of $W(x_1^*, \ldots, x_M^*)$:

$$\begin{aligned} Prob\,(W_N \leq z) &\approx Prob\,(W(x_1^*, \ldots, x_M^*) \leq z) \\ &\int \cdots \int \mathbf{1}\,(W(x_1^*, \ldots, x_M^*) \leq z)\, F^*\,(dx_1^*, \ldots, dx_M^*; x_1, \ldots, x_N)\,. \end{aligned}$$

Moreover, the resampling estimate of W_N's variance is $E_{F^*}W^2 - (E_{F^*}W)^2$, where both expectations are taken over the resampling distribution F^*:

$$\begin{aligned} E_{F^*}W^2 - (E_{F^*}W)^2 = &\int \cdots \int W(x_1^*, \ldots, x_M^*)^2 F^*\,(dx_1^*, \ldots, dx_M^*; x_1, \ldots, x_N) \\ &- \left[\int \cdots \int W(x_1^*, \ldots, x_M^*) F^*\,(dx_1^*, \ldots, dx_M^*; x_1, \ldots, x_N)\right]^2. \end{aligned} \tag{2}$$

- Instances of resampling distributions:

  - **Bootstrap:** resampled dataset $x_1^*, \ldots, x_N^*$ (same size as original dataset) are N iid draws (with replacement) from the original dataset $x_1, \ldots, x_N$. Specifically:

$$x_i^* = \begin{cases} x_1 & \text{w/prob } \frac{1}{N} \\ x_2 & \text{w/prob } \frac{1}{N} \\ \cdots & \cdots \\ x_N & \text{w/prob } \frac{1}{N} \end{cases} \qquad i = 1, \ldots, N.$$

- **Jackknife:** resampled dataset $x_1^*, \ldots, x_{N-1}^*$ is original sample $x_1, \ldots, x_N$ with one datapoint randomly omitted. Specifically:

$$x_1^*, \ldots, x_{N-1}^* = \begin{cases} x_2, \ldots, x_N & \text{w/prob } \frac{1}{N} \\ x_1, x_3, \ldots, x_N & \text{w/prob } \frac{1}{N} \\ x_1, x_2, x_4, \ldots, x_N & \text{w/prob } \frac{1}{N} \\ \cdots & \cdots \\ x_1, \ldots, x_{N-1} & \text{w/prob } \frac{1}{N}. \end{cases}$$

- **Subsampling:** resampled dataset $x_1^*, \ldots, x_M^*$ (with $M < N$, but $M \to \infty$ as $N \to \infty$, and $M/N \to 0$) is a random subsample (without replacement) of M datapoints from the original sample $x_1, \ldots, x_N$: that is,

$$x_1^*, \ldots, x_M^* = \begin{cases} x_1, \ldots, x_M & \text{w/prob } \frac{1}{\binom{N}{M}} \\ x_1, x_3, \ldots, x_{M+1} & \text{w/prob } \frac{1}{\binom{N}{M}} \\ x_1, x_4, \ldots, x_{M+2} & \text{w/prob } \frac{1}{\binom{N}{M}} \\ \cdots & \cdots \\ x_{N-M}, \ldots, x_N & \text{w/prob } \frac{1}{\binom{N}{M}}. \end{cases}$$

We will discuss the first and third in more detail.

1 Bootstrap

Just consider a simple example: x_1, x_2 are *i.i.d.* (μ, σ^2). Say (for simplicity) that the realized $x_1 = 1$ and $x_2 = 0$.

You make inference about the unknown μ by estimating μ using the sample average $\hat{\mu} \equiv \bar{x}_2 \equiv \frac{1}{2}(x_1 + x_2)$. For given x_1 and x_2, $\bar{x}_2 = \frac{1}{2}$.

Next, you want to obtain standard errors and confidence intervals for $\hat{\mu}$.

Asymptotic approach: use asymptotic approximation that $\hat{\mu} \overset{A}{\sim} N(\mu_0, \sigma^2/n)$. Since σ^2 is not known, we approximate using sample variance, so that our estimate of variance is $\frac{1}{2}\frac{1}{2}\sum_{i=1}^{2}(x_i - \bar{x}_2)^2 = \frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{8}$.

Accordingly, a 95% asymptotic confidence interval is then given by $\mu \in \left[\frac{1}{2} - \frac{1.96}{2},\ \frac{1}{2} + \frac{1.96}{2}\right]$.

Bootstrap approach: you estimate the variance of $\bar{x}_2$ by $E(\bar{x}_2^*)^2 - (E\bar{x}_2^*)^2$, where both expectations is taken with respect to the resampling distribution (and conditional on the original dataset x_1, x_2). Here, $\bar{x}_2^* \equiv \frac{1}{2}(x_1^* + x_2^*)$, where

$$x_i^* = \begin{cases} x_1 & \text{with prob } \frac{1}{2} \\ x_2 & \text{with prob } \frac{1}{2} \end{cases} \qquad i = 1, 2.$$

For the given values of $x_1 = 1,\ x_2 = 0$, we can explicitly derive the bootstrap estimate of variance:

x_1^*	x_2^*	$\bar{x}_2^*$	$(\bar{x}_2^*)^2$	Prob	$((\bar{x}^*)_2^2)^2$
1	0	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{16}$
0	0	0	0	$\frac{1}{4}$	0
1	1	1	1	$\frac{1}{4}$	1
0	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{16}$

Hence, $E\bar{x}_2^* = \frac{1}{2}$, $E(\bar{x}_2^*)^2 = \frac{3}{8}$ and the bootstrap variance estimate therefore equals $\frac{3}{8} - \left(\frac{1}{2}\right)^2 = \frac{1}{8}$, which coincides with asymptotic estimate of variance.

However, consider the case of estimating μ^2, using the sample average estimator $\hat{\mu}^2 = \bar{x}_2^2$. The asymptotic variance is obtained using the Delta method, yielding

$$AV(\hat{\mu}^2) = (2\mu)^2\sigma^2/n \approx 1^2 \cdot \frac{1}{8}.$$

The bootstrap variance is (see table above)

$$BV(\hat{\mu}^2) = E((\bar{x}^*)_2^2)^2 - [E(\bar{x}_2^*)^2]^2 = \frac{9}{32} - \left[\frac{3}{8}\right]^2 = \frac{9}{64}.$$

After obtaining estimates of standard error, bootstrap confidence intervals can be formed. There are different ways to do this, which we will look at below.

Use of simulation: One limitation of the bootstrap, however, is that it can be computationally intractable when N becomes large.

Example: $x_1, \ldots, x_N \sim$ iid (μ, σ^2). N is finite but large. What is bootstrap estimate of the variance of sample mean $\hat{\mu} \equiv \bar{x}_N$?

F^*, the bootstrap resampling distribution, is the discrete *multinomial* distribution with points of support at the N points $x_1, \dots, x_N$, each with probability $\frac{1}{N}$ (ie. an "N-sided die" with faces reading $x_1, \dots, x_N$) with mean $Ex^* = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}_N$. Accordingly, the bootstrap variance estimate is the sample variance of the

$$V_N^* \equiv E(\bar{x}_N^* - \bar{x}_N)^2 \tag{3}$$

(where x_N is taken as non-random). When N is big, this is not analytically tractable to calculate.

However, it can be approximated by *simulation*:

1. Draw B resampled datasets using the F^* resampling distribution: for each $b = 1, \dots, B$, draw $x_{1,b}^*, \dots, x_{N,b}^*$ from F^*. For each resampled dataset b, calculate the sample mean $\bar{x}_{N,b}^* \equiv \frac{1}{N}\sum_{i=1}^{N} x_{i,b}^*$.

2. Approximate the expectation in the expression for V_N^* by the sample variance (across the B resampled datasets) of $\bar{x}_{N,b}^*$:

$$V_N^{*,s} \equiv \frac{1}{B}\sum_{b=1}^{B}\left(\bar{x}_{N,b}^* - \bar{x}_N\right)^2 \tag{4}$$

   By the LLN, Eq. (4) $\xrightarrow{p}$ Eq. (3), as $B \to \infty$, for a fixed N, and for all realizations of $(X_1, \dots, X_N)$.

Consistency of bootstrap: usual criterion for validity of using bootstrap. Consider a statistic $W_n = W(x_1, \dots, x_N)$. Assume that it (suitably normalized and "blown-up") has a nondegenerate limiting distribution: $W_n \xrightarrow{d} J$. The requirement for consistency is that the bootstrapped version $\tilde{W}_n^* \xrightarrow{d} J$ also. That is, *the non-bootstrapped $\tilde{W}_n$ and bootstrapped statistics $\tilde{W}_n^*$ have the same limiting distribution.* This is not that easy to confirm in practice.

Counterexample: One well-known counterexample[1] is when $X_1, \dots, X_n \sim U[0, \theta]$. Obviously, the sample maximum $X_{(n)} \equiv \max(X_1, \dots, X_n) \xrightarrow{p} \theta$. Consider bootstrapping the distribution of $T_n \equiv \frac{n(\theta - X_{(n)})}{\theta}$. The bootstrapped version is: $T_n^* \equiv \frac{n(X_{(n)} - X_{(n)}^*)}{X_{(n)}}$

[1] Bickel and Freedman, "On the Consistency of Bootstrap Estimates", *Annals of Statistics*, 1981.

where $X_1^*, \ldots, X_n^*$ are iid draws from the multinomial distribution with support $(X_1, \ldots, X_n)$. Note that

$$\begin{aligned} P(T_n^* = 0) &= P(X_{(n)}^* = X_{(n)}) \\ &= 1 - \prod_{i=1}^{n} P(X_i^* < X_{(n)}) = 1 - \left(\frac{n-1}{n}\right)^n = 1 - (1 - \frac{1}{n})^n \\ &\to 1 - \exp(-1) \end{aligned}$$

which is ≈ 0.63.

On the other hand, we know that (for $k > 0$)

$$\begin{aligned} P(T_n \le k) &= P(X_{(n)} \ge \theta - \frac{\theta k}{n}) \\ &= 1 - P(X_{(n)} \le \theta - \frac{\theta k}{n})^n = 1 - (1 - \frac{k}{n})^n \\ &\to 1 - \exp(-k) \end{aligned}$$

which is the CDF of an exponential random variable, which evaluates to zero at $k = 0$. Hence T_n and T_n^* do not have the same limiting distribution, and bootstrap is not consistent here. ■

1.1 Bootstrap confidence intervals

Recall: consider a *pivotal* statistic $W_n(X_1, \ldots, X_n; \theta)$ with distribution G_n, which does not depend on θ. A size-$(1-\alpha)$ two-sided confidence interval is one such that

$$1 - \alpha = G_n(U) - G_n(L) = P(L \le W_n \le U \Leftrightarrow \underline{\theta} \le \theta \le \bar{\theta})$$

where we assume that the values U and L are chosen for the desired size, and we assume that the equation $L \le W_n \le U$ can be "inverted" to obtain the confidence interval $\underline{\theta} \le \theta \le \bar{\theta}$. We can set $U = G_n^{-1}(1 - \alpha/2)$ and $L = G_n^{-1}(\alpha/2)$.

The most-common test statistic used here is the T-statistic: $W_n = (\hat{\theta}_n - \theta)/\hat{\sigma}_n$, for which the associated confidence region is

$$\left[\hat{\theta}_n - \hat{\sigma}_n G_n^{-1}(1 - \alpha/2), \hat{\theta}_n - \hat{\sigma}_n G_n^{-1}(\alpha/2)\right]. \tag{5}$$

Problem is that for many cases, G_n is unknown, so we need to approximate it.

Asymptotic approach approximates G_n by $N(0,1)$.

Bootstrap approach approximates G_n by F_n^*, the bootstrap resampling distribution corresponding to the observed $X_1, \ldots, X_n$. Next we go over several common ways of constructing bootstrap confidence intervals.

1.1.1 Bootstrap "t-stat"

Inference is based on T-statistic $W_n = (\hat{\theta}_n - \theta)/\hat{\sigma}_n$. By plug-in principle:

$$G_n(x) = P(W_n \leq x) \longrightarrow G_n^*(x) = P_{F_n^*}(W_n^* = \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)/\hat{\sigma}_n^* \leq x)$$

where $\hat{\theta}_n^*$ is the bootstrapped estimate, and $\hat{\sigma}_n^*$ is the bootstrapped estimate of standard error. If you are using simulation to approximate $G_n^*(x)$, $\hat{\theta}_n^*$ and $\hat{\sigma}_n^*$ have to be computed for each resampled dataset. (More below.)

Accordingly, the bootstrap confidence interval, corresponding to Eq. (5), is

$$\left[\hat{\theta}_n - \hat{\sigma}_n G_n^{*-1}(1-\alpha/2), \hat{\theta}_n - \hat{\sigma}_n G_n^{*-1}(\alpha/2)\right].$$

Note: $\hat{\sigma}_n$ and $\hat{\sigma}_n^*$ are different. Very often, as we consider above, the variance $\hat{\sigma}_n = Var(\hat{\theta}_n)$ is itself estimated by bootstrap. Then $\hat{\sigma}_n^*$ denotes the bootstrapped version of the bootstrapped standard error $\hat{\sigma}_n$.

If you are using simulation to approximate G_n^*, then bootstrapping the bootstrap (also called the "nested bootstrap") can require a great deal of computer time. The idea is: in order to simulate G_n^*, you draw B resampled datasets. For each resampled dataset $b = 1, \ldots, B$, consisting of observations $x_{1,b}^*, \ldots, x_{n,b}^*$:

- Estimate $\hat{\theta}_{n,b}^*$.
- In order to estimate the standard error $\hat{\sigma}_{n,b}^*$, you resample M datasets using $F_{n,b}^*$, the bootstrap resampling distribution for the b-th resampled dataset $x_{1,b}^*, \ldots, x_{n,b}^*$.
  - For each resampled dataset (m, b), you estimate $\hat{\theta}_{n,b,m}^*$.

- Then approximate $\hat{\sigma}^*_{n,b} \approx \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\hat{\theta}^*_{n,b,m}\right)^2 - \left[\frac{1}{M}\sum_{m=1}^{M}\hat{\theta}^*_{n,b,m}\right]^2}$.

- Form resampled T-statistic for resampled dataset b as $W^*_{n,b} = (\hat{\theta}^*_{n,b} - \hat{\theta}_n)/\hat{\sigma}^*_{n,b}$.

Then approximate

$$G^*_n(x) \approx \frac{1}{B}\sum_{b=1}^{B}\mathbf{1}(W^*_{n,b} \leq x).$$

This requires a total of $B * M$ resampled datasets.

1.1.2 Bootstrap percentile

To avoid computational burden associated with boostrap t-stat, we can do the bootstrapped percentile, which is based on the unnormalized estimator $\hat{\theta}_n$. By plug-in principle:

$$G_n(x) = P_{F_n}(\hat{\theta} \leq x) \longrightarrow G^*_n(x) = P_{F^*_n}(\hat{\theta}^*_n \leq x)$$

where $\hat{\theta}^*_n$ is resampled estimator.

Accordingly, the bootstrap percentile confidence interval is

$$\left[G^{*-1}_n(\alpha/2), G^{*-1}_n(1-\alpha/2)\right].$$

As above, $G^*_n(x)$ can be simulated, but nested bootstrap is not necessary because there is no variance estimate here.

1.1.3 The hybrid bootstrap

This is based on the statistic: $W_n = \sqrt{n}(\theta_n - \theta)$. By the plug-in principle:

$$G_n(x) = P_{F_n}(\sqrt{n}(\hat{\theta} - \theta) \leq x) \longrightarrow G^*_n(x) = P_{F^*_n}(\sqrt{n}(\hat{\theta}^*_n - \hat{\theta}_n) \leq x)$$

where $\hat{\theta}^*_n$ is resampled estimator. Accordingly, the bootstrap confidence interval is

$$\left[\hat{\theta}_n - (1/\sqrt{n})G^{*-1}_n(1-\alpha/2), \hat{\theta}_n - (1/\sqrt{n})G^{*-1}_n(\alpha/2)\right].$$

As with the Bootstrap percentile, this avoids the computational burdens associated with the bootstrap t-stat method.

In each of these cases, the asymptotic validity of these confidence sets relies on the consistency of the underlying bootstrap estimates of the sampling distribution G_n.

2 Subsampling

To illustrate, start with an $n = 3$ Bernoulli experiment, with $x_1 = 1,\ x_2 = 0,\ x_3 = 0$. We take the subsample size as $M = 2$. Each subsampled dataset, then, is

$$x_1^*, x_2^* = \begin{cases} x_1, x_2 & \text{w/ prob } \frac{1}{3} \\ x_1, x_3 & \text{w/ prob } \frac{1}{3} \\ x_2, x_3 & \text{w/ prob } \frac{1}{3}. \end{cases}$$

The sample mean $\bar{x}_3$ is $\frac{1}{3}$. The asymptotic estimate of the variance is $\frac{1}{n}\bar{x}_3(1 - \bar{x}_3) = \frac{1}{3}\frac{1}{3}\frac{2}{3} = \frac{2}{27}$.

The subsample variance estimate is $E(\bar{x}_2^*)^2 - (E\bar{x}_2^*)^2$, where the expectation is taken over the subsampling distribution above. We can explicitly derive the distribution of $\bar{x}_2^*$:

x_1^*	x_2^*	$\bar{x}_2^*$	$(\bar{x}_2^*)^2$	Prob
1	0	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{3}$
1	0	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{3}$
0	0	0	0	$\frac{1}{3}$

Hence, the subsampling variance estimate is $\frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{18}$.

Simulating the subsample variance More generally, if N is large, the subsample variance will become difficult to calculate, due to the large number of subsamples (for given dataset size N and subsample size $M < N$, the total number of subsamples is $\binom{N}{M}$). However, as for the bootstrap, the subsample variance can also be approximated by simulation. The simulation procedure is simple:

1. For $b = 1, \ldots, B$ (where presumably $B << \binom{N}{M}$), randomly draw a M-datapoint subset $x_{1,b}^*, \ldots, x_{M,b}^*$ of the original dataset. For each subsampled dataset, calculate the sample mean $\bar{x}_{M,b}^* \equiv \frac{1}{M}\sum_{i=1}^{M} x_{i,b}^*$

2. Approximate the subsample variance by averaging over subsampled datasets:

$$\frac{1}{B}\sum_{b=1}^{B}(\bar{x}^*_{M,b})^2 - \left(\frac{1}{B}\sum_{b=1}^{B}\bar{x}^*_{M,b}\right)^2.$$

Validity of Subsampling We say that the subsampling procedure is valid when the subsampled distribution of $\tilde{W}^*_n \equiv M^\gamma\left(W(x^*_1,\ldots,x^*_M) - W_n\right)$ resembles that of $\tilde{W}_n \equiv n^\gamma\left(W_n - W_0\right)$ as n gets large; γ denotes the rate of convergence of W_n. Unlike for the bootstrap, it is often simple to establish this result: the standard theorem for subsampling states that all that is required is for the limiting distribution of $\tilde{W}_n$ to be nondegenerate.

Theorem 1 *(2.2.1 in Politis, Romano and Wolf (1999).) Assume $\tilde{W}_n$ has a nondegenerate limiting distribution with CDF J. Also assume $M/n \to 0$ and $M^\gamma/n^\gamma \to 0$. Then, letting $L^*_n(\cdot)$ denote the CDF of the subsampled statistic $\tilde{W}^*_n$, we have*

(i) $L^*_n(x) \xrightarrow{p} J(x)$, for all x where $J(\cdot)$ is continuous
(ii) $sup_x|L^*(x) - J(x)| \xrightarrow{p} 0$.
(iii) Asymptotically valid subsample p-values: for $\alpha \in (0,1)$, let $c^*_n(1-\alpha)$ and $c(1-\alpha)$ denote, respectively, the quantile functions of $L^*_n(x)$ and $J(x)$. Then

$$P(\tilde{W}_n \leq c^*_n(1-\alpha)) \to 1-\alpha; \quad \textit{as } n \to \infty.$$

■

Subsampling Hypothesis testing Another advantage of the subsampling approach is the ease in performing hypothesis tests. All we require is that the test statistic (suitably normalized) has a nondegenerate limiting distribution under the null hypothesis.

Assume that we have a test-statistic $T_n = T(x_1,\ldots,x_N)$ such that $T_n \xrightarrow{p} 0$ under H_0, and $\xrightarrow{p}> 0$ under the the alternative H_1 (one-sided test-statistics, as well as chi-squared, likelihood ratio test statistics satisfy this). Furthermore, assume that $n^\gamma T_n$ converges is distribution to some non-degenerate limiting distribution (so γ is the rate of convergence).

We construct a subsampled distribution for the normalized test statistic $n^\gamma T_n$; for each subsampled dataset k, we construct an analogous test statistic $M^\gamma T^*_{k,M} \equiv M^\gamma T(x^*_{k,1}, \ldots, x^*_{k,M})$. Let

$$G_{N,M}(z) \equiv \frac{1}{\binom{N}{M}} \sum_{i=1}^{\binom{N}{M}} \mathbf{1}\left(M^\gamma T^*_M \leq z\right)$$

denote the CDF of the set of subsampled test statistics (it is the proportion of the subsampled test stats which do not exceed z). Also, let $g_{N,M}(1-\alpha)$ denote the $1-\alpha$-th quantile of this CDF: $g_{N,M}(\tau) = \min(z : \ G_{N,M}(z) \geq \tau)$ for $0 \leq \tau \leq 1$.

Given this, a subsample size-α test of H_0 vs. H_1 obtains if we reject H_0 whenever $n^\gamma T_n > g_{N,M}(1-\alpha)$. In other words, we reject the null when the test statistic (normalized by its rate of convergence) calculated using the original data exceeds over $(1-\alpha)\%$ of the analogous subsampled test statistics.

Theorem 2 *(2.6.1 in Politis, Romano, Wolf (1999).)*

Assume that, under H_0, $n^\gamma T_n$ has limiting distribution with CDF G, and corresponding quantile function g. Let G^*_n and g^*_n denote CDF and quantile function of the subsampled test statistic $M^\gamma T^*_n$.

(i) Under H_0, $P(n^\gamma T_n > g^*_n(1-\alpha)) \to \alpha$; as $n \to \infty$
(ii) Under H_1, $P(n^\gamma T_n > g^*_n(1-\alpha)) \to 1$; as $n \to \infty$ ■

Given the previous theorem, part (i) is not surprising. However (ii) is interesting. The argument goes like this: note that the quantile function for $M^\gamma T^*_n$ is just M^γ times the quantile function for T^*_n, which we denote by $g^0_n(1-\alpha)$. Therefore

$$P(n^\gamma T_n > g^*_n(1-\alpha)) = P(n^\gamma T_n > m^\gamma g^0_n(1-\alpha)) = P(\frac{n^\gamma}{m^\gamma} T_n > g^0_n(1-\alpha)).$$

Since, by assumption, $T_n \xrightarrow{p} T > 0$, we have also that $g^0_n(1-\alpha) \xrightarrow{p} T$ (for all α). Because $\frac{n^\gamma}{m^\gamma} \to \infty$, asymptotic rejection probability is one.

As an example, consider an $N = 4$, $M = 3$ Bernoulli case, with $x_1 = 1, x_2 = 0, x_3 = 0, x_4 = 0$. We test $H_0 : p = \frac{1}{2}$ versus $H_1 : p \neq \frac{1}{2}$.

Consider the test statistic $T_n \equiv |\bar{X}_n - \frac{1}{2}|$, which converges to zero under H_0 but converges to something strictly > 0 under H_1. It turns out that under the null, $\sqrt{n}T_n$ converges to a non-degenerate distribution. However, its asymptotic distribution under the null is not easy to characterize. Hence, we wish to obtain critical regions for this test statistic using resampling techniques.

The test statistic in the original sample is $\sqrt{4}\frac{1}{4} = \frac{1}{2}$.

As before, we can derive the exact distribution of T_3^*, the subsampled test statistic formed from considering three-datapoints subsamples:

Dataset	$\bar{x}_3^*$	$\sqrt{3}T_3^*$	Prob
$\{x_1, x_2, x_3\}$	$\frac{1}{3}$	$\sqrt{3}\frac{1}{6}$	$\frac{1}{4}$
$\{x_2, x_3, x_4\}$	0	$\sqrt{3}\frac{1}{2}$	$\frac{1}{4}$
$\{x_1, x_2, x_4\}$	$\frac{1}{3}$	$\sqrt{3}\frac{1}{6}$	$\frac{1}{4}$
$\{x_1, x_3, x_4\}$	$\frac{1}{3}$	$\sqrt{3}\frac{1}{6}$	$\frac{1}{4}$

so that the CDF for $\sqrt{3}T_3^*$ is

$$G_{4,3}(z) = \begin{cases} 0 & z \in [0, \sqrt{3}\frac{1}{6}) \\ \frac{3}{4} & z \in [\sqrt{3}\frac{1}{6}, \sqrt{3}\frac{1}{2}) \\ 1 & z \in [\sqrt{3}\frac{1}{2}, +\infty). \end{cases}$$

Hence, an $\alpha = 0.25$ sized test would reject when $\sqrt{4}T_4 \geq \sqrt{3}\frac{1}{6} \approx 0.289$. Hence, for our assumed data, we would reject this test.