BUREAU OF THE CENSUS

STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-86/10

THE ASYMPTOTIC DISTRIBUTION OF THE LIKELIHOOD RATIO TEST FOR A CHANGE IN THE MEAN

by

John M. Irvine*
Bureau of the Census
Washington, D.C. 20233

Recommended by: David F. Findley

Report completed: March 25, 1986

Report issued: March 25, 1986

*Former ASA/NSF/Census Junior Fellow

The Asymptotic Distribution of The Likelihood Ratio Test for a Change in the Mean

Abbreviated Title: Likelihood Ratio Test for Change

John M. Irvine

Summary

A likelihood ratio test is one technique for detecting a shift in the mean of a sequence of independent normal random variables. If the time of the possible change is unknown, the asymptotic null distribution of the test statistic is extreme value, rather than the usual chi-square distribution. The asymptotic distribution is derived here under the null hypothesis of no change.

Keywords

asymptotic distribution, changepoint, extreme-value distribution, likelihood ratio test

1. Introduction

This discussion examines the problem of testing for a change in the mean of a sequence of independent normal variates, where the time of a possible change is unknown. The asymptotic distribution of the likelihood ratio statistic is derived. Shaban (1980) presents a bibliography of related problems.

Let Y_1, Y_2, ..., Y_N be independent normal random variables with means μ_i and equal variance σ^2. The inferential question under consideration is to test H_0: $\mu_i = \mu$ for all i; against H_1: $\mu_i = \mu_1$ for $i \leq \nu$ and $\mu_i = \mu_2$ for $i > \nu$ is unknown. Calculation of the likelihood ratio statistic λ for this problem is straight-forward and can be found in Hawkins (1977). Using a recursive integral equation, Hawkins produces a table of fractiles for the likelihood ratio statistic. Arguing heuristically, he asserts that $(-2 \log \lambda)^{1/2}$ converges to an extreme value distribution. Unfortunately the heuristic argument does not yield the correct normalizing constants for the distribution. The asymptotic result presented here resolves this question.

2. The Asymptotic Distribution

Let $Y_1, \ldots, Y_N$ be as in the previous section and consider the testing problem presented. First examine the case where σ^2 is known and, without loss of generality, let $\sigma^2 = 1$. Define

$$S_o = \sum_{i=1}^{N} (Y_i - \bar{Y})^2$$

$$S_1(\nu) = \sum_{i=1}^{\nu} (Y_i - \bar{Y}_1)^2 + \sum_{i=\nu+1}^{N} (Y_i - \bar{Y}_2)^2$$

where

$$\bar{Y} = (1/N) \sum_{i=1}^{N} Y_i$$

$$\bar{Y}_1 = (1/\nu) \sum_{i=1}^{\nu} Y_i$$

$$\bar{Y}_2 = [1/(N - \nu)] \sum_{i=\nu+1}^{N} Y_i$$

Note that S_o is the residual sum of squares under H_o and $S_1(\nu)$ is the residual sum of squares under H_1 given ν. Then the likelihood ratio statistic λ can be written as

(2.1) $$-2 \log \lambda = S_o - \inf_{1 \leq \nu < N} S_1(\nu)$$

When σ^2 unknown, the likelihood ratio statistic is

$$\lambda = \inf_{1 \leq \nu < N} \left[\frac{S_o}{S_1(\nu)}\right]^{-N/2}$$

For σ^2 known the asymptotic null distribution of $-2 \log \lambda$ is given by:

Theorem 2.1: Let $-2 \log \lambda$ be as in equation (2.1) and suppose H_o is true. Then

$$\lim_{N \to \infty} P\left[(-2 \log \lambda)^{1/2} < (2 \log \log N)^{1/2} + \frac{\log \log \log N}{2(2 \log \log N)^{1/2}} + \frac{w}{(2 \log \log N)^{1/2}} \right] = \exp\left[\frac{-2e^{-w}}{\pi^{1/2}}\right]$$

Proof: Several lemmas are needed to establish the theorem.

Lemma 2.2: For the testing problem under consideration:

$$-2 \log \lambda = \max_{1 \leq \nu < N} \frac{1}{(1 - \nu/N)(\nu/N)} \left[\frac{\sum_{i=1}^{\nu} y_i}{N^{1/2}}\right]^2$$

where $y_i = Y_i - \bar{Y}$.

Proof of lemma: Follows directly from (2.1).

The next lemma will utilize the properties of the Ornstein-Uhlenbeck process and Brownian bridge.

Definition 2.3: A continuous stochastic process $U(t)$ is called an Ornstein-Uhlenbeck process if $U(t)$ is stationary, Markov and Gaussian with $E[U(t)] = 0$ and $\mathrm{Cov}\,[U(t), U(s)] = \exp(-|t - s|)$ for any real t and s.

Definition 2.4: A continuous Gaussian process $B°(t)$ on $[0,1]$ is called a Brownian bridge if $B°(t) = B(t) - tB(1)$ where $B(t)$ is a standard Brownian motion on $[0,1]$.

Lemma 2.5: Let $y_1, \ldots, y_N$ be defined as above and let

$$W_N^* = \sup_{s \in S_N^+} |U(s)|$$

where $U(s)$ is an Ornstein-Uhlenbeck process and S_N^+ is a set defined below. Then W_N^* has the same distribution as $(-2 \log \lambda)^{1/2}$

Proof of lemma:

The result will be established by comparing $(-2 \log \lambda)^{1/2}$ to a Brownian bridge, which is related to the Ornstein-Uhlenbeck process by a simple transformation. The definition of S_N^+ will be constructed in the process.

First note

$$\sum_{i=1}^{\nu} y_i = \sum_{i=1}^{\nu} e_i - (\nu/N) \sum_{i=1}^{N} e_i$$

where e_i ~ i.i.d. $N(0,1)$, $i=1, \ldots, N$. Since e_i is Gaussian, $\{N^{-1/2} \Sigma_1^{\nu} y_i,\ \nu = 1, \ldots, N-1\}$ and $\{B^o(t),\ t \varepsilon T_N\}$ have the same joint distribution where

$$T_N = \left\{ \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N} \right\}$$

Similarly, the two sets of random variables $\{[(1-t)(t)]^{1/2} B^o(t),\ t \varepsilon T_N\}$ and $\{[(1 - \nu/N)(\nu/N)]^{1/2} N^{-1/2} \Sigma^{\nu} y_i,\ \nu=1, \ldots, N-1\}$ have the same joint distribution.

For the Brownian bridge there exists an Ornstein-Uhlenbech process $U(s)$ on the real line satisfying

$$B^o(t) = [t\,(1-t)]^{1/2}\, U((1/2)\log(\frac{t}{1-t}))$$

Let

$$S_N = \{ s \mid s = 1/2 \log(\frac{n}{N-n}),\ n = 1, \ldots, N-1\}$$

Then $W_N = \sup_{t \varepsilon T_N} [(1-t)t]^{-1/2} |B^{O}(t)|$ has the same distribution as $\sup_{s \varepsilon S_N} |U(s)|$.

Define S_N^+ by

$$S_N^+ = \{s^+ \mid s^+ = s + 1/2 \log(N-1),\ s \varepsilon S_N\}$$

Then $\sup_{s \varepsilon S_N} |U(s)|$ and $\sup_{s \varepsilon S_N^+} |U(s)|$ have the same distribution. Therefore $(-2 \log \lambda)^{1/2}$ has the same distribution as $W_N^* = \sup_{s \varepsilon S_N} |U(s)|$.

Lemma 2.6: Let U(s) be an Ornstein-Uhlenbeck process defined for $s \geq 0$.

Then

$$\lim_{N \to \infty} P \left[\sup_{0<s<\log(N)} |U(s)| < (2 \log \log N)^{1/2} + \frac{\log \log \log N}{2(2 \log \log N)^{1/2}} + \frac{w}{(2 \log \log N)^{1/2}} \right] = \exp \left[\frac{-2e^{-w}}{\pi^{1/2}} \right]$$

Proof of Lemma:

Using some results from Darling and Erdos (1956), the lemma follows easily. Let

$$T(\alpha) = \sup\{t \mid U(\tau) < \alpha,\ 0 \leq \tau \leq t\}$$

In other words $T(\alpha)$ is the time at which $U(t)$ first crosses α.

Darling and Erdos (1956) show that

$$\lim_{\alpha \to \infty} P\left[T(\alpha) > \frac{(2\pi)^{1/2}}{\alpha} e^{\alpha^2/2} z\right] = e^{-z}$$

And if

$$\frac{(2\pi)^{1/2}}{\alpha} e^{\alpha^2/2} z = \log N$$

then, as $N \to \infty$,

$$\alpha = (2 \log \log N)^{1/2} + \frac{\log \log \log N}{2(2 \log \log N)^{1/2}} - \frac{\log (\pi^{1/2} z)}{(2 \log \log N)^{1/2}}$$

$$+ o\left(\frac{1}{(\log \log N)^{1/2}}\right)$$

Combining these facts yield

$$\lim_{N \to \infty} P \left[\sup_{0<s<\log(N)} U(s) < \alpha \right] = \lim_{N \to B} P \left[T(a) > \log N \right]$$

$$= e^{-z} \qquad\qquad = \exp\left[\frac{-e^{-w}}{\pi^{1/2}}\right]$$

where $w = -\log(\pi^{1/2} z)$. Using the above expression for α produces:

$$\lim_{N \to \infty} P \left[\sup_{0<s<\log(N)} |U(s)| < (2 \log \log N)^{1/2} + \frac{\log \log \log N}{2(2 \log \log N)^{1/2}} \right.$$

$$\left. + \frac{w}{(2 \log \log N)^{1/2}} \right] = e^{-2z} = \exp\left[\frac{-2e^{-w}}{\pi^{1/2}}\right]$$

The next lemma requires some additional notation:

$$S_K = 1/2 \log \quad K \left(\frac{N-1}{N-K}\right)$$

$N(\alpha)$ = an integer such that $U(S_K) < \alpha$, $\quad K < N(\alpha)$

$U(S_K) \geqq \alpha$ for $K = N(\alpha)$

$K(\alpha)$ is the integer such that $S_{K(\alpha)-1} < T(\alpha) \leqq S_{K(\alpha)}$

(Recall $T(\alpha)$ is the time $U(t)$ first crosses α).

$L(\alpha) = S_{K(\alpha)} - S_{K(\alpha)} - 1$

$$\mu(\alpha) = \frac{(2\pi)^{1/2}}{\alpha} e^{\alpha^2/2}$$

Lemma 2.7:

There exists $\alpha(N)$ a function of N such that

$$\lim_{N \to \infty} \frac{\alpha}{(2 \log \log N)^{1/2}} = 1$$

and $\log N(\alpha) / T(\alpha) \to 1$ as $N \to \infty$.

Proof of lemma:

The lemma will be established by showing that $P[K(\alpha) \leqq N(\alpha) \leqq K(\alpha+\varepsilon)] \to 1$ as $N \to \infty$, where ε is a suitable function of N. The lemma will follow from this expression. Two results from Darling and Erdos (1956) are used:

1. $\lim_{\alpha \to \infty} P[T(\alpha) > \mu(\alpha)y] = e^{-y}$

2. $|T(\alpha+\varepsilon) - T(\alpha)| \to 0$, as $N \to \infty$ where $\varepsilon = 1/\alpha^2$.

Note that $N(\alpha) \geqq K(\alpha)$ by definition, so it suffices to show

(2.6) $$P[N(\alpha) \leqq K(\alpha+\varepsilon)] \to 1$$

as $\varepsilon \to 0$, $\alpha \to \infty$. Let $\varepsilon = 1/\alpha^2$.

To establish (2.6), calculate:

$$q = P \quad N(\alpha) > K(\alpha+\varepsilon) \;\Big|\; T(\alpha) > \frac{\mu(\alpha)}{\alpha}$$

For $K > 2$ note $(N-K+1)/(N-1) \leqq K/(K-1)$ and $(N-1)/(N-K) \leqq K$.

Hence:

$$L(\alpha) = S_{K(\alpha)} - S_{K(\alpha) - 1}$$

$$= -1/2 \log \left(\frac{(K(\alpha)\ (N-1)}{N-K(\alpha)} \right) - 1/2 \log \left(\frac{(K(\alpha)-1)\ (N-1)}{N-K(\alpha)\ +\ 1} \right)$$

$$\leqq \log K(\alpha) - \log(K(\alpha) - 1)$$

$$\leqq \frac{2}{K(\alpha)}$$

Also

$$T(\alpha) \leqq S_{K(\alpha)}$$

$$= 1/2 \log \quad K(\alpha) \frac{N-1}{N-K(\alpha)}$$

$$\leqq \log K(\alpha)$$

$$\Rightarrow \quad 1/K(\alpha) < e^{-T(\alpha)}$$

So for $T(\alpha) > \mu(\alpha)/\alpha$ this implies

$$L(\alpha) \leqq 2 \exp \left[\frac{-\mu(\alpha)}{\alpha}\right]$$

$$= 2 \exp \left[- (2\pi)^{1/2}/\alpha \; e^{\alpha^2/2}\right]$$

$$= \psi(\alpha)$$

The event $N(\alpha) \geqq K(\alpha + \varepsilon)$ implies that $U(t)$ having reached $\alpha + \varepsilon$, has decreased below α in a time span less than $\psi(\alpha)$. Recalling that the Ornstein-Uhlenbeck process is stationary:

$$q \leqq P[U(\psi(\alpha)) < \alpha \mid U(0) = \alpha+\varepsilon]$$

The conditional distribution of $U(\psi(\alpha))$ given $U(0) = \alpha + \varepsilon$ is normal with mean $(\alpha + \varepsilon)\, e^{-\psi(\alpha)}$ and variance $1-e^{-2\psi(\alpha)} \leqq 2\psi(\alpha)$. Therefore

$$q \leqq (2\pi\xi^2)^{-1/2} \quad \exp(-\xi^2/2)$$

where

$$\xi = \frac{(\alpha+\varepsilon)\ e^{-\psi(\varepsilon)}}{\sigma(\alpha)} \geqq \frac{(\alpha+\varepsilon)\ e^{-\psi(\alpha)}}{(2\psi(\alpha))^{1/2}}$$

$$\geqq \frac{\varepsilon}{2(\psi(\alpha))^{1/2}} \qquad \text{for } \alpha > \alpha_o$$

Since $\varepsilon = 1/\alpha$, $\varepsilon/(\psi(\alpha))^{1/2} \to \infty$, implying $q \to 0$.

Setting $y=1/\alpha$ and using the result of Darling and Erdos (1956) we see that $P\ [T(\alpha) > \mu(\alpha)/\alpha] \to 1$ as $N \to \infty$. Hence $P\ [K(\alpha)<N(\alpha)<K(\alpha+\varepsilon)] \to 1$. Further, note that the result of Darling and Erdos implies that $T(\alpha) \to \infty$, as $\alpha \to \infty$.

Now consider for $\alpha \to \infty$ the equation:

$$\mu(\alpha)\ z = \log N + z$$

The solution is

$$\alpha = (2 \log \log N)^{1/2} + \frac{\log \log \log N}{2(2 \log \log N)^{1/2}} - \frac{\log(\pi^{1/2} z)}{(2 \log \log N)^{1/2}} + o\left(\frac{1}{(2 \log \log N)^{1/2}}\right)$$

$$\Rightarrow \qquad P\ [T(\alpha) > \log N] \to e^{-z} \qquad \text{as } N \to \infty$$

$$\Rightarrow \qquad T(\alpha) = \log N + 0_p(1)$$

Now recall

$$S_{K(\alpha)-1} \leqq T(\alpha) \leqq S_{K(\alpha)}$$

$$\Rightarrow (K(\alpha)-1) \frac{N+1}{N-K(\alpha)+1} \leqq e^{2T(\alpha)} \leqq K(\alpha) \frac{N-1}{N-K(\alpha)}$$

For large N, $(N-1)/(N-K(\alpha)) \doteq N/(N-K(\alpha))$. So

$$e^{2T(\alpha)} \leqq K(\alpha) \frac{N}{N-K(\alpha)}$$

$$\Rightarrow \qquad K(\alpha) \leqq e^{2T(\alpha)} / 1 + \frac{e^{2T(\alpha)}}{N}$$

Similarly

$$K(\alpha) \leqq e^{2T(\alpha)} / (1 + \frac{e^{2T(\alpha)}}{N}) + 1$$

So for large N

$$K(\alpha) = e^{2T(\alpha)} / (1 + \frac{e^{2T(\alpha)}}{N}) + \delta \qquad 0 \leqq \delta < 1$$

Substituting for $T(\alpha)$ implies that $|N(\alpha) - K(\alpha)| = 0_p(1)$. Recalling

$$S_K = 1/2 \log \quad K\left(\frac{N-1}{N-K}\right)$$

we obtain

$$S_{K(\alpha)} = \log K(\alpha) + 0_p(1)$$

And $T(\alpha) + \delta_1 = S_{K(\alpha)}$ for $0 \leqq \delta_1 < .1$, implying $T(\alpha) = \log K(\alpha) - \delta_1 + 0_p(1)$, for large N. Hence

$$\frac{\log K(\alpha)}{T(\alpha)+\delta_1+0_p(1)} \rightarrow \quad \text{as } N \rightarrow \infty$$

Recalling expression (2.6) we obtain, for $0 \leqq \delta_2 < 1$:

$$P\ [K(\alpha) \leqq N(\alpha) \leqq K(\alpha+\varepsilon)] \rightarrow 1$$

$$\Rightarrow \quad P\left[\frac{\log K(\alpha)}{T(\alpha)+\delta_1+0_p(1)}\ (T(\alpha)+\delta_1+0_p(1)) \leqq \log N(\alpha) \leqq \frac{\log N(\alpha)}{T(\alpha+\varepsilon)+\delta_2+0_p(1)}\right.$$

$$\left.(T(\alpha+\varepsilon)+\delta_2+0_p(1))\right] \rightarrow 1$$

$$\Rightarrow \quad P\left[1 + \frac{\delta_1}{T(\alpha)} + \frac{0_p(1)}{T(\alpha)} \leqq \frac{\log N(\alpha)}{T(\alpha)} \leqq \frac{T(\alpha+\varepsilon)}{T(\alpha)} + \frac{\delta_2}{T(a)} + \frac{0_p(1)}{T(\alpha)}\right] \rightarrow 1$$

Recall the result of Darling and Erdos (1956), namely that $|T(\alpha+\varepsilon) - T(\alpha)| \overset{P}{\to} 0$. This implies $T(\alpha+\varepsilon)/T(\alpha) \overset{P}{\to} 1$ and therefore

$$\frac{\log N(\alpha)}{T(\alpha)} \overset{P}{\to} 1 \qquad \text{as } N \to \infty$$

The preceeding lemmas enable us to establish Theorem 2.1, which states

$$\lim_{N\to\infty} P_{H_0}\left[(-2\log\lambda)^{1/2} \leqq (2\log\log N)^{1/2} + \frac{\log\log\log N}{2(2\log\log N)^{1/2}} + \frac{w}{(2\log\log N)^{1/2}}\right] = \exp\left[\frac{-2e^{-w}}{\pi^{1/2}}\right]$$

To see this, consider:

$$\lim_{N\to\infty} P\left[\sup_{\nu}\left[\frac{\nu}{N}\left(1-\frac{\nu}{N}\right)\right]^{-1/2} < a\right] = \lim_{N\to\infty} P\left[\sup_{s\in S_N^+} U(s) < a\right]$$

$$= \lim_{N\to\infty} P\left[N(\alpha) > N\right]$$

$$= \lim_{N\to\infty} P\left[T(\alpha) > \log N\right]$$

$$= \lim_{N\to\infty} P\left[\sup_{0<s<\log N} U(s) < \alpha\right]$$

$$= e^{-z}$$

where

$$\alpha = (2 \log \log N)^{1/2} + \frac{\log \log \log N}{2(2 \log \log N)^{1/2}} - \frac{\log (\pi^{1/2} z)}{(2 \log \log N)^{1/2}}$$

$$+ o\left(\frac{1}{(2 \log \log N)^{1/2}}\right)$$

Substituting $w = -\log(\pi^{1/2} z)$ and recalling

$$\sup_{\nu} \left[\frac{\nu}{N}\left(1-\frac{\nu}{N}\right)\right]^{-1/2} N^{-1/2} \left|\sum_{1}^{\nu} y_i\right| = (-2 \log \lambda)^{1/2}$$

and

$$\lim_{N \to \infty} P\left[\sup_{0<s<\log N} |U(s)| < \alpha\right] = e^{-2z}$$

yields the desired result.

For σ^2 unknown the same asymptotic distribution holds under H_0, as demonstrated by:

Theorem 2.7:

Let $Y_1, \ldots, Y_N \sim$ i.i.d. $N(\mu, \sigma^2)$ and let

$$\lambda = \min_{1 \leq \nu < N} \left[\frac{S_0}{S_1(\nu)} \right]^{-N/2}$$

Then

$$\lim_{N \to \infty} P\left[(-2 \log \lambda)^{1/2} \leqq (2 \log \log N)^{1/2} + \frac{\log \log \log N}{2(2 \log \log N)^{1/2}} \right.$$

$$\left. + \frac{w}{(2 \log \log N)^{1/2}} \right] = \exp\left[\frac{-2 e^{-w}}{\pi^{1/2}} \right]$$

Proof of theorem:

Essentially $-2 \log \lambda$ can be written as the sum of two terms, one of which behaves asymptotically like $-2 \log \lambda$ when σ^2 is known. The other term in 0_p ((log log N)/N) and can be ignored.

Note:

$$-2 \log \lambda = \sup_{1 \leq \nu < N} - N \log (S_1/S_0)$$

$$= \sup_{\nu} \left[\frac{S_0 - S_1}{N^{-1} S_0} \right] + \sup_{\nu} N \; 0 \left(\frac{S_0 - S_1}{S_0} \right)^2$$

Note $S_0 = N\hat{\sigma}^2_{H_0}$ is consistent for σ^2, implying

$$-2 \log \lambda = \sup_{\nu} \left[\frac{S_0 - S_1}{\sigma^2_{H_0}} \right] + 0_P \left(\frac{\log \log N}{N}\right)$$

$$= \sup_{\nu} \left[(1-\nu/N)(\nu/N)\right]^{-1} \left[\frac{\sum_{i=1}^{\nu} y_i}{s_{H_0} N^{1/2}} \right]^2 + 0_P\left(\frac{\log \log N}{N}\right)$$

The result now follows from Theorem 2.1.

Acknowledgements: This research was supported in part by NSF Grant No. SOC 76-15271 and the US Census Bureau under the American Statistical Association Fellowship Program. This work is based on sections of the author's Ph.D. dissertation (Yale 1982) and the author gratefully acknowledges Professors F.J. Anscombe, John Hartigan and David Pollard and Dr. David Findley for valuable advice during its preparation.

REFERENCES

Darling, D.A. and Erdos, P. (1956). A limit theorem for the maximum of normalized sums of independent random variables. Duke Math. Journal, 23, 143-155.

Hawkins, D.M. (1977). Testing a sequence of observations for a shift in location. J. Amer. Statist. Assoc., 72, 180-186.

Shaban, S.A. (1980). Change point problems and two-phase regression: an annotated bibliography. International Statist. Rev., 48, 83-93.