Change-Point Detection Of Weak Signals: How To Use Signal Correlation?

Yao Xie, David Siegmund

November 1, 2012

Abstract

We explore the problem of how to exploit known signal temporal correlation is detecting weak signal. We model the signal as a time series with a known signal correlation structure, and proposed a novel maximum score test (MST) for weak signal detection. The MST avoids the computational expensive inversion of covariance matrix in the maximum likelihood test. We develop analytic approximation of significant level and standardized power function of MST, which are shown to be very accurate by simulation. We compare the MST detector with the maximum likelihood detector without using signal correlation.

1 Introduction

Detection of the emergence of a signals in noisy background arises in many applications: gene mapping, disease surveillance, and wireless communications [KS09, MF05, QCPS08, LFP08, LLD08]. In these models, the emergence of a signal is abrupt: the signal appears from a certain time. Before the change-point, the observation is a process noise plus observation noise. After the change-point, the observation is a signal plus observation noise. The signal may have different power from the noise, and/or different autocorrelation structure. Our goal is to detect the presence of such a change-point, using both the power and (possibly) autocorrelation change.

When the signal is weak, we may not be able to detect signal only based on signal power. One such example is in wireless cognitive radio [TS08]: when the signal power is comparable to the noise power, the signal cannot be detected by using the change in power only. This effect is called "SNR wall". In such scenario, we would naturally ask a question: does signal temporal correlation structure help, if we have such a priori information, and how to use such a priori information?

In this paper, we explore the problem of how to exploit known signal temporal correlation is detecting weak signal. We also tackle the problem when signal correlation will be helpful. We proposed a novel maximum score test (MST), assuming a known signal correlation structure. We assume the signal dynamic follows a time series model: autoregressive moving-average model. The maximum likelihood test (MLT) counterpart has a computational issue, because the covariance matrix has to been inverted in evaluating the likelihood function, and this has to be done for all possible changepoint location. MST avoids the compute inversion of covariance matrix. We have analytic approximation of significant level and standardized power function of the proposed score test, which are shown to be very accurate by simulation. We compare the MST detector with the maximum likelihood detector without using signal correlation. We showed that the detection power of MST is higher than MLT when signal has fairly significant temporal correlation and signal-to-noise ratio (SNR) is small. When signal power is fairly strong and signal temporal correlation is not significant, MLT may be better. This fact suggests a better detector for a wider range of SNR and temporal correlation, would be a mixture method that switches between MST and MLT depending on the signal correlation and SNR.

The paper is organized as follows. Section 2 shows the model. The maximum score test and maximum likelihood test with and without employing the signal correlation, are presented in Section 3.3. Sections 4 and 5 contains our main theoretical results: approximate significance level and standardized power function. Some numerical examples are shown in 7. Finally, Section 8 concludes the paper.

2 Model

Consider a sequence of observations $y_l, l = 1 \cdots N$ with fixed sample size N. There are two possibilities: the first one is that there is no signal present and therefore the observation is due to white observation noise only; another possibility is that a signal appears at some time k: before k the observations are due to white noise, and after k the observations are signal plus noise. Assume the signal is colored, with a known autocorrelation structure but determined by unknown parameters (e.g., an autoregressive signal with known order but unknown model parameters.) The signal may also have different power from the signal.

In our problem, the test statistics form a two dimensional random fields, one dimension is a smooth random field, and the other dimension behaves like a random walk. The tail of maximum of the random field has been studied for several cases: for sequential change-point detection [Sie88], where the two dimensional random field has both dimensions behave like random field; the likelihood ratio test for a change-point in simple linear regression [KS89], where the two dimensional random fields has one dimension as random walk, and one dimension as a smooth random process. In these cases the random fields can be decomposed as a sum of two random processes, which enables a tractable tail approximation. Our problem also enjoys this property. A general discussion for cases where two dimensions are both smooth random process is presented in [NSY08].

Another related problem is maximum score test for spatial correlation [CS09]. The test statistic therein has a similar form to the test statistics in our problem, although the test statistics in [CS09] is a one dimensional smooth random field, and we have a one more dimension to vary over the change-point location.

We are interested in finding the approximate significant level of the maximum score test. This is an interesting case in that the random field $\{Z(k, \theta)\}$ has one dimension in k behaves like random walk, and in θ behaves like a smooth random process. Also $\{Z(k, \theta)\}$ is non-Gaussian. Our main result is the following Theorems. The method we described to obtain the significant level is quite general. It can be used to obtain the tail probability of maxima of random field, not necessarily Gaussian, where in one parameter subspace behaves like random walk, and in rest of the parameter subspace it is a smooth random process. In our paper we assume signals are real. However, these tests can be extended to complex signals with some efforts (complex signals are commonly used in communications systems).

The signal models for these two possibilities are

$$\text{Case 0:} \quad y_l = n_l, \quad l = 1, \cdots N, \tag{1}$$

$$\text{Case 1:} \quad \begin{cases} y_l = n_l, & l = 1, \cdots k; \\ y_l = x_l + n_k, & l = k+1, \cdots N. \end{cases} \tag{2}$$

where x_l, n_l are the signal and white noise samples at time l. We formulate this problem as the following hypothesis test:

$$\begin{array}{ll} H_0: & [y_1, \cdots, y_N]^T \sim \mathcal{N}(0, \sigma_0^2 I_N), \\ H_1: & [y_1, \cdots, y_k]^T \sim \mathcal{N}(0, \sigma_0^2 I_k), \\ & [y_{k+1}, \cdots y_N]^T \sim \mathcal{N}(0, \Sigma), \end{array} \tag{3}$$

where T denotes transpose, and I_k is the identity matrix with dimension k. The covariance matrix of the post change observation vector $\mathbf{y}_{k+1} \triangleq [y_{k+1}, \cdots, y_N]$ is given by the sum of the covariance of the observation noise and the covariance of the signal vector $\mathbf{x}_{k+1} = [x_{k+1}, \cdots, x_N]^T$:

$$\Sigma = \sigma_0^2 I_{N-k} + \tau V_{N-k,\theta}. \tag{4}$$

Here k is the change point, σ_0^2 is the known variance of the observation noise, $\tau V_{N-k,\theta} \in \mathbb{R}^{(N-k)\times(N-k)}$ is the covariance matrix of $\mathbf{x}_{k+1}$, τ characterizes the signal power, and θ is an unknown scalar parameter, whose value is in the interval $\theta \in (\theta_1, \theta_2)$. Assuming the sequence is wide sense stationary, and hence $V_{N-k,\theta}$ is symmetric. The function form of the covariance $V_{N-k,\theta}$ with argument θ is known, and amplitude of the change τ is unknown.

3 Detectors

Before we introduce the maximum score test detector, we will present two other commonly used detectors: the energy detector, and the maximum likelihood detector without employing signal correlation.

3.1 Energy Detector

The energy detector only compares the change in power in a chunk of data with the noise power σ_0^2:

$$\text{Energy detector:} \quad \frac{1}{N}\sum_{i=1}^{N} y_i^2 \geq b.$$

3.2 Maximum Likelihood (ML) Detector

3.2.1 ML Employing Signal Correlation

This detector will give an performance upper bound for the detectors we considered here. The log likelihood ratio of hypothesis H_1 versus H_0 is given by:

$$\mathcal{L}(\tau, k, \theta) = \frac{1}{2}\mathbf{y}_k^T \left(\frac{I}{\sigma_0^2} - \Sigma^{-1} \right) \mathbf{y}_k - \frac{1}{2} \log \left(\frac{|\Sigma|}{(\sigma_0^2)^{(N-k)}} \right). \tag{5}$$

where $\Sigma(\tau, k, \theta) = \sigma_0^2 I_{N-k} + \tau V_{N-k,\theta}$. The detection rule is given by

$$\text{(Maximum likelihood employing covariance:)} \quad \max_{\substack{0 \le k \le N - N_0, \\ \tau \ge 0, \theta \in \Theta}} \mathcal{L}(\tau, k, \theta) \ge b$$

for some threshold $b \in \mathbb{R}$.

To analyze this, we may again approximately decompose it as a Gaussian random walk. (However, the result for this decomposition is not so good.) (And we cannot decompose it as random walk by change-of-coordinate system). To simply probability a bit in significance level analysis, we may assume τ takes the true value, which gives an lower bound on significance level.

(Show its drift under H0 is negative and under H1 is positive.)

3.2.2 ML Employing No Signal Correlation

If we ignore the signal correlation and assumes $V_{N-k,\theta} = \tau I_{N-k}$, then the signal covariance matrix is given by $\Sigma = (\sigma_0^2 + \tau) I_{N-k}$. This assumes the signal differs from noise only in power. The signal power is unknown and is assumed to be within (τ_1, τ_2). Similar to [LFP08], we can show the generalized likelihood ratio is given by:

$$\begin{aligned}
\mathcal{L}_I(\tau, k) &= \frac{\tau}{2(\tau + \sigma_0^2)\sigma_0^2} \mathbf{y}_k^T \mathbf{y}_k - \frac{(N-k)}{2} \ln \left(\frac{\tau + \sigma_0^2}{\sigma_0^2} \right), \\
&= \frac{(N-k)\rho}{2\sigma_0^2(1+\rho)} \hat{\sigma} - \frac{(N-k)}{2} \ln (1 + \rho).
\end{aligned}$$

where $\rho = \frac{\tau}{\sigma_0}$ is the signal-to-noise ratio (SNR), and $\hat{\sigma}^2 = \frac{1}{N-k} \sum_{i=k+1}^{t} y_i^2$ is the maximum likelihood estimator of σ_0^2. The test is given by

$$\text{(Maximum likelihood ignoring covariance:)} \quad \max_{\substack{0 \le k \le N - N_0 \\ \tau \in [\tau_1, \tau_2], \theta \in \Theta}} \mathcal{L}_I(\tau, k) \ge b$$

3.3 Maximum Score Detector Employing Signal Correlation

The maximum likelihood test employing signal correlation, $\mathcal{L}$ in (5), has a major computational cost from computing the the matrix inversion Σ^{-1} for all possible values of k, θ, and τ. However, we can avoid this matrix inversion computational cost by considering the maximum score test. For this purpose, we consider the derivative of $\mathcal{L}$ with respect to τ and evaluate at $\tau = 0$ (only a partial score expansion), and keep the other two parameters k and θ. This yields:

$$\begin{aligned}
\left.\frac{\partial \mathcal{L}}{\partial \tau}\right|_{\tau=0} &= \left.\left[\frac{1}{2}\mathbf{y}_{k+1}^T \Sigma^{-1} V_{N-k,\theta} \Sigma^{-1} \mathbf{y}_{k+1} - \frac{1}{2}\mathrm{tr}(\Sigma^{-1} V_{N-k,\theta})\right]\right|_{\tau=0}. \\
&= \frac{1}{2\sigma_0^4}\mathbf{y}_{k+1}^T V_{N-k,\theta}\mathbf{y}_{k+1} - \frac{1}{2\sigma_0^2}\mathrm{tr}(V_{N-k,\theta}), \\
&= \frac{1}{2\sigma_0^4}\mathrm{tr}(V(\mathbf{y}_k\mathbf{y}_k^T - \sigma_0^2 I)) \qquad (6)
\end{aligned}$$

In the following $P\{\cdot\}$, $\mathbb{E}\{\cdot\}$ and $\mathrm{VAR}\{\cdot\}$ denote the probability, mean and variance under the null hypothesis. We can verify the mean of the derivative under the null hypothesis is zero: $\mathbb{E}\left\{\left.\frac{\partial \mathcal{L}}{\partial \tau}\right|_{\tau=0}\right\} = 0$. And the covariance under null hypothesis is given by

$$\begin{aligned}
\mathrm{VAR}\left\{\left.\frac{\partial \mathcal{L}}{\partial \tau}\right|_{\tau=0}\right\} &= \left(\frac{1}{2\sigma_0^4}\right)^2 \mathbb{E}\left\{\mathbf{y}_{k+1}^T V_{N-k,\theta}\mathbf{y}_{k+1}\right\}^2 - \left(\frac{1}{2\sigma_0^2}\mathrm{tr}(V_{N-k,\theta})\right)^2. \\
&= \frac{1}{2\sigma_0^4}\mathrm{tr}(V_{N-k,\theta}V_{N-k,\theta}^T). \qquad (7)
\end{aligned}$$

Derivations for this expression can be found in the Appendix B. The score statistics is given by $\left.\frac{\partial \mathcal{L}}{\partial \tau}\right|_{\tau=0}$ normalized by its mean and variance:

$$\begin{aligned}
Z(k,\theta) &\triangleq \frac{\frac{1}{\sigma_0^2}\mathbf{y}_{k+1}^T V_{N-k,\theta}\mathbf{y}_{k+1} - \mathrm{tr}(V_{N-k,\theta})}{\sqrt{2\mathrm{tr}(V_{N-k,\theta}V_{N-k,\theta}^T)}}. \qquad (8) \\
&= \mathrm{tr}\left[\frac{V_{N-k,\theta}}{\sqrt{2\mathrm{tr}(VV^T)}}\left(\frac{\mathbf{y}_{k+1}\mathbf{y}_{k+1}^T}{\sigma_0^2} - I_{N-k}\right)\right] \qquad (9)
\end{aligned}$$

Note that $Z(k,\theta)$ is a two dimensional random field in $k \in \{1, \cdots n\}$ and $\theta \in (\theta_1, \theta_2)$, with zero mean and unit variance. The second expression shows that $Z(k,\theta)$ can be interpreted as the inner product of the parameter space and the data.

The maximum score test rejects the null hypothesis when

$$\text{Maximum score test employing covariance} \max_{\substack{0\le k\le N-N_0, \\ \theta\in(\theta_1,\theta_2)}} Z(k,\theta) \ge b, \qquad (10)$$

for some threshold $b > 0$. The N_0 is chosen large enough so we accept H_1 at least after we have N_0 samples for the post-change distribution.

Remkark: Note that the maximum score test and maximum likelihood test, with they both ignore the signal covariance: $V = \tau I_{N-k}$, the two tests are the same (after normalized to have zero-mean and variance 1) - both rely on $\mathbf{y}_k^T\mathbf{y}_k$.

4 Approximation of Significance Level

4.1 Power detector

Under H_0, $\sum_{i=1}^N \frac{y_i^2}{\sigma_0^2}$ is distributed as χ_N^2. So the significance level is given by $P_0(\sum_{i=1}^N (y_i/\sigma_0)^2 \ge Nb/\sigma_0^2) = 1 - F(Nb/\sigma_0^2)$ where $F(\cdot)$ is the cumulative distribution function of the χ_N^2 distribution.

4.2 Maximum likelihood test without exploiting signal correlation

Theorem 1. *The significance level of a maximum likelihood test (refer to a number) is given by*

$$b^{1/2}e^{-b}\int_{\rho:\eta_1\leq\frac{1}{2}(\rho-\log(1+\rho))\leq\eta_2}\frac{1}{(1+\rho)^2}\frac{1}{\sqrt{\rho-\log(1+\rho)}}\frac{d\rho}{\sqrt{2\pi}}\quad (\text{ change of variable } \rho=\frac{\tau}{\sigma_0^2}).$$

The method we used to evaluate the significance level of this test is different from the method we used for max score test. Because for this case, the test statistics in k can be approximated as an exponential random walk - a special structure that we could employ.

4.3 Maximum score test exploiting signal correlation

If ignore the covariance structure of the signal, which is equivalent of assuming $V = I_{N-k}$:

$$\text{maximum score test without employing signal correlation:}\quad \frac{1}{2\sigma_0^4}\mathbf{y}_{k+1}^T y_{k+1}-\frac{N-k}{2\sigma_0}\gtrless b, \tag{11}$$

where b is a threshold.

4.3.1 Scalar Parameter

When the parameter θ is a scalar, we use a discrediting method described later on, to obtain the approximate significant level as

Theorem 2. *The asymptotic significance level (under H_0) of the maximum score test for large b and $\theta\in\mathbb{R}$ is given by*

$$P\left(\max_{\substack{0\leq k\leq N-N_0,\\ \theta\in(\theta_1,\theta_2)}} Z(k,\theta)\geq b\right)\sim$$
$$\frac{1}{\sqrt{\pi}}\sum_{k=0}^{N-N_0}\int_{\theta_1}^{\theta_2} g(k,\theta)|\gamma(k,\theta)|\frac{b^2\mu(k,\theta)}{2(N-k)}\nu\left(\sqrt{\frac{b^2\mu(k,\theta)}{N-k}}\right)d\theta \tag{12}$$

where

$$
\begin{aligned}
g(k,\theta) &= \frac{\exp\{-\xi_0(k,\theta)b + \psi(\xi_0(k,\theta))\}}{\sqrt{2\pi\, Var_{\xi_0}\{Z\}}}, \\
\xi_0(k,\theta) &> 0 \text{ is the solution of } tr\left[(I_{N-k} - 2\xi_0\tilde{V})^{-1}\tilde{V}\right] - tr(\tilde{V}) = b, \\
\psi(\xi) &= -tr(\xi\tilde{V}) - \frac{1}{2}\log|I - 2\xi\tilde{V}|, \\
\tilde{V} &= \frac{V_{N-k,\theta}}{\sqrt{2tr(V_{N-k,\theta}V_{N-k,\theta}^T)}}, \\
Var_{\xi_0}\{Z\} &= 2tr\left[(I_{N-k} - 2\xi_0\tilde{V})^{-1}\tilde{V}(I - 2\xi_0\tilde{V})^{-1}\tilde{V}\right], \\
\mu(k,\theta) &= (N-k)\left[\frac{tr\left(V_{N-k+1,\theta}V_{N-k+1,\theta}^T\right)}{tr\left(V_{N-k,\theta}V_{N-k,\theta}^T\right)} - 1\right], \\
\gamma(k,\theta) &= \frac{tr\left(V'_{N-k,\theta}V_{N-k,\theta}^T\right)}{tr\left(V_{N-k,\theta}V_{N-k,\theta}^T\right)}.
\end{aligned} \tag{13}
$$

The special function (given in Corollary 8.44 of [Sie85] or Page 112 of [SY07]):

$$
\begin{aligned}
\nu(x) &= 2x^{-2}\exp\left\{-2\sum_{n=1}^{\infty}\frac{1}{n}\Phi\left(-\frac{1}{2}x\sqrt{n}\right)\right\} \quad (x>0) \\
&\approx \frac{\frac{2}{x}\left[\Phi\left(\frac{x}{2}\right) - \frac{1}{2}\right]}{\frac{x}{2}\Phi(\frac{x}{2}) + \phi\left(\frac{x}{2}\right)}
\end{aligned} \tag{14}
$$

where $\phi(x)$ and $\Phi(x)$ are the unit normal density and distribution function, respectively.

Remarks:

- The accuracy of solution ξ_0 is very important to the final accuracy of the approximate level. It can be solved numerically with the initial value getting from second order approximation of the $\phi(\xi)$: initial value $\xi_0^{initial} = \frac{-1+\sqrt{1+16b\mathrm{tr}(\tilde{V}^3)}}{8\mathrm{tr}(\tilde{V}^3)}$.
- There is not term involving both $\mu(k,\theta)$, which is related to the random walk dimension, and $\gamma(k,\theta)$, which is related to the smooth process in parameter space θ. This is a consequence that the random field can be locally decoupled.
- This formula is also consistent with the Rice's formula. Since the random field can be decoupled, the smooth process in the parameter space has to be proportional to $\int_\theta \sqrt{-\frac{\partial^2\rho(\theta)}{\partial^2\theta}}d\theta = \int_\theta \sqrt{2\gamma^2(\theta)}d\theta = \sqrt{2}|\gamma(\theta)|$, where we have used the local covariance structure.

4.3.2 Vector parameter

Theorem 3. *The asymptotic approximate significance level for large b and $\theta \in \Theta \subset \mathbb{R}^d$ is given by*

$$P\left(\max_{\substack{0\le k\le N-N_0,\\ \theta\in\Theta}} Z(k,\theta) \ge b\right) \sim$$
$$\frac{1}{(2\pi)^{\frac{d}{2}}}\sum_{k=0}^{N-N_0}\int_{\theta\in\Theta}\frac{[b\xi_0(k,\theta)]^{\frac{d}{2}}}{\xi_0(k,\theta)}g(k,\theta)|H(k,\theta)|^{\frac{1}{2}}\frac{b^2\mu(k,\theta)}{2(N-k)}\nu\left(\sqrt{\frac{b^2\mu(k,\theta)}{N-k}}\right)d\theta \quad (15)$$

where the new quantity involved in the integral other than that are in Theorem 2 is the Hessian of the covariance matrix (essentially the Fisher information matrix)

$$\begin{aligned} H(k,\theta) &= -\left.\frac{\partial^2\mathbb{E}\left\{Z(k,\theta)Z(k,s)\right\}}{\partial^2 s}\right|_{s=\theta}, \\ &= \mathbb{E}\left\{\dot{Z}(k,\theta)\dot{Z}^T(k,\theta)\right\}. \end{aligned} \quad (16)$$

where the expectation is under the Null hypothesis, and the dot denotes derivative with respect to θ.

Remark: Note that for $d = 1$, $H(k,\theta) = 2\gamma^2(k,\theta)$, and under Gaussian assumption, $\xi_0(k,\theta) = b$, and then (21) reduces to (12).

5 Standardized Power Function

We will compare the standardized power function of the tests. If the significant level α as a function of threshold b is given by: $\alpha = f(b)$. For a fixed α, we can invert f to find the corresponding threshold: $b = f^{-1}(\alpha)$. Suppose the test statistic is T, and we are interested in

$$P_{\mathsf{H}_1}(T \ge f^{-1}(\alpha)).$$

A heuristic is that under H_1, most power comes from the probability at the true parameters: change point location k, power ξ, and correlation coefficients θ. And we can also approximate the test statistic using Gaussian distribution:

$$P_{\mathsf{H}_1}(T \ge f^{-1}(\alpha)) \approx 1 - \Phi\left(\frac{f^{-1}(\alpha) - E_{\mathsf{H}_1}\{T\}}{\sqrt{\mathrm{VAR}_{\mathsf{H}_0}(T)}}\right).$$

Some some simple calculation, we have the following quantities in the table:

Remark:

We can switch between the two methods based on which regime of signal power we are in (if we have an expect for the power of the signal).

Table 1: The Mean and Variance of Test Statistics Needed for Standardized Power Calculation

	$E_{\mathsf{H}_1}\{T\}$	$\text{VAR}_{\mathsf{H}_0}(T)$
Max Likelihood $V = I$	$\frac{\rho}{2(1+\rho)}[\text{tr}(I_{N-k}) + \rho\text{tr}(V_{N-k}(\theta))] - (N-k)\frac{1}{2}\log(1+\rho)$	$\frac{\rho^2}{2(1+\rho)^2}\text{tr}(I_{N-k})$
Max Score Test	$\frac{\rho}{\sqrt{2}}\sqrt{\text{tr}(V_{N-k}(\theta)^2)}$	1
Full Max Likelihood with V	$-\frac{1}{2}[\log\frac{\lvert\Sigma\rvert}{(\sigma_0^2)^{(N-k)}}] + \frac{1}{2}\text{tr}(\frac{\Sigma}{\sigma_0^2} - I)$	$\frac{1}{2}\text{tr}\left[(\frac{I}{\sigma_0^2} - \Sigma^{-1})^2\right]$

6 Sequential Analysis

Now we consider the sequential score test: the online version of the tests. The test repeats as a new sample is acquired, until the likelihood ratio reaches a threshold. The sequential maximum score test can be written as:

$$\max_{\substack{0\leq k\leq N-N_0, 0\leq N\leq M\\ \theta\in(\theta_1,\theta_2)}} Z_k^N(\theta) \geq b, \tag{17}$$

where M is the maximum number of samples.

The performance of the sequential test is usually evaluated using its significance level (average run length is also used sometimes as a less accurate performance metric under H_0), and the detection delay: the number of samples needed to reach threshold level when a change happens.

We study the significance level of the significance level following the same strategy. First, we'll find the local covariance of $Z_k^N(\theta)$. Note that, since the process is assumed to be stationary, decreasing k or increasing N has the same effect: increasing the number of post-change samples in our analysis. So follow a very similar analysis we had for the fixed-sample case, we have

$$\mathbb{E}\left\{Z_n^{N_1}(\theta_1) Z_m^{N_2}(\theta_2)\right\} = \frac{\text{tr}\left(V_{N_1-n,\theta_1} V_{N_1-n,\theta_2}^T\right)}{\left[\text{tr}\left(V_{N_1-n,\theta_1} V_{N-n,\theta_1}^T\right)\text{tr}\left(V_{N_2-m,\theta_2} V_{N_2-m,\theta_2}^T\right)\right]^{1/2}}. \tag{18}$$

assuming $m \leq n$, $N_2 \geq N_1$. Now follow a similar local covariance analysis, we have

$$\mathbb{E}\left\{Z_k^N(\theta) Z_{k-i}^{N+j}(\theta+\delta)\right\} \tag{19}$$

$$\approx 1 - \gamma^2(N,k,\theta)\delta^2 - \frac{\mu(N,k,\theta)}{2(N-k)} i - \frac{\mu(N,k,\theta)}{2(N-k)} j + o(\delta^2) + o(i+j). \tag{20}$$

where $i > 0$, $j > 0$, $\delta > 0$. This shows that by allowing N to vary, the test statistics becomes a three dimensional random field: two independent random walks on the sample size N and change point location k dimensions, and a smooth random field along in parameter space Θ independent of the other dimensions. Based on this observation, following the same procedure we had for deriving the significance level based on Mill's ratio, now we can decompose the score test statistics as a product of three terms, in stead of two terms. And we would expect the significance level to be given by:

$$P_\infty\left(\max_{\substack{0\le k\le N-N_0, 1\le N\le M\\ \theta\in\Theta}} Z_k^N(\theta)\ge b\right)\sim$$

$$\frac{1}{(2\pi)^{\frac{d}{2}}}\sum_{N=1}^{M}\sum_{k=0}^{N-N_0}\int_{\theta\in\Theta}\frac{[b\xi_0(k,\theta)]^{\frac{d}{2}}}{\xi_0(k,\theta)}g(k,\theta)|H(k,\theta)|^{\frac{1}{2}}\left[\frac{b^2\mu(k,\theta)}{2(N-k)}\nu\left(\sqrt{\frac{b^2\mu(k,\theta)}{N-k}}\right)\right]^2 d\theta \quad (21)$$

The expected detection delay is given by:

$$E_k^\theta[\max_{\theta',k'} Z_k'^N(\theta')]\approx E_k^\theta[Z_k^N(\theta)]=E_k^\theta[Z_k^N(\theta)-b]+b=E_k^\theta[N-k+1|N\ge k]E(\text{“increment”}). \quad (22)$$

the first equation is because under P_k^θ, the probability of max is dominated by the value at true parameter. The last equality is Wald's equation. So to find the expected detection delay, we need two quantities: E(“increment”) and $E_k^\theta[Z_k^N(\theta)-b]$, the expected overshoot.

7 Numerical Examples

7.1 Approximated significant level for AR(1)

The first example is for the autoregressive process of order 1 AR(1), with $\theta\in[0.1,0.5]$, with $N=100$, $N_0=3$. The AR(1) process evolves as

$$x_{l+1}=\theta x_l+\varepsilon_l.$$

The process noise ε_ls are i.i.d. normal with zero mean and unit variance. The covariance matrix for AR(1) is:

$$V=\begin{pmatrix} 1 & \theta & \theta^2 & \cdots & \theta^{N-k-1}\\ \theta & 1 & \theta & \cdots & \theta^{N-k-2}\\ & & \ddots & & \end{pmatrix}_{N\times k, N\times k}. \quad (23)$$

We choose $V(\theta)$ in these forms because the corresponding $V'(\theta)$s have simple expression. The constant times these $V(\theta)$ can be incorporated in τ.

The Monte Carlo results are from 1000 trails. All other parameters are the same as the example one.

Table I: Approximate Significant Level for AR(1)

b	Monte Carlo	Formula
3.5000	0.1170	0.1045
4.5000	0.0560	0.0518
5.5000	0.0250	0.0242
6.5000	0.0090	0.0112
7.0000	0.0050	0.0076

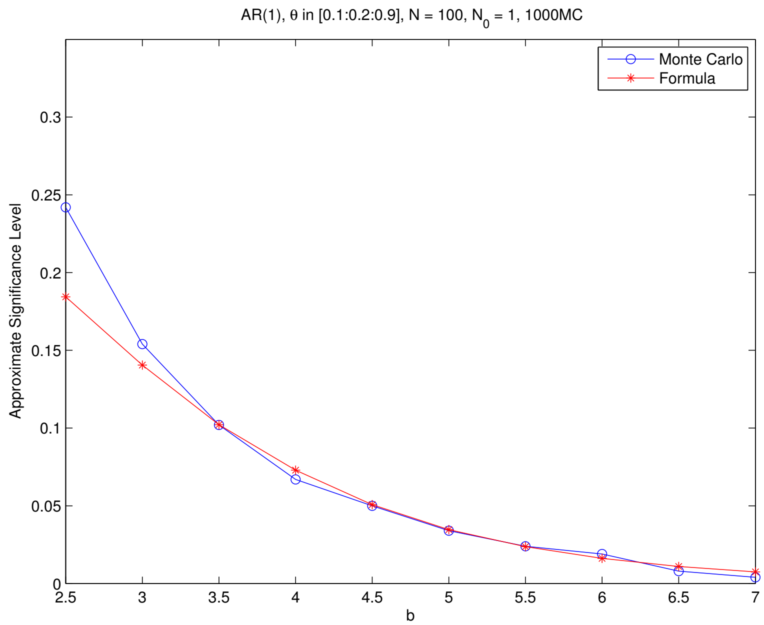

Figure 1: The approximate significance level versus b, from 1000 Monte Carlo simulations, and from our formula in Theorem 12.

7.2 Approximated significant level for ARMA(1,1)

The example for ARMA(1,1) demonstrate the performance of our approximation for higher dimensional parameter space case. ARMA(1,1) has parameters θ and ϕ and hence $d = 2$. The AR(1) process evolves as

$$x_{l+1} + \phi x_l = \theta \varepsilon_l + \varepsilon_{l+1}.$$

The covariance matrix of ARMA(1,1) process is:

$$V = \begin{pmatrix} 1+\theta^2-2\phi\theta & (\phi-\theta)(1-\phi\theta) & \phi(\phi-\theta)(1-\phi\theta) & \cdots \\ (\phi-\theta)(1-\phi\theta) & 1+\theta^2-2\phi\theta & (\phi-\theta)(1-\phi\theta) & \cdots \\ \phi(\phi-\theta)(1-\phi\theta) & \ddots & \ddots & \ddots \end{pmatrix}_{N\times k, N\times k} . \quad (24)$$

Table I: Approximate Significant Level for ARMA(1, 1)

b	Monte Carlo	Formula
3.5000	0.0780	0.0667
4.5000	0.0340	0.0341
5.0000	0.0260	0.0235
6.0000	0.0120	0.0109
6.5000	0.0060	0.0074

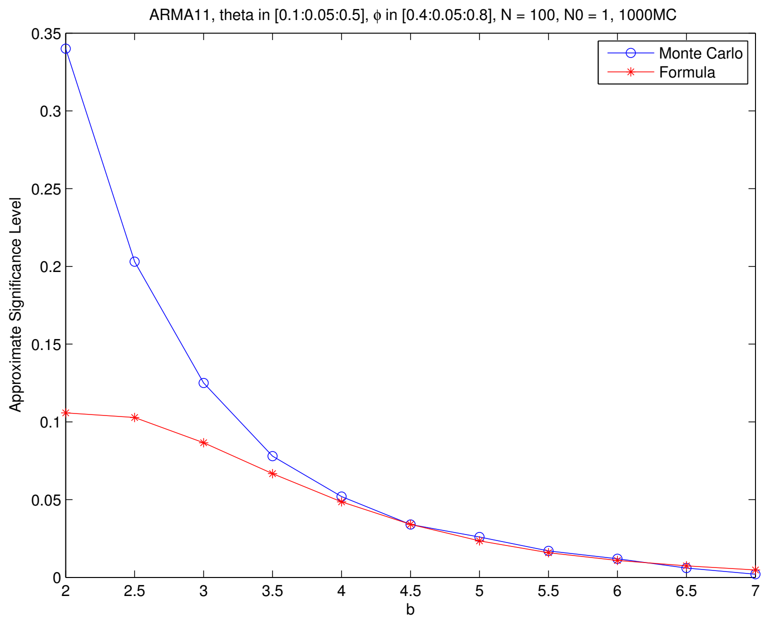

Figure 2: The approximate significance level versus b, from 1000 Monte Carlo simulations, and from our formula in Theorem 12.

7.3 Comparison of Standardized Power Function

We obtain their power: the probability that their test statistics exceed the threshold under H_1, using 1000 Monte Carlo trails. We simulate 1000 AR(1) sequences with $N = 100$, $k_0 = 50$, $\tau = 1$, $\theta_0 = 0.5$. For the maximum score test, the parameter set is $\theta \in [0.3, 0.8]$. The maximum of the score is searched within this set with a meshgrid size of 0.1.

The ROC curves: significant level versus power, is shown in Fig. **??**. Note that the score test has much better ROC performance than the other two tests when the true parameter lies in the assumed set.

8 Conclusions

We present a novel maximum score test (MST) for weak signal detection by exploiting the temporal correlation of the signal. Besides the merit of incorporating signal correlation, our score detector also avoids the computational complexity of covariance matrix inversion incurred by the corresponding maximum likelihood statistic assuming signal correlation. We provide a theoretical approximation to the p-value of the score detector, which can be used to determine the threshold efficiently. We demonstrate that our approximation is quite accurate, and that our score detector has an advantage when the signal is weak and the temporal correlation is not too weak. We studied the significance level and standardized power function of the proposed score test, compared those with a maximum likelihood test that does employ the signal covariance. We showed that the performance of MST is better than the maximum likelihood test (MLT) counterpart, when the signal has a strong correlation in time, and when the signal-to-noise ratio (SNR) is small. Otherwise the MLT is better than MST. One possible reason is because MST amplifies noise and has a higher significance level. This indicates a mixture method which switches between MST and MLT depending on the signal correlation and SNR, may detect signal in a wide range of scenarios.

A Proof of Theorems

A.1 Scalar Parameter

We will go through the following main steps in obtaining the asymptotic in Theorem 2:

(1) Write the probability in (12), which is essentially P{the random field $Z(k, \theta)$ exceed threshold at some point}, as a sum of probabilities of the last hitting events: P{randomly field "last" goes above threshold at (k_0, θ_0)}, running k_0 and θ_0 over the entire supports of k and θ.

(2) Rewrite the last hitting probability as product of two parts: part I, the probability that the random field go above the threshold b at the last hitting time, and part II, the conditional probability that the "future" random field values are all below the threshold.

(3) Approximate part I by using the cummulant generating function of the random field (since normal approximation to $Z(k, \theta)$ may not sufficient.), which leads to $g(k, \theta)$ in (12).)

(4) Approximate part II by employing the covariance structure of random field, which enables us to decompose the two dimensional random fields as a sum of two one dimensional Gaussian random process.

(5) Plug in the above results, sum and integrate over the k and θ spaces.

A.1.1 Last Hitting Time Formula

First, discretize the parameter $\theta \in (\theta_1, \theta_2)$, $k \in \{1, \cdots N - N_0\}$ by rectangular mesh grid of size of $\frac{\Delta}{\sqrt{N}}$ times 1, where $\Delta > 0$ is a small number. The size of the mesh is chosen to balance the difference in the order of the variance in these two coordinates. Then the significance level can be approximated as

$$P\left\{\max_{(i,j)\in D} Z\left(i, j\frac{\Delta}{\sqrt{N}}\right) \geq b\right\} \tag{25}$$

where the index set

$$D = \left\{(i,j) : 0 \leq i \leq N, \theta_1 \leq j\frac{\Delta}{\sqrt{N}} \leq \theta_2\right\} \tag{26}$$

covers the parameter space. Let the index set $J(i_0, j_0)$ denotes everything to the "future" of the current index (i_0, j_0) (upper and to the right of (i_0, j_0)) in the random field.

$$J(i_0, j_0) = \{(i,j) \in D : j \geq j_0, \quad \text{or} \quad j = j_0 \quad \text{and} \quad i \geq i_0.\} \tag{27}$$

Using the "last hitting time" decomposition, we can rewrite (25) as

$$
\begin{aligned}
& P\left\{\max_{(i,j)\in D} Z\left(i, j\frac{\Delta}{\sqrt{N}}\right) \geq b\right\} \\
\sim & \sum_{(i_0,j_0)\in D} P\left\{Z\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right) \geq b, \max_{(i,j)\in J(i_0,j_0)} Z\left(i, j\frac{\Delta}{\sqrt{N}}\right) \leq b\right\} \\
= & \sum_{(i_0,j_0)\in D} \int_0^\infty P\left\{Z\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right) = b + \frac{x}{b}\right\} \frac{dx}{b} \cdot \\
& P\left\{\max_{(i,j)\in J(i_0,j_0)} Z\left(i, j\frac{\Delta}{\sqrt{N}}\right) \leq b \middle| Z\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right) = b + \frac{x}{b}\right\}. \qquad (28)
\end{aligned}
$$

Next we will find approximations for Part I, the probability $P\left\{Z\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right) = b + \frac{x}{b}\right\} \frac{dx}{b}$, and Part II, the conditional probability $P\left\{\max_{(i,j)\in J(i_0,j_0)} Z\left(i, j\frac{\Delta}{\sqrt{N}}\right) \leq b \middle| Z\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right) = b + \frac{x}{b}\right\}$, respectively.

A.1.2 Skewness Correction

Note that $Z(k, \theta)$ is a quadratic function of the normal distributed data vector, so normal distribution is less likely to be a good approximation for $Z(k, \theta)$. Also, since the part II in (28) is smaller than Part I, getting an accurate approximate for Part II is important. The normal approximation is better for the mean than for the tail of the distribution. So for a better approximation we can use change-of-measure to shift the mean of the measure to the threshold to use the normal approximation, and also use the complete cumulative generating function. The details are given in in Appendix C, we have:

$$
P\left\{Z\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right) = b + \frac{x}{b}\right\} \frac{dx}{b} \approx g\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right) \exp\left(-\frac{\xi_0}{b}x\right) \frac{dx}{b}. \qquad (29)
$$

where g and ξ_0 are defined in (16).

A.1.3 Local Analysis of Covariance

We will show that the local covariance under H_0 has no cross term in the change-point location k and the parameter l. Due to this property, the two dimensional random field can be decomposed as sum of two independent random processes in k and θ. The random processes in k behaves like a random walk, in θ is like a smooth random process. Finally this leads to a tractable expression for the significance level.

Consider the covariance $\mathbb{E}\{Z(n, \theta_1)Z(m, \theta_2)\}$ for scores at change-points n and m, and with θ_1 and θ_2, respectively. Assume the covariance matrix associated with $\mathbf{y}_n$ is V_{N-n,θ_1}, and that associated with $\mathbf{y}_m$ is V_{N-m,θ_2}, respectively. Also assume $n > m$, so the dimension of the covariance matrix for $\mathbf{y}_m$ is larger than that for $\mathbf{y}_n$. Also note that $\mathbf{y}_n$ and $\mathbf{y}_m$ have overlapping samples, so V_{N-n,θ_1} is actually a sub-block matrix of V_{N-m,θ_2}. Based on these observations, after some derivations given in Appendix D, we find the covariance of $Z(k, \theta)$

under H_0 to be

$$\mathbb{E}\left\{Z(n,\theta_1)Z(m,\theta_2)\right\} = \frac{\text{tr}\left(V_{N-n,\theta_1}V_{N-n,\theta_2}^T\right)}{\left[\text{tr}\left(V_{N-n,\theta_1}V_{N-n,\theta_1}^T\right)\text{tr}\left(V_{N-m,\theta_2}V_{N-m,\theta_2}^T\right)\right]^{1/2}}. \tag{30}$$

A special case is when $\theta_1 = \theta_2$, $n = m$, then (30) becomes $\mathbb{E}\left\{Z(n,\theta_1)^2\right\} = 1$, which is consistent with the unit variance of $Z(k,\theta)$.

To study the local covariance of the random field, set $\theta_1 = \theta$, $\theta_2 = \theta + \delta$, $n = k$ and $m = k - i$, $i = 1, \cdots, k-1$ in (30). We have to run the index k for the change point backwards, because the smaller the k, the more post-change samples we have, and hence the larger the dimension of the covariance matrix $V_{N-n,\theta}$. Assume δ and i are small relative to θ and k, respectively. We will expand everything in terms of θ, k, δ and i. If use the first order approximation (keeping only the first order terms), the local covariance is give by

$$\mathbb{E}\left\{Z(k,\theta)Z(k-i,\theta+\delta)\right\} \approx 1 - \gamma^2(k,\theta)\delta^2 - \frac{\mu(k,\theta)}{2(N-k)}i + o(\delta^2) + o(i). \tag{31}$$

Details for (31) can be found in Appendix E. Here

$$\gamma(k,\theta) = \frac{\text{tr}\left(V'_{N-k,\theta}V_{N-k,\theta}^T\right)}{\text{tr}\left(V_{N-k,\theta}V_{N-k,\theta}^T\right)}, \tag{32}$$

The prime f' denotes $\frac{\partial f(\theta)}{\partial\theta}$. Note that $\gamma(k,\theta)$ is independent of i, and it can be interpreted as the sensitivity of the Frobenius norm of $V_{N-k,\theta}$ at certain value of θ. The $\mu(\theta_0, k_0)$ defined in (16), is the ratio of the average sum-of-squares of the "additional" terms when we decrease k by a small amount, over the average sum-of-squares of the "original" terms, in the covariance matrix $V_{N-k,\theta}$.

A.1.4 Local Random Field Decomposition

Note that the first order expansion of the covariance does not have cross product terms. This implies that if we assume $Z(k,\theta)$ to be Gaussian, then it can be decomposed as a sum of two independent 1-D random process. Using the local covariance we just found, we have the following Lemma:

Lemma 4. *Assume $\xi \to \infty$, $b \to \infty$, $n \to \infty$, with $\frac{\xi}{b} \sim 1$ and $\frac{b}{\sqrt{N}} \sim d$ where $d > 0$ is some constant. The discretized process $b\left[Z\left(k-i, \theta + \frac{\Delta}{\sqrt{N}}j\right) - \xi\right]$, $i \in \mathbb{Z}, j \geq 0$, conditioned on $Z(k,\theta) = \xi$ can be written as sum of two independent processes:*

$$\left\{ b\left[Z\left(k-i,\theta+\frac{\Delta}{\sqrt{N}}j\right)-\xi\right]\middle| Z(k,\theta)=\xi\right\} = S_i + V_j, \tag{33}$$

where $S_i = \sum_{l=1}^{i} a_l$, with $a_l \sim \mathcal{N}\left(-\frac{\mu(k,\theta)}{2(N-k)}b^2, \frac{\mu(k,\theta)}{N-k}b^2\right)$, and $V_j = \sqrt{2}\gamma\frac{b}{\sqrt{N}}\Delta jV - \gamma^2\frac{b^2}{N}\Delta^2 j^2$ with $V \sim \mathcal{N}(0,1)$.

Though the conclusion is similar to Lemma 1 in [KS89], proof is not given in therein. For completeness we present the proof in the Appendix F.

By Lemma 4, using the techniques in [Sie88] and [KS89], we have the conditional probability can be written in terms of decomposed random processes: (add one more step in the following)

$$\begin{aligned} & P\left\{ \max_{(i,j)\in J(i_0,j_0)} b\left[Z\left(i, j\frac{\Delta}{\sqrt{N}}\right) - Z\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right)\right] \leq -x \,\middle|\, Z\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right) = b + \frac{x}{b}\right\} \\ \approx\ & P\left\{\max_{i\geq 1} S_i \leq -x\right\} P\left\{\max_{i\leq 0} S_i + \max_{j\geq 1} V_j \leq -x\right\} \end{aligned} \tag{34}$$

A similar argument for (34) can be found in [Sie88] and [KS89]. (I have the details for these derivations but did not type them out. γ disappear in the process of derivation.)

A.1.5 Limit by Shrinking Δ

Put this together with approximation for Part I in 29, the approximate significant level (28) becomes

$$\begin{aligned} & P\left\{\max_{(i,j)\in D} Z\left(i, j\frac{\Delta}{\sqrt{N}}\right) \geq b\right\} \\ \approx & \sum_{(i_0,j_0)\in D} g\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right)\frac{\Delta}{\sqrt{N}} \cdot \\ & \frac{\sqrt{N}}{\Delta b}\int_0^\infty \exp\left(-\frac{\xi_0}{b}x\right) P\left\{\max_{i\geq 1} S_i \leq -x\right\} P\left\{\max_{i\leq 0} S_i + \max_{j\geq 1} V_j \leq -x\right\} dx. \end{aligned} \tag{35}$$

The following Lemma, which is an extension of Lemma 2 in [KS89], enables us to find an expression for integration over x in (28):

Lemma 5. *Assume $x_1, x_2, \cdots$ are i.i.d. $\mathcal{N}(-\mu_1, \sigma_1^2)$, with $\mu_1 > 0$. Define the random walk $S_0 = 0$, $S_i = \sum_{l=1}^i x_l$, $i = 1, 2, \cdots$, and the smooth varying random process $V_j = \beta\Delta j V - \frac{\beta^2}{2}\Delta^2 j^2$, for some constants $\Delta > 0$, $\beta > 0$. As $\Delta \to 0$, for some constant $\alpha > 0$, we have*

$$\begin{aligned} & \frac{1}{\Delta}\int_0^\infty e^{-\alpha x} P\left\{\max_{i\geq 1} S_i \leq -x\right\} P\left\{\max_{i\leq 0} S_i + \max_{j\geq 1} V_j \leq -x\right\} dx \\ & \xrightarrow{\Delta\to 0} \frac{|\beta|}{\sqrt{2\pi}}\left(\frac{2\mu_1^2}{\sigma_1^2}\right)\nu\left(\frac{2\mu_1}{\sigma_1}\right). \end{aligned} \tag{36}$$

where $\nu(x)$ is defined in (14).

Finally, using Lemma 5 for (35) with $\alpha = \frac{\xi_0}{b}$, $\beta = \sqrt{2}\gamma\frac{b}{\sqrt{N}}$, $\mu_1 = \frac{\mu(k,\theta)}{2(N-k)}b^2$ and $\sigma_1^2 = \frac{\mu(k,\theta)}{N-k}b^2$ we have the approximate significance level

$$\frac{1}{2\sqrt{\pi}}\sum_{(i_0,j_0)\in D} g\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right)\cdot\frac{b^2\mu(i_0, j_0\frac{\Delta}{\sqrt{N}})}{N-i_0}\nu\left(\sqrt{\frac{b^2\mu\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right)}{N-i_0}}\right)\gamma\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right)\frac{\Delta}{\sqrt{N}}. \tag{37}$$

As $\Delta \to 0$ this expression is a Riemann sum, which leads to the final expression in Theorem 2.

A.2 Proof for Vector Parameter Case

The techniques used for this proof are different. The proof consists of several parts. First, we will use change-of-measure to recenter the process mean to the threshold so that the tail probability becomes a much higher probability. Second, in Part A.2.2 by conditioning on the local sigma field, we can focus on calculating the boundary crossing probability using local central limit theorem in Part A.2.1, which turns out to be deterministic. Finally, in Part A.2.4 we will compute a Mill's ratio type of expression, by decompose the process as random walk plus smooth Gaussian field.

A.2.1 Change of measure

Introduce an exponential family: $dF_{\xi_{k,\theta}} = \exp\{\xi_{k,\theta} Z(k,\theta) - \psi(\xi_{k,\theta})\} dF$, where F is the probability measure under H_0, ψ is the cumulant generating function of $Z(k,\theta)$. Rewrite the expression for significance level as

$$\begin{aligned} & P\left(\max_{\substack{0\le k\le N-N_0,\\ \theta\in\Theta}} Z(k,\theta) \ge b\right) \\ = \; & \mathbb{E}\left\{\frac{\sum_k \int_{\theta\in\Theta} e^{\xi Z(k,\theta)-\psi(\xi)} d\theta}{\sum_l \int_{s\in\Theta} e^{\xi Z(l,s)-\psi(\xi)} ds}; \max Z(k,\theta) \ge b\right\}. \quad (38) \\ = \; & \sum_k \int_{\theta\in\Theta} \mathbb{E}\left\{\frac{e^{\xi Z(k,\theta)-\psi(\xi)}}{\sum_l \int_{s\in\Theta} e^{\xi Z(l,s)-\psi(\xi)} ds}; \max Z(k,\theta) \ge b\right\} d\theta. \quad (39) \end{aligned}$$

where $\mathbb{E}\{X; A\} = \mathbb{E}\{X I_A\}$. Change the measure dF for to dF_ξ, then (39) becomes:

$$\begin{aligned} & \sum_k \int_{\theta\in\Theta} \mathbb{E}_k^\theta \left\{\frac{1}{\sum_l \int_{s\in\Theta} e^{\xi Z(l,s)-\psi(\xi)} ds}; \max Z(k,\theta) \ge b\right\} \\ = \; & \sum_k \int_{\theta\in\Theta} \exp\{\psi(\xi) - \xi b\} \\ & \cdot \mathbb{E}_k^\theta \left\{\frac{e^{\xi \max_{l,s}[Z(l,s)-Z(k,\theta)]} \cdot e^{-\xi[Z(k,\theta)-b+M_{k,\theta}]}}{\sum_l \int_{\theta\in\Theta} e^{\xi Z(l,s)-\psi(\xi)} ds}; \quad Z(k,\theta) - b + \max_{l,s}[Z_{l,s} - Z_{k,\theta}] \ge 0\right\} \end{aligned}$$

where we have just used some algebra.

A.2.2 Localization

Now consider the expectation with $\mathbb{E}_k^\theta\{\cdot\}$ in the above equation. Define

$$\begin{aligned} M_{k,\theta} &= \max_{l,s}[Z(l,s) - Z(k,\theta)] \\ S_{k,\theta} &= \sum_l \int_{\theta\in\Theta} e^{\xi Z(l,s)-\psi(\xi)} ds. \end{aligned} \quad (40)$$

Then we have

$$\mathbb{E}_k^\theta \left\{ \frac{M_{k,\theta}}{S_{k,\theta}} e^{-\xi(Z_{k,\theta}-b+\frac{\log M_{k,\theta}}{\xi})}; \quad Z_{k,\theta} - b + \frac{\log M_{k,\theta}}{\xi} \geq 0 \right\}. \tag{41}$$

Conditioning the integrant inside the expectation on the σ-field formed by the local random field $\mathcal{F}_{k,\theta}$, we have the above equation equals to

$$\mathbb{E}_k^\theta \left\{ \mathbb{E}_k^\theta \left\{ \frac{M_{k,\theta}}{S_{k,\theta}} e^{-\xi(Z_{k,\theta}-b+\frac{\log M_{k,\theta}}{\xi})}; \quad Z_{k,\theta} - b + \frac{\log M_{k,\theta}}{\xi} \geq 0 | \mathcal{F}_{k,\theta} \right\} \right\} \tag{42}$$

We can show that the ratio

$$\frac{M_{k,\theta}}{S_{k,\theta}} \approx \frac{\hat{M}_{k,\theta}}{\hat{S}_{k,\theta}}, \tag{43}$$

where $\hat{M}$ and $\hat{S}$ are the maximum and sum formed by local terms so they are completely determined by $\mathcal{F}_{k,\theta}$. So the above expression can be approximate as:

$$\mathbb{E}_k^\theta \left\{ \frac{\hat{M}_{k,\theta}}{\hat{S}_{k,\theta}} \mathbb{E}_k^\theta \left\{ e^{-\xi(Z_{k,\theta}-b+\frac{\log M_{k,\theta}}{\xi})}; \quad Z_{k,\theta} - b + \frac{\log M_{k,\theta}}{\xi} \geq 0 | \mathcal{F}_{k,\theta} \right\} \right\} \tag{44}$$

Choose $\xi = \xi_{k,\theta}$ such that:

$$\psi'(\xi_{k,\theta}) = b. \tag{45}$$

Under this choice we have $\mathbb{E}_k^\theta \{Z_{k,\theta}\} = b$.

We can show that

$$\frac{\log M_{k,\theta}}{\xi_{k,\theta}} \ll Z_{k,\theta} - b \tag{46}$$

the term $\frac{\log M_{k,\theta}}{\xi_{k,\theta}}$ can be ignored and the above expression is approximately:

$$\mathbb{E}_k^\theta \left\{ \frac{\hat{M}_{k,\theta}}{\hat{S}_{k,\theta}} \mathbb{E}_k^\theta \left\{ e^{-\xi_{k,\theta}(Z_{k,\theta}-b)}; \quad Z_{k,\theta} - b \geq 0 | \mathcal{F}_{k,\theta} \right\} \right\} \tag{47}$$

A.2.3 Local central limit theorem

Use local central limit theorem, we can assume $Z_{k,\theta} - b$, under measure $dF_{\xi_{k,\theta}}$, has mean 0 and variance $\mathrm{Var}_{\xi_{k,\theta}}(Z_{k,\theta}) = \psi''(\xi_{k,\theta})$ (we have found this expression before). Then the conditional expectation of the overshoot:

$$\begin{aligned} &\mathbb{E}_k^\theta \left\{ e^{-\xi_{k,\theta}(Z_{k,\theta}-b)}; \quad Z_{k,\theta} - b \geq 0 | \mathcal{F}_{k,\theta} \right\} \\ &\sim \int_0^\infty e^{-\xi_{k,\theta} x} \frac{e^{-\frac{x^2}{2\mathrm{VAR}_{\xi_{k,\theta}}(Z_{k,\theta})}}}{\sqrt{2\pi \mathrm{VAR}_{\xi_{k,\theta}}(Z_{k,\theta})}} dx, \\ &= \frac{1}{\sqrt{2\pi \mathrm{VAR}_{\xi_{k,\theta}}(Z_{k,\theta})}} \end{aligned} \tag{48}$$

Using the local central limit result, the approximated significant level become

$$\sum_k \int_\theta \frac{e^{\psi(\xi_{k,\theta})} - \xi_{k,\theta} b}{\sqrt{2\pi \text{VAR}_{\xi_{k,\theta}}(Z_{k,\theta})}} \frac{1}{\xi_{k,\theta}} \mathbb{E}_k^\theta \left\{ \frac{\hat{M}_{k,\theta}}{\hat{S}_{k,\theta}} \right\} d\theta. \tag{49}$$

A.2.4 Expectation of Mill's Ratio

Now we focus on finding the expectation of Mill's ratio. Because the local field covariance matrix has not cross product term, we can assume the local difference of the field can be written as the increments in the random walk and the smooth field dimensions, respectively:

$$Z_{l,s} - Z_{k,\theta} \approx (Z_{l,\theta} - Z_{k,\theta}) + (Z_{k,s} - Z_{k,\theta}). \tag{50}$$

Then

$$\begin{aligned} & \mathbb{E}_k^\theta \left\{ \frac{\hat{M}_{k,\theta}}{\hat{S}_{k,\theta}} \right\} \\ &= \mathbb{E}_k^\theta \left\{ \frac{e^{\max_l \xi_{k,\theta}[Z_{l,\theta} - Z_{k,\theta}]}}{\sum_l e^{\xi_{k,\theta}[Z_{l,\theta} - Z_{k,\theta}]}} \cdot \frac{e^{\max_s \xi_{k,\theta}(Z_{k,s} - Z_{k,\theta})}}{\int_s e^{\xi_{k,\theta}[Z_{k,s} - Z_{k,\theta}]}} \right\} \end{aligned} \tag{51}$$

The second ratio in the above expression is due to non-Gaussian smooth random field. We can show in the Appendix H that it is approximately

$$\frac{e^{\max_l \xi_{k,\theta}[Z_{l,\theta} - Z_{k,\theta}]}}{\sum_l e^{\xi_{k,\theta}[Z_{l,\theta} - Z_{k,\theta}]}} = (b\xi_{k,\theta})^{d/2} \frac{|H_{k,\theta}|}{(2\pi)^{d/2}}. \tag{52}$$

Then the above ratio becomes: $(b\xi_{k,\theta})^{d/2} \frac{|H_{k,\theta}|}{(2\pi)^{d/2}} \mathbb{E}_k^\theta \left\{ \frac{e^{\max_l \xi_{k,\theta}[Z_{l,\theta} - Z_{k,\theta}]}}{\sum_l e^{\xi_{k,\theta}[Z_{l,\theta} - Z_{k,\theta}]}} \right\}$ We show in the Appendix I that the expectation term is

$$\mathbb{E}_k^\theta \left\{ \frac{e^{\max_l \xi_{k,\theta}[Z_{l,\theta} - Z_{k,\theta}]}}{\sum_l e^{\xi_{k,\theta}[Z_{l,\theta} - Z_{k,\theta}]}} \right\} = \frac{b^2 \mu(k,\theta)}{2(N-k)} \nu \left(\sqrt{\frac{b^2 \mu(k,\theta)}{N-k}} \right). \tag{53}$$

Combine these result then we have the approximated significant level for the vector parameter case given in Theorem 3.

B Variance of The Likelihood Derivative

Let y_i denote the ith element of y_k, let V denote $V_{N-k,\theta}$, and V_{ki} be the kth row ith column element of V.

$$\begin{aligned} \mathbb{E}\left\{\mathbf{y}_k^T V \mathbf{y}_k\right\}^2 &= \mathbb{E}\left\{\text{tr}(V y_k y_k^T V y_k y_k^T)\right\} \\ &= \mathbb{E}\left\{ \sum_{k,j,i,l} V_{ki} V_{jl} y_i y_j y_l y_k \right\}. \end{aligned} \tag{54}$$

Under H_0, $\mathbf{y}_k \sim \mathcal{N}(0, \sigma_0^2 I_{N-k})$, and the sum can be computed as several cases:

(1) When $i = j = k = l$, then

$$\mathbb{E}\left\{\sum_i V_{ii}^2 y_i^4\right\} = 3\sigma_0^4 \sum_i V_{ii}^2, \tag{55}$$

(2) When $i = j$, $k = l$, and $i \neq k$,

$$\mathbb{E}\left\{\sum_{i\neq k} V_{ki} V_{ik} y_k^2 y_i^2\right\} = \sigma^4 \left[\mathrm{tr}(VV) - \sum_i V_{ii}^2\right]. \tag{56}$$

(3) When $k = j, i = l, k \neq i$

$$\mathbb{E}\left\{\sum_{k\neq i} V_{ki}^2 y_k^2 y_i^2\right\} = \sigma^4 \left[\mathrm{tr}(VV^T) - \sum_i V_{ii}^2\right]. \tag{57}$$

(4) When $k = i, j = l, k \neq j$

$$\mathbb{E}\left\{\sum_{k\neq j} V_{kk} V_{jj} y_k^2 y_j^2\right\} = \sigma^4 \left[\sum_{k\neq j} V_{kk} V_{jj}\right]. \tag{58}$$

Combine these cases together we have

$$\begin{aligned}
\mathbb{E}\left\{\mathbf{y}_k^T V \mathbf{y}_k\right\}^2 &= 3\sigma_0^4 \sum_i V_{ii}^2 + 2\sigma_0^4 \left[\mathrm{tr}(VV^T) - \sum_i V_{ii}^2\right] + \sigma_0^4 \left[\sum_{k\neq j} V_{kk} V_{jj}\right] \\
&= 2\sigma_0^4 \mathrm{tr}(VV^T) + \sigma_0^4 \sum_i V_{ii}^2 + \sigma_0^4 \left[\sum_{k\neq j} V_{kk} V_{jj}\right] \\
&= \sigma_0^4 \left[2\mathrm{tr}(VV^T) + \{\mathrm{tr}(V)\}^2\right].
\end{aligned} \tag{59}$$

Hence the variance is given by

$$\begin{aligned}
\mathrm{VAR}\left\{\left.\frac{\partial l}{\partial \tau}\right|_{\tau=0}\right\} &= \left(\frac{1}{2\sigma_0^4}\right)^2 \mathbb{E}\left\{\mathbf{y}_k^T V \mathbf{y}_k\right\}^2 - \left(\frac{\mathrm{tr}(V)}{2\sigma_0^2}\right)^2 \\
&= \frac{1}{2\sigma_0^4}\mathrm{tr}(VV^T).
\end{aligned} \tag{60}$$

C Approximation Distribution Considering Skewness

We would like to get $P\left\{Z\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right) = b + \frac{x}{b}\right\}\frac{dx}{b}\cdot$. In the following we denote $Z\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right)$ as Z. The normal approximation of $P\left(Z = b + \frac{x}{b}\right)$ deviates the true distribution in the tail of that distribution, but gives a good approximation close to the mean of that distribution. Considering this, we will use measure transformation to re-center the point-of-interest to the mean of a measure, find normal approximation under that measure, and then transform the measure back.

Define the cumulant generating function of Z to be $\psi(\xi) = \log \mathbb{E}\left\{\exp(\xi Z)\right\}$. We can choose a $\xi_0 > 0$ such that

$$\psi'(\xi_0) = b. \tag{61}$$

Then use exponential embedding to construct a new probability measure dF_{ξ_0} by

$$dF_{\xi_0} = \exp\left\{\xi_0 Z - \psi(\xi_0)\right\} dF. \tag{62}$$

where dF is the distribution function of Z. We can rewrite the threshold crossing probability by change of measure:

$$\begin{aligned} P\left\{Z = b + \frac{x}{b}\right\} &= \mathbb{E}_{\xi_0}\left\{\frac{1}{\exp\left[\xi_0 Z - \psi(\xi_0)\right]}; Z = b + \frac{x}{b}\right\} \\ &= \exp\left\{\psi(\xi_0) - \xi_0\left(b + \frac{x}{b}\right)\right\} P_{\xi_0}\left\{Z = b + \frac{x}{b}\right\} \end{aligned} \tag{63}$$

where $\mathbb{E}_{\xi_0}$ and P_{ξ_0} denotes the expectation and probability under the new measure dF_{ξ_0}. One can verify that under this new measure $\mathbb{E}_{\xi_0}\left\{Z\right\} = b$:

$$\mathbb{E}_{\xi_0}\left\{Z\right\} = \mathbb{E}\left\{Z \exp\left[\xi_0 Z - \psi(\xi_0)\right]\right\} = e^{-\psi(\xi_0)} \left.\frac{\partial e^{\psi(\xi)}}{\partial \xi}\right|_{\xi=\xi_0} = e^{-\psi(\xi_0)} e^{\psi(\xi_0)} \Psi'(\xi_0) = b. \tag{64}$$

Hence under the new measure the mean of Z equals to its level ($\frac{x}{b}$ is a small overshoot). Now we will use normal approximation for $P_{\xi_0}\left\{Z = b + \frac{x}{b}\right\}$, then use (63) to get back to the original probability. We only need the variance of Z under dF_{ξ_0}, which can be found from the cumulant generating function: $\mathrm{Var}_{\xi_0}\left\{Z\right\} = \left.\frac{\partial^2 \psi(\xi)}{\partial^2 \xi}\right|_{\xi=\xi_0}$. We will keep upto the third order in the Taylor expansion of $\psi(\xi)$, which corrects for the skewness of the distribution. Also, also use this truncated expansion in solving ξ_0 from (61).

By the multivariate normal assumption of $\mathbf{y}_k$, we can find an expression for $\Psi(\xi)$. Rewrite $Z(k, \theta) = \frac{\mathbf{y}_k^T \tilde{V} \mathbf{y}_k}{\sigma_0^2} - \mathrm{tr}(\tilde{V})$, where $\tilde{V}$ is defined in (16). Then

$$\begin{aligned} \psi(\xi) &= \log \mathbb{E}\left\{e^{\xi\left(\frac{\mathbf{y}_k^T \tilde{V} \mathbf{y}_k}{\sigma_0^2} - \mathrm{tr}(\tilde{V})\right)}\right\} = -\xi \mathrm{tr}(\tilde{V}) + \log \mathbb{E}\left\{e^{\xi \frac{\mathbf{y}^T \tilde{V} \mathbf{y}}{\sigma_0^2}}\right\} \\ &= -\xi \mathrm{tr}(\tilde{V}) + \log\left\{\frac{1}{(2\pi\sigma_0^2)^{\frac{N-k}{2}}} \int \exp\left\{\xi \frac{\mathbf{y}^T \tilde{V} \mathbf{y}}{\sigma_0^2} - \frac{\mathbf{y}^T \mathbf{y}}{2\sigma_0^{2(N-k)}}\right\} d\mathbf{y}\right\} \\ &= -\xi \mathrm{tr}(\tilde{V}) - \frac{1}{2} \log\left|I - 2\xi \tilde{V}\right|. \end{aligned} \tag{65}$$

The first order derivative is:

$$\begin{aligned} \frac{\partial \psi(\xi)}{\partial \xi} &= -\mathrm{tr}(\tilde{V}) + \mathrm{tr}\left[\left(I - 2\xi \tilde{V}\right)^{-1} \tilde{V}\right] \\ &= \sum_{k=1}^{\infty} (2\xi)^k \mathrm{tr}(\tilde{V}^{k+1}) \\ &\approx 2\xi \mathrm{tr}(\tilde{V}^2) + 4\xi^2 \mathrm{tr}(\tilde{V}^3) \\ &= \xi + 4\xi^2 \mathrm{tr}(\tilde{V}^3), \end{aligned} \tag{66}$$

where we have used the fact that $\mathrm{tr}(\tilde{V}^2) = \mathrm{tr}\left(\frac{VV^T}{2\mathrm{tr}(VV^T)}\right) = \frac{1}{2}$.

(one step missing here, can prove the norm of V decrease on the order of N?)

By setting (66) to b, we have $4\xi_0^2 \mathrm{tr}(\tilde{V}^3) + \xi_0 - b = 0$, from which we can solve for the positive root $\xi_0 = \frac{-1+\sqrt{1+16b\mathrm{tr}(\tilde{V}^3)}}{8\mathrm{tr}(\tilde{V}^3)}$. And also the variance:

$$\mathrm{Var}_{\xi_0}\{Z\} = \left.\frac{\partial^2 \psi(\xi)}{\partial^2 \xi}\right|_{\xi=\xi_0} \tag{67}$$

Plug everything back to (63), noting that $\exp\left\{-\frac{x^2/b^2}{2\mathrm{Var}_{\xi_0}\{Z\}}\right\} \approx 1$, we have

$$\begin{aligned} P\left\{Z\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right) = b + \frac{x}{b}\right\} &\approx \exp\left\{\psi(\xi_0) - \xi_0\left(b + \frac{x}{b}\right)\right\}\frac{1}{\sqrt{\mathrm{Var}_{\xi_0}\{Z\}}} \\ &= g\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right)\exp\left(-\frac{\xi_0}{b}x\right). \end{aligned} \tag{68}$$

D Covariance

Denote the covariance matrices for $\mathbf{y}_m$ $\mathbf{y}_n$: V_{N-m,θ_2}, V_{N-n,θ_1}, as $V_2(\theta_2)$ and $V_1(\theta_1)$, respectively. Rewrite $V_2(\theta_2)$ as

$$\mathbf{y}_m = \begin{bmatrix} \mathbf{y}_\Delta \\ \mathbf{y}_n \end{bmatrix}, \quad \text{and} \quad V_{N-m,\theta_2} = \begin{bmatrix} V_{11}(\theta_2) & V_{12}(\theta_2) \\ V_{21}(\theta_2) & V_{N-n,\theta_2} \end{bmatrix}. \tag{69}$$

From now on we will dropped the dependence of θ_1 and θ_2 and recover them when needed:

$$\mathbb{E}\{Z(n,\theta_1)Z(m,\theta_2)\} = \frac{1}{2\sigma_0^4\left[\mathrm{tr}(V_1V_1^T)\mathrm{tr}(V_2V_2^T)\right]^{1/2}}\left[\mathbb{E}\left\{\mathbf{y}_n^T V_1 \mathbf{y}_n \mathbf{y}_m^T V_2 \mathbf{y}_m\right\} - \mathbb{E}\left\{\mathbf{y}_n^T V_1 \mathbf{y}_n\right\}\mathbb{E}\left\{\mathbf{y}_m^T V_2 \mathbf{y}_m\right\}\right], \tag{70}$$

The first term in the above expression can be written as

$$\mathbb{E}\left\{\mathbf{y}_n^T V_1 \mathbf{y}_n \mathbf{y}_m^T V_2 \mathbf{y}_m\right\} = \mathbb{E}\left\{(\mathbf{y}_\Delta^T V_{11}\mathbf{y}_\Delta + \mathbf{y}_n^T V_1 \mathbf{y}_n + \mathbf{y}_\Delta^T V_{12}\mathbf{y}_n + \mathbf{y}_n^T V_{21}\mathbf{y}_\Delta)(\mathbf{y}_n^T V_1 \mathbf{y}_n)\right\}. \tag{71}$$

(1) The cross terms vanish $\mathbb{E}\left\{(\mathbf{y}_\Delta^T V_{12}\mathbf{y}_n)(\mathbf{y}_n^T V_1 \mathbf{y}_n)\right\} = \mathbb{E}\left\{(\mathbf{y}_n^T V_{21}\mathbf{y}_\Delta)(\mathbf{y}_n^T V_1 \mathbf{y}_n)\right\} = 0$ because $\mathbb{E}\{\mathbf{y}_\Delta\} = 0$.

(2) The first term, $\mathbb{E}\left\{(\mathbf{y}_\Delta^T V_{11}\mathbf{y}_\Delta)(\mathbf{y}_n^T V_1 \mathbf{y}_n)\right\} = \mathbb{E}\left\{(\mathbf{y}_\Delta^T V_{11}\mathbf{y}_\Delta)\right\}\mathbb{E}\left\{(\mathbf{y}_n^T V_1 \mathbf{y}_n)\right\}$, because $\mathbf{y}_\Delta$ and $\mathbf{y}_n$ are independent under H_0.

(3) The second term $\mathbb{E}\left\{\mathbf{y}_n^T V_1 \mathbf{y}_n \mathbf{y}_n^T V_1 \mathbf{y}_n\right\}$ has similar form to the variance expression we just derived. Following a similar procedure, we have $\mathbb{E}\left\{\mathbf{y}_n^T V_1 \mathbf{y}_n \mathbf{y}_n^T V_1 \mathbf{y}_n\right\} = \sigma_0^4[2\mathrm{tr}(V_1V_1^T) + \mathrm{tr}(V_1)\mathrm{tr}(V_1)]$.

If we plug the block partition of y_m and V_2, the second term can be simplified as

$$\mathbb{E}\left\{\mathbf{y}_\Delta^T V_{11}\mathbf{y}_\Delta + \mathbf{y}_n^T V_1 \mathbf{y}_n\right\}\mathbb{E}\left\{\mathbf{y}_n^T V_1 \mathbf{y}_n\right\} = \sigma_0^4 \mathrm{tr}(V_1(\theta_1))\mathrm{tr}(V_1(\theta_2)) + \mathbb{E}\left\{\mathbf{y}_\Delta^T V_{11}(\theta_2)\mathbf{y}_\Delta\right\}\mathbb{E}\left\{\mathbf{y}_n^T V_1(\theta_1)\mathbf{y}_n\right\}. \quad (72)$$

Combine these results, after some cancellation, we have

$$\begin{aligned}
\mathbb{E}\left\{Z_n(\theta_1)Z_m(\theta_2)\right\} &= \frac{1}{2\sigma_0^4\left[\mathrm{tr}(V_1(\theta_1)V_1(\theta_1)^T)\mathrm{tr}(V_2(\theta_2)V_2(\theta_2)^T)\right]^{1/2}} \cdot \\
&\quad \left[\mathbb{E}\left\{(\mathbf{y}_\Delta^T V_{11}(\theta_2)\mathbf{y}_\Delta)\right\}\mathbb{E}\left\{(\mathbf{y}_n^T V_1(\theta_1)\mathbf{y}_n)\right\} + \right. \\
&\quad 2\sigma_0^4 \mathrm{tr}(V_1(\theta_2)V_1(\theta_1)^T) + \sigma_0^4\mathrm{tr}(V_1(\theta_2))\mathrm{tr}(V_1(\theta_1)) - \sigma_0^4\mathrm{tr}(V_1(\theta_2))\mathrm{tr}(V_1(\theta_1)) \\
&\quad \left. -\mathbb{E}\left\{\mathbf{y}_\Delta^T V_{11}(\theta_2)\mathbf{y}_\Delta\right\}\mathbb{E}\left\{\mathbf{y}_n^T V_1(\theta_1)\mathbf{y}_n\right\}\right], \\
&= \frac{\mathrm{tr}(V_1(\theta_2)V_1(\theta_1)^T)}{\left[\mathrm{tr}(V_1(\theta_1)V_1(\theta_1)^T)\mathrm{tr}(V_2(\theta_2)V_2(\theta_2)^T)\right]^{1/2}}. \qquad (73)
\end{aligned}$$

E Local Covariance Analysis

Examine the numerator and denominator of the covariance expression (73) separately. Recall that we have set $\theta_1 = \theta$, $\theta_2 = \theta + \delta$, $n = k$ and $m = k - i$, $i = 1, \cdots, k-1$ in (73). Expand functions depending on θ:

(1) The numerator:

$$\begin{aligned}
\mathrm{tr}(V_1(\theta+\delta)V_1(\theta)) &\approx \mathrm{tr}(V_1(\theta)V_1(\theta)^T) + \delta\mathrm{tr}(V_1'(\theta)V_1(\theta)^T) \\
&= \mathrm{tr}(V_1(\theta)V_1(\theta)^T)(1 + \delta\gamma(\theta, k)). \qquad (74)
\end{aligned}$$

where $V_1'(\theta) = \left.\frac{\partial V_1(\theta)}{\partial \theta}\right|_{\theta=\theta}$.

(2) The denominator, we can write V_2 using the sub-block matrices defined in (69)

$$\begin{aligned}
&\mathrm{tr}(V_2(\theta_2)V_2(\theta_2)^T) = \\
&\mathrm{tr}(V_{11}(\theta_2)V_{11}(\theta_2)^T) + \mathrm{tr}(V_{12}(\theta_2)V_{12}(\theta_2)^T) + \mathrm{tr}(V_{21}(\theta_2)V_{21}(\theta_2)^T) + \mathrm{tr}(V_1(\theta_2)V_1(\theta_2)),
\end{aligned}$$

and expand each term separately about θ_0, for example:

$$\mathrm{tr}(V_1(\theta_2)V_1(\theta_2)) \approx \mathrm{tr}(V_1(\theta_0)V_1(\theta_0)^T) + 2\delta\mathrm{tr}(V_1^T(\theta_0)V_1'(\theta_0)).$$

And similarly for other terms. Then the denominator can be written as (all the arguments are θ_0 hence we have dropped it)

$$\left[\mathrm{tr}(V_1(\theta_1)V_1(\theta_1)^T)\mathrm{tr}(V_2(\theta_2)V_2(\theta_2)^T)\right]^{1/2} = \mathrm{tr}(V_1V_1^T)\sqrt{1+2\delta a}\sqrt{1+b}. \qquad (75)$$

where

$$a = \frac{\mathrm{tr}(V_1'V_1^T) + \mathrm{tr}(V_{12}'V_{12}^T) + \mathrm{tr}(V_{21}'V_{21}^T) + \mathrm{tr}(V_{11}'V_{11}^T)}{\mathrm{tr}(V_1V_1^T) + \mathrm{tr}(V_{12}V_{12}^T) + \mathrm{tr}(V_{21}V_{21}^T) + \mathrm{tr}(V_{11}V_{11}^T)}, \qquad (76)$$

$$\begin{aligned} b &= \frac{\mathrm{tr}(V_{12}V_{12}^T)+\mathrm{tr}(V_{21}V_{21}^T)+\mathrm{tr}(V_{11}V_{11}^T)}{\mathrm{tr}(V_1V_1^T)}. \\ &= \frac{2i}{(N-k)}\frac{\frac{1}{2i(N-k)}\left[\mathrm{tr}(V_2V_2^T)-\mathrm{tr}(V_1V_1^T)\right]}{\frac{1}{(N-k)^2}\mathrm{tr}(V_1V_1^T)}. \end{aligned} \tag{77}$$

The expressions $\mathrm{tr}(V_{11}V_{11}^T)$ and $\mathrm{tr}(V_1'V_1^T)$ are sums of i^2 terms, when i is very small relative to $N-k$, these terms can be neglected from a and b. Also, $\mathrm{tr}(V_{12}V_{12}^T)$ and $\mathrm{tr}(V_{21}V_{21}^T)$ are sums of $i(N-k)$ "off-diagonal" terms, which are smaller compared with $\mathrm{tr}(V_1V_1)^T$, which is a sum of $(N-k)^2$ "diagonal terms". Similar argument holds for the numerator of a. Hence when i is small relative to $N-k$, we have:

$$\begin{aligned} a &\approx \frac{\mathrm{tr}(V_1^TV_1')}{\mathrm{tr}(V_1^TV_1)} = \gamma(k,\theta) \\ b &\approx \frac{i}{N-k}\mu(k,\theta). \end{aligned} \tag{78}$$

In obtaining b we have used the approximation to replace the average over increment i by average over increment 1:

$$\frac{\frac{1}{i(N-k)}\left[\mathrm{tr}(V_2V_2^T)-\mathrm{tr}(V_1V_1^T)\right]}{\frac{1}{(N-k)^2}\mathrm{tr}(V_1V_1^T)} \approx \frac{\frac{1}{N-k}\left[\mathrm{tr}(V_{N-k+1,\theta}V_{N-k+1,\theta}^T)-\mathrm{tr}(V_{N-k,\theta}V_{N-k,\theta}^T)\right]}{\frac{1}{(N-k)^2}\mathrm{tr}(V_{N-k,\theta}V_{N-k,\theta}^T)} \triangleq \mu(k,\theta).$$

Now use (74), (75), (78) and the Taylor expansion $\frac{1}{\sqrt{1+x}} \approx 1-\frac{1}{2}x+o(x)$, we have the first order approximation for covariance

$$\begin{aligned} \mathbb{E}\left\{Z(k,\theta)Z(k-i,\theta+\delta)\right\} &\approx \frac{\mathrm{tr}(V_1V_1^T)(1+\delta\gamma)}{\mathrm{tr}(V_1V_1^T)\sqrt{1+2\delta a}\sqrt{1+b}} \\ &\approx [1+\delta\gamma(k,\theta)]\left[1-\frac{1}{2}2\delta\gamma(k,\theta)\right]\left[1-\frac{1}{2}\frac{i}{N-k}\mu(k,\theta)\right] \\ &= \left[1-\gamma^2(k,\theta)\delta^2\right]\left[1-\frac{1}{2}\frac{i}{N-k}\mu(k,\theta)\right]. \end{aligned} \tag{79}$$

The last line (79) leads to (31).

F Proof of Lemma 4

Proof. Since $Z(k,\theta)$ is zero mean with unit variance for all k and θ, under the normal distribution assumption, the conditional distribution of $Z\left(k+i,\theta+\frac{\Delta}{\sqrt{N}}j\right)$ given $Z(k,\theta)=\xi$ is a Gaussian random variable with mean and variance determined by the their means, variances, and covariance. Using (79), the conditional mean is $\sigma_{12}\sigma_{11}^{-1}Z(k,\theta)$, i.e., $\left(1-\gamma^2\delta^2-\frac{\mu}{2(N-k)}i\right)\xi$, and the conditional variance is $\sigma_{11}-\sigma_{12}\sigma_{22}^{-1}\sigma_{21} = 1-\left(1-\gamma^2\delta^2-\frac{\mu i}{2(N-k)}\right)^2 \approx 2\gamma^2\delta^2+\frac{\mu i}{N-k}$. So the conditional distribution $b\left\{Z\left(k+i,\theta+\frac{\Delta}{\sqrt{N}}j\right)-\xi\right\}$ given $Z(k,\theta)=\xi$ is Gaussian with mean $-\gamma^2\Delta^2\frac{b^2}{N}j^2 - \frac{\mu b^2}{2(N-k)}i$ and variance $2\gamma^2\Delta^2\frac{b^2}{N}j^2 + \frac{\mu b^2}{N-k}i$. Note that there is no cross terms in i and j. So the conditional distribution can be written as sum of two random processes: S_i and V_j, where S_l is a sum of Gaussian random variables $S_i = \sum_{l=1}^{l} a_i$ where $a_i \sim \mathcal{N}\left(-\frac{\mu(\theta)}{2}b^2, \mu(\theta)b^2\right)$, and $V_j = \sqrt{2}\gamma\Delta\frac{b}{\sqrt{N}}jV - \gamma^2\Delta^2\frac{b^2}{N}j^2$ where $V \sim \mathcal{N}(0,1)$. □

G Proof of Lemma 5

Proof. The statement can be treated for $V > 0$ and $V < 0$ separately. Note that V_j is quadratic function of j and it can be rewritten as $V_j = -\frac{\beta^2\Delta^2}{2}\left(j - \frac{V}{\Delta\beta}\right)^2 + \frac{V^2}{2}$.

(1) For $V > 0$ and $\beta > 0$, $\max_{j\geq 1} V_j = \frac{V^2}{2}$, which happens at $j = \frac{V}{\Delta\beta}$. However $\max_{i\geq 0} S_i \geq 0$, so $P\left\{\max_{i\geq 0} S_i + \max_{j\geq 1} V_j \leq -x, V > 0\right\} = 0$ all $x > 0$.

(2) For $V < 0$ and $\beta < 0$, $\max_{j\geq 1} V_j = \beta\Delta V - \frac{\beta^2}{2}\Delta^2$, which happens at $j = 1$. Then the probability in Lemma 5 becomes

$$
\begin{aligned}
& \frac{1}{\Delta}\int_0^\infty e^{-\alpha x} P\left\{\max_{i\geq 1} S_i \leq -x\right\} P\left\{\max_{i\geq 0} S_i + \beta\Delta V - \frac{\beta^2}{2}\Delta^2 \leq -x, V < 0\right\} dx \\
= & \int_0^\infty e^{-\alpha\Delta y} P\left\{\max_{i\geq 1} S_i \leq -\Delta y\right\} \\
& \quad P\left\{\max_{i\geq 0} S_i + \beta\Delta V - \frac{\beta^2}{2}\Delta^2 \leq -\Delta y, V < 0\right\} dy, \quad \text{(change of variable } x = \Delta y) \\
= & \int_0^\infty P\left\{\max_{i\geq 1} S_i \leq -\Delta y\right\} \int_{-\infty}^{-y+\frac{\Delta}{2}} P\left\{\max_{i\geq 0} S_i \leq -\Delta(y - \beta v) + \frac{\beta^2}{2}\Delta^2\right\} \phi(v) dv dy \\
\xrightarrow{\Delta\to 0} & \left(P\left\{\max_{i\geq 0} S_i \leq 0\right\}\right)^2 \int_0^\infty \int_{-\infty}^{y} \phi(v) dv dy \\
= & \frac{\beta}{\sqrt{2\pi}}\left(P\left\{\max_{i\geq 0} S_i \leq 0\right\}\right)^2. \qquad (80)
\end{aligned}
$$

where $\left(P\left\{\max_{i\geq 0} S_i \leq 0\right\}\right)^2 = \frac{1}{\sqrt{2\pi}}\left(\frac{2\mu^2}{\sigma^2}\right)\nu\left(\frac{2\mu}{\sigma}\right)$ (given in Corollary 8.44 of [Sie85]).

For $\beta < 0$, using parallel argument we can get the approximation is about $\frac{-\beta}{\sqrt{2\pi}}\left(P\left\{\max_{i\geq 0} S_i \leq 0\right\}\right)^2$. Hence the conclusion in Lemma 5.

□

H Proof of Mill's ratio for smooth field

Suppose the $\max_s \xi_{k,\theta}[Z_{k,s} - Z_{k,\theta}]$ is achieved at $s = \hat{s}$, then

$$\begin{aligned} &\frac{e^{\max_l \xi_{k,\theta}[Z_{l,\theta}-Z_{k,\theta}]}}{\sum_l e^{\xi_{k,\theta}[Z_{l,\theta}-Z_{k,\theta}]}} \\ &= \frac{e^{\xi_{k,\theta}[Z_{k,\hat{s}}-Z_{k,\theta}]}}{\int_\theta e^{\xi_{k,\theta}[Z_{k,s}-Z_{k,\hat{s}}]ds} \cdot e^{\xi_{k,\theta}[Z_{k,\hat{s}}-Z_{k,\theta}]}} \\ &= \frac{1}{\int_\theta e^{\xi_{k,\theta}[Z_{k,s}-Z_{k,\hat{s}}]}ds} \end{aligned} \tag{81}$$

Now we can expand the difference by Taylor's expansion:

$$Z_{k,s} - Z_{k,\hat{s}} \approx \dot{Z}_{k,\hat{s}}(s - \hat{s}) - \frac{1}{2}(s - \hat{s})bH(k, \hat{s})(s - \hat{s}). \tag{82}$$

where we have used $\dot{Z}_{k,\hat{s}} = 0$, and approximated the random Hessian matrix $\ddot{Z}_{\hat{s}}$ by its expectation. To find this expectation, first note that

$$\begin{aligned} &\mathbb{E}_k^\theta \left\{Z_{k,s}\right\} = \mathbb{E}\left\{e^{\xi_{k,\theta}Z_{k,\theta}-\psi(\xi_{k,\theta})}Z_{k,s}\right\} \\ &= \mathbb{E}\left\{e^{\xi_{k,\theta}Z_{k,\theta}-\psi(\xi_{k,\theta})}\mathbb{E}\left\{Z_{k,s}|Z_{k,\theta}\right\}\right\} \end{aligned} \tag{83}$$

If we approximate $Z_{k,s}$ as a Gaussian random field, then the conditional expectation is just $Z_{k,\theta}R_k(\theta, s)$ where R is the covariance function. Hence the above equation is:

$$R_k(\theta, s)\mathbb{E}\left\{e^{\xi_{k,\theta}[Z_{k,\theta}-\psi(\xi_{k,\theta})]}Z_{k,\theta}\right\} = bR_k(\theta, s). \tag{84}$$

Hence the expectation of the Hessian matrix under the tilted measure is:

$$\begin{aligned} &\mathbb{E}_k^\theta \left\{\ddot{Z}_{k,\theta}\right\} \approx \mathbb{E}_k^\theta \left\{\frac{Z(k, \theta + h) - 2Z(k, \theta) + Z(k, \theta - h)}{h^2}\right\} \\ &= \frac{1}{h^2}[bR_k(\theta, \theta + h) - 2bR(\theta, \theta) + bR(\theta, \theta - h)] \end{aligned}$$

Let $h \to 0$ and set $s = \theta$, then we have

$$\mathbb{E}_k^\theta \left\{\ddot{Z}_{k,\theta}\right\} = b\frac{\partial^2 R_k \theta, s}{\partial^2 s}\bigg|_{s=\theta} \triangleq -bH(k, \theta). \tag{85}$$

Then (81) becomes:

$$\begin{aligned} &\left(\int_\theta e^{\xi_{k,\theta}[Z_{k,s}-Z_{k,\hat{s}}]}ds\right)^{-1} \\ &= \left(\int_\theta e^{-\frac{1}{2}(s-\hat{s})^T b\xi_{k,\theta}H(k,\hat{s})}(s - \hat{s})ds\right)^{-1} \\ &= \frac{(b\xi_{k,\theta})^{d/2}\sqrt{|H(k, \hat{s})|}}{(2\pi)^{d/2}}. \end{aligned} \tag{86}$$

In the final expression we use $|H_{k,\hat{s}}| \approx |H_{k,\theta}|$ for each θ.

I Proof of Mill's ratio for random walk

The argument of this proof is adapted from [SY00]. By an invariant argument,

$$\begin{aligned}
&\mathbb{E}_k^\theta\left\{\frac{e^{\max_l \xi_{k,\theta}[Z_{l,\theta}-Z_{k,\theta}]}}{\sum_l e^{\xi_{k,\theta}[Z_{l,\theta}-Z_{k,\theta}]}}\right\} \\
&\approx \frac{1}{2M+1}\sum_{l=-M}^{M}\mathbb{E}_k^\theta\left\{\frac{e^{\max_l \xi_{k,\theta}(Z_{l,\theta}-Z_{k,\theta})}}{\sum_l e^{\xi_{k,\theta}[Z_{l,\theta}-Z_{k,\theta}]}}\right\}
\end{aligned} \tag{87}$$

Use a change of measure, introduce a new measure Q that $\frac{dQ}{dF_{-M}^\theta} = \sum_{l=-M}^{M}\frac{dF_l^\theta}{dF_{-M}^\theta}$. Then the above equation becomes:

$$\begin{aligned}
&\frac{1}{2M+1}\mathbb{E}_{-M}^\theta\left\{e^{\max_{0\le l\le 2M+1}\xi_{k,\theta}(Z_{l,\theta}-Z_{k,\theta})}\right\} \\
&= \frac{1}{2M+1}\mathbb{E}_{-M}\left\{e^{\max_{0\le l\le 2M+1}\xi_{k,\theta}S_l}\right\}
\end{aligned} \tag{88}$$

where $S_l = \sum_{i=1}^{l} a_i$ with a_i i.i.d. normal random variable with mean $-b\xi_{k,\theta}\frac{\mu(k,\theta)}{2(N-k)}$ and variance $\xi_{k,\theta}^2\frac{\mu(k,\theta)}{N-k}$. Then the above equation becomes:

$$\begin{aligned}
&\frac{1}{2M+1}\mathbb{E}_{-M}^\theta\left\{e^{\xi_{k,\theta}\max_{0\le l\le 2M+1}S_l}\right\} \\
&\sim \frac{1}{2M+1}\int_0^\infty \mathbb{P}_{-M}^\theta\left\{\max_{0\le l\le 2M+1}S_l \ge x\right\}e^{\xi_{k,\theta}x}dx \\
&\sim \int_0^\infty e^{-\sigma\xi_{k,\theta}(x/\sigma)}\mathbb{P}_{-M}^\theta\left\{\max_l \frac{S_l}{\sigma}\ge\frac{x}{\sigma}\right\} \\
&\sim \frac{\sigma^2\xi_{k,\theta}}{2}\nu(\sigma\xi_{k,\theta}) \\
&= \frac{\xi_{k,\theta}^2\mu(k,\theta)}{2(N-k)}\nu\left(\xi_{k,\theta}^2\sqrt{\frac{\mu(k,\theta)}{N-k}}\right).
\end{aligned} \tag{89}$$

J Approximate Significant Level of Maximum Likelihood Test

The significant level of this changepoint detection test is given by:

$$P_0(m_0 < T \le N - N_0) \tag{90}$$

Introduce a new measure, which is a mixture of the old measure:

$$Q = \int P_\xi d\xi/\sqrt{2\pi}. \tag{91}$$

Then the log-likelihood is given by:

$$L_{N-k} = \frac{dQ}{dP_0} = \int \exp[l_{N-k}(\xi) - l_{N-k}(0)]d\xi/\sqrt{2\pi}$$

Expand $l_{N-k}(\xi)$ around $\hat{\sigma}^2$, we have:

$$l_{N-k}(\xi) = l_{N-k}(\hat{\sigma}^2) + \frac{1}{2}(\xi - \hat{\sigma}^2)^2 \ddot{l}_{N-k}(\hat{\sigma}^2).$$

Plug this in the integral, we have

$$\begin{aligned} L_{N-k} &\approx \exp[l_{N-k}(\hat{\sigma}^2) - l_{N-k}(0)] \int \exp[\frac{\ddot{l}_{N-k}(\hat{\sigma}^2)}{2}(\xi - \hat{\sigma}^2)^2]d\xi/\sqrt{2\pi} \\ &= \frac{\exp(\Lambda_{N-k})}{\sqrt{\ddot{l}_{N-k}(\hat{\sigma}^2)}}. \end{aligned}$$

Use change-of-measure, we have

$$\begin{aligned} P_0(m_0 \leq T \leq N - N_0) &= \int E_\xi\{L_{N-T}^{-1}; m_0 < T \leq N - N_0\}d\xi/\sqrt{2\pi} \\ &\approx b^{1/2}e^{-b} \int E_\xi\{b^{-1/2}\sqrt{\ddot{l}_T(\hat{\sigma}^2)}e^{-(\Lambda_{N-T}-b)}; m_0 < T \leq N - N_0\}d\xi/\sqrt{2\pi} \quad (22) \end{aligned}$$

Define $\eta_2 = \frac{b}{m_0}$, and $\eta_1 = \frac{b}{N-N_0}$, then by strong law of large number we have

$$P_\xi(m_0 \leq T \leq N - N_0) \to \begin{cases} 1, & \text{if } \eta_1 < J(\xi, \sigma_0^2) < \eta_2, \\ 0, & \text{Otherwise.} \end{cases}$$

where $J(\xi, \sigma_0^2)$ is the increment of the random walk formed by likelihood ratio:

$$\begin{aligned} J(\xi, \sigma_0^2) &= E_\xi\left\{l_1(\xi) - l_1(0)\right\} \\ &= E_\xi\left\{\frac{\xi y_N^2}{2(\xi + \sigma_0^2)\sigma_0^2} + \frac{1}{2}\ln\left(\frac{\sigma_0^2}{\xi + \sigma_0^2}\right)\right\} \\ &= \frac{\xi}{2\sigma_0^2} + \frac{1}{2}\ln(\frac{\sigma_0^2}{\sigma_0^2 + \xi}). \end{aligned}$$

Also by law of large number:

$$P_\xi\left\{\lim_b b^{-1}[\ddot{l}_T(\hat{\sigma}^2)] = -I(\xi)/J(\xi, \sigma_0^2)\right\} = 1.$$

where the Fisher information is given by

$$I(\xi) = -E_\xi\{\ddot{l}_1(\xi)\} = \frac{1}{2(\xi + \sigma_0^2)^2}.$$

Hence the significance level (92) is:

$$P_0(m_0 \le T \le N - N_0) \approx$$
$$b^{1/2}e^{-b}\int_{\xi:\eta_1\le J(\xi,\sigma_0^2)\le\eta_2}\sqrt{\frac{-I(\xi)}{J(\xi,\sigma_0^2)}}\nu(\xi)d\xi/\sqrt{2\pi}, \tag{93}$$

where $\nu(\xi) = \lim_{b\to\infty} E_\xi\{e^{-(\Lambda_{N-T}-b)}\} \approx \frac{1}{1+\rho}$, $\rho = \frac{\xi}{\sigma_0^2}$ is the signal-to-noise ratio. Hence the significant level is given by:

$$\begin{aligned}
&P_0(m_0 \le T \le N - N_0)\\
&\approx b^{1/2}e^{-b}\int_{\xi:\eta_1\le J(\xi,\sigma_0^2)\le\eta_2}\frac{1}{1+\rho}\sqrt{\frac{1}{\sigma_0^4(1+\rho)^2(\rho-\log(1+\rho))}}\frac{d\xi}{\sqrt{2\pi}},\\
&= b^{1/2}e^{-b}\int_{\rho:\eta_1\le\frac{1}{2}(\rho-\log(1+\rho))\le\eta_2}\frac{1}{(1+\rho)^2}\frac{1}{\sqrt{\rho-\log(1+\rho)}}\frac{d\rho}{\sqrt{2\pi}} \quad (\text{ change of variable } \rho = \frac{\xi}{\sigma_0^2}).
\end{aligned}$$

K Find The Special Function

Next we find the special function $\nu(\xi) = \lim_{b\to\infty} E_\xi\{e^{-(\Lambda_{N-T}-b)}\}$. Under measure P_ξ for $\xi \neq \sigma_0^2$, the first approximation of the behavior of Λ_{N-k} is similar to the behavior of the random walk $l_{N-k}(\xi) - l_{N-k}(0)$:

$$\Lambda_{N-k} \approx \left[\frac{\xi}{2(\xi+\sigma_0^2)\sigma_0^2}\sum_{i=k+1}^{N} y_i^2 + \frac{(N-k)}{2}\ln\left(\frac{\sigma_0^2}{\xi+\sigma_0^2}\right)\right], \text{under } P_0.$$

We make an approximation that the change point only been detected with even number of samples: $N-k$ is even. Hence we can write the random walk as:

$$\Lambda_{N-k} \approx \sum_{i=1}^{(N-k)/2}\left[\rho x_i - \ln(1+\rho)\right] \triangleq S_n, \quad n \triangleq \frac{N-k}{2},$$

where $x_i = \frac{y_{2(i+k)-1}^2 + y_{2(i+k)}^2}{2(\xi+\sigma_0^2)}$, $\rho = \frac{\xi}{\sigma_0^2}$ is the signal-to-noise ratio. Under P_ξ, y_i's are i.i.d. with normal distribution with zero mean and variance $\xi + \sigma_0^2$, x_is are i.i.d. with exponential distribution with mean 1. We will show that the overshoot has exponential distribution (due to memoryless property of exponential distribution):

$$
\begin{aligned}
& P\{S_T - b \geq x; T \leq 2n\} \\
= & \sum_{l=1}^{2n} P\{S_1 \leq b, \cdots S_{l-1} \leq b, S_l \geq b + x, T = l\} \\
= & \sum_{l=1}^{n} \int_{-\infty}^{b} P\{S_1 < b, \cdots, S_{l-2} < b, S_{l-1} \in dy, S_l \geq b + x\} \\
= & \sum_{l=1}^{n} \int_{-\infty}^{b} P\{S_1 < b, \cdots, S_{l-2} < b, S_{l-1} \in dy\} P\{[\rho x_l - \ln(1+\rho)] \geq b + x - y\} \\
= & e^{-\frac{x}{\rho}} \sum_{l=1}^{n} \int_{-\infty}^{b} P\{S_1 < b, \cdots, S_{l-2} < b, S_{l-1} \in dy\} P\{[\rho x_l - \ln(1+\rho)] \geq b - y\} \\
= & e^{-\frac{x}{\rho}} \sum_{l=1}^{n} \int_{-\infty}^{b} P\{S_1 < b, \cdots, S_{l-2} < b, S_{l-1} < b, S_l \geq b\} \\
= & e^{-\frac{x}{\rho}} P(T \leq 2n).
\end{aligned}
$$

where we have used the following fact:

$$P\{[\rho x_l - \ln(1+\rho)] \geq b + x - y\} = e^{-\frac{x}{\rho}} P\{[\rho x_l - \ln(1+\rho)] \geq b - y\}.$$

So the overshoot $S_\tau - k_i$ and τ are independent. Hence the conditional distribution:

$$P\{S_T - b \in du | T \leq 2n\} = \frac{1}{\rho} e^{-\frac{u}{\rho}}.$$

Let $n \to \infty$, then we have $E_\xi\{e^{-(\Lambda_{N-T}-b)}\} \approx E_\xi\{E\{e^{-(S_T-b)}|T \leq 2n\}\} = \frac{1}{1+\rho}$.
(The moment generating function of the random walk increment is $Z(\theta) = \frac{1}{(1+\rho)^\theta(1-\theta\rho)}$.)

References

[CS09] J. Chen and D. Siegmund. A maximum score test for spatial correlation. *submitted to Biometrica*, 2009.

[KS89] H.-J. Kim and D. Siegmund. The likelihood ratio test for a change-point in simple linear regression. *Biometrica*, 76(3):409–423, 1989.

[KS09] Yoshinobu Kawahara and Masashi Sugiyama. Change-point detection in time-series data by direct density-ratio estimation. In *Proceedings of the SIAM International Conference on Data Mining, SDM 2009, April 30 - May 2, 2009, Sparks, Nevada, USA*, pages 389–400. SIAM, 2009.

[LFP08] L. Lai, Y. Fan, and H. V. Poor. Quickest detection in cognitive radio: a sequential change detection framework. *IEEE GLOBECOM 2008*, 2008.

[LLD08] H. Li, C. Li, and D. Dai. Quickest spectrum sensing in cognitive radio. *IEEE GLOBECOM 2008*, 2008.

[MF05] A. Mohammad-Djafari and O. Féron. Bayesian approach to change points detection in time series. *International Journals of Imaging Systems and Technology*, 16:215–221, Sept 2005.

[NSY08] Y. Nardi, D. Siegmund, and B. Yakir. The distribution of maxima of approximately Gaussian random fields. *The Annuals of Statistics*, 36(3):1375–1403, 2008.

[QCPS08] Z. Quan, S. Cui, H. V. Poor, and A. H. Sayed. Collaborative wideband sensing for cognitive radio. *IEEE Signal Processing Magazine*, pages 60–73, Nov. 2008.

[Sie85] D. Siegmund. *Sequential Analysis: Tests and Confidence Intervals.* Springer Series in Statistics. Springer, 1985.

[Sie88] D. Siegmund. Approximate tail probabilities for the maxima of some random fields. *The Annuals of Probability*, 16(2):487–501, 1988.

[SY00] D. Siegmund and B. Yakir. Tail probabilities for the null distribution of scanning statistics. *Bernoulli*, 6(2):191–213, 2000.

[SY07] D. Siegmund and B. Yakir. *The statistics of gene mapping.* Springer, 2007.

[TS08] R. Tandra and A. Sahai. SNR walls for signal detection. *IEEE Journal of Special Topics in Signal Processing*, 2(1):4–17, Feb. 2008.