Testing Homogeneity in Gamma Mixture Models

XIN LIU

Rockefeller University

CRISTIAN PASARICA

Columbia University

YONGZHAO SHAO

Columbia University

ABSTRACT. This paper characterizes the asymptotic behaviour of the likelihood ratio test statistic (LRTS) for testing homogeneity (i.e. no mixture) against gamma mixture alternatives. Under the null hypothesis, the LRTS is shown to be asymptotically equivalent to the square of Davies's Gaussian process test statistic and diverges at a log log *n* rate to infinity in probability. Based on the asymptotic analysis, we propose and demonstrate a computationally efficient method to simulate the null distributions of the LRTS for small to moderate sample sizes.

Key words: empirical process, gamma mixture model, Gaussian process test, likelihood ratio test, loss of identifiability

1. Introduction

Gamma distributions are frequently used in reliability theory to model hazard rates. When there are multiple hazard factors, it is natural to consider gamma mixture models (Seal, 1969; Nelson, 1982). For instance, Slud (1997) proposed a two-component exponential mixture model $(1-p)\lambda_0 e^{-\lambda_0 x} + p\lambda e^{-\lambda x}$ $(\lambda_0 \leq \lambda)$ to test imperfect debugging in software reliability. A question of interest is to determine whether the observations have the single gamma distribution $\lambda_0 e^{-\lambda_0 x}$ or a mixture distribution, i.e. to test the hypothesis:

$$H_0 : \lambda_0 e^{-\lambda_0 x} \quad \text{against} \quad H_1 : (1-p)\lambda_0 e^{-\lambda_0 x} + p\lambda e^{-\lambda x}, \qquad \lambda > \lambda_0, \quad 0 < p \leq 1, \tag{1}$$

where $\lambda_0 > 0$ is known, λ and p are unknown parameters. The standard practice for such a test is to employ the likelihood ratio test (LRT). In general *regular* cases, the likelihood ratio test statistic (LRTS) has an asymptotic chi-square type null distribution. However, because of loss of identifiability of the null distribution, the homogeneity test in mixture models is one of the *nonregular* cases and the classical asymptotics theory does not apply. By loss of identifiability of the null distribution, we mean that parameters representing the null distribution are not unique. For example, the null distribution in (1) corresponds to the parameter set $\{\lambda = \lambda_0, p \in [0, 1]\} \cup \{p = 0, \lambda > \lambda_0\}$. Under loss of identifiability of the null hypothesis, the asymptotic behaviour of the LRTS is difficult to characterize. Moreover, when the range of some parameters is unbounded, e.g. (λ_0, ∞) for λ in (1), the LRTS may diverge to infinity instead of converging to any finite limiting distribution. In fact, Hartigan (1985) discovered the divergence of the LRTS for testing homogeneity in normal mean mixture models with an unbounded mean parameter. It is not hard to see that Hartigan's observation may also be true for gamma mixture models. This divergence of the LRTS causes major difficulties both in analysing the LRT and in obtaining stable simulation results for the null distributions of the LRTS (Lindsay, 1995).

Because of the undetermined asymptotic behaviour of the LRTS for testing homogeneity in mixture models, the Gaussian process test (GPT) introduced by Davies (1977) is sometimes recommended as an alternative method to the LRT (Cheng & Traylor, 1995). The GPT is generally applicable for testing hypotheses where nuisance parameters appear only under the alternative hypothesis. Suppose $\{f(\cdot, \phi, \xi); (\phi, \xi) \in \Theta\}$ is a family of densities which are assumed known except for parameters (ϕ, ξ), and $f(\cdot, \phi, \xi_1) = f(\cdot, \phi, \xi_2)$ for any ξ_1 and ξ_2 iff $\phi = \phi_0$. Then in the hypothesis testing problem

$$H_0\colon f(x, \phi_0, \xi) \quad \text{against} \quad H_1\colon f(x, \phi, \xi), \qquad (\phi, \xi) \in \Theta,$$

the Gaussian process test statistic (GPTS) can be written as $M_n = \sup_{(\phi,\xi)\in\Theta} S_n(\phi, \xi)$, where, $S_n(\phi_0, \xi) = 0$ and for $\phi \neq \phi_0$,

$$S_n(\phi, \xi) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{f(X_i, \phi, \xi)/f(X_i, \phi_0) - 1}{\sqrt{\mathrm{E}[f(X_i, \phi, \xi)/f(X_i, \phi_0) - 1]^2}}. \tag{2}$$

In equation (2), the expectation 'E' is taken under the true distribution. As discussed in Cheng & Traylor (1995), the GPT is easier to conduct than the LRT. Titterington *et al.* (1985) also considered the GPT for (1), but the null distribution of the GPTS was not discussed. Characterizing the asymptotic behaviour of the GPTS has been an open problem.

In this paper, we characterize the asymptotic behaviour of the LRTS and GPTS for testing homogeneity against the two-component gamma mixture alternative, i.e.,

$$H_0\colon g_{\lambda_0}(x) \quad \text{against} \quad H_1\colon p g_{\lambda}(x) + (1-p) g_{\lambda_0}(x), \qquad \lambda > \lambda_0, \quad 0 < p \leq 1, \tag{3}$$

where $g_\lambda(x) = (\Gamma(\kappa))^{-1} \lambda^\kappa x^{\kappa-1} e^{-x\lambda}$ is the gamma density with a known shape parameter κ. λ_0 is known, λ and p are unknown. In section 2, we present our main results of asymptotic distributions of the LRTS and GPTS. First, we derive the asymptotic behaviour of the GPTS under H_0:

$$\lim_{n\to\infty} \mathrm{P}\{M_n^2 - \log\log n + \log(16\pi^2/\kappa) \leq x\} = \exp(-e^{-x/2}). \tag{4}$$

Then we show that under the null hypothesis, the LRTS $2\Lambda_n$ and M_n^2 are asymptotically equivalent, i.e. $2\Lambda_n = M_n^2 + o_P(1)$. Consequently, (4) yields the asymptotic distribution of the LRTS by replacing M_n^2 with $2\Lambda_n$. Thus, the LRTS diverges to infinity at a $\log\log n$ rate, which is also the rate conjectured by Hartigan (1985) for the normal mean mixture models. Proofs of results in section 2 can be found in the appendix.

In section 3, we discuss an important issue in practice: how to simulate the null distribution of the LRTS for a given sample size. When sample sizes are small, the simulations can be done easily using the Monte Carlo method. However, as sample size increases, the computation involved in simulation becomes very intensive. One might expect that the results obtained in section 2 can be used directly to simulate the null distributions of the LRTS for reasonably large sample sizes. Unfortunately, according to our simulations, $2\Lambda_n - \log\log n + \log(16\pi^2/\kappa)$ is still far from converging to its asymptotic extreme distribution for sample size as large as 5000. On the other hand, some results in section 2 suggest that $2\Lambda_n$ can be approximated by the square of the supremum of some Gaussian process indexed by a restricted parameter set. This approximation works quite well according to our simulation study in section 3. Since the supremum of a Gaussian process with known covariance matrix is relatively easy to simulate, this method greatly reduces the amount of computation required for simulating the null distribution of the LRTS.

2. Asymptotic distributions of the GPTS and LRTS

This section considers asymptotic distributions of the GPTS and LRTS for the homogeneity test in the gamma mixture model (3). First, we prove our main result regarding the asymptotic distribution of M_n in theorem 1. To do so, we establish the asymptotic equivalence of the GPT with the supremum of a stochastic process. We show that this stochastic process is stationary, then derive the asymptotic distribution of the GPTS using results from Leadbetter *et al.* (1983). Next, in theorem 2, we prove the asymptotic equivalence of the LRTS with the GPTS and obtain the asymptotic distribution of LRTS by theorem 1. In the rest of this section, we give a brief outline of proofs of theorems 1 and 2. Technical details of proofs of the lemmas and theorems can be found in the appendix.

We list some frequently used notations for ease of reference. Without loss of generality, we assume $\lambda_0 = 1$ and write $g(x)$ for $g_1(x)$. Denote the mixture density by $g(x, p, \lambda) \triangleq pg_\lambda(x) + (1-p)g(x)$ and by $G(x)$ the CDF of $g(x)$. Let $X_1, \ldots, X_n$ be i.i.d. random samples from $G(x)$ and $X_{1,n}, \ldots, X_{n,n}$ be their order statistics. Then $\{U_i = G(X_i) : 1 \le i \le n\}$ are i.i.d. uniform random variables. We denote by $F_n(u)$ their empirical CDF. Define $x \vee y \equiv \max(x, y)$, $\log_{(2)} n \triangleq \log\log n$ and $\log_{(3)} n \triangleq \log\log\log n$ for large enough n. Let

$$Y(x,\lambda) \triangleq (2\lambda-1)^{\kappa/2}(e^{-x(\lambda-1)} - \lambda^{-\kappa}) \quad \text{and} \quad Z(x,\lambda) \triangleq (2\lambda-1)^{\kappa/2}e^{-x(\lambda-1)}.$$

2.1. *Asymptotics for the Gaussian process test*

According to Davies (1977), the GPTS for (3) is $M_n \triangleq \sup_{\lambda \ge 1} S_n(\lambda)$, where:

$$\begin{aligned} S_n(\lambda) &\triangleq \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{g_\lambda(X_i)/g(X_i)-1}{\sqrt{\mathrm{E}(g_\lambda(X_i)/g(X_i)-1)^2}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y(X_i,\lambda) \\ &= n^{1/2}\int_0^\infty Y(x,\lambda)d(F_n(G(x))-G(x)) \\ &= n^{1/2}\int_0^\infty Z(x,\lambda)d(F_n(G(x))-G(x)) \end{aligned}$$

In a suitable probability space, the uniform empirical process $\{\alpha_n(u) = n^{1/2}(F_n(u) - u), 0 \le u \le 1\}$ can be approximated by a Brownian bridge $\{B_n(u) = W_n(u) - W_n(1)u, 0 \le u \le 1\}$, where $\{W_n(u), 0 \le u \le 1\}$ is a standard Brownian motion (Csörgõ *et al.*, 1986). Then, $S_n(\lambda)$ can be expressed as

$$\begin{aligned} S_n(\lambda) &= \int_0^\infty Z(x,\lambda)d\Big(W_n(G(x)) - W_n(1)G(x) + [\alpha_n(G(x)) - B_n(G(x))]\Big) \\ &= H_n(\lambda) - W_n(1)(\sqrt{2\lambda-1}/\lambda)^\kappa + R_n(\lambda), \end{aligned} \qquad (5)$$

where

$$H_n(\lambda) \triangleq \int_0^\infty Z(x,\lambda)dW_n(G(x)), \qquad R_n(\lambda) \triangleq \int_0^\infty Z(x,\lambda)\,d[\alpha_n(G(x)) - B_n(G(x))].$$

Among the three parts in (5) making up $S_n(\lambda)$, H_n is the most important part in determining the asymptotic distribution of the GPTS. After the transformation, $\lambda = \exp(s) + 1/2$, $H_n(\exp(s) + 1/2)$ becomes a stationary Gaussian process. Then results on stationary Gaussian process obtained by Leadbetter *et al.* (1983) allow us to derive the limiting distribution of $\sup_{0 \le s \le T} H_n(\exp(s) + 1/2)$. That is the following lemma.

Lemma 1

$\{H_n(\exp(s)+1/2), 0 \le s \le \infty\}$ is a stationary Gaussian process, and

$$\lim_{T\to\infty} P\left\{A_T\left[\sup_{0\le s\le T} H_n(\exp(s)+1/2) - A_T\right] + \log(4\pi/\sqrt{\kappa}) \le x\right\} = \exp(-e^{-x}), \tag{6}$$

where $A_T = (\log T)^{1/2}$.

The asymptotic equivalence of the supremum of H_n and M_n can be established if the remaining two parts in (5) have a smaller order than the supremum of H_n. We prove this by considering them in the three intervals of λ^κ: $[1, \log n]$, $[\log n, n(\log n)^{-4}]$, $[n(\log n)^{-4}, \infty)$. That is the following lemma.

Lemma 2

In a suitable probability space,

$$\sup_{\lambda^\kappa \in [1,\log n]\cup[n(\log n)^{-4},\infty)} S_n(\lambda) \vee 0 = O_P((\log_{(3)} n)^{1/2}),$$

$$\sup_{\lambda^\kappa \in [\log n, n(\log n)^{-4}]} |S_n(\lambda) - H_n(\lambda)| = O_P((\log n)^{-1}).$$

By lemma 1, it is easy to show that $\sup_{\lambda^\kappa\in[\log n, n(\log n)^{-4}]} H_n(\lambda)$ diverges to infinity at the $(\log_{(2)} n)^{1/2}$ rate. Then, lemma 2 yields that with probability going to one

$$M_n = \sup_{\lambda^\kappa\in[\log n, n(\log n)^{-4}]} S_n(\lambda) = \sup_{\lambda^\kappa\in[\log n, n(\log n)^{-4}]} H_n(\lambda) + O_P((\log n)^{-1}).$$

The asymptotic distribution of M_n can be directly derived using lemma 1. We summarize our results so far in the following theorem:

Theorem 1

Under the null hypothesis of homogeneity, the GPTS $M_n = \sup_{\lambda\ge 1} S_n(\lambda) \vee 0$ for (3) satisfies

$$\lim_{n\to\infty} \mathrm{P}\left\{\sqrt{\log_{(2)} n}\left(M_n - \sqrt{\log_{(2)} n}\right) + \log(4\pi/\sqrt{\kappa}) \le x\right\} = \exp(-e^{-x}).$$

Moreover, the asymptotic distribution of M_n can be also expressed as

$$\lim_{n\to\infty} P\left\{M_n^2 - \log_{(2)} n + \log(16\pi^2/\kappa) \le x\right\} = \exp(-e^{-x/2}).$$

2.2. Asymptotics for the likelihood ratio test

In the following, we consider the LRT for (3) and prove that the LRTS has the same asymptotic distribution as that of M_n^2. The LRTS is $2\Lambda_n = 2\sup_{0\le p\le 1, 1\le\lambda<\infty} L_n(p,\lambda)$, where

$$L_n(p,\lambda) = \sum_{i=1}^n \log[1 + p(l(X_i,\lambda) - 1)] = \sum_{i=1}^n \log[1 + D(p,\lambda)Y(X_i,\lambda)],$$

and the likelihood ratio function is $l(x,\lambda) \triangleq g_\lambda(x)/g(x) = \lambda^\kappa e^{-x(\lambda-1)}$, $D(p,\lambda) \triangleq p(\lambda/\sqrt{2\lambda-1})^\kappa$.

We first show that, to derive the limiting distribution of the LRTS, it suffices to maximize $L_n(p,\lambda)$ for $\lambda \in [\lambda^*, \kappa/X_{1,n}]$ for some constants $\lambda^* > 1$. Given $\lambda^* > 1$, by Liu & Shao (2002, theorem 3.1), when $1 \le \lambda \le \lambda^*$, the maximum of the log-likelihood function converges to the supremum of the square of some centered Gaussian process. Consequently, $\sup_{\lambda\in[1,\lambda^*], p\in[0,1]} L_n(p,\lambda) \vee 0 = O_P(1)$. The partial derivative of the log-likelihood function with respect to λ is

$$\frac{\partial L_n}{\partial \lambda}(p,\lambda) = \sum_{i=1}^{n} \frac{(\kappa - X_i\lambda)\lambda^{\kappa-1}e^{-X_i(\lambda-1)}}{1+p(\lambda^{\kappa}e^{-X_i(\lambda-1)}-1)}.$$

When $\lambda > \kappa/X_{1,n}$, $(\partial L_n/\partial\lambda)(p,\lambda) < 0$ and $L_n(p,\lambda)$ is a decreasing function in λ. Since the LRTS diverges to infinity in probability, it suffices to maximize $L_n(p,\lambda)$ for $\lambda \in [\lambda^*, \kappa/X_{1,n}]$.

Next, we approximate the log-likelihood function by the Taylor expansion and prove the asymptotic equivalence between the LRTS and GPTS. The following lemmas are useful. Define $P_nY^2(\lambda) \triangleq n^{-1}\sum_{i=1}^{n} Y^2(X_i,\lambda)$, $I_n \triangleq [\log n, n(\log n)^{-4}]$ and $I_n(\lambda^*) \triangleq [(\lambda^*)^{\kappa}, \log n] \cup [n(\log n)^{-4}, (\kappa/X_{1,n})^{\kappa}]$ for $\lambda^* \geq 1$.

Lemma 3

There exist constants $\lambda^* \geq 1$ *such that,* $\sup_{\lambda^{\kappa}\in I_n\cup I_n(\lambda^*), L_n(p,\lambda)>0} p\lambda^{\kappa} = O_P(1)$.

Lemma 4

There exist constants $\lambda^* \geq 1$ *such that* $\sup\limits_{\lambda^{\kappa}\in I_n\cup I_n(\lambda^*)} 1/P_nY^2(\lambda) = O_P(1)$. *Moreover, when* $\lambda^{\kappa} \in I_n$, $P_nY^2(\lambda) = 1 + O_P((\log n)^{-1/2})$.

For $c_0 > 1$, we have the inequality, $\log(1+x) \leq x - x^2/(4c_0)$, for $-1 < x \leq c_0$. Note that $p(l(x,\lambda)-1) = D(p,\lambda)Y(x,\lambda) \leq p\lambda^{\kappa}$. In the inequality above, let $c_0 = p\lambda^{\kappa}$

$$L_n(p,\lambda) \leq \sqrt{n}D(p,\lambda)S_n(\lambda) - nD^2(p,\lambda)P_nY^2(\lambda)/(4p\lambda^{\kappa}). \tag{7}$$

Choose λ^* such that lemmas 3 and 4 hold. Equation (7) yields

$$\begin{aligned}\sup_{\lambda^{\kappa}\in I_n(\lambda^*), p\in[0,1]} L_n(p,\lambda)\vee 0 &\leq \sup_{\lambda^{\kappa}\in I_n(\lambda^*), L_n(p,\lambda)>0} p\lambda^{k}(S_n(\lambda)\vee 0)^2/P_nY^2(\lambda)\\ &\leq \sup_{\lambda^{\kappa}\in I_n(\lambda^*), L_n(p,\lambda)>0} p\lambda^{k}\cdot \sup_{\lambda^{\kappa}\in I_n(\lambda^*)}(S_n(\lambda)\vee 0)^2\cdot \sup_{\lambda^{\kappa}\in I_n(\lambda^*)} 1/P_nY^2(\lambda)\\ &= O_P((\log_{(3)} n)^{1/2}).\end{aligned}$$

Because the LRTS diverges to infinity at the $\log_{(2)} n$ rate, it suffices to maximize $L_n(p,\lambda)$ for $\lambda^{\kappa} \in I_n$. For the right-hand side of (7) to be positive, we have $D(p,\lambda) \leq 4n^{-1/2} p\lambda^{k}S_n(\lambda)/P_nY^2(\lambda)$. Then lemma 3 and 4 yield

$$\begin{aligned}\sup_{\lambda^{\kappa}\in I_n, L_n(p,\lambda)>0} D(p,\lambda) &= O(n^{-1/2}) \sup_{\lambda^{\kappa}\in I_n, L_n(p,\lambda)>0} p\lambda^{\kappa}\cdot \sup_{\lambda^{\kappa}\in I_n} S_n(\lambda)\cdot \sup_{\lambda^{\kappa}\in I_n} 1/P_nY^2(\lambda)\\ &= O_P(n^{-1/2}(\log_{(2)} n)^{1/2}).\end{aligned}$$

Note that $|Y(x,\lambda)| = O(1)\lambda^{\kappa/2} = O(n^{1/2}(\log n)^{-2})$ for $\lambda^{\kappa} \in I_n$. Therefore,

$$\sup_{\lambda^{\kappa}\in I_n, L_n(p,\lambda)>0} D(p,\lambda)Y(x,\lambda) = O_P((\log n)^{-1}),$$

and the positive part of the log-likelihood function has the Taylor expansion for $\lambda^{\kappa} \in I_n$

$$2L_n(p,\lambda)\vee 0 = \left[2\sqrt{n}D(p,\lambda)S_n(\lambda) - nD^2(p,\lambda)P_nY^2(\lambda)\right]\vee 0 + o_P(1). \tag{8}$$

Maximizing the right-hand side of (8) by $D(p,\lambda)$ yields

$$2\Lambda_n = \sup_{\lambda^{\kappa}\in I_n}(S_n(\lambda)\vee 0)^2/P_nY^2(\lambda) + o_P(1) = \sup_{\lambda^{\kappa}\in I_n}(S_n(\lambda)\vee 0)^2 + o_P(1).$$

By lemmas 1 and 2, the supremum of $S_n(\lambda)$ is positive and attained in the interval I_n. Therefore, we prove $2\Lambda_n = M_n^2 + o_P(1)$. By theorem 1, we have the following theorem for the asymptotic distribution of the LRTS.

Theorem 2

The LRTS and the GPTS for (3) *satisfy* $2\Lambda_n = M_n^2 + o_P(1)$ *and*

$$\lim_{n\to\infty} \mathrm{P}\left\{2\Lambda_n - \log_{(2)} n + \log(16\pi^2/\kappa) \le x\right\} = \exp(-e^{-x/2}).$$

3. Simulating the null distribution of the LRTS

From a practitioner's point of view, one can decide whether to reject or accept the null hypothesis in the hypothesis testing problem (3) provided that the null distribution of the LRTS is available for a given sample size n. When n is small, the null distribution of $2\Lambda_n$ can be easily simulated using the Monte Carlo method. However, there are difficulties in this approach as sample size increases because of the extensive computation required for large sample sizes. As suggested by theorem 2, another way of simulation is to approximate the distribution of $2\Lambda_n - \log\log n + \log(16\pi^2/\kappa)$ by the extreme distribution for large n. Unfortunately, our simulation results indicate that this direct approximation is quite poor even for a sample size as large as 5000.

In this section, we describe an approach to simulate the null distribution of the LRTS $2\Lambda_n$ for large values of n. The basic idea of this approach is to approximate the distribution of $2\Lambda_n$ by that of the square of the supremum of some Gaussian process. By Liu & Shao (2002), the restricted LRTS, $\sup_{0\le p\le 1, 1\le\lambda\le T} 2L_n(p,\lambda)$, converges in distribution to $\sup_{1\le\lambda\le T}(\tilde{W}_\lambda \vee 0)^2$, where $\{\tilde{W}_\lambda, 1\le\lambda\le T\}$ is a centred Gaussian process with the covariance kernel

$$\mathrm{E}(\tilde{W}_{\lambda_1}\tilde{W}_{\lambda_2}) = \left(\frac{\sqrt{2\lambda_1 - 1}\sqrt{2\lambda_2 - 1}}{\lambda_1 + \lambda_2 - 1}\right)^{\kappa}, \quad 1 \le \lambda_1, \ \ \lambda_2 \le T.$$

The same Gaussian process is also the limit of the restricted GPTS $\{S_n(\lambda), 1\le\lambda\le T\}$ as n tends to infinity. An important property of the Gaussian process $\tilde{W}_\lambda$ as shown in lemma 1 is that after the reparameterization $W_t \triangleq \tilde{W}_{\exp(t)+1/2}$, W_t becomes a *stationary* Gaussian process with covariance kernel:

$$\mathrm{E}(W_{t_1}W_{t_2}) = \left(\frac{\exp((t_1 - t_2)/2) + \exp((t_2 - t_1)/2)}{2}\right)^{-\kappa}.$$

We approximate the supremum of the stationary Gaussian process $\{W_t, 0\le t\le T\}$ using the maximum of a discrete Gaussian process $\{W_{Ti/m}, i = 0, 1, \dots, m\}$. The discrete Gaussian process can be simulated using a Choleski decomposition of the covariance matrix and multiplying the result with an $(m+1)$-dimensional standard normal random variable. These steps can be efficiently implemented in a MATLAB program for m as large as 2000. Moreover, as confirmed by our simulations, the discrete Gaussian process with $m = 1100$ approximates well $\{W_t, 0\le t\le T\}$ when $T\le 11$.

We have conducted Monte Carlo simulations for the LRTS for (3) with $\lambda_0 = 1$ for many sample sizes n ranging from 100 to 5000. We also simulated $\sup_{0\le t\le T}(W_t \vee 0)^2$ for various values of T using the discrete process with $m = 1100$. The empirical CDFs of $2\Lambda_n$ and $\sup_{0\le t\le T}(W_t\vee 0)^2$ for every n and T are obtained from 5000 independent replicates. On a Pentium III 660 PC with 128 MHz RAM, the computational time for simulating the LRTS ranges from 2 h for $n = 100$ to 100 h for $n = 5000$. It takes only 15 min to simulate the GPTS, however.

For each given n, we found T_n, the value of T minimizing the Kolmogorov–Smirnov distance between the empirical CDF of $2\Lambda_n$ and $\sup_{0\le t\le T}(W_t\vee 0)^2$. We also found T_n, which minimizes the distance between their upper tail over the 95th percentile. There is no significant difference between these two values. We use the distance of the upper tail distribution because it is the area

Table 1. *Comparison of percentiles of the LRTS, SGP and fitted SGP*

Sample size n	T_n	$\hat{T}_n$	LRTS 95%	LRTS 97.5%	LRTS 99%	SGP 95%	SGP 97.5%	SGP 99%	Fitted SGP 95%	Fitted SGP 97.5%	Fitted SGP 99%
100	6.96	6.96	5.27	6.37	8.06	5.24	6.49	8.27	5.24	6.49	8.27
250	7.63	7.64	5.41	6.60	8.07	5.42	6.64	8.49	5.42	6.66	8.51
500	7.75	8.20	5.39	6.70	8.65	5.44	6.67	8.61	5.54	6.76	8.68
750	8.71	8.52	5.55	6.90	8.97	5.65	6.89	8.69	5.61	6.83	8.69
1000	8.98	8.75	5.62	7.04	8.77	5.68	6.96	8.69	5.66	6.89	8.69
1500	9.25	9.06	5.67	7.02	9.03	5.71	6.97	8.71	5.68	6.96	8.69
2000	9.08	9.30	5.73	7.00	8.44	5.68	6.96	8.69	5.71	6.97	8.71
3000	9.77	9.61	5.78	7.11	9.16	5.82	7.06	8.74	5.79	7.05	8.73
4000	10.25	9.85	5.75	7.33	9.10	5.90	7.12	8.99	5.85	7.07	8.84
5000	9.50	10.02	5.66	7.08	9.33	5.77	7.00	8.73	5.86	7.07	8.86

of main interest in hypothesis testing problems. The largest value of the Kolmogorov–Smirnov distance between the upper tails of the two distributions is 1.98 per cent.

From the proof of theorem 2, the supremum of the log-likelihood function is achieved when $\lambda \in [\log n, n(\log n)^{-4}]$. It suggests that, if the distribution of $\sup_{0\le t\le T_n}(W_t \vee 0)^2$ is close to that of the LRTS, T_n should be at the order of $\log n$. In fact, the plot of T_n against $\log n$ indicates a very strong linear pattern with an R-square value 0.91. Thus, for any reasonably large sample size n, our recommended value of T_n is given by the regression equation: $\hat{T}_n = 3.27 + 0.8 \log n$.

Table 1 presents the 95, 97.5 and 99 per cent percentiles of the LRTS $2\Lambda_n$, the simulated Gaussian process (SGP), $\sup_{0\le t\le T_n}(W_t \vee 0)^2$, and the fitted SGP, $\sup_{0\le t\le \hat{T}_n}(W_t \vee 0)^2$.

Comparing the CDFs of the LRTS, we observe an overall increase of values of quantiles with sample sizes as excepted. Because the variances of quantiles of the LRTS and SGP increase as the corresponding percentage increases, the quantiles of both Gaussian processes, $\sup_{0\le t\le T_n}(W_t \vee 0)^2$ and $\sup_{0\le t\le \hat{T}_n}(W_t \vee 0)^2$, fit that of the LRTS best for 95 per cent percentile and worst for 99 per cent percentile. The residual plot of the regression model (17) also confirms that the regression error increases as the sample size increases.

Acknowledgements

This work has been partially supported by grant HG00008 from the National Human Genome Research Institute. We are indebted to the editor, the associate editor and the referees for their careful reading of the first version of the manuscript and for their very helpful suggestions in the organization of the paper as well as the simulation section.

References

Cheng, R. C. H. & Traylor, L. (1995). Non-regular maximum likelihood problems (with discussion). *J. R. Statist. Soc. B* **57**, 3–44.

Csörgö, M., Csörgö, S., Horváth, L. & Mason, D. M. (1986). Weighted empirical and quantile processes. *Ann. Probab.* **14**, 31–85.

Csörgö, M. & Révész, P. (1981). *Strong approximations in probability and statistics*. Academic Press, New York-London.

Davies, R. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64**, 247–254.

Hartigan, J. A. (1985). A failure of likelihood ratio asymptotics for normal mixtures. In *Proc. Berkeley conference in honor of Jerzy Neyman and Jack Kiefer* (eds L. Le Cam & A. Olshen, Vol II. Wasworth Advanced Books, Monterey, CA.

Komlos, J., Major, P. & Tusnady, G. (1975). An approximation of partial sums of independent r.v.s and the sample distribution function. *Z. Wahr. verw. Geb.* **32**, 111–131.

Leadbetter, M. R., Lindgren, G. & Rootzen, H. (1983). *Extremes and related properties of random sequences and processes*. Springer, New York.

Lindsay, B. G. (1995). *Mixture models: theory, geometry and applications*. NSF-CBMS regional conference series in probability and statistics, Vol. 5. IMS.

Liu, X. & Shao, Y. (in press). Asymptotics for the likelihood ratio test under loss of identifiability. *Ann. Statist.*

Nelson, W. (1982). *Applied life data analysis*. Wiley, New York.

Seal, H. (1969). *Stochastic theory of a risk business*. Wiley, New York.

Slud, E. (1997). Testing for imperfect debugging in software reliability. *Scand. J. Statist.* **24**, 555–572.

Titterington, D. M., Smith, A. F. M. & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*. Wiley, New York.

Received September 2000, in final form February 2002

Xin Liu, Laboratory of Statistical Genetics, Rockefeller University, 1230 York Avenue, New York, NY 10021, USA.
E-mail: liuxin@linkage.rockefeller.edu

Appendix

Assume that $U_1, \dots, U_n$ are i.i.d. uniformly distributed random variables. Denote its empirical CDF by $F_n(u) = n^{-1} \sum_{i=1}^{n} 1_{[0,u]}(U_i)$ $(0 \le u \le 1)$. A useful strong approximation result for the uniform empirical process $\alpha_n(u) = n^{1/2}(F_n(u) - u)$ is the following proposition from Csörgö *et al.* (1986).

Proposition 1 (*Csörgö et al.,* 1986, *theorem* 2.1)
On a suitable probability space we have

$$\sup_{c/n \le u \le 1-c/n} n^{v} \left| \frac{\alpha_n(u) - B_n(u)}{(u(1-u))^{1/2-v}} \right| = O_P(1), \quad \textit{as } n \to \infty, \tag{9}$$

where $\{B_n(u); 0 \le u \le 1\}$ *is a Brownian bridge,* $0 \le v < 1/2$ *and* $0 < c < \infty$.

Denote by $U_{1,n}, \dots, U_{n,n}$ the order statistic of $U_1, \dots, U_n$. Both $nU_{1,n}$ and $n(1 - U_{n,n})$ converges in distribution to standard exponential random variables. Then (9) yields,

$$\sup_{U_{1,n} \le u \le U_{n,n}} n^{v} \left| \frac{\alpha_n(u) - B_n(u)}{(u(1-u))^{1/2-v}} \right| = O_P(1). \tag{10}$$

The law of the iterated logarithm for Brownian bridge implies that (see also Csörgö *et al.*, 1986),

$$\sup_{U_{1,n} \le u \le U_{n,n}} (u(1-u))^{-1/2} B_n(u) = O_P((\log_{(2)} n)^{1/2}). \tag{11}$$

Equations (10) and (11) yield

$$\sup_{U_{1,n} \le u \le U_{n,n}} (u(1-u))^{-1/2} |\alpha_n(u)| = O_P((\log_{(2)} n)^{1/2}). \tag{12}$$

The Hungarian construction (Komlos *et al.*, 1975) implies that on a suitable probability space, we have

$$\sup_{0 \le u \le 1} |\alpha_n(u) - B_n(u)| = O_P(n^{-1/2} \log n). \tag{13}$$

In our proof, the Brownian bridge $\{B_n(u); 0 \le u \le 1\}$ is constructed on some suitable probability space such that equations (10)–(13) are all satisfied. We assume that B_n has the decomposition $\{B_n(u) = W_n(u) - W_n(1)u, 0 \le u \le 1\}$, where $\{W_n(u), 0 \le u \le 1\}$ is a standard Brownian motion.

Proof of lemma 1. Since $\int_0^\infty Z^2(x,\lambda)\,dG(x) = 1$, we have $\mathrm{E}H_n^2(\lambda) = 1$. Denote by $\rho(\cdot,\cdot)$ the covariance kernel of $H_n(e^s + 1/2)$, i.e. $\rho(s_1, s_2) = \mathrm{E}H_n(e^{s_1} + 1/2)H_n(e^{s_2} + 1/2)$. Then,

$$\begin{aligned}\rho(s_1, s_2) &= \int_0^\infty Z(x, e^{s_1} + 1/2)Z(x, e^{s_1} + 1/2)\,dG(x)\\ &= \left[1 + \frac{1}{2}\left(e^{(s_1-s_2)/4} - e^{-(s_1-s_2)/4}\right)^2\right]^{-\kappa}\\ &= 1 - \kappa(s_1 - s_2)^2/8 + o((s_1 - s_2)^2).\end{aligned}$$

It yields that $H_n(e^s + 1/2)$ is a stationary Gaussian process. Theorem 8.2.7 in Leadbetter *et al.* (1983) yields lemma 1.

Proof of Lemma 2. We prove the lemma by considering each of following intervals of λ^κ:

$$[1, \log n], [\log n,\ n(\log n)^{-4}],\ [n(\log n)^{-4}, n],\quad [n, \infty).$$

Case (1) When $\lambda^\kappa \in [\log n, n(\log n)^{-4}]$. Since $W_n(1)$ is a standard normal random variable, $W_n(1)(\sqrt{2\lambda - 1}/\lambda)^\kappa = O_P((\log n)^{-1/2})$. Note that $Z(x,\lambda)$ is a monotone function of x. $R_n(\lambda)$ can be bounded using Hungarian construction (13) as follows:

$$\begin{aligned}R_n(\lambda) &= O(1)\sup_x |\alpha(G(x)) - B_n(G(x))| \sup_x Z(x,\lambda)\\ &= O_P(n^{-1/2}\log n)\lambda^{\kappa/2} = O_P(n^{-1/2}(\log n)^{-1}).\end{aligned} \tag{14}$$

Then (5) yields $\sup_{\lambda^\kappa \in [\log n, n(\log n)^{-4}]} |S_n(\lambda) - H_n(\lambda)| = O_P((\log n)^{-1/2})$.

Case (2) When $\lambda^\kappa \in [1, \log n]$, similar to the proof of the first case, we can show $\sup_{\lambda^\kappa \in [1,\log n]} |S_n(\lambda) - H_n(\lambda)| = O_P(1)$. In lemma 1, let $T = \log[(\log n)^{1/\kappa} - 1/2] - \log(1/2)$. Then $T = O(\log_{(2)} n)$ and $\sup_{\lambda^\kappa \in [1,\log n]} H_n(\lambda) = O_P((\log_{(3)} n)^{1/2})$. Thus, $\sup_{\lambda^\kappa \in [1,\log n]} S_n(\lambda) \vee 0 = O_P((\log_{(3)} n)^{1/2})$.

Case (3) When $\lambda^\kappa \in [n(\log n)^{-4}, n]$, we first estimate the order of $R_n(\lambda)$.

$$\begin{aligned}R_n(\lambda) &= \int_{X_{1,n}}^{X_{n,n}} [\alpha_n(G(x)) - B_n(G(x))]\,d(-Z(x,\lambda)) + \int_{X_{n,n}}^\infty [\alpha_n(G(x)) - B_n(G(x))]\,d(-Z(x,\lambda))\\ &\quad + \int_0^{X_{1,n}} \sqrt{n}[F_n(G(x)) - G(x)]\,d(-Z(x,\lambda)) + \int_0^{X_{1,n}} B_n(G(x))\,dZ(x,\lambda).\end{aligned}$$

We denote the four parts above by Δ_1, Δ_2, Δ_3 and Δ_4, respectively. In (10), let $v = 1/4$. We have uniformly for $X_{1,n} \le x \le X_{n,n}$ in probability

$$|\alpha_n(G(x)) - B_n(G(x))| = O_P(n^{-1/4})[G(x)(1 - G(x))]^{1/4}.$$

Note that $G(x) = O(1)x^\kappa$. We can then bound Δ_1 as follows:

$$|\Delta_1| = O_P(n^{-1/4}) \int_{X_{1,n}}^{X_{n,n}} x^{\kappa/4}\lambda^{1+\kappa/2}e^{-x(\lambda-1)}dx = O_P(n^{-1/4})\lambda^{\kappa/4} = O_P(1).$$

When $x \ge X_{n,n}$, $e^{x(\lambda-1)} > n^k$ with probability going to one for any given $k > 0$. Then (13) directly yields $\Delta_2 = o_P(1)$.

When $0 \leq x < X_{1,n}$, we have $F_n(x) = 0$ and $-Z(x, \lambda)$ is an increasing function of x. Thus, $\Delta_3 < 0$.

Last, when $0 \leq x \leq X_{1,n}$, we have $B_n(G(x)) = O_P(1)G^{1/2}(X_{1,n}) = O_P(1)X_{1,n}^{\kappa/2}$. Thus,

$$|\Delta_4| = O_P(1) \int_0^{X_{1,n}} X_{1,n}^{\kappa/2} \lambda^{1+\kappa/2} e^{-x(\lambda-1)} dx = O_P(n^{1/2}) X_{1,n}^{\kappa/2} = O_P(1).$$

Therefore, we prove $R_n(\lambda) \vee 0 = O_P(1)$. Lemma 1 yields $\sup_{\lambda^\kappa \in [n(\log n)^{-4}, n]} H_n(\lambda) = O_P((\log_{(3)} n)^{1/2})$. Finally, we obtain

$$\sup_{\lambda^\kappa \in [n(\log n)^{-4}, n]} S_n(\lambda) \vee 0 = O_P((\log_{(3)} n)^{1/2}).$$

Case (4) When $\lambda^\kappa \geq n$, $G(x) = O(x^\kappa)$ and $d(-Z(x, \lambda))/dx = O(1)\lambda^{1+\kappa/2} e^{-x(\lambda-1)}$. Straightforward calculation shows

$$\begin{aligned} S_n(\lambda) &= n^{1/2} \int_0^\infty Z(x, \lambda) d[F_n(G(x)) - G(x)] \\ &= n^{1/2} \int_0^\infty [F_n(G(x)) - G(x)] d(-Z(x, \lambda)) \\ &= O_P(n^{1/2}) \int_0^\infty \lambda^{1+\kappa/2} x^\kappa e^{-x(\lambda-1)} dx = O_P(1). \end{aligned}$$

This completes the proof of lemma 2.

Proof of theorem 1. In lemma 1, let $T_n = \log[n^{1/\kappa}(\log n)^{-4/\kappa} - \frac{1}{2}] - \log[(\log n)^{1/\kappa} - \frac{1}{2}]$. Then,

$$A_{T_n} = (\log_{(2)} n)^{1/2} + O((\log n)^{-1/2}).$$

We have

$$\begin{aligned} &\lim_{n\to\infty} \mathrm{P}\left\{M_n^2 - \log_{(2)} n + \log(16\pi^2/\kappa) \leq x\right\} \\ &= \lim_{n\to\infty} \mathrm{P}\left\{\left[\sup_{0\leq s\leq T_n} H_n(s)\right]^2 - \log_{(2)} n + 2\log(4\pi/\sqrt{\kappa}) \leq x\right\} \\ &= \lim_{n\to\infty} \mathrm{P}\left\{\left[\sup_{0\leq s\leq T_n} H_n(s) + A_{T_n}\right]\left[\sup_{0\leq s\leq T_n} H_n(s) - A_{T_n}\right] + 2\log(4\pi/\sqrt{\kappa}) + o_P(1) \leq x\right\} \\ &= \lim_{n\to\infty} \mathrm{P}\left\{2A_{T_n}\left[\sup_{0\leq s\leq T_n} H_n(s) - A_{T_n}\right] + 2\log(4\pi/\sqrt{\kappa}) + o_P(1) \leq x\right\} \\ &= \lim_{n\to\infty} \mathrm{P}\left\{A_{T_n}\left[\sup_{0\leq s\leq T_n} H_n(s) - A_{T_n}\right] + \log(4\pi/\sqrt{\kappa}) \leq x/2\right\} \\ &= \exp(-e^{-x/2}). \end{aligned}$$

Proof of lemma 3. To make notations simple, we define $t(p, \lambda) = p\lambda^\kappa$ and $x_0(p, \lambda) = \log(p\lambda^\kappa)/(\lambda - 1)$. When no confusion might occur, we simply write $t(p, \lambda)$ and $x_0(p, t)$ as t and x_0, respectively. Without loss of generality, we assume $t \geq 2$ (otherwise, we already have $t = O_P(1)$). Note that $pl(x, \lambda) = e^{-(x-x_0)(\lambda-1)}$. When $x \leq x_0$,

$$1 + p(l(x, \lambda) - 1) \leq 1 + e^{-(x-x_0)(\lambda-1)} \leq 2e^{-(x-x_0)(\lambda-1)}. \tag{15}$$

Applying the inequalities $\log(1 + x) \leq x$ $(x > -1)$ and (15) to $L_n(p, \lambda)$, we have

$$n^{-1}L_n(p,\lambda) = \left(\int_{x_0}^{\infty} + \int_0^{x_0}\right) \log(1 + p(l(X_i,\lambda) - 1))\, dF_n(G(x))$$
$$\leq \int_{x_0}^{\infty} (e^{-(x-x_0)(\lambda-1)} - p)\, dF_n(G(x)) + \int_0^{x_0} [\log 2 + (x_0 - x)(\lambda - 1)]\, dF_n(G(x)). \quad (16)$$

Note that $F_n(G(x)) = O_P(1)G(x)$, we have

$$\int_{x_0}^{\infty} (e^{-(x-x_0)(\lambda-1)} - p)\, dF_n(G(x)) = -p + (p-1)F_n(G(x_0)) + \int_{x_0}^{\infty} F_n(G(x))\, d(-e^{-(x-x_0)(\lambda-1)})$$
$$\leq -p + O_P(1)\int_{x_0}^{\infty} G(x)\, d(-e^{-(x-x_0)(\lambda-1)})$$
$$= -p + O_P(1)\left(G(x_0) + \int_{x_0}^{\infty} e^{-(x-x_0)(\lambda-1)} dG(x)\right)$$
$$= -p + O_P(1)[G(x_0) + p(1 - G(\lambda x_0))].$$

Note that $(x_0 - x)(\lambda - 1) \leq \log t$ for $x \leq x_0$. The second integral in (16) can be bounded by $O_P(1)G(x_0)\log t$:

$$n^{-1}L_n(p,\lambda) \leq -p + O_P(1)[p(1 - G(\lambda x_0)) + G(x_0)\log t].$$

Since $G(x) = O_P(1)x^{\kappa}$ and $\lambda^{-\kappa} = p/t$, we have $G(x_0) = O_P(1)[\log t/(\lambda-1)]^{\kappa} = O_P(1)p(\log t)^{\kappa}/t$. Note that $\lambda x_0 \geq \log t$, we have $1 - G(\lambda x_0) \leq 1 - G(\log t)$. Therefore,

$$n^{-1}L_n(p,\lambda) \leq p\{-1 + O_P(1)[1 - G(\log t) + (\log t)^{\kappa}/t]\}. \quad (17)$$

The right-hand side of (17) is decreasing in t when t is large and it tends to the negative value $-p$ as $t \to \infty$. Therefore, when $L_n(p,\lambda) > 0$, $t = O_P(1)$.

Proof of lemma 4. First, we bound the integral $\int G(x)\, d(-Y^2(x,\lambda))$. Note that,

$$d(-Y^2(x,\lambda)) = -(\lambda-1)(2\lambda-1)^{\kappa/2} e^{-x(\lambda-1)} Y(x,\lambda)\, dx.$$

Let $\tilde{x} = \kappa \log\lambda/(\lambda - 1)$. Then $Y(\tilde{x},\lambda) = 0$. $d(-Y^2(x,\lambda))/dx$ is positive, when $x < \tilde{x}$; negative, when $x > \tilde{x}$. We can bound $|d(-Y^2(x,\lambda))/dx|$ and $G(x)$ as follows according to $x \leq \tilde{x}$ or $x > \tilde{x}$:

$$|d(-Y^2(x,\lambda))/dx| = \begin{cases} O(1)\lambda^{1+\kappa} e^{-2(\lambda-1)x} \\ O(1)\lambda e^{-(\lambda-1)x} \end{cases} \qquad G(x) = \begin{cases} O(1)x^{\kappa}, & \text{if } x \leq \tilde{x}, \\ O(1), & \text{if } x > \tilde{x}. \end{cases}$$

Then, for $0 < a < b < \infty$, we have

$$I(a,b) = \int_a^b G(x)|d(-Y^2(x,\lambda))| = \begin{cases} O(1)\tilde{G}(2(\lambda-1)b), & \text{if b} \leq \tilde{\text{x}}, \\ O(1)e^{-(\lambda-1)a}, & \text{if a} > \tilde{\text{x}} \end{cases}$$

where $\tilde{G}(x)$ is the CDF of a gamma density: $\tilde{G}(x) = \int_0^x (\Gamma(\kappa+1))^{-1} y^{\kappa+1} e^{-y} dy$.

Using integration by parts, we can write $P_n Y^2(\lambda)$ as,

$$P_n Y^2(\lambda) = \int_0^{\infty} Y^2(x,\lambda)\, dG(x) + \int_0^{\infty} Y^2(x,\lambda)\, d(F_n(G(x)) - G(x))$$
$$= 1 - (\sqrt{2\lambda-1}/\lambda)^{2\kappa} + \left[\left(\int_0^{X_{1,n}} + \int_{X_{n,n}}^{\infty}\right) + \int_{X_{1,n}}^{X_{n,n}}\right][F_n(G(x)) - G(x)]\, d(-Y^2(x,\lambda))$$
$$= 1 - (\sqrt{2\lambda-1}/\lambda)^{2\kappa} + \Delta_5 + \Delta_6. \quad (18)$$

Assuming that λ^* is sufficiently large, we bound Δ_5 and Δ_6 in the following two intervals of λ^{κ}.

Case (1) When $\lambda^\kappa \in [(\lambda^*)^\kappa, n(\log n)^{-4}]$. Obviously $X_{1,n} \le \tilde{x} \le X_{n,n}$. Since $\tilde{G}(x) = O_P(1)x^{\kappa+1}$, we have

$$|\Delta_5| = O_P(1)[I(0, X_{1,n}) + I(X_{n,n}, \infty)] = O_P(1)[\tilde{G}((\lambda-1)X_{1,n}) + e^{-(\lambda-1)X_{n,n}}]$$
$$= O_P(1)(\lambda X_{1,n})^{\kappa+1} + O_P((\log n)^{-1}) = O_P((\log n)^{-1}).$$

Equation (12) implies $|\alpha_n(G(x))| = O_P((\log_{(2)} n)^{1/2})(G(x))^{1/2} = O_P((\log_{(2)} n)^{1/2})x^{\kappa/2}$ uniformly for $X_{1,n} \le x \le X_{n,n}$. Note that, for $x \ge 0$,

$$|d(-Y^2(x,\lambda))/dx| = O_P(1)(\lambda^{1+\kappa}e^{-2(\lambda-1)x} + \lambda e^{-(\lambda-1)x}).$$

Then Δ_6 can be bounded as follows:

$$|\Delta_6| = O_P(n^{-1/2}(\log_{(2)} n)^{1/2}) \int_0^\infty x^{\kappa/2}(\lambda^{1+\kappa}e^{-2(\lambda-1)x} + \lambda e^{-(\lambda-1)x})dx$$
$$= O_P(n^{-1/2}(\log_{(2)} n)^{1/2})\lambda^{\kappa/2} = O_P((\log n)^{-1}).$$

It is then clear that, when $\lambda^\kappa \in [\log n, n(\log n)^{-4}]$, $P_nY^2(\lambda) = 1 + O_P((\log n)^{-1})$. When $\lambda^\kappa \in [(\lambda^*)^\kappa, \log n]$, in probability going to one $P_nY^2(\lambda) \ge [1-(2\lambda^*-1)^\kappa/(\lambda^*)^{2\kappa}]/2$.

Case (2) When $\lambda^\kappa \in [n(\log n)^{-4}, \kappa/X_{1,n}]$. The results in Csörgö & Révész (1981) implies that uniformly for $10/n \le G(x) \le 1$

$$B_n(G(x)) = O_P(1)G^{1/2}(x)[\log_2(nG(x))]^{1/2} = O_P(1)(nG(x))^{-1/2}[\log_2(nG(x))]^{1/2}\sqrt{n}G(x).$$

Since $x^{-1/2}(\log_2 x)^{1/2}$ can be arbitrarily small as x tends to infinity, there exists a sequence $\{t_{1,n} > 10, n = 1, 2, \ldots\}$ such that $t_{1,n} = O_P(1)$ and when $nG(x) \ge t_{1,n}$, $|B_n(G(x))| \le \sqrt{n}G(x)/4$. In (10), let $v = 0$, we have $|\alpha_n(G(x)) - B_n(G(x))| = O_P(1)G^{1/2}(x)$ for $X_{1,n} \le x \le X_{n,n}$. Similarly, we can find a sequence $\{t_{2,n} > nG(X_{1,n}), n = 1, 2, \ldots\}$ such that $t_{2,n} = O_P(1)$ and when $nG(x) \ge t_{2,n}$, $|\alpha_n(G(x)) - B_n(G(x))| \le \sqrt{n}G(x)/4$. Let $t_{3,n} = \max(t_{1,n}, t_{2,n})$. Then $t_{3,n} = O_P(1)$ and when $nG(x) \ge t_{3,n}$,

$$|F_n(G(x)) - G(x)| \le n^{-1/2}|\alpha_n(G(x)) - B_n(G(x))| + n^{-1/2}|B_n(G(x))| \le G(x)/2.$$

Therefore, $G(x)/2 \le F_n(G(x)) \le 3G(x)/2$ for $x \ge G^{-1}(t_{3,n}/n)$. Since $nG(\tilde{x}/2) > (n/\lambda^\kappa)(\log\lambda)^{\kappa/2} > t_{3,n}$, we have $\tilde{x} \ge 2G^{-1}(t_{3,n}/n)$ and,

$$P_nY^2(\lambda) = \int_0^\infty F_n(x)\,d(-Y^2(x,\lambda))$$
$$\ge \frac{1}{2}\int_{G^{-1}(t_{3,n}/n)}^{\tilde{x}/2} G(x)\,d(-Y^2(x,\lambda)) - \frac{3}{2}\int_{\tilde{x}}^\infty G(x)\,dY^2(x,\lambda).$$

It is easy to prove that when $x \le \tilde{x}/2$, for some constants c_1, $G(x)\,d(-Y^2(x,\lambda))/dx \ge c_1\lambda^{\kappa+1}x^\kappa e^{-2(\lambda-1)x}$. Let $t_{4,n} = (2\lambda-1)G^{-1}(t_{3,n}/n)$. Then, there exist constants $c_2 > 0$ such that

$$P_nY^2(\lambda) \ge c_1\int_{G^{-1}(t_{3,n}/\lambda)}^{\tilde{x}/2} x^\kappa\lambda^{\kappa+1}e^{-(2\lambda-1)x}dx - O_P(1)\int_{\tilde{x}}^\infty \lambda e^{-(\lambda-1)x}dx$$
$$\ge c_2[\tilde{G}((2\lambda-1)\tilde{x}/2) - \tilde{G}(t_{4,n})] - O_P(1)e^{-(\lambda-1)\tilde{x}}$$
$$\ge c_2[1-\tilde{G}(t_{4,n})] - o_P(1). \tag{19}$$

When $0 \le x \le 1/2$, $G^{-1}(x) = O(1)x^{1/\kappa}$. Since $t_{3,n} = O_P(1)$, we have $t_{4,n} = O_P(n^{-1/\kappa})\,X_{1,n}^{-1} = O_P(1)$. Consequently, $1/(1-\tilde{G}(t_{4,n})) = O_P(1)$. Then (19) yields $P_nY^2(\lambda) \ge c_2\,(1-\tilde{G}(t_{4,n}))/2$ with probability going to one. Let $c_n = \min([1-(2\lambda^*-1)^\kappa/(\lambda^*)^{2\kappa}]/2, c_2(1-\tilde{G}(t_{4,n}))/2)$. Then we prove that $P_nY^2(\lambda) \ge c_n$ with probability going to one. Note that $1/c_n = O_P(1)$. This completes the proof of lemma 2.

Proof of theorem 2. We have already shown in section 2 that

$$\sup_{\substack{L_n(p,\lambda)>0 \\ \lambda^{\kappa}\in[\log n, n(\log n)^{-4}]}} D(p,\lambda)Y(x,\lambda) = O_P((\log n)^{-1}).$$

Applying Taylor expansion $\log(1+x) = x - (1/2 + O(x))x^2$ and lemma 4 to $L_n(p,\lambda)$ for $\lambda^{\kappa} \in [\log n, n(\log n)^{-4}]$ yields

$$\begin{aligned} 2L_n(p,\lambda) \vee 0 &= \left\{ 2\sqrt{n}D(p,\lambda)S_n(\lambda) - [1 + O_P((\log n)^{-1})]nD^2(p,\lambda)P_nY^2(\lambda) \right\} \vee 0 \\ &= \left[2\sqrt{n}D(p,\lambda)S_n(\lambda) - nD^2(p,\lambda) \right] \vee 0 + o_P(1). \end{aligned} \quad (20)$$

Obviously, (20) and the arguments in section 2 yield $2\Lambda_n \le M_n^2 + o_P(1)$. To prove $2\Lambda_n = M_n^2 + o_P(1)$, it suffices to show that there exists a point (p', λ') such that $(\lambda')^{\kappa} \in [\log n, n(\log n)^{-4}]$ and $2L_n(p', \lambda') = M_n^2 + o_P(1)$. In fact, let λ' be the point maximizing $S_n(\lambda)$, i.e. $S_n(\lambda') = M_n$, and p' be the solution of the equation $D(p', \lambda') = n^{-1/2}M_n$. By lemma 1, with probability going to one, $(\lambda')^{\kappa} \in [\log n, n(\log n)^{-4}]$. It is clear that $2L_n(p', \lambda') = M_n^2 + o_P(1)$. Therefore, $2\Lambda_n = M_n^2 + o_P(1)$. This completes the proof of theorem 2.