The Annals of Applied Statistics
2010, Vol. 4, No. 2, 1081–1104
DOI: 10.1214/09-AOAS294

EXACT ASYMPTOTIC DISTRIBUTION OF CHANGE-POINT MLE FOR CHANGE IN THE MEAN OF GAUSSIAN SEQUENCES

BY STERGIOS B. FOTOPOULOS[1], VENKATA K. JANDHYALA[1] AND ELENA KHAPALOVA

Washington State University

We derive exact computable expressions for the asymptotic distribution of the change-point mle when a change in the mean occurred at an unknown point of a sequence of time-ordered independent Gaussian random variables. The derivation, which assumes that nuisance parameters such as the amount of change and variance are known, is based on ladder heights of Gaussian random walks hitting the half-line. We then show that the exact distribution easily extends to the distribution of the change-point mle when a change occurs in the mean vector of a multivariate Gaussian process. We perform simulations to examine the accuracy of the derived distribution when nuisance parameters have to be estimated as well as robustness of the derived distribution to deviations from Gaussianity. Through simulations, we also compare it with the well-known conditional distribution of the mle, which may be interpreted as a Bayesian solution to the change-point problem. Finally, we apply the derived methodology to monthly averages of water discharges of the Nacetinsky creek, Germany.

1. Introduction. While modeling time-ordered data, one is concerned about the parameters of the model being dynamically stable. One way of addressing the dynamic instability of the model parameters is to model the time dependence of parameters through a possible change at an unknown time-point so that the parameters remain stable both before and after the unknown change-point. Clearly, the methodology is extremely important from a practical point of view, mainly because the changes in phenomena observed over time usually occur unannounced, such as change in the quality characteristic of a manufacturing process, changes in water or air quality overtime, changes in the pattern of stock market indices and so on. The change-point problem allows modelers to detect the presence of any such unknown change-points and further capture them through either point or interval estimates. Such modeling has found applications from all areas of scientific endeavor, including environmental monitoring, global climatic changes, quality control, reliability, financial and econometric time series, and medicine, to name

Received December 2008; revised September 2009.
[1]Supported by NSF Grant DMS-0806133.
Key words and phrases. Ladder epochs, likelihood ratio, maximum likelihood estimate, random walk with negative drift.

a few. For examples of real life applications, see Braun and Müller (1998) for application of change-point methods in DNA segmentation and bioinformatics; Fearnhead (2006), Ruggieri et al. (2009) for applications in geology; Perreault et al. (2000a, 2000b) for application in hydrology; Jarušková (1996) for applications in meteorology; Fealy and Sweeney (2005) and DeGaetano (2006) for applications in climatology; Kaplan and Shishkin (2000) and Lebarbier (2005) for applications in signal processing; Andrews and Ploberger (1994), and Hansen (2000) for applications in econometrics; and Lai (1995), Wu, Cheng and Jeng (2005) and Zou, Qiu and Hawkins (2009) for applications in statistical process control. Even though there are recent advances in addressing multiple changes in scientific phenomena [see Fearnhead (2006), Fearnhead and Liu (2007), Girón, Moreno and Casella (2007) and Seidou and Ouarda (2007)], the classical change-point literature is most well developed in the case of a single unknown change-point in time-ordered processes.

Classical change-point methods involve two fundamental inferential problems, detection and estimation. Under the likelihood-based approach, the detection part is addressed through likelihood ratio statistics and their asymptotic sampling distributions. Maximum likelihood estimation of an unknown change-point first begins with obtaining the mle as a point estimate. Interval estimates of any desired level, which are preferred over point estimates, can be constructed around the mle, provided distribution theory for the mle is available. However, distribution theory for a change-point mle can be analytically intractable, particularly when no smoothness conditions are assumed regarding the amount of change. In contrast, advances in the Bayesian approach to change-point methodology have been occurring at a faster pace. Ever since Markov chain Monte Carlo (MCMC) methods were seen as a tool for overcoming the computational complexities in Bayesian analysis, there has been rapid progress in the overall development of this important methodological tool, and advances in Bayesian change-point analysis have not lagged behind.

While the classical change-point problem dates back to Page (1955), there has been a large amount of literature on the problem covering both detection and estimation aspects. One may consult the monographs of Brodsky and Darkhovsky (1993, 2000), Basseville and Nikiforov (1993), Csörgő and Horváth (1997), Chen and Gupta (2000) and Wu (2005), as well as a rich collection of references in these monographs for a comprehensive account of various approaches to inference on change-point problems. In reviewing the literature in terms of both theory and applications, it becomes clear that the detection aspect of the change-point problem attracted greater attention than its counterpart of estimation. Perhaps this has not been accidental, in that asymptotic theory for change-point estimators is technically a more challenging problem than deriving asymptotic distribution theory for change detection statistics. In an attempt to make estimation of the unknown change-point more accessible to practitioners, the main purpose of this paper is to derive exact computable expressions for the asymptotic distribution of the maximum likelihood estimate (mle) of the unknown change-point when a change occurs abruptly in the mean only of a Gaussian process.

Asymptotic distribution theory for the change-point mle in the abrupt case was first initiated by Hinkley (1970, 1971, 1972). While Hinkley (1970) derived the asymptotic theory for the change-point mle in a fairly general setup, the distribution was not in a computable form, and was primarily technical in nature. It turns out that Hinkley (1970) computed the distribution for change in the mean of a normal distribution only through certain approximations. While Hu and Rukhin (1995) provided a lower bound for the probability of the mle being in error of capturing the true change-point, Jandhyala and Fotopoulos (1999) and Fotopoulos and Jandhyala (2001) derived upper and lower bounds and also suggested two approximations for the asymptotic distribution of the change-point mle. Similarly, Borovkov (1999) also provided only upper and lower bounds for the distribution of the change-point mle. Thus, despite the attempts of various authors, the problem of deriving computable expressions for the asymptotic distribution of the change-point mle remained unsolved to date. It is particularly striking that exact computable expressions for the asymptotic distribution of the change-point mle have not been derived in the literature for even selected distributions of the underlying process such as the Gaussian and exponential distributions.

Tackling this important problem, we derive in this article exact computable expression for the distribution of the change-point mle when a change occurs in the mean only of a univariate or multivariate Gaussian process. The derived asymptotic distribution is not only exact but is also quite elegant and can be computed in a simple and straightforward manner. In fact, the result we derive demonstrates that the second suggested approximation in Jandhyala and Fotopoulos (1999) is the exact solution to the problem, in the Gaussian case. It should be pointed out that the distribution we derive assumes that the parameters of the distribution before and after the change-point are known. However, this should not pose difficulties, since Hinkley [(1972), page 520], in a theorem has shown that the asymptotic distribution of the change-point mle remains the same even for unknown parameter scenarios. From a practical point of view, this asymptotic equivalence result is extremely important. In practice, apart from the change-point being unknown, the parameters before and after the change-point also invariably remain unknown. The problem of deriving the distribution of the change-point mle when the parameters are unknown is the one that practitioners would be most interested, as opposed to the distribution of the change-point mle for the case when the parameters are known. There is no a priori reason to believe that the distributions of the change-point mle for the known and unknown cases be asymptotically equivalent. It is in this sense that the asymptotic equivalence result of Hinkley (1972) plays a key role for practitioners. One only needs to examine whether this asymptotic property holds well for reasonable sample sizes, and for this we carry out a simulation study in Section 4.

Since the exact solution derived in the paper assumes Gaussianity, it is tempting to explore robustness of this exact computable expression when the true process deviates from Gaussianity. If the derived result is indeed robust to such departures, then it can be applied more widely than merely Gaussian processes. While a simulation study covering a wide class of non-Gaussian families of distributions may be of interest for practitioners, in this paper we pursue a limited robustness study by performing large scale simulations wherein the error process is assumed to be symmetric and follows the t-distribution, or asymmetric and follows the standardized chi-square distribution. In both cases, we change the degrees of freedom from being small to large, so that one approaches Gaussianity as the degrees of freedom become large.

Hinkley's approach to deriving distribution of the change-point mle is perceived as the unconditional approach in the literature. Against this, Cobb (1978) proposed a conditional approach to the distribution of the change-point mle, wherein the distribution of the mle is derived by conditioning upon sufficient information on either side of the unknown change-point. Since the exact distribution of the unconditional mle is now available, it is relevant to compare the conditional and unconditional distributions in terms of their performance, including robustness properties. Thus, we have also included Cobb's conditional distribution in our simulations. As pointed out by Cobb (1978), since the conditional distribution of the change-point mle can also be interpreted as the Bayesian posterior for the change-point under a uniform prior on the unknown change-point, the comparisons between the two distributions have a broader appeal than what might appear at first glance.

Finally, we apply the methodology derived in the paper to multivariate analysis of hydrological data. The data, previously analyzed in a univariate setup by Gombay and Horváth (1997), represents averages of log transformed water discharges for the Nacetinsky creek for the months of February, July and August during the years 1951–1990. The bivariate and trivariate change-point analysis shows that a significant increase has occurred in the water discharges, whereas the univariate change-point analyses show no significant changes in the mean water flows.

The organization of the paper is as follows. In Section 2 we present some general background regarding the change-point mle and its asymptotic distribution. Then, we state the main theorem in Section 3, and the proof of the theorem is presented in Appendix A. While Section 4 consists of empirical assessment of the performance of derived theory for the case of known and unknown parameters, Section 5 contains the multivariate change-point analysis of the Nacetinsky creek data. Finally, Section 6 concludes the paper with a discussion.

2. Distribution of the mle.

Let $Y_1, Y_2, \dots, Y_n$, $n \geq 1$, be a sequence of real-valued independent time ordered random variables defined on a probability space (Ω, F, P). Let there be a natural number $\tau_n \in \{1, 2, \dots, n-1\}$ such that $Y_1, Y_2, \dots, Y_{\tau_n}$ have a common distribution F_1, whereas the subsequent observations $Y_{\tau_n+1}, Y_{\tau_n+2}, \dots, Y_n$ have a common distribution F_2 with $F_1 \neq F_2$. Here, the change-point τ_n is an unknown parameter and should be estimated. The likelihood function of τ_n is given by $p_n(Y; \tau_n) = \prod_{i=1}^{\tau_n} f_1(Y_i) \prod_{i=\tau_n+1}^{n} f_2(Y_i)$, where the functions f_1 and f_2 are densities of F_1 and F_2, respectively, with respect to some dominating measure $\mu(F_1, F_2 \ll \mu)$. In the sequel we assume that the densities f_1 and f_2 are known, perhaps through known parameters. Following Hinkley (1970), the mle $\hat{\tau}_n$ may be expressed as

$$\hat{\tau}_n = \underset{1 \le j \le n-1}{\arg\max} \sum_{i=1}^{j} a(Y_i), \tag{2.1}$$

where $a(Y_i) = \log\{f_1(Y_i)/f_2(Y_i)\}, i = 1, \ldots, n-1$. For establishing distribution theory, it is convenient to work with $\hat{\tau}_n - \tau_n \in \{-\tau_n + 1, \ldots, n - \tau_n - 1\}$ instead of $\hat{\tau}_n$. Hence, we have

$$\xi_n = \hat{\tau}_n - \tau_n = \underset{-\tau_n+1 \le j \le n-\tau_n-1}{\arg\max} \sum_{i=1}^{\tau_n+j} a(Y_i), \tag{2.2}$$

where the maximizer is a result of the following two-sided random walk $\Gamma(\cdot)$:

$$\Gamma_n(j;\tau_n) = \begin{cases} \sum_{i=1}^{j} a(Y_i^*) = \sum_{i=1}^{j} X_i^* = S_j^*, & j \in \{1, \ldots, n-\tau_n-1\}, \\ 0, & j = 0, \\ -\sum_{i=1}^{-j} a(Y_i) = \sum_{i=1}^{-j} X_i = S_{-j}, & j \in \{-1, \ldots, -\tau_n+1\}. \end{cases} \tag{2.3}$$

Here, $\{Y, Y_i : i \ge 1\}$ and $\{Y^*, Y_i^* : i \ge 1\}$ are two independent sequences with independent and identical copies on $(\mathbf{R}, \mathrm{R})$ such that Y is distributed according to F_1, and Y^* is distributed according to F_2. Note that X and X^* are real valued random variables defined on $\mathbf{R}$. Also note that when $F_1 \ne F_2$,

$$\begin{aligned} E(X) &= -\int_S \log\{f_1(x)/f_2(x)\} f_1(x)\mu(dx) = -K(f_1, f_2) \\ &= -E_{f_1}\{a(Y)\} < 0 \quad \text{and} \\ E(X^*) &= \int_S \log\{f_1(x)/f_2(x)\} f_2(x)\mu(dx) = -K(f_2, f_1) \\ &= E_{f_2}\{a(Y^*)\} < 0, \end{aligned} \tag{2.4}$$

where K is the usual Kullback–Leibler information. It can be seen that (2.4) is also related to the entropy function, which in many instances is used for measuring the distinctness of probabilities. We assume that $P(X > 0) > 0$. For $\theta > 0$, let

$$\phi(\theta) = E\{\exp(\theta X)\} \quad \text{and} \quad \psi(\theta) = E\{\exp(\theta X^*)\}. \tag{2.5}$$

Note that $\phi(\theta) = \psi(1-\theta)$. Moreover, $\phi(\theta) \le 1, \forall \theta \in [0, 1]$, since

$$\begin{aligned} \phi(\lambda) &= \int_S f_1(x)\{f_1(x)/f_2(x)\}^{-\lambda}\mu(dx) = \int_S f_1^{1-\lambda}(x) f_2^{\lambda}(x)\mu(dx) \\ &\le \left\{\int_S f_1(x)\mu(dx)\right\}^{1-\theta} \left\{\int_S f_2(x)\mu(dx)\right\}^{\theta} = 1. \end{aligned} \tag{2.6}$$

It is known that when $E(X) < 0$, $P(X > 0) > 0$ and $\vartheta = \sup\{\theta > 0 : \phi(\theta) \le 1\}$, the asymptotic behavior of the tail for the ultimate maximum, $M = \sup\{S_n : n \in \mathbf{N}\}$, can be described by the following three cases:

(i) $\vartheta = 0$, the tail has a polynomial form (sub-exponential case),
(ii) $\vartheta > 0$ and $\phi(\vartheta) < 1$ an intermediate case,
(iii) $\vartheta > 0$ and $\phi(\vartheta) = 1$ the Cramér's case.

Now, in a sequence of observations for which $F_1 \neq F_2$, the μ-derivatives also satisfy $f_1 \neq f_2$. From (2.6), it is clear that the choice of ϑ greater than zero for which (iii) is satisfied is $\vartheta = 1$, the unity. Consequently, it follows that X satisfies Cramér's condition. Furthermore, merely noting that $\psi(\vartheta) = \phi(1 - \vartheta)$, it follows that X^* also satisfies Cramér's condition. This observation implies that $\vartheta = \vartheta^* = 1$, in Proposition 1 of Jandhyala and Fotopoulos (1999) for general distributions including Gaussian random variables.

It also follows that $\phi(\theta) < 1, \forall\theta \in (0, 1)$ and that ϕ is strictly convex on $\theta \in (0, 1)$. This suggests that $\phi(\theta)$ attains its minimum at a unique $\theta_0 \in (0, 1)$ such that $\phi(\theta_0) = \inf_{\theta \in (0,1)} \phi(\theta) < 1$. This firmly establishes that assumptions 1–3 in Jandhyala and Fotopoulos (1999) are no more required and that they hold naturally whenever $F_1 \neq F_2$, and $P(X > 0) > 0$ are satisfied.

In this paper we are interested in deriving the distribution of the limiting variable ξ_∞, by letting $n \to \infty$ in such a way that $\tau_n \to \infty$ and $n - \tau_n \to \infty$. In this regard, it has been shown that ξ_∞ is a proper random variable and $\xi_n \to \xi_\infty$ a.s. [see, e.g., Fotopoulos and Jandhyala (2001)].

We begin by stating a theorem found in Fotopoulos (2009). For all purposes, this result is a restatement of Theorem 2 in Jandhyala and Fotopoulos (1999).

THEOREM 2.1. *Let $F_1 \neq F_2$ and $P(X > 0) > 0$. Then, the probability distribution of ξ_∞ is given by*

$$P(\xi_\infty = j) = \begin{cases} P(T_1^+ = \infty)\Big\{P(T_1^- > -j) \\ \qquad\qquad - \displaystyle\int_{0+}^{\infty} P(M^* \ge x) P(T_1^- > -j \cap S_{-j} \in dx)\Big\}, \\ \qquad j \le -1, -2, \dots, \\ P(T_1^+ = \infty) P(T_1^{*+} = \infty), \qquad j = 0, \\ P(T_1^{*+} = \infty)\Big\{P(T_1^{*-} > j) \\ \qquad\qquad - \displaystyle\int_{0+}^{\infty} P(M \ge x) P(T_1^{*-} > j \cap S_j^* \in dx)\Big\}, \\ \qquad j = 1, 2, \dots, \end{cases}$$

where $T_1^+ := \inf\{j > 0 : S_j > 0\}$, $T_1^- := \inf\{j > 0 : S_j \le 0\}$ and $M := \max_{0 \le n} S_n$, and M^*, T_1^{*+} and T_1^{*-} are defined in a similar manner.

The convergence rate of the above asymptotic result is of interest for purposes of both theory and practice. Knowledge about the convergence rate allows one to judge the appropriateness of the sample size and other ancillary parameters for which the asymptotic distribution can be utilized for finite sample sizes without committing disproportional errors. In this regard, both Borovkov (1999) and Jandhyala and Fotopoulos (2001) derived important results that establish the convergence rate applicable to Theorem 2.1. We state here some relevant facts from these articles and then formulate a theorem without proof that establishes a bound for the total variation distance between the finite sample and infinite sample distributions of the change-point mle.

From Theorem 2 of Jandhyala and Fotopoulos (2001), we have

$$\sup_{B\in \mathsf{B}_{\tau_n,n}} |P(\xi_n \in B) - P(\xi_\infty \in B)| = P(\xi_\infty \le -\tau_n \text{ or } \xi_\infty \ge n-\tau_n),$$

where $\mathsf{B}_{\tau_n,n}$ is the Borel σ-field defined on $\mathbf{Z}_{\tau_n,n} \equiv \{-\tau_n+1,\dots,0,\dots, n-\tau_n-1\}$. Then, as argued in Jandhyala and Fotopoulos (2001), upon augmenting $\mathsf{B}_{\tau_n,n}$ into the Borel σ-filed on $\mathbf{Z}$, it follows that the total variation distance between ξ_n and ξ_∞ defined by

$$d_{\mathrm{TV}}(\xi_n,\xi_\infty) = \sup_{B\in\mathsf{B}} |P(\xi_n\in B) - P(\xi_\infty \in B)|$$

may be seen to yield

$$d_{\mathrm{TV}}(\xi_n,\xi_\infty) = P(\xi_\infty \le -\tau_n \text{ or } \xi_\infty \ge n-\tau_n). \tag{2.7}$$

The following theorem, which provides a bound for $d_{\mathrm{TV}}(\xi_n,\xi_\infty)$, follows immediately upon applying (2.7) into Theorem 1 of Borovkov (1999).

THEOREM 2.2. *Let $F_1 \ne F_2$ and $P(X>0)>0$. Let ξ_n and ξ_∞ be the centered random variables of the change-point mle for finite and infinite samples, respectively. Then, the total variation distance between ξ_n and ξ_∞ admits the inequality given by*

$$d_{\mathrm{TV}}(\xi_n,\xi_\infty) \le 4\max\{\phi(\theta_0)^{\tau_n}, \phi(\theta_0)^{n-\tau_n}\},$$

where $\phi(\theta_0) = \inf_{\theta\in(0,1)} \phi(\theta) < 1$.

Theorem 2.2 clearly establishes a geometric rate of convergence as ξ_n approaches ξ_∞, asymptotically. The above result is more friendly from a computational point of view than Theorem 3 of Jandhyala and Fotopoulos (2001).

While Theorem 2.1 provides the probability distribution of ξ_∞, the expressions therein are still only of technical interest. The main problem is that, as far as we know, a computable expression for the distribution function $M(x)$ [or $M^*(x)$] is not available in the literature. Clearly, the behavior of $1-M(x)$ (or $1-M^*(x)$) depends upon the characteristics of the underlying distributions f_1 and f_2, in study.

Moreover, the term $P(T_1^+ = \infty)$ that appears in both Theorems 2.1 and 2.2 may also be unavailable for computation unless we know the exact distribution of S_n, for all $n \in N$. Thus, the determination of an exact expression for the distribution of M for any general distribution is beyond analytical scope, and consequently, an exact computable form for the probability distribution $P(\xi_\infty = j)$, $j \in \mathbf{Z}$, in Theorem 2.1 is also analytically not tractable. To this extent, in this paper we shall concentrate on developing the analysis by assuming that the underlying process is of Gaussian type.

3. Asymptotic distribution of the mle under Gaussian processes. We shall establish the main theorem regarding computationally accessible distribution of ξ_∞ first under the univariate Gaussian case. Subsequently, we shall illustrate how the univariate case itself can be directly applied to the more general multivariate setup.

3.1. *The univariate Gaussian case.* We begin by assuming that the underlying process is univariate Gaussian, and the means before and after the change-point are given by μ_1, μ_2, wherein we let $\mu_1 \neq \mu_2$. We do assume that the standard deviation σ is known and remains the same throughout the sampling period. Clearly, the likelihood ratios in (2.1) may then be expressed as

$$
\begin{aligned}
X &= -a(Y) = \log\{f_2(Y)/f_1(Y)\} \\
&= \log\left\{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(Y-\mu_2)^2/2\sigma^2} \Big/ \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(Y-\mu_1)^2/2\sigma^2}\right\} \\
&=_{\mathsf{D}} -\frac{(\mu_1-\mu_2)^2}{2\sigma^2} - \frac{(\mu_1-\mu_2)}{\sigma} Z,
\end{aligned}
\tag{3.1}
$$

where $Z \sim N(0,1)$, and, similarly,

$$
X^* =_{\mathsf{D}} -\frac{(\mu_1-\mu_2)^2}{2\sigma^2} + \frac{(\mu_1-\mu_2)}{\sigma} Z^*,
\tag{3.2}
$$

where $Z^* \sim N(0,1)$, and is independent of Z. Note that in this case, the random variables X and X^* are both identically distributed with means $E(X) = E(X^*) = -\eta^2/2 < 0$ and variances $\operatorname{var}(X) = \operatorname{var}(X^*) = \eta^2$, where $\eta = \frac{|\mu_1-\mu_2|}{\sigma}$ represents the standardized amount of change. Hence, it is sufficient to confine our analysis to only one side of the random walk $\Gamma(\cdot)$.

Under the formulation in (3.1), it can be seen that $S_n =_{\mathsf{D}} -n\frac{\eta^2}{2} - \eta\sqrt{n}Z$, where again $Z \sim N(0,1)$. Note [Asmussen (1987), Corollary 4.4] that when $E(X) < 0$, the ladder height distribution given by $G_+(dx) = P(S_{T_1^+} \in dx \cap T_1^+ < \infty)$ is defective. Thus, $\|G_+\| = P(T_1^+ < \infty) < 1$ and $\frac{1}{E(T_1^-)} = 1 - \|G_+\| = P(T_1^+ = \infty) = P(M = 0)$. We shall now state our main theorem, which provides a computable expression for the distribution of ξ_∞. The computability of the terms in the expression will be demonstrated in the discussion following the theorem. The proof of the theorem is presented in Appendix A. Subsequent to the theorem, we state a corollary, which establishes a closed form computable expression for the bound in Theorem 2.2.

THEOREM 3.1. *Suppose that the time-ordered sequence* $Y_1, Y_2, \dots, Y_n$, $n \geq 1$, *is such that* $Y_i \sim N(\mu_1, \sigma^2), i = 1, \dots, \tau_n$, *and* $Y_i \sim N(\mu_2, \sigma^2)$, $i = \tau_n + 1, \dots, n$. *Then, the probability distribution of* ξ_∞ *is given by*

$$P(\xi_\infty = k) = \begin{cases} (1 - \|G_+\|)\big(q_{|k|} - \|G_+\|\tilde{q}_{|k|}\big), & k = \pm 1, \pm 2, \dots, \\ (1 - \|G_+\|)^2, & k = 0, \end{cases}$$

where $1 - \|G_+\| = \exp\{-\sum_{j=1}^{\infty} \frac{1}{j} \bar{\Phi}(\eta\sqrt{j}/2)\}$ *and* $q_k = E\{I(T_1^- > k)\}$, $\tilde{q}_k = E\{e^{-S_k} I(T_1^- > k)\}$, $k = 1, 2, \dots$ *and* $q_0 = \tilde{q}_0 = 1$.

It is fairly straightforward to state the bound in Theorem 2.2 for the Gaussian case. Specifically, it follows that the total variation distance in the Gaussian case admits

$$d_{\mathrm{TV}}(\xi_n, \xi_\infty) \leq 4 \max\left\{\exp\left(-\frac{\eta^2 \tau_n}{8}\right), \exp\left(-\frac{\eta^2 (n - \tau_n)}{8}\right)\right\}. \tag{3.3}$$

3.2. *The multivariate Gaussian case.* Here, we let $\{Y, Y_i : i \in \mathbf{N}\}$ be a sequence of time-ordered independent Gaussian elements defined on R^d, the d-dimensional Euclidean space with $f(x; \mu_{d\times 1}, \Sigma_{d\times d})$ denoting the corresponding probability density function. In the sequel, mainly for convenience, we represent the parameter only as (μ, Σ) by dropping the respective dimension subscripts. Let the parameter (μ, Σ) change from its initial value of (μ_1, Σ) to (μ_2, Σ), at some unknown index point $\tau_n \in \{1, 2, \dots, n-1\}$, with mean vectors $\mu_1, \mu_2 \in \Theta$, and common variance-covariance matrix Σ. For reason of convenience, we assume that Σ is positive definite and the mean vectors satisfy $\mu_1 \neq \mu_2$.

The functional $\langle x, y\rangle$ denotes the usual inner product and the extended seminorm is defined if there exists a covariance operator Σ such that $\|x\|_\Sigma^2 = \langle \Sigma x, x\rangle$. Then, we may write $Y =_{\mathsf{D}} \mu_1 + \Sigma^{1/2}\mathbf{Z}$ for all data before the change-point, where $\mathbf{Z}$ is a d-variate standard normal vector. Consequently, the random variable $X = -\ln f(Y; \mu_1, \Sigma)/f(Y; \mu_2, \Sigma)$ is expressed as

$$\begin{aligned} X &= \tfrac{1}{2}\{\langle \Sigma^{-1}(Y - \mu_1), Y - \mu_1\rangle - \langle \Sigma^{-1}(Y - \mu_2), Y - \mu_2\rangle\} \\ &=_{\mathsf{D}} -\tfrac{1}{2}\|\mu_1 - \mu_2\|_{\Sigma^{-1}}^2 - \|\mu_1 - \mu_2\|_{\Sigma^{-1}} Z, \end{aligned} \tag{3.4}$$

where Z now stands for the standard normal random variable with mean zero and variance one.

Similarly, for data after the change-point, we have $Y =_{\mathrm{D}} \mu_2 + \Sigma^{1/2}\mathbf{Z}^*$, where $\mathbf{Z}^*$ is the d-variate standard normal vector, and in this case, we obtain

$$
\begin{aligned}
X^* &= \ln f(Y; \mu_1, \Sigma)/f(Y; \mu_2, \Sigma) \\
&=_{\mathrm{D}} -\tfrac{1}{2}\|\mu_1 - \mu_2\|^2_{\Sigma^{-1}} + \|\mu_1 - \mu_2\|_{\Sigma^{-1}} Z^*,
\end{aligned}
\tag{3.5}
$$

where Z^* is univariate standard normal independent of Z. Upon letting $\eta = \|\mu_1 - \mu_2\|_{\Sigma^{-1}}$ represent the amount of standardized change in the means, it should be clear that the multivariate case translates itself into a corresponding univariate case with η as defined above.

4. Performance of the distribution of the change-point mle. In this section we wish to assess the performance of the derived asymptotic distribution in two different ways. First, we investigate the equivalence result of Hinkley (1972) and, second, we compare the derived distribution of the mle with the conditional distribution of mle as derived by Cobb (1978).

4.1. *Distribution of the change-point mle for known and unknown parameters.* The assumption of known parameters does not apply in practice, and it is common that they must be estimated from the data. While Hinkley (1972) has shown asymptotic equivalence of change-point mle under both known and estimated cases, its applicability to sample sizes of practical interest requires empirical evidence. This issue is perhaps even more important in the multivariate case, mainly because the multivariate case involves estimation of many more parameters. As discussed in Sections 2 and 3, for comparing the closeness of two distributions, we find it convenient to utilize the total variation distance measure, which for discrete random variables X and Y is given by $d_{\mathrm{TV}}(X, Y) = \frac{1}{2}\sum_{i\in\mathbf{Z}} |P(X = i) - P(Y = i)|$.

Simulations are performed by letting the parameter choices for sample size and true change-point be as follows: $n = 40$, $\tau = 20$; $n = 60$, $\tau = 20$; $n = 60$, $\tau = 30$; $n = 100$, $\tau = 20$; $n = 100$, $\tau = 30$; $n = 100$, $\tau = 40$ and $n = 100$, $\tau = 50$. For each of the above cases, the choice of values for η are set at $\eta = 1.0, 1.5, 2.0, 2.5$. The results for univariate and bivariate cases based on 500,000 simulations for each individual scenario are presented in Tables 1 and 2, respectively. As one might expect, the situation of known parameters yields excellent agreement with the theoretical distribution in both tables, irrespective of the sample size as well as the location of the change-point. When parameters are estimated, the univariate case (Table 1) shows very good to extremely good agreement with the theoretical distribution. The values, for even the bivariate case (Table 2), show very good agreement except when η is very small ($\eta = 1$).

TABLE 1
Total variation distances of known and estimated empirical distributions (based on 500,000 simulations) from theoretical distribution of change-point mle in the univariate case

n	τ	$\eta = 1$		$\eta = 1.5$		$\eta = 2$		$\eta = 2.5$	
		Known	**Est.**	**Known**	**Est.**	**Known**	**Est.**	**Known**	**Est.**
100	20	0.0106	0.0665	0.0070	0.0264	0.0033	0.0139	0.0014	0.0082
100	30	0.0113	0.0493	0.0065	0.0205	0.0032	0.0104	0.0021	0.0057
100	40	0.0112	0.0437	0.0065	0.0189	0.0033	0.0091	0.0020	0.0050
100	50	0.0109	0.0412	0.0068	0.0176	0.0040	0.0082	0.0022	0.0044
60	20	0.0105	0.0721	0.0070	0.0298	0.0033	0.0155	0.0014	0.0086
60	30	0.0112	0.0641	0.0065	0.0271	0.0032	0.0133	0.0021	0.0076
40	20	0.0104	0.0852	0.0070	0.0383	0.0033	0.0191	0.0014	0.0105

4.2. *Unconditional change-point mle against Cobb's conditional mle.* Cobb (1978) derived conditional distribution of the change-point mle by conditioning upon sufficient observations around the true change-point, which according to Cobb (1978) is also equivalent to the Bayesian posterior when the prior on the unknown change-point is uniform. If δ denotes the number of data points to be considered on either side of $\hat{\tau}_n$, then Cobb's conditional solution for $l \in \{-\delta, \dots, \delta\}$ is given by

$$P(\hat{\tau}_n - \tau_n = l | Y_{\hat{\tau}_n-\delta+1}, \dots, Y_{\hat{\tau}_n+\delta}) \cong p_n(Y; \hat{\tau}_n + l) \Big/ \sum_{l=-\delta}^{\delta} p_n(Y; \hat{\tau}_n + l). \tag{4.1}$$

TABLE 2
Total variation distances of known and estimated empirical distributions (based on 500,000 simulations) from theoretical distribution of change-point mle in the bivariate case

n	τ	$\eta = 1$		$\eta = 1.5$		$\eta = 2$		$\eta = 2.5$	
		Known	**Est.**	**Known**	**Est.**	**Known**	**Est.**	**Known**	**Est.**
100	20	0.0108	0.0991	0.0066	0.0376	0.0035	0.0197	0.0018	0.0126
100	30	0.0110	0.0718	0.0065	0.0281	0.0034	0.0153	0.0016	0.0099
100	40	0.0119	0.0624	0.0070	0.0252	0.0044	0.0135	0.0017	0.0075
100	50	0.0121	0.0595	0.0076	0.0236	0.0040	0.0126	0.0016	0.0075
60	20	0.0107	0.1140	0.0066	0.0466	0.0035	0.0248	0.0018	0.0157
60	30	0.0107	0.1006	0.0065	0.0410	0.0034	0.0218	0.0016	0.0146
40	20	0.0105	0.1383	0.0065	0.0647	0.0035	0.0350	0.0018	0.0233

The method of choosing δ is clearly detailed in Cobb (1978). It is then relevant to compare the unconditional distribution of the mle derived in Section 3 with the above conditional solution. Also, we investigate the robustness of the exact limiting distribution for departures from normality through simulations, limiting the study to the univariate framework only. Here, incorporating both symmetric and asymmetric distributions, the error structures are modeled by the standardized t_ν and χ^2_ν distributions.

For simplicity, we let only $\eta = 1.0$ and $\eta = 2.5$, and then perform simulations for all the choices of sample sizes and true change-points considered in Section 4.1. The choices of ν under t_ν-distribution were $\nu = 5, 10, 20$ and they were $\nu = 1, 5, 20$ under χ^2_ν-distribution. Note that while implementing Cobb's conditional solution, we determined the value of δ so that the error rate detailed in Cobb (1978) is close to 10^{-5}. To save space, we present the computed distributions (based on 50,000 simulations) in the form of figures only, and that too only for the case of $n = 100$, $\tau = 50$. Figure 1(a–c) correspond to the cases of normal, t_5 and χ^2_1 distributions when $\eta = 1.0$, and Figure 1(d–f) correspond to the same cases when $\eta = 2.5$.

For the remaining cases, we summarized the computed distributions through Bias and mean square error (MSE), and to save space, we only describe the salient features of these computations. It can be seen from Figure 1(a) that in the normal case, the unconditional distributions under both known and estimated cases are almost identical and they closely agree with the theoretical distribution even when change is small with $\eta = 1.0$. While the distributions of cmle under known and estimated cases are also quite identical to each other, there is more spread in the cmle, with the probability at the true change-point being substantially smaller than that of the unconditional mle. It is clear from Figure 1(b) and (c) that robust to deviations from normality is quite pronounced even when degrees of freedom under t_5 and χ^2_1 distributions are small. Moving on to $\eta = 2.5$, we find from Figure 1(d–f) that, overall, there is greater robustness and even better agreement between known and estimated solutions.

Though not presented, the Bias and MSE values show some differences from known case to the estimated case, mainly when η is small ($\eta = 1.0$). The robustness for large changes ($\eta = 2.5$) is extremely good throughout the computations, thus depicting good tail behavior for large changes under both t and χ^2 distributions. Also, extreme behavior is noticed for the estimated case when $\eta = 1.0$ and $n = 100$, $\tau = 20$. In this case, Cobb's cmle shows somewhat smaller MSE values than the mle, though only marginally. For all other parameter choices, the mle performs better in terms of MSE values.

Finally, we noticed that the behavior of MSE values for mle in the known case are lower than the corresponding theoretical MSE values and that the MSE values increase with the sample size. This behavior can be explained by the fact that the theoretical distribution derived for infinite samples possesses infinite domain, whereas the domain under finite samples is truncated by the sample size. This

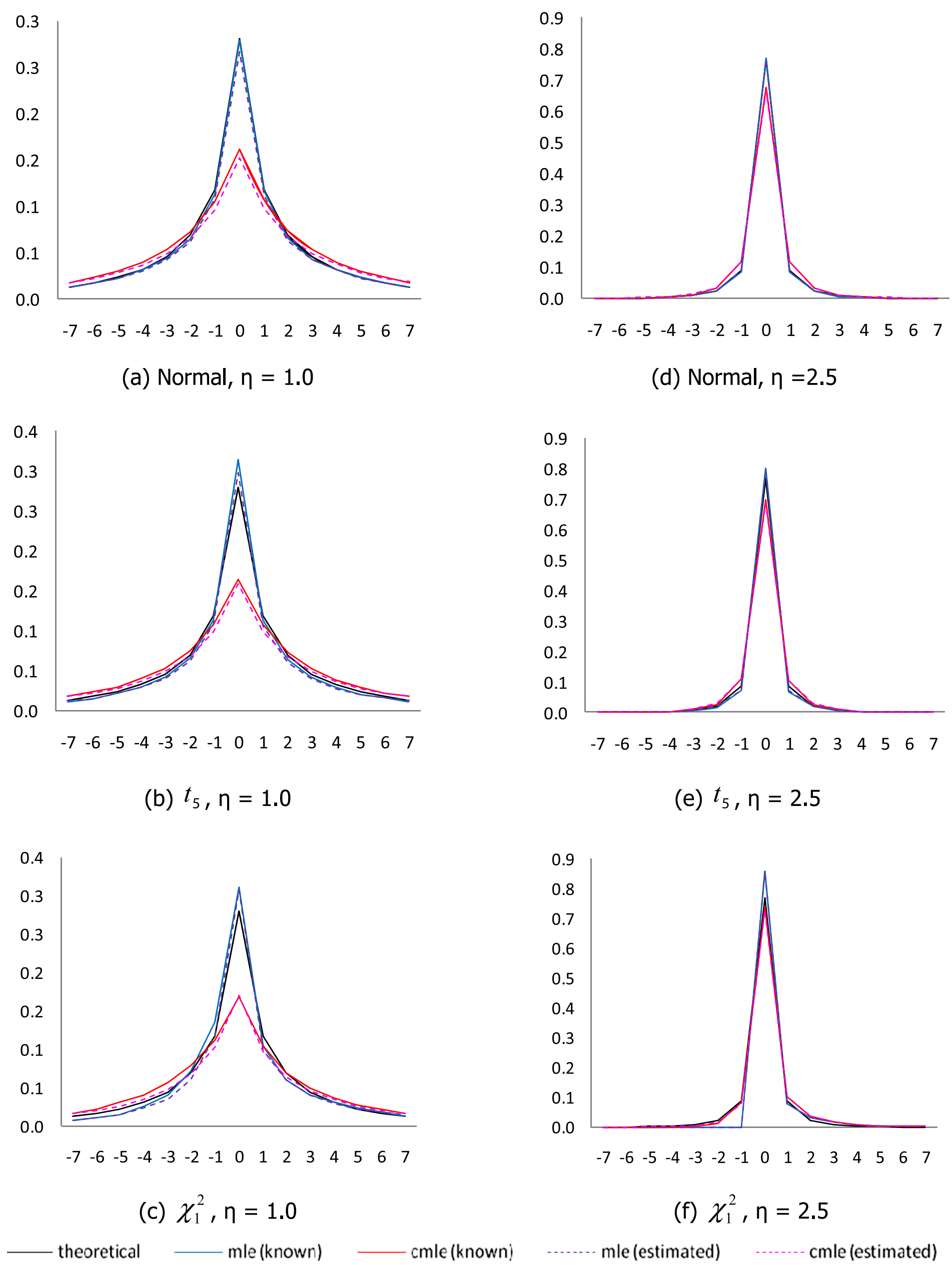

FIG. 1. *Plots of theoretical mle, empirical mle (known), empirical mle (estimated), empirical cmle (known) and empirical cmle (estimated) distributions of the centered change-point when* $n = 100, \tau = 50$ *under normal* (a); t_5 (b) *and* χ_1^2 (c) *when* $\eta = 1.0$; *and normal* (d); t_5 (e) *and* χ_1^2 (f) *when* $\eta = 2.5$.

truncation effect for finite samples is found to be most pronounced when $n = 40$. The same argument also explains why MSE values in both tables increase with increasing sample sizes.

5. Multivariate change-point analysis of water discharges at Nacetinsky creek. The Nacetinsky is a small creek in the German part of the Ergebirge Mountains. Gombay and Horváth (1997) analyzed the monthly averages of water discharges for the Nacetinsky creek during the years 1951–1990 and found that the lognormal distribution appropriately models the monthly average discharges in the creek. Consequently, applying the log transformation, they applied likelihood ratio based change detection methodology in a univariate framework for detecting changes in mean only as well as changes in the variance only of the normal distribution for the transformed data. When changes were detected, they obtained point estimates of the unknown change-point by the value at which the likelihood ratio was maximum. In detecting the change points, Gombay and Horváth (1997) found that the change-detection methodology under independence was applicable for the monthly water discharges.

We revisited the monthly data and first analyzed the data in a univariate setup, mainly for detecting changes in mean only or variance only of the transformed data. Applying the respective likelihood ratio change-detection statistics (B.2) and (B.4) in Appendix B, we found no evidence of change in either the mean or in the variance for almost all months. We were then interested to learn whether bivariate or multivariate analyses might convey a different message than what has been learned from the univariate analysis. One can expect significant covariances in the water discharges among various months within a year, and it is of interest to know whether such covariances contribute significantly as one pursues change-detection and estimation. To this extent, we found that a multivariate analysis of the data for the months of February, July and August yields some interesting results.

Change-point analysis, whether at the univariate level or at the multivariate level, involves two parts, namely, change-detection and change-point estimation whenever a change-point is detected. The focus of this paper clearly is on estimation, where we derive computable expressions for the asymptotic distribution of the change-point mle. Change-detection is not pursued in the theoretical part of this paper. However, change-detection precedes change-point estimation for the analysis of data. Keeping this in mind, we first present analysis and results from change-detection in Appendix B, and only results from change-point estimation will be emphasized in this section. Once again, our analysis in both detection and estimation is based on log transformed water discharges data for the months of February, July and August as reported in Figure 2.

To proceed with the formulation, let Y_i represent the log transformed monthly water discharges at the Nacetinsky creek for the months of February, July and August for the for the ith year, $i = 1, \dots, 40$, so that in this case the dimension $d = 3$, and the sample size $n = 40$. We begin modeling the data by assuming that $Y_1, \dots, Y_n$ are independent and that $Y_i \sim N(\mu^{(i)}, \Sigma), i = 1, \dots, n$.

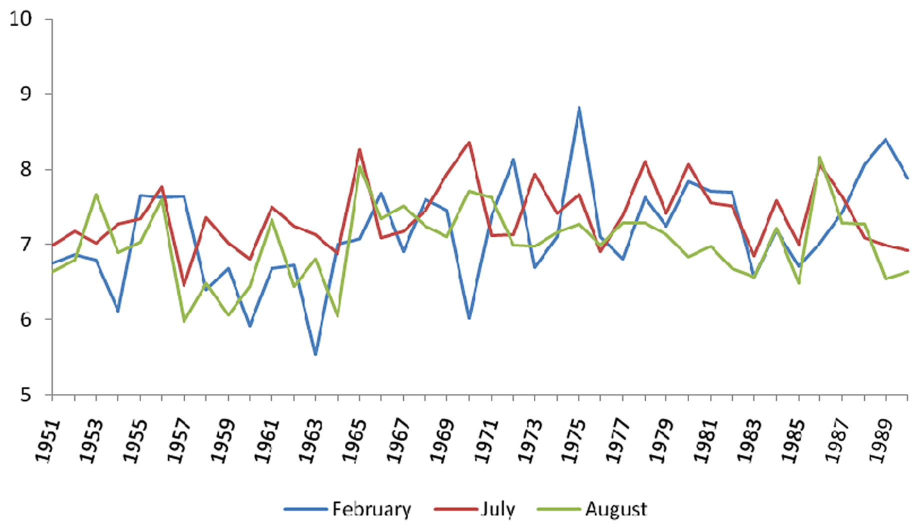

FIG. 2. *Time series plot of log transformed data on mean monthly water discharges of the Nacetinsky creek for the months of February, July and August for the years* 1951–1990.

Under the change-point setup with τ_n as the unknown change-point, one lets $\mu^{(i)} = \mu_1, i = 1, \ldots, \tau_n$ and $\mu^{(i)} = \mu_2, i = \tau_n + 1, \ldots, n$.

With the above as the basic setup, one can first apply change-detection methodology, and this has been done comprehensively in Appendix B. Basically, it has been found that the bivariate tests for Feb–Jul, and Feb–Aug pairs as well as the multivariate test for all the three months, were found to be significant even though none of the univariate tests showed significance. The bivariate and multivariate analyses resulted in the change-point mle being $\hat{\tau}_n = 14$, so that a change in water discharges occurred subsequent to the year 1964. The analysis in the Appendix was quite supportive of the assumptions of both Gaussianity and independence.

We shall now implement the theoretical distribution derived in Section 3 to the data in Figure 2 under the bivariate and trivariate cases. Based on $\hat{\tau}_n = 14$, we estimated the values of η to be $\hat{\eta}_{FJ} = 1.47$, $\hat{\eta}_{FA} = 1.52$ and $\hat{\eta}_{FJA} = 1.60$. Visualizing these as known values, we implemented the theoretical distribution for each of the three cases. We found the period 1960–1968 to yield confidence levels of 94.8%, 95.6% and 96.5%, respectively. Simulations suggest that the same period under both bivariate and trivariate estimated cases with true parameter values set at $\eta = 1.51$ and $n = 40$, $\tau = 14$ yields a confidence level of 90%. Applying the conditional distribution of Cobb (1978) for the same data with an error rate of approximately 10^{-5}, we found that 95% coverage probability for Feb–Jul is the period 1963–1971, for Feb–Aug the period is 1963–1969, and for Feb–Jul–Aug the period is obtained as 1963–1967. Clearly, for this particular data, Cobb's cmle seems to yield shorter confidence interval than the unconditional mle. However, under repeated samples for data of the same size with the true parameters set at

$\eta = 1.51$ and $n = 40$, $\tau = 14$, we found that the period 1960–1968 under Cobb's cmle yields a coverage probability of 88% under both bivariate and trivariate cases, thus showing a similar performance as the mle on average.

6. Discussion. Asymptotic distribution of the change-point mle is quite complicated and an exact computable expression for the distribution of the mle has not been derived in the literature to date, even though Hinkley (1970, 1971, 1972) published his seminal work more than three decades back. Assuming the parameters before and after the unknown change-point to be known, this investigation establishes an exact and yet computationally attractive form for the asymptotic distribution of the change-point mle, thus far not available in the literature.

To have a better understanding of its performance, we carried out an empirical study to compare the distribution under known parameters with the case where the nuisance parameters remain unknown. We also compare the derived distribution with the conditional distribution of Cobb (1978) as well as assessing the robustness of the derived distribution for departures from normality. Simulations have shown good agreement between known and estimated cases except for the case where parameters are estimated and amount of change is relatively small. Also, both mle and cmle are quite robust to deviations from normality, for the most part.

We have applied the derived change-point estimation methodology to compute the asymptotic distribution under both mle and cmle methods for the log transformed data on annual mean discharges for the months of February, July and August for the Nacetinsky creek for the years 1951–1990. At first it may appear that sample size of $n = 40$ may be somewhat small for asymptotics to apply. However, simulations under the estimated case for samples of this size show excellent accuracy in the univariate case (Table 1, $\eta = 1.5$) and good accuracy in the bivariate case (Table 2, $\eta = 1.5$). Detection methodology for this data set under univariate setup yields no significance for the presence of a change-point for any of the three months. However, change-detection under the multivariate setup shows significance for Feb–Jul and Feb–Aug in the bivariate case and also for the trivariate case of Feb–Jul–Aug.

In summary, the methodology proposed in this article appears quite useful for practitioners in all areas, mainly because it is readily computable, and it is quite robust to deviations from the assumption Gaussianity. Also, sample size does not seem to be a serious concern while implementing the asymptotic result. In terms of future directions, it would be of interest to derive such computationally feasible distributions for other distributions such as exponential and Weibull in the continuous case and binomial and Poisson in the discrete case.

APPENDIX A

Proof of Theorem 3.1. The proof of the theorem essentially follows upon applying the following three lemmas into Theorem 2.1.

The following lemma is well known [see, e.g., Shiryaev et al. (1994)], and will be given without proof. It should be noted that even though the original result was given for the continuous Brownian motion, the same can be applied for a random walk with negative drift. This lemma addresses the fundamental issue of establishing the distributions of M (and M^*) in a simple exponential form, thereby making the integrals in Theorem 2.1 analytically tractable.

LEMMA 1. *Let the random walk* $\{S_n, n \geq 0\}$ *be as specified in* (2.3). *Then, for* $x \geq 0$,

$$P\Big(\max_{m \leq n} S_m \leq x\Big) = \Phi\left(\frac{x + n\eta^2/2}{\sigma\sqrt{n}}\right) - e^{-x}\Phi\left(\frac{-x + n\eta^2/2}{\sigma\sqrt{n}}\right) \to 1 - e^{-x}$$
$$= P(M \leq x) \qquad \textit{as } n \to \infty.$$

The following remark, which provides the complementary probability for M for strictly positive values ($x > 0$), plays an important role in the proof of the theorem.

REMARK. Note that $P(M \geq x) = P(M \geq x | M > 0) P(M > 0) = \|G_+\| e^{-x}$, $x > 0$.

The next lemma provides an analytical and convenient expression for $P(T_1^- > n \cap S_{-n} \in dx)$. As can be seen from the proof of Lemma 3, this lemma is critical for carrying out the integrals in Theorem 2.1 in a fully analytical manner.

LEMMA 2. *Let the random walk* $\{S_n, n \geq 0\}$ *be as specified in* (2.3). *Then, for* $x \geq 0$,

$$P(T_1^- > n \cap S_n \in dx) = \eta^{-1} E\left\{(T_1^- > n - 1) \cap \varphi\left(\frac{x - S_{n-1} + \eta^2/2}{\eta}\right)\right\}, \qquad n \geq 1.$$

PROOF. In light of (3.1), we have that, for $x > 0$,

$$\begin{aligned} &P\{T_1^- > n \cap S_n \in (0, x]\} \\ &\quad = P\left\{\bigcap_{j=0}^{n-1} (S_j > 0) \cap S_n \in (0, x]\right\} \\ &\quad = E\left[I\left\{\bigcap_{j=0}^{n-1} (S_j > 0)\right\} P\big(X_n \in (-S_{n-1}, x - S_{n-1}] | \mathsf{F}_{n-1}\big)\right] \\ &\quad = E\left[I\left\{\bigcap_{j=0}^{n-1} (S_j > 0)\right\}\right. \qquad \text{(A.1)} \end{aligned}$$

$$\times P\biggl(Z_n \in \biggl(\frac{-S_{n-1}+\eta^2/2}{\eta}, \frac{x-S_{n-1}+\eta^2/2}{\eta}\biggr]\bigg|\mathsf{F}_{n-1}\biggr)\biggr]$$

$$= E\biggl[I(T_1^- > n-1)\cap\biggl\{\Phi\biggl(\frac{x-S_{n-1}+\eta^2/2}{\eta}\biggr) - \Phi\biggl(\frac{-S_{n-1}+\eta^2/2}{\eta}\biggr)\biggr\}\biggr], \qquad n \geq 1.$$

Thus, differentiating (A.1) with respect to x, the proof of Lemma 2 is now in order. □

The next lemma provides a manageable expression for the second term in Theorem 2.1.

LEMMA 3. *The following holds*:

$$\int_{0+}^{\infty} P(M^* \geq x)P(T_1^- > n \cap S_n \in dx) = \|G_+^*\|E\{e^{-S_n}I(T_1^- > n)\}, \qquad n \geq 1.$$

PROOF. Using Lemma 2, and the remark following Lemma 1, we note that

$$\begin{aligned}
&\int_{0+}^{\infty} P(M^* \geq x)P(T_1^- > n \cap S_n \in dx)\\
&\quad = \eta^{-1}\|G_+^*\|E\biggl\{I(T_1^- > n-1)\int_{0+}^{\infty} e^{-x}\varphi\biggl(\frac{x-S_{n-1}+\eta^2/2}{\eta}\biggr)\,dx\biggr\}\\
&\quad = \|G_+^*\|E\{I(T_1^- > n-1)e^{-S_n}I(\eta Z_n > -S_{n-1}+\eta^2/2)\}\\
&\quad = \|G_+^*\|E\{e^{-S_n}I(T_1^- > n)\}, \qquad n \geq 1.
\end{aligned}$$

□

Remarks regarding computational aspects of expressions in Theorem 3.1. Here, we first address computational issues of the two sequences $\{q_n : n \geq 1\}$ and $\{\tilde{q}_n : n \geq 1\}$ that appear in Theorem 3.1. Set $b_n = P(S_n > 0)$ and $\tilde{b}_n = E\{e^{-S_n}I(S_n > 0)\}$, for $n \geq 1$. From Feller (1971), Volume II, page 416, and Chover, Ney and Wainger (1973), it is well known that the generating function of the sequences $\{q_n : n \geq 1\}$ and $\{\tilde{q}_n : n \geq 1\}$, respectively, satisfy the following relationships:

$$\text{(A.2)} \qquad \sum_{n=1}^{\infty} s^n q_n = \exp\biggl\{\sum_{n=1}^{\infty}\frac{s^n b_n}{n}\biggr\} \quad \text{and} \quad \sum_{n=1}^{\infty} s^n \tilde{q}_n = \exp\biggl\{\sum_{n=1}^{\infty}\frac{s^n \tilde{b}_n}{n}\biggr\}.$$

Note that the second equation in (A.2) appears in Chover, Ney and Wainger (1973) as a type of a Laplace transform. In addition, both the equations in (A.2) may be obtained iteratively as simple consequences of the Weiner–Hopf factorization. In particular, the Leibnitz rule yields the following iterative relations, and thus enables one to compute $\{q_n : n \geq 1\}$ and $\{\tilde{q}_n : n \geq 1\}$:

$$nq_n = \sum_{j=0}^{n-1} b_{n-j} q_j \quad \text{and} \quad n\tilde{q}_n = \sum_{j=0}^{n-1} \tilde{b}_{n-j} \tilde{q}_j, \tag{A.3}$$
$$n = 1, 2, \ldots, \text{ and } \tilde{q}_0 = q_0 = 1.$$

Note that, in the Gaussian case, $b_n = \bar{\Phi}(\eta\sqrt{n}/2)$ and $\tilde{b}_n = e^{n\eta^2}\bar{\Phi}(3\eta\sqrt{n}/2)$, $n \geq 1$.

Next, we demonstrate that the probabilities in Theorem 3.1 sum to one, and then provide an expression for the variance of the limiting distribution.

From Hinkley (1970), and the remark after Lemma 1 above, it follows that

$$P(\xi_\infty > 0) = P(M^* > M, M^* > 0) = \int_{0+}^{\infty} P(M < x) P(M^* \in dx)$$
$$= \int_{0+}^{\infty} (1 - \|G_+\| e^{-x}) \|G_+\| e^{-x}\, dx = 1 - (1 - \|G_+\|)^2/2.$$

Since $P(\xi_\infty = 0) = (1 - \|G_+\|)^2$, and ξ_∞ is symmetric, the claim that the probabilities for ξ_∞ sum to one follows immediately. The following expression for the variance may be derived in a somewhat tedious but straightforward manner:

$$\operatorname{Var}(\xi_\infty) = 2\{B''(1) + (B'(1))^2\}$$
$$- 2\exp\bigl(-B(1) + \tilde{B}(1)\bigr)\bigl(1 - \exp(-B(1))\bigr)\{\tilde{B''}(1) + (\tilde{B}'(1))^2\},$$

where $B(1) = \sum_{n=1}^{\infty} b_n/n$, $B'(1) = \sum_{n=1}^{\infty} b_n$, $B''(1) = \sum_{n=1}^{\infty} nb_n$ and $\tilde{B}(1)$, $\tilde{B}'(1)$ and $\tilde{B''}(1)$ are defined upon $\tilde{b}_n$, $n \geq 1$, in a similar manner.

APPENDIX B

Change-point detection for Nacetinsky water discharges. We first formulate the following hypotheses that test for the presence of an unknown change-point in the mean vector of the data series:

$$\begin{aligned} &H_0 : \mu^{(1)} = \cdots = \mu^{(n)} = \mu_1 \quad \text{vs.} \\ &H_a : \mu^{(1)} = \cdots = \mu^{(\tau)} = \mu_1 \neq \mu^{(\tau+1)} = \cdots = \mu^{(n)} = \mu_2, \end{aligned} \tag{B.1}$$

where $\tau \in \{1, \ldots, n-1\}$ is the unknown change-point. Asymptotic theory of the generalized likelihood ratio statistic for testing the above hypothesis has been well addressed in the literature and the limiting result may be found in Csörgő and Horváth (1997). It may be shown that the twice log-likelihood ratio statistic for testing the above hypothesis is

$$U_n = \max_{1 \leq t \leq n-1} n \log(|\hat{\boldsymbol{\Sigma}}_n| / |\hat{\boldsymbol{\Sigma}}_t|), \tag{B.2}$$

where $\hat{\boldsymbol{\Sigma}}_t = n^{-1}\{\sum_{i=1}^{t}(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_{1,t})(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_{1,t})^T + \sum_{i=t+1}^{n}(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_{2,t})(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_{2,t})^T\}$, $\hat{\boldsymbol{\mu}}_{1,t} = t^{-1}\sum_{i=1}^{t}\mathbf{Y}_i$ and $\hat{\boldsymbol{\mu}}_{2,t} = (n-t)^{-1}\sum_{i=t+1}^{n}\mathbf{Y}_i, t = 1,\dots,n$. The asymptotic distribution of the above statistic is based upon $W_n = (2\log\log nU_n)^{1/2} - (2\log\log n + \frac{p}{2}\log\log\log n - \log\Gamma(p/2))$, where p denotes the number of parameters that change under the alternative hypothesis, and in this case we have $p = d = 3$. The limiting distribution of W_n is given by the following double exponential form:

$$\lim_{n\to\infty} P[W_n \le t] = \exp(-2e^{-t}). \tag{B.3}$$

The p-value is obtained based on a two-sided critical region of the above limiting distribution. When a test is significant, the maximum likelihood estimator of the unknown change-point τ is obtained as the argument at which U_n attains its maximum. In principle, we may apply the above procedure for the data of each month individually with $p = 1$, and also for data on each pair of months with $p = 2$. The results of the tests for all cases are presented in Table 3. Clearly, all univariate tests are not significant. Among the bivariate tests, the pair July–August is not significant, whereas the other two pairs yield significance. The multivariate test for all three months is also significant. The significance based upon the biviariate and multivariate tests takes into account the covariance structure in the data and hence should be believed more so than the univariate tests where no significance is found. The change-point mle is obtained as $\hat{\tau} = 14$.

At this point, we need to investigate the validity of the main assumptions, namely, constancy of the covariance matrix, Gaussianity and independence over time. The investigation regarding the covariance matrix requires that we compute the deviation vector $D_i, i = 1,\dots,40$, from the estimated mean for each observation, taking into account the differences in the means before and after the estimated change-point. It is of interest then to know whether the covariance structure of the deviations remained constant throughout the sampling period. The generalized log-likelihood ratio statistic for the constancy of the covariance matrix over time

TABLE 3
The statistic W for change in mean for various months and their p-values

Months	W	p**-value**	$\hat{\tau}$
Feb	2.74	0.1206	15
Jul	1.86	0.2674	14
Aug	2.29	0.1825	14
Feb–Jul	3.59	0.0539	14
Feb–Aug	3.76	0.0455	14
Jul–Aug	1.90	0.2593	14
Feb–Jul–Aug	3.78	0.0448	14

against the alternative that the covariance matrix has changed at an unknown time is given by

$$U_n^* = \max_{1\le t\le n-1} \log\{|\hat{\mathbf{\Sigma}}_{1:n}|^n / (|\hat{\mathbf{\Sigma}}_{1:t}|^t |\hat{\mathbf{\Sigma}}_{t+1:n}|^{(n-t)})\}, \tag{B.4}$$

where $|\hat{\mathbf{\Sigma}}_{1:t}|$ and $|\hat{\mathbf{\Sigma}}_{t+1:n}|$ are the usual estimators of the covariance matrix based on the first t and last $n-t$ deviations, respectively. The limiting distribution of U_n^* is obtained through the distribution of W_n^*, where W_n^* is defined upon U_n^* in an analogous manner. It follows that p, the number of parameters that change in this case, is given by $p = d(d+1)/2$. The p-values for the univariate, bivariate and multivariate tests are reported in Table 4. Clearly, all tests are insignificant except the multivariate test. However, the significance is not particularly relevant since the change-point mle of 3 obtained in this case implies no change in the covariance structure, for all practical purposes. Thus, there is no evidence in the data against the assumption of stationarity of the covariance matrix. Utilizing the estimated change-point ($\hat{\tau} = 14$), estimates for the mean vector before and after the change-point as well as the pooled estimator of the common covariance matrix are then obtained as $\hat{\mu}_{1\hat{\tau}} = (6.738, 7.137, 6.725)$, $\hat{\mu}_{2\hat{\tau}} = (7.383, 7.483, 7.166)$ and

$$\hat{\Sigma}_{\hat{\tau}} = \begin{bmatrix} 0.365 & -0.032 & -0.029 \\ -0.032 & 0.161 & 0.104 \\ -0.029 & 0.104 & 0.211 \end{bmatrix}.$$

It remains to be seen whether the assumptions of Gaussianity and independence over time are valid. We can verify this by utilizing the deviation vectors D_i, $i = 1, \dots, 40$, and the covariance matrix $\hat{\Sigma}_{\hat{\tau}}$ found above. Specifically, if D_i is multivariate normal, then it is well known that $d_i^2 = \|D_i\|^2_{\hat{\Sigma}_{\hat{\tau}}^{-1}}$ is approximately chi-square with 3 degrees of freedom $i = 1, \dots, 40$. The same can be applied for the bivariate case also with the degrees of freedom being 2 in this case. Thus, one only needs to verify whether $d_i^2, i = 1, \dots, 40$ form a sample from the corresponding chi-square distribution. Upon applying the Anderson–Darling statistic, we found

TABLE 4
The statistic W for change in variance for various months and their p-values

Months	W	p**-value**	$\hat{\tau}$
Feb	3.18	0.0796	3
Jul	1.91	0.2556	5
Aug	1.39	0.3929	2
Feb–Jul	3.02	0.0927	3
Feb–Aug	2.28	0.1842	2
Jul–Aug	2.32	0.1788	2
Feb–Jul–Aug	4.26	0.0278	3

the p-value for the three months case to be 0.185. The corresponding p-values for Feb–Jul, Feb–Aug and Jul–Aug pairs were 0.244, 0.250 and 0.10, respectively. In the univariate case, we applied the Anderson–Darling test for the deviations for each individual month and found the p-values to be 0.927, 0.530 and 0.177, respectively. Thus, the assumption of Gaussianity seems quite appropriate at each of the univariate, bivariate and multivariate levels.

As for independence over time, we first tested each of the three deviation series for significance of both autocorrelations and partial autocorrelations up to the first twenty lags. The ACF and PACF plots for each individual series showed no evidence of significant correlations. We then computed the cross-correlations for each pair and found that these were also not significant and, thus, there was no indication that the assumption of independence over time was in violation. Overall, the change-point model with estimated parameters may be seen to fit the data quite well.

Acknowledgments. The authors thank the Editor Michael Stein, the Associate Editor and two anonymous referees for their in-depth comments and suggestions that led to a substantial improvement in both content and presentation of the paper. We are especially thankful to Professor Daniela Jarušková for providing us the data on Nacetinsky creek.

REFERENCES

ANDREWS, D. W. K. and PLOBERGER, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* **62** 1383–1414. MR1303238

ASMUSSEN, S. (1987). *Applied Probability and Queues*. Wiley, New York. MR0889893

BASSEVILLE, M. and NIKIFOROV, I. V. (1993). *Detection of Abrupt Changes: Theory and Application*. Prentice Hall, Englewood Cliffs, NJ. MR1210954

BOROVKOV, A. A. (1999). Asymptotically optimal solutions in the change-point problem. *Theory Probab. Appl.* **43** 539–561.

BRAUN, J. V. and MÜLLER, H.-G. (1998). Statistical methods for DNA sequence segmentation. *Statist. Sci*. **13** 142–162.

BRODSKY, B. E. and DARKHOVSKY, B. S. (1993). *Nonparametric Methods in Change-point Problems*. Springer, New York. MR1228205

BRODSKY, B. E. and DARKHOVSKY, B. S. (2000). *Non-parametric Statistical Diagnosis*: *Problems and Methods*. *Mathematics and Its Applications* **509**. Kluwer Academic, Dordrecht.

CHEN, J. and GUPTA, A. K. (2000). *Parametric Statistical Change Point Analysis*. Birkhäuser, New York. MR1761850

CHOVER, J., NEY, P. and WAINGER, S. (1973). Functions on probability measures. *J. Anal. Math.* **26** 255–302. MR0348393

COBB, G. W. (1978). The problem of the Nile: Conditional solution to a change-point problem. *Biometrika* **65** 243–251. MR0513930

CSÖRGŐ, M. and HORVÁTH, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley, New York.

DEGAETANO, A. T. (2006). Attributes of several methods for detecting discontinuities in temperature series: Prospects for a hybrid homogenization procedure. *J. Climate* **9** 1646–1660.

FEALY, R. and SWEENEY, J. (2005). Detection of a possible change point in atmospheric variability in the North Atlantic and its effect on Scandinavian glacier mass balance. *Int. J. Climatol.* **25** 1819–1833.

FEARNHEAD, P. (2006). Exact and efficient Bayesian inference for multiple change-point problems. *Stat. Comput.* **16** 203–213. MR2227396

FEARNHEAD, P. and LIU, Z. (2007). On-line inference for multiple change points problems. *J. Roy. Statist. Soc. Ser. B* **69** 589–605. MR2370070

FELLER, W. R. (1971). *An Introduction to Probability Theory and Its Applications, Vol. II*. Wiley, New York. MR0270403

FOTOPOULOS, S. B. and JANDHYALA, V. K. (2001). Maximum likelihood estimation of a change-point for exponentially distributed random variables. *Statist. Probab. Lett.* **51** 423–429. MR1820801

FOTOPOULOS, S. B. (2009). The geometric convergence rate of the classical change-point estimate. *Statist. Probab. Lett.* **79** 131–137. MR2483529

GIRÓN, F. J., MORENO, E. and CASELLA, G. (2007). Objective Bayesian analysis of multiple changepoints for linear models (with discussion). In *Bayesian Statistics 8* (J. M. Bernardo, M. J. Bayarri and J. O. Berger, eds) 227–252. Oxford Univ. Press, Oxford. MR2433195

GOMBAY, E. and HORVÁTH, L. (1997). An application of the likelihood method to change-point detection. *Environmetrics* **8** 459–467.

HANSEN, B. E. (2000). Testing for structural change in conditional models. *J. Econometrics* **97** 93–115. MR1788819

HINKLEY, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **57** 1–17. MR0273727

HINKLEY, D. V. (1971). Inference about the change-point from cumulative sum tests. *Biometrika* **58** 509–523. MR0312623

HINKLEY, D. V. (1972). Time ordered classification. *Biometrika* **59** 509–523. MR0368317

HU, I. and RUKHIN, A. L. (1995). A lower bound for error probability in change-point estimation. *Statist. Sinica* **5** 319–331. MR1329301

JANDHYALA, V. K. and FOTOPOULOS, S. B. (1999). Capturing the distributional behavior of the maximum likelihood estimator of a change-point. *Biometrika* **86** 129–140. MR1688077

JANDHYALA, V. K. and FOTOPOULOS, S. B. (2001). Rate of convergence of the maximum likelihood estimate of a change-point. *Sankhyā Ser. A* **63** 277–285. MR1897454

JARUŠKOVÁ, D. (1996). Change-point measurement in meteorological measurement. *Mon. Weather Rev.* **124** 1535–1543.

KAPLAN, A. Y. and SHISHKIN, S. L. (2000). Application of the change-point analysis to the investigation of the brain's electrical activity. In *Non-Parametric Statistical Diagnosis: Problems and Methods* (B. E. Brodsky and B. S. Darkhovsky, eds.) 333–388. Kluwer, Dordrecht. MR1862475

LAI, T. L. (1995). Sequential change-point detection in quality control and dynamical systems. *J. Roy. Statist. Soc. Ser. B* **57** 613–658. MR1354072

LEBARBIER, L. (2005). Detecting multiple change-points in the mean of Gaussian process by model selection. *Sign. Proc.* **85** 717–736.

PAGE, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika* **42** 523–526. MR0072412

PERREAULT, L., BERNIER, J., BOBÉE, B. and PARENT, E. (2000a). Bayesian change-point analysis in hydrometeorological time series. Part 1. Normal model revisited. *J. Hydrol.* **235** 221–241.

PERREAULT, L., BERNIER, J., BOBÉE, B. and PARENT, E. (2000b). Bayesian change-point analysis in hydrometeorological time series. Part 2. Comparison of change-point models and forecasting. *J. Hydrol.* **235** 242–263.

RUGGIERI, E., HERBERT, T., LAWRENCE, K. T. and LAWRENCE, C. E. (2009). Change point method for detecting regime shifts in paleoclimatic time series: Application to $\delta^{18}O$ time series of the Plio-Pleistocene. *Paleoceanography* **24** PA1204, DOI:10.1029/2007PA001568.

SEIDOU, O. and OUARDA, T. B. M. J. (2007). Recursion-based multiple changepoint detection in multiple linear regression and application to river streamflows. *Water Resour. Res.* **43**, DOI:10.1029/2006WR005021.

SHIRYAEV, A. N., KABANOV, Y. M., KRAMKOV, D. O. and MELNIKOV, A. V. (1994). Towards the theory of pricing of options of both European and American types, II, continuous time. *Theory Probab. Appl.* **39** 61–102.

WU, Y. (2005). *Inference for Change-Point and Post-Change Means After a CUSUM Test*. *Lecture Notes in Math.* **180**. Springer, New York. MR2142337

WU, Q.-Z., CHENG, H.-Y. and JENG, B.-S. (2005). Motion detection via change-point detection for cumulative histograms of ratio images. *Pattern. Recog. Lett*. **26** 555–563.

ZOU, C., QIU, P. and HAWKINS, D. (2009). Nonparametric control chart for monitoring profiles using change point formulation and adaptive smoothing. *Statist. Sinica* **19** 1337–1357. MR2536159

S. B. FOTOPOULOS
E. KHAPALOVA
DEPARTMENT OF MANAGEMENT AND OPERATIONS
WASHINGTON STATE UNIVERSITY
PULLMAN, WASHINGTON 99164-4736
USA
E-MAIL: fotopo@wsu.edu
elena_k@wsu.edu

V. K. JANDHYALA
DEPARTMENT OF STATISTICS
WASHINGTON STATE UNIVERSITY
PULLMAN, WASHINGTON 99164-3113
USA
E-MAIL: jandhyala@wsu.edu