1 Introduction

Suppose $F_1, F_2, \ldots, F_k$ are probability distributions. Then the probability distribution

$$p_1F_1 + p_2F_2 + \cdots + p_kF_k \ ,$$

where $p_1 + p_2 + \cdots + p_k = 1$, is called a *finite mixture* distribution. Observations from this mixture distribution can be thought of as being drawn from F_i with probability p_i, $i = 1, 2, \ldots, k$. We call $F_1, F_2, \ldots, F_k$ the *components* of the mixture.

Finite mixture distributions have enjoyed much attention over the last few decades for various reasons. One of these is that when testing the hypothesis of one component against the alternative of two components (a *test of homogeneity*), usually well-behaved test statistics exhibit strange behaviour when the hypothesis is true. For instance, when testing the hypothesis of $\mathcal{N}(0, 1)$ against the alternative of

$$(1 - p)\mathcal{N}(0, 1) + p\mathcal{N}(\theta, 1)$$

where both p and θ are unknown, it is shown in an early paper (Hartigan (1985)) that the likelihood ratio based on an independent and identically distributed sample tends to infinity in probability (albeit very slowly) as the sample size increases. The main problem is one of *non-indentifiability*; when $p = 0$, θ can be anything, and when $\theta = 0$, p can be anything and we still just have $\mathcal{N}(0, 1)$. The approach in Hartigan (1985) is to firstly suppose that θ is known and fixed, and just test $p = 0$. This 'sub-problem' is more or less standard with a reasonably well-behaved likelihood ratio statistic, which can be approximated in terms of the standardised *score* statistic. However, to obtain the likelihood ratio for the *full* problem we have to maximise over a *stochastic process* indexed by θ. It is the maximum of the *score process* that tends to infinity in probability, which then gives the result.

Few authors pursued this problem directly; most imposed conditions on the range of θ. If θ is restricted to a finite interval, then the statistic no longer tends to infinity in probability under the hypothesis; it behaves like the maximum of a certain Gaussian process on the interval, which is bounded in probability. There exist many generalisations based on this approach, but each involves a parameter restriction of some sort to prevent this 'divergent' behaviour of the statistic under the hypothesis. Possibly the most general result of this type is in Dacunha-Castelle & Gassiat (1997), where mixtures of general parametric families satisfying certain differentiability conditions are considered. However there is still an effective parameter space restriction; the collection of score functions needs to be a Donsker class, which ensures that the score process converges to a tight Gaussian limit process. However this condition is not satisfied even in the simple problem examined in Hartigan (1985).

In Bickel & Chernoff (1993) the composite hypothesis version of the Hartigan problem is considered: the test of the hypothesis of $\mathcal{N}(\theta, 1)$ with θ unknown against the alternative of

$$(1-p)\mathcal{N}(\theta_1, 1) + p\mathcal{N}(\theta_2, 1)$$

for unknown p, θ_1 and θ_2. Rather than considering the likelihood ratio statistic Λ_n directly they show that the maximum of the *efficient*(composite-hypothesis) *score* process M_n behaves asymptotically like the maximum of a stationary standardised Gaussian process over a *slowly growing interval.* So M_n tends to infinity in probability, but the quantity

$$\sqrt{\log\log n}\left(M_n - \sqrt{\log\log n}\right) + \log 2\pi$$

has an asymptotic Gumbel extreme value distribution, with cumulative distribution function $G(x) = \exp\{-e^{-x}\}$ (whether or not M_n and Λ_n are asymptotically equivalent is not addressed). The method of proof utilises various known results and tools, including the so-called Hungarian construction of Komlós *et al.* (1975) for approximating the $U(0, 1)$ empirical process with a Brownian Bridge. The phenomenon that guides the whole analysis is that the efficient score process behaves as we would expect over a certain range but then degrades rapidly to zero outside this range; in turn this 'range of activity' grows slowly with the sample size.

We provide a generalisation of the Bickel & Chernoff (1993) result motivated by a problem in neurobiology. A model for the duration of ion-channel openings is that there are 3 states: 'closed' ↔ 'open1' ↔ 'open2', and times in each state follow a continuous time Markov chain. However the detection mechanism cannot distinguish between the two open states. It is shown in Colquhoun & Hawkes (1981) that the distribution of durations of detectable openings is then a two component mixture of exponential distributions with density

$$(1-p)\theta_1 e^{-\theta_1 x} + p\theta_2 e^{-\theta_2 x}\ .$$

A simpler model is that there are just two states, 'closed' and 'open', in which case open times would follow a single exponential distribution. So a test to distinguish these models would be exactly a test of homogeneity, that is a test of the hypothesis of one exponential component against the alternative of a mixture of two exponential components.

Following Bickel & Chernoff (1993) we in fact provide a limit theorem for the asymptotic distribution of the maximum of the efficient score process for testing the hypothesis of G_θ against an alternative of

$$(1-p)G_{\theta_1} + pG_{\theta_2}$$

where G_θ is a member of a general one-parameter exponential family, that is the density is of the form

$$g_\theta(x) = e^{\theta t(x) - K(\theta)} g_0(x) \ .$$

The development is again guided by the fact that the efficient score process behaves nicely within a certain range of activity and then outside this range degrades rapidly to zero. We simultaneously cover the simple-hypothesis score and composite-hypothesis efficient score versions of the test for homogeneity by studying a third *approximate score* process.

Before embarking on the main development, we briefly discuss results from Lindsay (1995), which provide among other things results for tests of homogeneity when the sample space is finite, using geometric arguments. We provide a heuristic extension of these ideas to the infinite sample space case, in particular discussing the concept of the *arc length* of a one-parameter family of standardised functions, and its effect on the behaviour of the maximum of the empirical process indexed by this family.

The main development is in essentially three stages: a stage involving *approximation theorems*, a stage involving *Gaussian processes* and a *final assembly* stage.

The first stage concerns the approximation of the the approximate score process $\{S_n(\theta) \,|\, \theta \in \Theta, n \in \mathbb{N}\}$ by a *sequence* of Gaussian processes $\{Z_n(\theta) \,|\, \theta \in \Theta, n \in \mathbb{N}\}$. This is facilitated by an improvement of the Komlós *et al.* (1975) approximation theorem provided by Csorgo *et al.* (1986), which nicely explains the degradation phenomenon by approximating the $U(0,1)$ empirical process by a Brownian Bridge over the interval $\left[\frac{1}{n}, 1-\frac{1}{n}\right]$. The 'range of activity' of the efficient score process corresponds to this interval; the 'range of degradation' corresponds to the complementary end-segments $\left(0, \frac{1}{n}\right)$ and $\left(1-\frac{1}{n}, 1\right)$. Use of the result requires that the exponential family has a monotone likelihood ratio property, and that the densities satisfy a certain tail condition.

The second stage studies the maxima of the sequence of Gaussian processes

$$\widetilde{M}_n = \sup_{\theta \in \Theta} \widetilde{Z}_n(\theta) \ ,$$

where $\widetilde{Z}_n(\theta) = Z_n(\theta) + c_n(\theta)\widetilde{X}$, $c_n(\theta)$ is a certain sequence of functions and $\widetilde{X} \sim \mathcal{N}(0,1)$ independently of $Z_n(\theta)$. The addition of the extra term here gives the resultant sequence of Gaussian processes nice covariance functions; this is an adaptation of a trick used in Bickel & Chernoff (1993). The extra term is only significant for small values of θ, and so its overall effect is negligible. A limit theorem for $\widetilde{M}_n$ is obtained by extending results from Hüsler (1990, 1995). Results there pertain to the maximum over a growing interval of a *single* Gaussian process which is not necessarily stationary, not necessarily standardised but has a smooth covariance function:

$$\sup_{\theta \in \Theta_n} \{X_0(\theta) \,|\, \theta \in \mathbb{R}\} \ .$$

We extend the result to the maxima of a *sequence* of such processes whose covariance functions don't change too dramatically as n changes:

$$\sup_{\theta \in \Theta_n} \{X_n(\theta) \,|\, \theta \in \mathbb{R}, n \in \mathbb{N}\} \ .$$

We see that the limiting distribution of the the maxima of such a sequence of Gaussian processes is the same as that of the maximum of a *single* Gaussian process over a slowly growing interval, generalising precisely the result in Bickel & Chernoff (1993). In fact we have that the quantity

$$\sqrt{2\log T_n}\left(\widetilde{M}_n - \sqrt{2\log T_n}\right) + \log 2\pi$$

has an asymptotic Gumbel extreme value distribution, where in most cases the quantity T_n can be written as

$$T_n = \frac{\hat{\phi}(b_n) - \hat{\phi}(a_n)}{2} \ ,$$

where $a_n = G_0^{-1}(1/n)$ and $b_n = G_0^{-1}(1 - 1/n)$ are the lower and upper $1/n$ quantiles of the true distribution, and $\hat{\phi}(\cdot)$ is the maximum likelihood estimate of the variance-stabilised transformed parameter based on a single observation. The numerator is basically the 'expected sample range,' transformed onto the scale of the variance-stabilised parameter. Since $T_n \sim \sqrt{2 \log n}$ when $G_0 = \mathcal{N}(0, 1)$, this indeed reduces to the Bickel & Chernoff (1993) result.

The final assembly stage just shows that the limit theorem for $\widetilde{M}_n$ also applies to M_n, as well as to the maxima of the score and efficient score processes from the simple- and composite-hypothesis versions of the original testing problem. The steps there follow closely the corresponding steps in Bickel & Chernoff (1993).

There are two sources of error in using the limit theorem to approximate the finite-sample distribution:

1. the error in approximating the distribution of the maximum of the empirical score process with the distribution of the maximum of a Gaussian process;

2. the error in approximating the distribution of the maximum of a Gaussian process with its limit distribution.

It is well known that the rate of convergence of the maxima of Gaussian processes to their limiting distribution is very slow, so the second source of error above is likely to be appreciable. However errors due to the first source may be small. We use computer simulations to show that the sampling distribution of the maximum of the *Studentised* version of the score process, is very close to that of the maximum of a certain Gaussian process. A test which simulates Gaussian processes to compute approximate p-values is thus very accurate while using much less computing resources than a straight Monte-Carlo approximation.

2 Framework and Notation

We now discuss the general framework and define some notation that is used throughout what follows. We give a general discussion of testing for homogeneity for mixtures of parametric families, where the parameter is possibly vector valued, even though our main results that follow pertain only to one-parameter families.

Let λ be a fixed σ-finite measure on $(\mathbb{R}^k, \mathcal{B}^k)$, where $\mathbb{R}$ is the real line and $\mathcal{B}$ the Borel σ-field, for some fixed integer $k \geq 1$. Usually λ is either Lebesgue measure on (possibly a subset of) $\mathbb{R}^k$ or counting measure on $\{0, 1, 2\ldots\}^k$. Let $\mathcal{F}_\lambda$ denote the set of all probability distributions on $(\mathbb{R}^k, \mathcal{B}^k)$ dominated by λ. Let $\mathcal{F} = \{F_\theta \,|\, \theta \in \Theta\} \subset \mathcal{F}_\lambda$ denote a parametric family of such distributions, with $\Theta \subset \mathbb{R}^d$. For any measurable function $g\colon (\mathbb{R}^k, \mathcal{B}^k) \to (\mathbb{R}, \mathcal{B})$, we write $F_\theta g = \int g\, dF_\theta$ for the expectation. For the n-fold product measure of F_θ on $(\mathbb{R}^{kn}, \mathcal{B}^{kn})$ we write P_θ. For any measurable function $X\colon (\mathbb{R}^{kn}, \mathcal{B}^{kn}) \to (\mathbb{R}, \mathcal{B})$, we again write $P_\theta X = \int X\, dP_\theta$ for the expectation. For any $B \in \mathcal{B}$, $\{X \in B\}$ is short for the inverse image $\big\{\omega \in \mathbb{R}^{kn} \,|\, X(\omega) \in B\big\} \in \mathcal{B}^{kn}$. For any $A \in \mathcal{B}^{kn}$, $1_A\colon \mathbb{R}^{kn} \to \{0, 1\}$ is an indicator function: $1_A(\omega) = 1$ if $\omega \in A$, 0 otherwise. For any $B \in \mathcal{B}$ write $P_\theta(X \in B)$ as shorthand for $P_\theta 1_{\{X \in B\}}$.

Given any $0 \leq p \leq 1$ and any two elements $\eta, \theta \in \Theta$ we can consider the *mixture distribution*

$$G(p, \eta, \theta) = (1 - p)F_\eta + pF_\theta \ .$$

We refer to the values η and θ as the *components* of the mixture, and p as the *mixing proportion*. We refer to $\mathcal{F}$ as the *generating family*. We are essentially interested in testing the hypothesis of one component versus the alternative of two components, that is a *test of homogeneity*. On the surface this appears to be a usual parametric testing problem. However, beneath this deceptive exterior lies a complicated testing problem.

The central issue is one of *identifiability*. Suppose $\mathcal{G} = \{G(\gamma) \,|\, \gamma \in \Gamma\}$ is a general parametric family of distributions. A particular distribution $G_0 \in \mathcal{G}$ is said to be identifiable under the parametrisation $\gamma \mapsto G(\gamma)$ if there is only one value $\gamma_0 \in \Gamma$ such that $G_0 = G(\gamma_0)$, that is for any other $\gamma \neq \gamma_0$ in the parameter space Γ, $G(\gamma) \neq G_0$.

Referring to the mixture distribution $G(p, \eta, \theta)$ above, the single distribution F_β may be represented as

$$\begin{aligned} F_\beta &= G(0,\beta,\theta) \text{ for any } \theta, \\ &= G(1,\eta,\beta) \text{ for any } \eta \text{ or} \\ &= G(p,\beta,\beta) \text{ for any } p. \end{aligned}$$

So in the Euclidean representation of this mixture model, the single distribution F_β is represented as the set

$$S_0(\beta) = \{(0,\beta,\theta) \,|\, \theta \in \Theta\} \cup \{(1,\eta,\beta) \,|\, \eta \in \Theta\} \cup \{(p,\beta,\beta) \,|\, 0 \le p \le 1\} \,. \quad (1)$$

So specifying the hypothesis in terms of the parameters is unusual. A test of homogeneity is testing $p = 0$ or $p = 1$ or $\eta = \theta$. This causes a problem when we come to consider hypothesis testing asymptotics. The usual approach is as follows: expand the log-likelihood ratio as a function of the parameters in a Taylor series about the true value and evaluate it at the maximum likelihood estimates.

$$L_n(\hat{p},\hat{\eta},\hat{\theta}) = L_n(p_0,\eta_0,\theta_0) + \begin{pmatrix} \hat{p}-p_0 \\ \hat{\eta}-\eta_0 \\ \hat{\theta}-\theta_0 \end{pmatrix}^T L_n'(p_0,\eta_0,\theta_0) + \cdots$$

This relies on the fact that the three differences above all get small. However, what are the true values (p_0,η_0,θ_0) when there is only a single component? Also, when F_β is the true distribution, how do the maximum likelihood estimates perform? Does $\hat{p} \xrightarrow{P} 0$? Does $\hat{\eta}$ or $\hat{\theta} \xrightarrow{P} \beta$? In fact Redner (1981) showed that $(\hat{p},\hat{\eta},\hat{\theta})$ is eventually in a neighbourhood of the set $S_0(\beta)$ in (1) above. However, this still does not answer all of our questions.

The reason why there is a problem is because the Euclidean representation does not capture the true geometry of the model. For example, in a nice model like $\mathcal{A} = \{\mathcal{N}(\mu,\sigma^2) \,|\, \mu \in \mathbb{R}, \sigma > 0\}$, the Euclidean representation $\Theta = \mathbb{R} \times (0,\infty)$ actually does capture the geometry of the normal location-scale model. There are really only two 'dimensions' to the problem, we could only move away from, say $\mathcal{N}(0,1)$ toward any of the other distributions in $\mathcal{A}$ in one of two 'orthogonal' directions, that is by changing the mean and/or the variance. And for two distributions $\mathcal{N}(\mu_1,\sigma_1^2)$ and $\mathcal{N}(\mu_2,\sigma_2^2)$ in $\mathcal{A}$ that are 'close together', the Euclidean distance between the points (μ_1,σ_1) and (μ_2,σ_2) is also small. However, if we consider our mixture model

$$\mathcal{G} = \{G(p,\eta,\theta) = (1-p)F_\eta + pF_\theta \,|\, 0 \le p \le 1\}$$

it is clear that we can find distributions that are obviously 'close together' but such that the Euclidean distance between their parameter vectors in the parameter space is large. Indeed there are many distant points corresponding to the *same* distribution.

We could consider a reparametrisation, replacing p with $\pi = p|\eta - \theta|$, or indeed anything of the form

$$\pi = p\, d(F_\eta, F_\theta) \ ,$$

where $d(\,\cdot\,,\,\cdot\,)$ is any measure of distance between distributions; in particular it must equal zero if and only if the two distributions are the same. Then $\pi = 0$ if and only if there is one component. Then we would have

$$G^*(\pi, \eta, \theta) = \left(1 - \frac{\pi}{d(F_\eta, F_\theta)}\right) F_\eta + \frac{\pi}{d(F_\eta, F_\theta)} F_\theta$$

and the model can now be expressed as

$$\mathcal{G} = \{G^*(\pi, \eta, \theta) \,|\, 0 \le \pi \le d(F_\eta, F_\theta), \eta, \theta \in \Theta\} \ .$$

We now have that $G^*(\pi, \eta, \theta) = F_\beta$ implies $\pi = 0$ and $\eta = \beta$. However θ is still a problem. Indeed if F_β is the true value, then the expansion of the log-likelihood would look like

$$L_n(\hat\pi, \hat\eta, \hat\theta) = L_n(0, \beta, \theta_0) + (\hat\pi, \hat\eta - \beta, \hat\theta - \theta_0) L_n'(0, \beta, \theta_0) + \cdots$$

So then presumably, we would have $\hat\pi \xrightarrow{P} 0$, $\hat\eta \xrightarrow{P} \beta$, that is, the parameters could be consistently estimated. However, it seems that θ cannot be consistently estimated, since it vanishes when the hypothesis is true. We call it then a *vanishing parameter*.

Before we proceed, we refine slightly the reparametrisation π. Note that now the parameter space is of the form

$$\{(\pi, \eta, \theta) \,|\, 0 \le \pi \le d(F_\eta, F_\theta), \eta, \theta \in \Theta\} \ ,$$

which depends on the choice of distance chosen. We can get around this by noticing that we are only interested in testing $\pi = 0$, so we can replace it with something which is like π for small π, but does not have an upper bound that depends on the distance $d(\,\cdot\,,\,\cdot\,)$. One such choice is

$$r(p, \eta, \theta) = -\log(1 - p) d(F_\eta, F_\theta) \ .$$

This then gives us a mixture model with only one vanishing parameter with a nice parameter space as follows:

$$\mathcal{M} = \{M_d(r,\eta,\theta)\,|\,r \geq 0\,;\,\eta,\theta \in \Theta\}\ ,$$

where

$$M_d(r,\eta,\theta) \quad = \quad [1 - p(r,\eta,\theta)]F_\eta + p(r,\eta,\theta)F_\theta$$

and

$$p(r,\eta,\theta) \quad = \quad 1 - e^{-r/d(F_\eta,f_\theta)}\ .$$

We use this notation because this parametrisation is much like a *polar co-ordinate system*; η is like an origin, r is like a radius and θ is like an angle, or direction. When the radius parameter $r = 0$, the direction parameter θ vanishes, and we are just at the origin η. Otherwise, we are a radial distance r away from η in the direction θ.

We consider both simple and composite versions of the test for homogeneity. The composite version is just a test that $r = 0$ against the alternative that $r > 0$, where both components are regarded as unknown nuisance parameters. This can also be phrased as follows: suppose we have observations modelled as independent and identically distributed random variables $X_1, X_2, \ldots, X_n$ with common distribution F, assumed to be a member of $\mathcal{M}$. Then we are testing that $F \in \mathcal{F}$.

The simple test of homogeneity is a test of $r = 0$ against the alternative that $r > 0$, where the contaminating component θ is regarded as an unknown nuisance parameter, but the component 'being contaminated', $\eta = \eta_0$, is regarded as known. We can also phrase this by assuming that F is a member of $\mathcal{M}_{\eta_0} = \{M_d(r,\eta_0,\theta)\,|\,r \geq 0, \theta \in \Theta\}$, where $\eta_0 \in \Theta$ is regarded as fixed and known. Then we are testing that $F = F_{\eta_0}$. This test is often described in the literature as the test of homogeneity in the *contamination model*. Note that in both the simple and composite tests of homogeneity, the contaminating component θ vanishes under the hypothesis.

Much of the literature addresses the problem of deriving the asymptotic distribution of the log-likelihood ratio statistic for both the simple and composite (full) problems. We take the following approach. For each fixed θ_0 we identify random variables $S_n^\dagger(\theta_0)$ (simple sub-problem) and $S_n^*(\theta_0)$ (composite sub-problem) that are asymptotically equivalent to various optimal test statistics (including the likelihood-ratio statistic) in each sub-problem. These then become stochastic processes $\left\{S_n^\dagger(\theta) \mid \theta \in \Theta\right\}$ and $\{S_n^*(\theta) \mid \theta \in \Theta\}$ in the setting of the (full) simple and composite problems where θ is allowed to vary. The test statistics we consider for the (full) simple and composite problems are the maxima $M_n^\dagger = \sup\left\{S_n^\dagger(\theta) \mid \theta \in \Theta\right\}$ and $M_n^* = \sup\{S_n^*(\theta) \mid \theta \in \Theta\}$ respectively. Note that these maxima may or may not be asymptotically equivalent to the likelihood ratio statistics for the full problems (although see Liu & Shao (2001)).

3 The four testing problems

We now describe four different versions of the test of homogeneity in a two-component mixture model. The four versions have different nuisance parameters known and unknown.

In all cases we start with a parameteric family $\mathcal{F} = \{F_\theta \,|\, \theta \in \Theta\}$ for some Euclidean parameter space $\Theta \in \mathbb{R}^d$ for some positive integer d. Later we restrict attention to the case $d = 1$, but for the moment keep it general.

We define the mixture distribution

$$M(r, \eta, \theta) = [1 - p(r, \eta, \theta)]F_\eta + p(r, \eta, \theta)F_\theta \ ,$$

where

$$p(r, \eta, \theta) = 1 - \exp\{-r/d(F_\eta, F_\theta)\}$$

and $d(\,\cdot\,,\,\cdot\,)$ is a metric defined on $\mathcal{F}$ (we see later that the choice of metric is relatively unimportant, as long as it is a true metric, in that it equals zero only if the two distributions are the same and it is smoothly differentiable in the parameters η and θ).

We assume that we have $X_1, X_2, \ldots, X_n$ independent and identically distributed random variables with common distribution

1. $M(r, \eta_0, \theta_0)$ for some unknown $r \geq 0$ and known $\eta_0, \theta_0 \in \Theta$ (the simple sub-problem);

2. $M(r, \eta, \theta_0)$ for some unknown $r \geq 0$, some unknown $\eta \in \Theta$ and known $\theta_0 \in \Theta$ (the composite sub-problem);

3. $M(r, \eta_0, \theta)$ for some unknown $r \geq 0$, known $\eta_0 \in \Theta$ and some unknown $\theta \in \Theta$ (the full simple problem);

4. $M(r, \eta, \theta)$ for some unknown $r \geq 0, \eta, \theta \in \Theta$ (the full composite problem).

In each case we are testing the hypothesis that $r = 0$ (against the alternative that $r > 0$), that is a test of one component against two components; a test of homogeneity.

The first two sub-problems are studied as preludes to their full versions. The third problem has most of the technical difficulty of the fourth problem without the added complication of the extra nuisance parameter.

We define here the following quantities which are referred to below. The densities are $f_\theta = dF_\theta/d\lambda$, $m(r,\eta,\theta) = dM(r,\eta,\theta)/d\lambda$. The log-likelihood is

$$L_n(r,\eta,\theta) = \sum_{i=1}^{n} \log m(r,\eta,\theta)(X_i) \ .$$

When $r = 0$ or $\eta = \theta$ we sometimes write $L_n(0,\eta,\theta)$ or $L_n(r,\eta,\eta)$ as just $L_n(\eta) = \sum_{i=1}^{n} \log f_\eta(X_i)$.

Define

$$\begin{aligned}
\dot{\ell}(\,\cdot\,;\eta,\theta) &= \left.\frac{\partial \log m(r,\eta,\theta)}{\partial r}\right|_{r=0} , \\
&= f_\theta/f_\eta - 1 \ , \\
v(\eta,\theta) &= F_\eta\left(\dot{\ell}(\,\cdot\,;\eta,\theta)^2\right) = F_\eta\left(f_\theta^2/f_\eta^2\right) - 1
\end{aligned}$$

Also define the partial derivatives $f'_{\eta j} = \partial f_\theta/\partial\theta_j|_{\theta=\eta}$ for $j = 1,2,\ldots,d$ and define the j-th element of the d-vector $u(\eta)$ as $f'_{\eta j}/f_\eta$. Finally define $C(\eta)$ as the d-vector with j-th element

$$F_\eta\left[\dot{\ell}(\,\cdot\,;\eta,\theta) f'_{\eta j}/f_\eta\right] = F_\eta\left[f_\theta f'_{\eta j}/f_\eta^2\right]$$

and $I(\eta)$ as the d-by-d matrix with (i,j)-th element

$$F_\eta\left(f'_{\eta i} f'_{\eta j}/f_\eta^2\right) \ ,$$

for $i,j = 1,2,\ldots,d$.

3.1 The simple sub-problem

Since both η_0 and θ_0 are fixed, r only acts through the function $p(r,\eta_0,\theta_0)$. Also, $M(0,\eta_0,\theta_0) = F_{\eta_0}$ for each θ_0. So to test $r = 0$ we may as well regard $r = r(p)$ and write

$$M(r(p),\eta_0,\theta_0) = (1-p)F_{\eta_0} + pF_{\theta_0} \ .$$

The test statistic we consider is the standardised (Rao-simple-hypothesis) score statistic

$$S_n^\dagger(\eta_0,\theta_0) = [nv(\eta_0,\theta_0)]^{-1/2} \sum_{i=1}^{n} \dot{\ell}(X_i;\eta_0,\theta_0) \ .$$

We have immediately that the Studentised score statistic satisfies

$$\begin{aligned} T_n^\dagger(\eta_0, \theta_0) &= \frac{\sum_{i=1}^n \dot\ell(X_i; \eta_0, \theta_0)}{\sqrt{\sum_{i=1}^n \dot\ell(X_i; \eta_0, \theta_0)^2}} \\ &= S_n^\dagger(\eta_0, \theta_0) + o_p(1) \ . \end{aligned}$$

If certain regularity conditions hold we can write the log-likelihood ratio as

$$\Lambda_n(\eta_0, \theta_0) = \sup_p L_n(r(p), \eta_0, \theta_0) - L_n(\eta_0) \quad = \quad \frac{1}{2}\left\{0 \vee S_n^\dagger(\eta_0, \theta_0)\right\}^2 + o_p(1)$$

(see for example Hartigan (1985) or Dacunha-Castelle & Gassiat (1997) for a derivation).

The statistics $S_n^\dagger(\eta_0, \theta_0)$, $T_n^\dagger(\eta_0, \theta_0)$ and $\Lambda_n(\eta_0, \theta_0)$ each have various optimality properties. In particular, under a local asymptotic normality condition any statistic differing from $S_n^\dagger(\eta_0, \theta_0)$ by a quantity tending to zero in probability is (locally) asymptotically most powerful in the sense of Choi *et al.* (1996).

3.2 The composite sub-problem

The random variable we consider is the *efficient score* given by

$$S_n^*(\eta_0, \theta_0) = [nv^*(\eta_0, \theta_0)]^{-1/2} \sum_{i=1}^n \dot\ell^*(X_i; \eta_0, \theta_0) \ ,$$

where $\dot\ell^*(\,\cdot\,; \eta, \theta) = \dot\ell(\,\cdot\,; \eta, \theta) - C(\eta, \theta)^T I(\eta) u(\eta)$ and $v^*(\eta, \theta) = v(\eta, \theta) - C(\eta)^T I(\eta)^{-1} C(\eta)$.

Since the efficient score depends on the unknown value of the nuisance parameter, this is not a statistic; but under a local asymptotic normality condition any test statistic that differs from the efficient score by a quantity tending to zero in probability is (locally) *asymptotically uniformly most powerful* in the sense of Choi *et al.* (1996).

Under regularity conditions it can be shown that the (composite-hypothesis) Rao score statistic $S_n^\dagger(\hat\eta_0)$, where

$$\hat\eta_0 = \arg\max_{\eta \in \Theta} \sum_{i=1}^n \log f_\eta(X_i)$$

is the maximum likelihood estimate under the hypothesis, satisfies

$$S_n^\dagger(\hat{\eta}_0, \theta_0) \quad = \quad S_n^*(\eta_0, \theta_0) + o_p(1) \ .$$

In fact $\hat{\eta}_0$ can be replaced with any any *asymptotically efficient* estimator of η_0 under the hypothesis $r = 0$, that is any estimator $\widetilde{\eta}_0$ such that $n^{1/2}(\widetilde{\eta}_0 - \eta_0]) \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, I(\eta_0)^{-1})$ when the true distribution is F_{η_0}. A derivation of this follows closely the proof of proposition 4(i), section 4 of Hall & Mathiason (1990). It also follows by continuity that the corresponding Studentised score statistic $T_n^\dagger(\hat{\eta}_0, \theta_0) = S_n^*(\eta_0, \theta_0) + o_p(1)$.

Under regularity conditions it can also be shown that the log-likelihood ratio for the composite sub-problem

$$\Lambda_n^*(\theta_0) = \sup_{r \geq 0, \eta \in \Theta} L_n(r, \eta, \theta_0) - \sup_{\eta \in \Theta} L_n(0, \eta, \theta_0)$$

satisfies

$$\Lambda_n^*(\theta_0) = \frac{1}{2} \{0 \vee S_n^*(\eta_0, \theta_0)\}^2 + o_p(1)$$

(see for example Ghosh & Sen (1985) or Dacunha-Castelle & Gassiat (1997) for a derivation).

3.3 The full simple problem

We consider the test statistic

$$M_n^\dagger(\eta_0) = \sup_{\theta \in \Theta} S_n^\dagger(\eta_0, \theta) \ ,$$

when the true distribution of the X_i's is F_{η_0}. We give some motivation for this based on earlier discussion of the simple sub-problem.

The log-likelihood ratio statistic for the simple sub-problem where θ_0 is fixed satisfies

$$\Lambda_n(\eta_0, \theta_0) = \frac{1}{2} \left\{0 \vee S_n^\dagger(\eta_0, \theta_0)\right\}^2 + R_n^\dagger(\eta_0, \theta_0) \ ,$$

where $S_n^\dagger(\eta_0, \theta_0)$ is the standardised score statistic for the simple sub-problem and

$$P_{\eta_0}\left(\left|R_n^\dagger(\eta_0, \theta_0)\right| > \varepsilon\right) \to 0$$

for any $\varepsilon > 0$ as $n \to \infty$.

The log-likelihood ratio for the *full* simple problem is exactly

$$\begin{aligned}\Lambda_n^\dagger(\eta_0) &= \sup_{\theta\in\Theta} \Lambda_n^\dagger(\eta_0, \theta) \\ &= \sup_{\theta\in\Theta} \left[\frac{1}{2} \left\{ 0 \vee S_n^\dagger(\eta_0, \theta) \right\}^2 + R_n^\dagger(\eta_0, \theta) \right] .\end{aligned}$$

If

$$P_{\eta_0} \left(\sup_{\theta\in\Theta} \left| R_n^\dagger(\eta_0, \theta_0) \right| > \varepsilon \right) \to 0 \tag{2}$$

for any $\varepsilon > 0$ as $n \to \infty$, then we can say

$$\Lambda_n^\dagger(\eta_0) = \frac{1}{2} \left\{ 0 \vee M_n^\dagger \right\}^2 + o_p(1) \ ,$$

where $M_n^\dagger = \sup_{\theta\in\Theta} S_n^\dagger(\theta)$. The condition (2) is difficult to show, but an unpublished (at the time of writing) manuscript (Liu & Shao (2001)) is claimed to have a proof of this in the case where $\mathcal{F} = \{\mathcal{N}(\theta, 1) \,|\, \theta \in \mathbb{R}\}$.

Whether (2) holds or not, the statistic $M_n^\dagger$ is worth studying in its own right. A large value of $S_n^\dagger(\theta)$ is evidence against the pure distribution F_{η_0} and suggests a contamination from the distribution F_θ. Also $S_n^\dagger(\theta)$ has the same asymptotic distribution for each θ. So if there is a suggestion of a (small) contamination of any kind, $S_n^\dagger(\theta)$ should be large for some θ, making the maximum $M_n^\dagger$ large.

3.4 The full composite problem

For the full composite testing problem we study the asymptotic distribution of the random variable

$$M_n^*(\eta_0) = \sup_{\theta\in\Theta} S_n^*(\eta_0, \theta) \ ,$$

when the true distribution of the X_i's is F_{η_0}.

The log-likelihood ratio for the composite sub-problem where θ_0 is fixed satisfies

$$\Lambda^*(\theta_0) = \frac{1}{2} \left\{ 0 \vee S_n^*(\theta_0, \eta_0) \right\}^2 + R_n^*(\theta_0, \eta_0) \ ,$$

where $S_n^*(\theta_0, \eta_0)$ is the efficient score for the composite sub-problem and for each $\varepsilon > 0$, as $n \to \infty$,

$$P_{\eta_0}\left(|R_n^*(\theta_0, \eta_0)| > \varepsilon\right) \to 0 \ .$$

The log-likelihood ratio for the full composite problem is

$$\begin{aligned} \Lambda_n^* &= \sup_{\theta \in \Theta} \Lambda_n^*(\theta) \\ &= \sup_{\theta \in \Theta} \left[\frac{1}{2} \left\{0 \vee S_n^*(\theta, \eta_0)\right\}^2 + R_n^*(\theta_0, \eta_0) \right] \ . \end{aligned}$$

If for any $\varepsilon > 0$ we have

$$P_{\eta_0}\left(\sup_{\theta \in \Theta} |R_n^*(\eta_0, \theta_0)| > \varepsilon \right) \to 0 \tag{3}$$

as $n \to \infty$, then we can say

$$\Lambda_n^* \ = \ \frac{1}{2} \left\{0 \vee M_n^*\right\}^2 + o_p(1) \ ,$$

where $M_n^* = \sup_{\theta \in \Theta} S_n^*(\theta)$. Again, the relation (3) is difficult to show, but if the Liu & Shao (2001) result is true, then it should be able to be extended to the composite case with little difficulty. Just as for $M_n^\dagger$ in the full simple problem, the statistic M_n^* is worth studying in its own right, regardless of whether (3) holds or not.

4 Geometric Interpretation

We now present a geometric interpretation of the problem, based on the presentation given in Lindsay (1995), culminating in the definition of a certain property of a one-parameter family of standardised functions, the *arc length*. We firstly set up the geometric framework.

We are interested in the behaviour of statistics of the form

$$M_n = \sup_{\theta\in\Theta} S_n(\theta) \quad \text{and} \quad \sup_{\theta\in\Theta} \{0 \vee S_n(\theta)\} = \{0 \vee M_n\} \ ,$$

where

$$S_n(\theta) = n^{-1/2}\sum_{i=1}^{n} s_\theta(X_i) \ ,$$

$X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables with common distribution F and $\mathcal{S} = \{s_\theta \,|\, \theta \in \Theta\}$ is a one-parameter family of *standardised random variables*, that is functions satisfying $Fs_\theta = 0$ and $Fs_\theta^2 = 1$ for all $\theta \in \Theta \subset \mathbb{R}$. We restrict attention to parameter sets Θ that are intervals.

In particular,

$$\mathcal{S} \subset L_2(F) = \left\{a \,\middle|\, Fa^2 < \infty\right\} \ ,$$

which is an inner-product space, with inner-product $\langle a, b\rangle = Fab$ and norm $\|a\| = \langle a, a\rangle^{1/2} = \{Fa^2\}^{1/2}$. Since the members of $\mathcal{S}$ have mean zero and variance one, $\langle s_\theta, s_\eta\rangle$ gives the correlation of between $s_\theta(X)$ and $s_\eta(X)$ when $X \sim F$.

4.1 Finite dimensional case

Suppose for a moment that $\mathcal{S}$ lies in a d-dimensional linear subspace of $L_2(F)$, for some positive integer d; this occurs for *any* $\mathcal{S}$ if the sample space $\mathcal{X}$ is finite, for example when $\mathcal{X} = \{0, 1, 2, \ldots, d\}$. Then there exists an orthonormal basis $b_1, b_2, \ldots, b_d$ satisfying $\langle b_i, b_j\rangle = 1\{i = j\}$, such that for each $\theta \in \Theta$, and each $x \in \mathcal{X}$,

$$s_\theta(x) = \sum_{j=1}^{d} \langle s_\theta, b_j\rangle b_j(x) \ .$$

In particular, each function s_θ can be identified with a point $(s_{\theta 1}, s_{\theta 2}, \ldots, s_{\theta d}) \in \mathbb{R}^d$, where each $s_{\theta j} = \langle s_\theta, b_j \rangle$.

Moreover,

$$
\begin{aligned}
\langle s_\eta, s_\theta \rangle &= F s_\eta s_\theta \\
&= F \left(\sum_{i=1}^{d} s_{\eta i} b_i \right) \left(\sum_{j=1}^{d} s_{\theta j} b_j \right) \\
&= \sum_{i=1}^{d} \sum_{j=1}^{d} s_{\eta i} s_{\theta j} \langle b_i, b_j \rangle \\
&= \sum_{j=1}^{d} s_{\eta j} s_{\theta j} \ ,
\end{aligned}
$$

so the $L_2(F)$ inner product of the functions s_η and s_θ is given by the usual Euclidean dot-product of the corresponding vectors in $\mathbb{R}^d$. Also note that since $\|s_\theta\| = \sum_{j=1}^{d} s_{\theta j}^2 = 1$ for all $\theta \in \Theta$, $\mathcal{S}$ can (so long as it is "smooth" in some sense) be identified with a *curve* on the unit sphere in $\mathbb{R}^d$.

Now we can write

$$
\begin{aligned}
S_n(\theta) &= n^{-1/2} \sum_{i=1}^{n} s_\theta(X_i) \\
&= n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{d} s_{\theta j} b_j(X_i) \\
&= \sum_{j=1}^{d} s_{\theta j} \left(n^{-1/2} \sum_{i=1}^{n} b_j(X_i) \right) \\
&= \sum_{j=1}^{d} s_{\theta j} B_{nj} = \langle s_\theta, B_n \rangle \ ,
\end{aligned}
$$

where $(B_{n1}, B_{n2}, \ldots, B_{nd})$ is a random point in $\mathbb{R}^d$ corresponding to the random function given by

$$
x \mapsto B_n(x) = \sum_{j=1}^{d} B_{nj} b_j(x) \ .
$$

So $n^{1/2}B_n$ is a random d-vector whose components are sums of independent and identically distributed standardised, uncorrelated random variables. So $B_n \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, I_d)$, the d-dimensional Normal distribution with zero mean and covariance matrix equal to the d-by-d identity matrix.

The projection of B_n onto the line $\mathcal{T}_\theta = \{ts_\theta \mid t \in \mathbb{R}\}$ passing through the origin and the point corresponding to s_θ is given by

$$\begin{aligned} \text{Proj}\,(B_n|\mathcal{T}_\theta) &= \arg\min_{v\in\mathcal{T}_\theta} \|B_n - v\| \\ &= \langle s_\theta, B_n\rangle s_\theta \ . \end{aligned}$$

The projection of B_n onto the *positive ray* $\mathcal{C}_\theta = \{cs_\theta \mid c \geq 0\}$ is given by

$$\text{Proj}\,(B_n|\mathcal{C}_\theta) = \{0 \vee \langle s_\theta, B_n\rangle\}\, s_\theta \ .$$

Thus the *length* of this projection is just $\{0 \vee \langle s_\theta, B_n\rangle\} = \{0 \vee S_n(\theta)\}$. Define now the *tangent cone* as

$$\mathcal{C} = \{cs_\theta \mid c \geq 0, \theta \in \Theta\} = \cup_{\theta\in\Theta}\mathcal{C}_\theta \ .$$

Then (assuming that $\mathcal{C}$ is closed) we have that

$$\text{Proj}\,(B_n|\mathcal{C}) = \arg\min_{v\in\mathcal{C}} \|B_n - v\| \ .$$

We can write

$$\begin{aligned} \min_{v\in\mathcal{C}} \|B_n - v\| &= \min_{\theta\in\Theta}\left(\min_c \|B_n - cs_\theta\|\right) \\ &= \min_{\theta\in\Theta} \|B_n - \text{Proj}\,(B_n|\mathcal{C}_\theta)\| \end{aligned}$$

Now for each θ, $\text{Proj}\,(B_n|\mathcal{C}_\theta)$ is either the zero vector or is orthogonal to B_n so we always have $\|\text{Proj}\,(B_n|\mathcal{C}_\theta)\|^2 + \|B_n - \text{Proj}\,(B_n|\mathcal{C}_\theta)\|^2 = \|B_n\|^2$. So minimising $\|B_n - \text{Proj}\,(B_n|\mathcal{C}_\theta)\|$ over θ is equivalent to maximising $\|\text{Proj}\,(B_n|\mathcal{C}_\theta)\|$, and so the projection onto the whole tangent cone $\text{Proj}\,(B_n|\mathcal{C})$ is just the 'longest' of the projections $\{\text{Proj}\,(B_n|\mathcal{C}_\theta) \mid \theta \in \Theta\}$, and

$$\begin{aligned} \|\text{Proj}\,(B_n|\mathcal{C})\| &= \max_{\theta\in\Theta} \|\text{Proj}\,(B_n|\mathcal{C}_\theta)\| \\ &= \max_{\theta\in\Theta} \{0 \vee S_n(\theta)\} \\ &= \{0 \vee M_n\} \ . \end{aligned}$$

Thus the asymptotic distribution of $\{0 \vee M_n\}$ is exactly the distribution of the length of the projection of a random point B_0 with distribution $\mathcal{N}_d(0, I_d)$ onto the tangent cone $\mathcal{C}$. This is exactly the distribution of $\{0 \vee M_0\}$, where

$$M_0 = \sup\{S_0(\theta) = \langle s_\theta, B_0\rangle \,|\, \theta \in \Theta\} \ ,$$

is the supremum of a standardised Gaussian process indexed by Θ with correlation function $\rho(\theta, \eta) = \langle s_\theta, s_\eta \rangle$.

A geometric treatment of this situation is given in Lindsay (1995). An important geometric property described in the development is the *arc length* $A(\mathcal{S})$ of the curve created where the tangent cone $\mathcal{C}$ intersects the unit sphere in $\mathbb{R}^d$. The arc length $A(\mathcal{S})$ is shown to be equal to the integral

$$A(\mathcal{S}) = \int_\Theta \sqrt{F\left(\frac{\partial s_\theta}{\partial \theta}\right)^2} \, d\theta \ .$$

It is shown that for the special case where the tangent set lies in a plane, that

$$P\left(M_0^2 \geq t\right) = \frac{1}{2} P(\chi_1^2 \geq t) + \frac{A(\mathcal{S})}{2\pi} P(\chi_2^2 \geq t) \ .$$

Then it is explained that in the general case, where the tangent set is not restricted to the plane, that the above equality becomes an inequality, with the left hand side is less than or equal to the right hand side. The inequality is sharp in the sense that the ratio of the two sides tends to 1 as $t \to \infty$.

4.2 The infinite dimensional case

Now consider the situation where $\mathcal{S} = \{s_\theta \,|\, \theta \in \Theta\}$ is not necessarily contained in a d-dimensional linear subspace of $L_2(F)$. In this case we identify each function s_θ with the *sequence* $s_{\theta 1}, s_{\theta 2}, \ldots$, where $s_{\theta j} = \langle s_\theta, b_j \rangle$ and $b_1, b_2, \ldots$ is an orthonormal basis for all of $L_2(F)$. So each s_θ is now identified with a point on the unit sphere in $\mathbb{R}^\infty$, or more precisely with a point on the unit sphere in

$$\ell_2 = \left\{ (a_1, a_2, \ldots) \in \mathbb{R}^\infty \,\middle|\, \sum_j a_j^2 < \infty \right\} ,$$

the space of all square-summable real sequences. As before, the $L_2(F)$ inner product can be written as

$$\langle s_\eta, s_\theta \rangle = \sum_j s_{\eta j} s_{\theta j} \ ,$$

the ℓ_2 dot-product. Also we can again write

$$\begin{aligned} S_n(\theta) &= n^{-1/2}\sum_{i=1}^{n} s_\theta(X_i) \\ &= n^{-1/2}\sum_{i=1}^{n}\sum_j s_{\theta j} b_j(X_i) \\ &= \sum_j s_{\theta j}\left(n^{-1/2}\sum_{i=1}^{n} b_j(X_i)\right) \\ &= \sum_j s_{\theta j} B_{nj} \ . \end{aligned}$$

However we need to be careful how we interpret this. Naïvely we could interpret this as the ℓ_2 dot-product of s_θ and a "random sequence"

$$B_n = (B_{n1}, B_{n2}, \ldots) \ ,$$

but this is not possible because B_n is not an ℓ_2-sequence. Although for any other $L_2(F)$ function a with ℓ_2 sequence $a_1, a_2, \ldots$, the "dot-product"

$$\sum_j a_j B_{nj} = n^{1/2}\sum_{i=1}^{n} a(X_i)$$

is well defined, it would appear that the "norm" of B_n

$$\begin{aligned} \|B_n\|^2 &= \sum_j B_{nj}^2 \\ &= \sum_j \left\{ n^{1/2}\sum_{i=1}^{n} b_j(X_i) \right\}^2 \end{aligned}$$

is infinite, as it is the sum of squares of an infinite sequence of uncorrelated (although highly dependent) standardised random variables.

We can, however, still interpret this as a random point in $\mathbb{R}^\infty$ and think of the length of its "projection" onto the tangent cone

$$\mathcal{C} = \{cs_\theta \,|\, c \geq 0, \theta \in \Theta\}$$

as we do in the finite case, since we can always define this as just $\sup_\theta \{0 \vee S_n(\theta)\} = \{0 \vee M_n\}$.

We can also consider the arc length $A(\mathcal{S})$, defined in exactly the same way as in the finite case. It may be that results similar to that of Lindsay (1995) would carry over to the infinite-dimensional case, so long as the arc length $A(\mathcal{S})$ is finite.

4.3 Infinite arc length

The methods that we develop in subsequent sections pertain only to cases where the arc length is *infinite*. In fact we can provide at this stage another interpretation of what it means for a one-parameter family of mean-zero, variance-one functions to have an infinite arc length.

Suppose that $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables with common distribution F. Suppose $\mathcal{S} = \{s_\theta \,|\, \theta \in \Theta\}$ is a one-parameter family of functions such that for all $\theta \in \Theta$, $Fs_\theta = 0$ and $Fs_\theta^2 = 1$. Define the empirical process $S_n(\theta) = n^{-1/2}\sum_{i=1}^n s_\theta(X_i)$ and consider the statistic

$$M_n = \sup_{\theta \in \Theta} S_n(\theta) \ .$$

Then under certain conditions we would expect

$$M_n \begin{cases} = O_p(1) & \text{if } A(\mathcal{S}) < \infty \ , \\ \xrightarrow{P} \infty & \text{if } A(\mathcal{S}) = \infty \ . \end{cases}$$

The heuristic for this is as follows. Define the transformation $\tau \colon \overline{\Theta} \to [0, A(\mathcal{S})]$ (where $\overline{\Theta}$ denotes the closure of the interval Θ) via

$$\tau(\theta) = \int_\Theta 1\{\eta \leq \theta\} \sqrt{F\left(\frac{\partial s_\eta}{\partial \eta}\right)^2} \, d\eta \ ,$$

and define θ_t implicity via $\tau(\theta_t) = t$. Then the stochastic process

$$\{Z_n(t) = S_n(\theta_t) \,|\, 0 \leq t \leq A(\mathcal{S})\}$$

has mean zero, variance one for each t. Further, it has a *standardised derivative*:

$$Z_n'(t) = \frac{\partial Z_n(t)}{\partial t} = n^{-1/2} \sum_{i=1}^{n} \frac{\partial s_{\theta_t}(X_i)}{\partial \theta}$$

$$= n^{-1/2} \sum_{i=1}^{n} \left. \frac{\partial s_\theta(X_i)}{\partial \theta} \right|_{\theta=\theta_t} \frac{d\theta_t}{dt} .$$

Now if we can differentiate inside the integral sign,

$$F\left(\frac{\partial s_\theta}{\partial \theta}\right) = \frac{\partial}{\partial \theta} F s_\theta = 0 ,$$

since $F s_\theta = 0$ for each θ. Implicit differentiation gives

$$\frac{d\theta_t}{dt} = \frac{1}{\tau'(\theta_t)}$$

$$= \frac{1}{\sqrt{F\left(\left.\frac{\partial s_\theta}{\partial \theta}\right|_{\theta=\theta_t}\right)^2}} .$$

So $Z_n'(t)$ has mean zero and variance

$$Var Z_n'(t) = Var\left(n^{-1/2} \sum_{i=1}^{n} \left. \frac{\partial s_\theta(X_i)}{\partial \theta} \right|_{\theta=\theta_t} \frac{d\theta_t}{dt}\right)$$

$$= \left(n^{-1/2} \frac{d\theta_t}{dt}\right)^2 n \, Var\left(\left. \frac{\partial s_\theta(X_1)}{\partial \theta} \right|_{\theta=\theta_t}\right)$$

$$= 1 .$$

If the arc length $A(\mathcal{S})$ is finite, and if the correlation structure of the process is nice enough, then we imagine that the maximum M_n should be like $M_0 = \sup_t Z_0(t)$, where $\{Z_0(t) \,|\, 0 \le t \le A(\mathcal{S})\}$ is a standardised, differentiable Gaussian process with standardised derivative with the same correlation structure as $\{Z_n(t) \,|\, 0 \le t \le A(\mathcal{S})\}$. Also M_0 should be bounded in probability.

However if the arc length is infinite, then the best we can hope for is that there is a sequence of slowly growing sub-intervals

$$\{\Theta_n \,|\, n \in \mathbb{N}\}$$

with the properties $\Theta_n \subset \Theta_{n+1}$ and $\cup_n \Theta_n = \Theta$, such that the restricted maximum

$$\widetilde{M}_n = \sup_{\theta \in \Theta_n} S_n(\theta)$$

behaves like the supremum of a standardised Gaussian process with standardised derivative over an interval of length $A(\mathcal{S}_n) \to \infty$, where $\mathcal{S}_n = \{s_\theta \,|\, \theta \in \Theta_n\}$. Then we would have $\widetilde{M}_n = O_p \left(\log A(\mathcal{S}_n)\right)^{1/2}$ (see subsection 7 or Hüsler (1990, 1995)). This would mean $M_n \geq \widetilde{M}_n \xrightarrow{P} \infty$. This in fact agrees with what we show in later sections.

5 Overview of results

5.1 Literature Review

We review relevant results on the full simple and composite testing problems.

An early paper by Hartigan (1985) examines the log-likelihood ratio statistic for the simple testing problem, where $\mathcal{F} = \{\mathcal{N}(\theta, 1) \,|\, \theta \in \mathbb{R}\}$. It is shown there that for each finite-dimensional sub-problem obtained by fixing θ, the log-likelihood ratio statistic

$$\Lambda_n(\theta) = \sup_p L_n(p, \theta)$$

is equivalent to the Studentised score statistic (the difference goes to zero in probability for each fixed θ). Then by examining the correlation structure of the Studentised score process, it is shown that the supremum of this process is not bounded in probability. It is in fact conjectured to be $O_p\left(\log\log n\right)^{1/2}$. This is exactly the type of behaviour we expect when the set of standardised score functions has *infinite arc length*.

Ghosh & Sen (1985) examine a general class of composite infinite-dimensional testing-for-homogeneity problems. Under restrictions, including a strong separation condition and that the parameter space Θ is a compact set, they show that the log-likelihood ratio statistic

$$\Lambda_n = \sup_{\theta\in\Theta} \Lambda_n(\theta)$$

is asymptotically equivalent to the maximum of the standardised efficient score process, which in turn has the same asymptotic distribution as the maximum of its limiting Gaussian process. Similar results are obtained by Davies (1977, 1987). In many cases where the parameter space is one-dimensional, the restriction to a compact parameter set is effectively a restriction that the set of standardised (efficient) scores have *finite arc length*.

Later Dacunha-Castelle & Gassiat (1997) generalised further the class of problems studied in Ghosh & Sen (1985), removing the separation condition and considering more general hypotheses regarding the number of components in mixtures. They imposed strong differentiability conditions, and still effectively restricted the parameter space Θ to be a compact set, more precisely that the set of standardised score functions

$$\mathcal{S}_\Theta \;=\; \{s_\theta \,|\, \theta \in \Theta\}$$

be a Donsker class. This is a rather technical condition, which controls the size of the set of functions $\mathcal{S}_\Theta$ in some sense. It also ensures that the empirical process of score functions converges weakly to a tight Gaussian process. In certain one-dimensional examples this is effectively a restriction that the set of standardised (efficient) scores have finite arc length.

In Bickel & Chernoff (1993) the problem considered in Hartigan (1985) is revisited. It is shown there that if $\Theta = \mathbb{R}$, not restricted to a compact set, then the maximum of the standardised efficient score process has asymptotic distribution the same as

$$\sup_{|\theta| \le \sqrt{(\log n)/2}} \tilde{S}_0(\theta) \ ,$$

where $\left\{ \tilde{S}_0(\theta) \,|\, \theta \in \mathbb{R} \right\}$ is a certain standardised *stationary* Gaussian process. Extreme value theory for stationary Gaussian processes then gives that the maximum can be written as

$$a_n + \frac{G_n - \log \sqrt{2}\pi}{a_n} \ ,$$

where $a_n = (\log \log n)^{1/2}$, and G_n has asymptotic Gumbel distribution, with distribution function $x \mapsto \exp\{-e^{-x}\}$. This extends the result of Hartigan (1985), and agrees even more with our heuristic regarding the maximum when the set of standardised efficient scores has *infinite arc length*

The bulk of what follows constitutes a generalisation of the Bickel & Chernoff (1993) result to a general one-parameter exponential family.

5.2 Formulation and Notation

We wish to derive the asymptotic distributions of certain score statistics for testing for homogeneity in the two component mixture model

$$(1-p)G_{\omega_0} + pG_{\omega_1}$$

generated by a general one-parameter exponential family $\mathcal{G} = \{G_\omega \,|\, \omega \in \Omega \subset \mathbb{R}\}$ with densities of the form

$$g_\omega(x) = \exp\{\eta(\omega)t(x) - B(x) - A(\omega)\} \ . \tag{4}$$

We consider the simple and composite testing problems when the true distribution is G_{ω_0}.

We call $\eta(\cdot)$ and $t(\cdot)$ in (4) respectively the canonical parameter and the (accompanying) canonical statistic. Such a representation is not unique, for let $t^*(x) = a + bt(x)$ be a linear transformation of the canonical statistic, or $\eta^\dagger(\omega) = c + d\eta(\omega)$ a linear transformation of the canonical parameter. Then we can write

$$\begin{aligned} g_\omega(x) &= \exp\{\eta(\omega)[t^*(x) - a]/b - B(x) - A(\omega)\} \\ &= \exp\left\{\frac{\eta(\omega)}{b} t^*(x) - B(x) - [A(\omega) + a\eta(\omega)/b]\right\} \\ &= \exp\{\eta^*(\omega)t^*(x) - B(x) - A^*(\omega)\} \end{aligned} \tag{5}$$

or

$$\begin{aligned} g_\omega(x) &= \exp\big\{[\eta^\dagger(\omega) - c]t(x)/d - B(x) - A(\omega)\big\} \\ &= \exp\left\{\eta^\dagger(\omega)\frac{t(x)}{d} - [B(x) + ct(x)/d] - A(\omega)\right\} \\ &= \exp\big\{\eta^\dagger(\omega)t^\dagger(x) - B^\dagger(x) - A(\omega)\big\} \ . \end{aligned}$$

For any choice of canonical statistic $t(\cdot)$ we can arrive at a most convenient representation as follows. Start with any representation as in (4). Define the linear transformation

$$\theta(\omega) = \eta(\omega) - \eta(\omega_0) \ . \tag{6}$$

Now $\theta = 0$ corresponds to $\omega = \omega_0$. We assume this transformation admits an inverse $\theta \mapsto \omega(\theta)$. Write the density as

$$\begin{aligned} g_\omega(x) &= \exp\{[\eta(\omega) - \eta(\omega_0) + \eta(\omega_0)]t(x) - B(x) - A(\omega)\} \ , \\ &= \exp\{\theta(\omega)t(x) - B^*(x) - A(\omega)\} \end{aligned}$$

where $B^*(x) = B(x) + \eta(\omega_0)t(x)$. Define now

$$\begin{aligned} g^*_\theta(x) &= g_{\omega(\theta)}(x) \\ &= \exp\{\theta t(x) - B^*(x) - K(\theta)\} \ , \end{aligned}$$

where $K(\theta) = A(\omega(\theta))$. Denoting by G^*_θ the probability measure with density g^*_θ, we see that $G^*_\theta = G_{\omega(\theta)}$, and the family of distributions is now $\mathcal{G} =$

$\{G^*_\theta \,|\, \theta \in \Theta = \theta(\Omega)\}$. We call this the most convenient representation of the family, based on the choice of canonical statistic $t(\cdot)$. Note that in this form the function $K(\cdot)$ is the cumulant generating function of the random variable $T = t(X)$ where $X \sim G^*_0 = G_{\omega_0}$.

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with common distribution G_{ω_0}. The statistics are $\sup_{\theta\in\Theta} S^\dagger_n(\theta)$ and $\sup_{\theta\in\Theta} S^*_n(\theta)$ where

$$S^\dagger_n(\theta) \quad = \quad (v_\theta n)^{-1/2}\sum_{i=1}^{n}\left(e^{\theta t(X_i)-K(\theta)} - 1\right) \tag{7}$$

where

$$v_\theta \quad = \quad e^{K(2\theta)-2K(\theta)} - 1 \ , \tag{8}$$

and

$$S^*_n(\theta) \quad = \quad (v^*_\theta n)^{-1/2}\sum_{i=1}^{n}\left(e^{\theta t(X_i)-K(\theta)} - 1 - \frac{(\mu_\theta - \mu_0)(t(X_i) - \mu_0)}{\sigma_0^2}\right) \tag{9}$$

where

$$v^*_\theta = e^{K(2\theta)-2K(\theta)} - 1 - \frac{(\mu_\theta - \mu_0)^2}{\sigma_0^2} \ , \tag{10}$$

and $\mu_\theta = K'(\theta)$ and $\sigma^2_\theta = K''(\theta)$ are the mean and variance of $T = t(X)$ if $X \sim G^*_\theta$. These both only depend on X_i through $T_i = t(X_i)$, and θ only through the function K and its derivatives. Thus the *distribution* of these statistics when $X_1 \sim G^*_0$ only depends on the *distribution* of the canonical statistic $t(X_1)$ when $X_1 \sim G^*_0$. So for any two families whose canonical statistics have the same distribution, the distributions of the mixture model score statistics will be the same for each family.

If $X \sim G^*_\theta$ then the canonical statistic $T = t(X)$ has distribution F_θ with density

$$f_\theta(t) = \exp\{\theta t - G(t) - K(\theta)\} \ ,$$

for some other function $G(\cdot)$. So the collection of distributions

$$\mathcal{F} = \{F_\theta \,|\, \theta \in \Theta\}$$

is also an exponential family. We call this the *linear* exponential family (see Barndorff-Nielsen (1978), page 113) generated by the distribution F_0 with density $f_0(t) = e^{-G(t)}$, since the canonical statistic is linear in the observations, and $\theta = 0$ corresponds to F_0.

Now the canonical statistics, and thus the mixture model score statistics, have the same distribution for these two families. In some extreme cases, we need to use a 'method-of-sieve'-type approach, and consider an 'increasing sequence' $\mathcal{F}_n = \{F_\theta \,|\, \theta \in \Theta_n\}$, where $\{\Theta_n \,|\, n \in \mathbb{N}\}$ is a sequence of intervals such that for each $n \in \mathbb{N}$ $\Theta_n \subset \Theta_{n+1}$. Thus we suppose we have such a sequence $\{\Theta_n \,|\, n \in \mathbb{N}\}$ and define the statistics

$$M_n^\dagger = \sup_{\theta \in \Theta_n} S_n^\dagger(\theta) \quad \text{and} \quad M_n^* = \sup_{\theta \in \Theta_n} S_n^*(\theta) \tag{11}$$

in terms of them. We note that in many of the examples we consider, $\Theta_n \equiv \Theta$. The specification of $\{\Theta_n \,|\, n \in \mathbb{N}\}$ is influenced by the outcome of the first stage of the procedure which we describe later.

It is convenient at this stage to define the following quantities in relation to exponential families. See Barndorff-Nielsen (1978, 1980) for more information. Define the maximum likelihood estimate based on one observation $\hat{\theta}(x)$ as the solution of $\mu_\theta = x$, Define the variance-stabilising transformation as $\phi(\theta) = \int_0^\theta \sigma_\eta \, d\eta$ and another transformation we label $\tau(\theta) = \phi(2\theta)/2$, and also define $\hat{\phi}(x) = \phi(\hat{\theta}(x))$. Note that $\phi(\cdot)$ is defined on all of $\widetilde{\Theta}$, but $\tau(\cdot)$ is only defined for $\theta \in \Theta = \{\theta/2 \,|\, \theta \in \Theta\}$.

Now as detailed below we approximate both $M_n^\dagger$ and M_n^* with a third quantity $M_n = \sup_{\theta \in \Theta_n} S_n(\theta)$, where

$$S_n(\theta) = n^{-1/2} \sum_{i=1}^{n} e^{\theta X_i - \frac{1}{2} K(2\theta)} ,$$

which behaves much like both $S_n^\dagger(\theta)$ and $S_n^*(\theta)$ for large θ. Of primary interest is the sequence of arc lengths $A(\mathcal{S}_n)$, where $\mathcal{S}_n = \{s_\theta \,|\, \theta \in \Theta_n\}$.

It turns out that

$$F_0 \left(\frac{\partial s_\theta}{\partial \theta} \right)^2 = K''(2\theta) ,$$

so we have that

$$A(\mathcal{S}_n) = \int_{\inf \Theta_n}^{\sup \Theta_n} \sqrt{K''(2\theta)} \, d\theta = \tau(\sup \Theta_n) - \tau(\inf \Theta_n) .$$

Our methods that follow *only pertain* to cases where the arc lengths tend to infinity. Write $\tau(\Theta_n)$ for the image of Θ_n under the transformation τ. We assume in all that follows that

$$\limsup_{n\to\infty} \tau(\Theta_n) - \liminf_{n\to\infty} \tau(\Theta_n) = \infty \ . \tag{12}$$

This condition characterises the mixture problems that our methods apply to; those with *infinite arc length*.

5.3 Examples

We now present some examples of exponential families with densities of the form (4), derive for each a convenient generating distribution F_0 and accompanying full and restricted linear exponential families $\widetilde{\mathcal{F}}$ and $\mathcal{F}$ respectively. We also indicate what restrictions are required for a sequence of parameter spaces $\{\Theta_n \,|\, n \in \mathbb{N}\}$ to satisfy (12). We cover mixtures of Normal means, Gamma shape parameters, Gamma scale parameters (which includes mixtures of Normal variances), Beta shape parameters, Poisson means and Negative Binomial failure probabilities. We also examine mixtures of Binomial success probabilties, which don't ordinarily satisfy our condition (12), but we indicate an artificial version of the Binomial problem that does satisfy (12), namely where the number of trials increases with the sample size.

5.3.1 Normal mean mixtures

Let the family be

$$\mathcal{G} = \left\{ G_\mu = \mathcal{N}(\mu, \sigma_0^2) \,|\, \mu \in \mathbb{R} \right\} ,$$

where the variance σ_0^2 is known, and suppose that the true value is μ_0.

The density can be written as

$$g_\mu(x) \quad = \quad \exp\left\{ \frac{\mu x}{\sigma_0^2} + \log \frac{1}{\sqrt{2\pi}} - \log \sigma_0 - \frac{1}{2}\frac{x^2}{\sigma_0^2} - \frac{1}{2}\frac{\mu^2}{\sigma_0^2} \right\} \tag{13}$$

The term involving both μ and x is the leading term $\mu x/\sigma_0^2$. So the canonical statistic $t(x)$ has to be a linear function of x, and we should choose it so that F_0, the distribution of $t(X)$ when $X \sim \mathcal{N}(\mu_0, \sigma_0^2)$, is the most convenient possible. We list some possibilities:

- $t(x) = x$, $F_0 = \mathcal{N}(\mu_0, \sigma_0^2)$;
- $t(x) = x/\sigma_0$, $F_0 = \mathcal{N}(\mu_0/\sigma_0, 1)$;
- $t(x) = (x - \mu_0)/\sigma_0$, $F_0 = \mathcal{N}(0, 1)$.

The third choice seems the most convenient. Now

$$\int e^{\theta x}\, dF_0(x) = e^{\theta^2/2} < \infty$$

for every $\theta \in \mathbb{R}$, so the full parameter space $\widetilde{\Theta} = \mathbb{R}$, and the restricted space $\Theta = \mathbb{R}$ also. So the cumulant generating function $K(\theta) = \theta^2/2$, and F_θ is the distribution with density f_θ, given by

$$\begin{aligned} f_\theta(x) &= e^{\theta x - K(\theta)} f_0(x) \\ &= (2\pi)^{-1/2} \exp\{x\theta - \theta^2/2 - x^2/2\} \\ &= (2\pi)^{-1/2} \exp\{(x-\theta)^2/2\} \ , \end{aligned}$$

which is the density of $\mathcal{N}(\theta, 1)$.

Thus to solve the problem for $\mathcal{G}$, for any values of the known variance σ_0^2 and the true value μ_0, we need only solve it for the linear family

$$\mathcal{F} = \{\mathcal{N}(\theta, 1) \,|\, \theta \in \mathbb{R}\} \ ,$$

when the true value is $\theta = 0$.

Note that $K'(\theta) = \theta$, so $K''(\theta) \equiv 1$. Thus the variance stablising transformation is $\phi(\theta) = \int_0^\theta d\theta = \theta$ and $\tau(\theta) = \phi(2\theta)/2 = \theta$ also.

So any sequence of families

$$\mathcal{F}_n = \{\mathcal{N}(\theta, 1) \,|\, \theta \in \Theta_n\}$$

where the length of the intervals Θ_n tends to infinity satisfies (12). Thus our methods do not apply to the case of $\Theta_n \subset I$, a fixed bounded interval.

5.3.2 Gamma shape mixtures

For future reference we define the Gamma(α, λ) distribution as that distribution on $(0, \infty)$ with density

$$g(x) = x^{\alpha-1} e^{-\lambda x} \lambda^\alpha / \Gamma(\alpha) \ .$$

Let the (full) family be $\widetilde{\mathcal{G}} = \{G_\alpha = \text{Gamma}(\alpha, \lambda_0) \,|\, \alpha > 0\}$, for some known value λ_0. Suppose that the true distribution is G_{α_0}. Write the density as

$$g_\alpha(x) = \exp\{(\alpha - 1)\log x - \lambda_0 x + \alpha \log \lambda_0 - \log \Gamma(\alpha)\} \ .$$

The leading term tells us that the canonical statistic needs to be a linear function of $\log x$. A sensible choice is

$$t(x) = \log x + \log \lambda_0 = \log \lambda_0 x \ .$$

In this case, when $X \sim G_{\alpha_0}$,

$$\begin{aligned} P(t(X) \le t) &= P(\log \lambda_0 X \le t) \\ &= P(X \le e^t/\lambda_0) \\ &= G_{\alpha_0}(e^t/\lambda_0) \ . \end{aligned}$$

So the density of $t(X)$ is

$$\begin{aligned} f_0(t) &= g_{\alpha_0}(e^t/\lambda_0)e^t/\lambda_0 \\ &= \exp\big\{t\alpha_0 - e^t - \log \Gamma(\alpha_0)\big\} \end{aligned}$$

which does not depend on λ_0. This is the density of a distribution F_0 which is positive on the whole of $\mathbb{R}$. Now,

$$\int e^{\theta x}\, dF_0(x) = \Gamma(\alpha_0 + \theta)/\Gamma(\alpha_0) \ ,$$

which is finite for $\alpha_0 + \theta > 0$, so the full linear family is

$$\widetilde{\mathcal{F}} = \left\{ F_\theta \,\middle|\, \theta \in \widetilde{\Theta} = (-\alpha_0, \infty) \right\} \ ,$$

where F_θ is the distribution with density

$$\begin{aligned} f_\theta(x) &= e^{\theta x - K(\theta)} f_0(x) \\ &= \exp\{\theta x - \log \Gamma(\alpha_0 + \theta) + \log \Gamma(\alpha_0)\}\, f_0(x) \\ &= \exp\{(\theta + \alpha_0)x - e^x - \log \Gamma(\theta + \alpha_0)\} \ . \end{aligned}$$

This is the distribution of $\log \lambda_0 X$ when $X \sim \text{Gamma}(\alpha_0 + \theta, \lambda_0)$.

The restricted family is

$$\mathcal{F} = \left\{ F_\theta \,\middle|\, \theta \in \Theta = \widetilde{\Theta}/2 = (-\alpha_0/2, \infty) \right\} ,$$

which corresponds to a restricted family of Gamma distributions

$$\mathcal{G} = \{\text{Gamma}(\alpha, \lambda_0) \,|\, \alpha > \alpha_0/2\} .$$

Now the cumulant generating function $K(\theta) = \log\Gamma(\alpha_0 + \theta) - \log\Gamma(\alpha_0)$. So $K'(\theta) = \psi(\alpha_0 + \theta)$ and $K''(\theta) = \psi'(\alpha_0 + \theta)$, where $\psi(x)$ and $\psi'(x)$ are the first and second derivatives of $\log\Gamma(x)$. A closed form does not exist for ψ' in general, however we can approximate $\psi'(x)$ for very small and very large x as follows:

$$\psi'(x) \approx \begin{cases} x^{-2} & \text{for small } x \ , \\ x^{-1} & \text{for large } x \end{cases} \tag{14}$$

(see Jensen (1995), Appendix A.2 and Abramowitz & Stegun (1964), Section 6.4 for details). So for large θ,

$$\begin{aligned} \phi(\theta) &= \int_0^\theta \sqrt{K''(\eta)}\, d\eta \\ &\approx \int_0^\theta (\alpha_0 + \eta)^{-1/2}\, d\eta \\ &= 2[(\alpha_0 + \theta)^{1/2} - \alpha_0^{1/2}] \\ &\sim 2\theta^{1/2} \end{aligned}$$

as $\theta \to \infty$ and for θ near $-\alpha_0$,

$$\begin{aligned} \phi(\theta) &= \int_0^\theta \sqrt{K''(\eta)}\, d\eta \\ &\approx \int_0^\theta (\alpha_0 + \eta)^{-1}\, d\eta \\ &= \log(\alpha_0 + \theta) - \log\alpha_0 \end{aligned}$$

as $\alpha_0 + \theta \downarrow 0$.

So

$$\tau(\theta) = \tfrac{1}{2}\phi(2\theta) \begin{cases} \to \infty & \text{as } \theta \to \infty \\ \to -\infty & \text{as } \theta \to -\alpha_0/2 \ . \end{cases}$$

Thus any sequence of intervals $\{\Theta_n \mid n \in \mathbb{N}\}$ such that either $\inf \Theta_n \to -\alpha_0/2$ or $\sup \Theta_n \to \infty$ will yield sequences of families $\mathcal{F}_n = \{F_\theta \mid \theta \in \Theta_n\}$ that satisfy (12).

5.3.3 Gamma scale mixtures

Now let $\widetilde{\mathcal{G}} = \{G_\lambda = \text{Gamma}(\alpha_0, \lambda) \mid \lambda > 0\}$ for some known α_0, and consider mixtures of λ when λ_0 is the true value. Write the density as

$$g_\lambda(x) \quad = \quad \exp\{-\lambda x + \alpha_0 \log \lambda + (\alpha_0 - 1) \log x - \log \Gamma(\alpha_0)\} \ .$$

The leading term here tells us that the canonical statistic has to be a linear function of x. A convenient choice is $t(x) = \lambda_0 x$, since when $X \sim G_{\lambda_0}$, $t(X) = \lambda_0 X \sim G_1 = \text{Gamma}(\alpha_0, 1)$. We have

$$\begin{aligned} \int e^{\theta x}\, dG_1(x) &= \int e^{\theta x} x^{\alpha_0 - 1} e^{-x}\, dx/\Gamma(\alpha_0) \\ &= \int x^{\alpha_0 - 1} e^{-(1-\theta)x}\, dx/\Gamma(\alpha_0) \\ &= \frac{1}{(1-\theta)^{\alpha_0}} \quad \text{only if } 1 - \theta > 0 \ . \end{aligned}$$

So the full linear family is

$$\widetilde{\mathcal{F}} = \left\{ F_\theta \,\middle|\, \theta \in \widetilde{\Theta} = (-\infty, 1) \right\} ,$$

where F_θ has density

$$\begin{aligned} f_\theta(x) &= e^{\theta x - K(\theta)} g_1(x) \\ &= e^{\theta x} (1-\theta)^{\alpha_0} x^{\alpha_0 - 1} e^{-x}/\Gamma(\alpha_0) \\ &= g_{1-\theta}(x) \ . \end{aligned}$$

So $F_\theta = G_{1-\theta}$ and so the full linear family is just the family itself:

$$\begin{aligned} \widetilde{\mathcal{F}} &= \left\{ F_\theta \,\middle|\, \theta \in \widetilde{\Theta} \right\} \\ &= \{G_{1-\theta} \mid \theta < 1\} \\ &= \{G_\lambda \mid \lambda > 0\} \ . \end{aligned}$$

However, the restricted parameter space is $\Theta = \left\{\theta/2 \,\middle|\, \theta \in \widetilde{\Theta}\right\} = (-\infty, \frac{1}{2})$, so restricted family is

$$\begin{aligned}\mathcal{F} &= \{F_\theta \,|\, \theta \in \Theta\} \\ &= \{G_{1-\theta} \,|\, \theta < 1/2\} \\ &= \{G_\lambda \,|\, \lambda > 1/2\} \,.\end{aligned}$$

This corresponds to restricting the original family $\widetilde{\mathcal{G}}$ to $\mathcal{G} = \{\text{Gamma}(\alpha_0, \lambda) \,|\, \lambda > \lambda_0/2\}$.

So to solve the problem for $\mathcal{G}$, for any known value of the shape parameter α_0 and any true value λ_0, we need only restrict attention to the case when the known value is $\lambda_0 = 1$. However we note that not all mixtures are allowed; we must restrict attention to $\lambda > \lambda_0/2$.

So we can handle the following (maximal) simple and composite versions of the original problem: if $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables with common unknown distribution G, we test the hypothesis $H\colon G = \text{Gamma}(\alpha_0, \beta_0)$ against the alternative $A\colon G = (1-p)\text{Gamma}(\alpha_0, \beta_0) + p\text{Gamma}(\alpha_1, \beta_0)$, where $\beta_0 > 0$ is known, $\alpha_1 > \alpha_0/2$ and one of the following cases is true:

1. α_0 is known (simple problem);
2. α_0 is unknown (composite problem).

The cumulant generating function is $K(\theta) = -\alpha_0 \log(1-\theta)$, so $K'(\theta) = \alpha_0(1-\theta)^{-1}$ and $K''(\theta) = \alpha_0(1-\theta)^{-2}$. So

$$\begin{aligned}\phi(\theta) &= \int_0^\theta \frac{\alpha_0^{1/2}}{1-\theta}\, d\theta \\ &= -\alpha_0^{1/2} \log(1-\theta)\end{aligned}$$

and

$$\tau(\theta) = -\alpha_0^{1/2} \log(1-2\theta)/2 \begin{cases} \to \infty & \text{as } \theta \to \frac{1}{2} \\ \to -\infty & \text{as } \theta \to -\infty \,. \end{cases}$$

So any sequence of intervals $\{\Theta_n \,|\, n \in \mathbb{N}\}$ such that either $\inf \Theta_n \to -\infty$ or $\sup \Theta_n \to \frac{1}{2}$ will yield sequences of families $\mathcal{F}_n = \{F_\theta \,|\, \theta \in \Theta_n\}$ that satisfy (12).

5.3.4 Normal variance mixtures

Let $\mathcal{G} = \{G_{\sigma^2} = \mathcal{N}(\mu_0, \sigma^2) \,|\, \sigma > 0\}$ for some known value μ_0, and consider mixtures of σ^2 when the true distribution is $G_{\sigma_0^2}$. The density can be written as

$$g_{\sigma^2}(x) = \exp\left\{ -\frac{1}{2\sigma^2}(x-\mu_0)^2 - \log \sigma\sqrt{2\pi} \right\} ,$$

so the canonical statistic must be a linear function of $(x-\mu_0)^2$. A convenient choice is

$$t(x) = \frac{(x-\mu_0)^2}{2\sigma_0^2} ,$$

since then if $X \sim G_{\sigma_0^2} = N(\mu_0, \sigma_0^2)$, then

$$t(X) = \frac{(X-\mu_0)^2}{2\sigma_0^2} \sim \tfrac{1}{2}\chi_1^2 = \text{Gamma}(1/2, 1) .$$

So this is in fact a special case of Gamma scale mixtures from the previous section, with $\alpha_0 = 1/2$.

In fact, writing the density as

$$g_{\sigma^2}(x) = \exp\{\eta(\sigma^2)t(x) - B(x) - A(\sigma^2)\} ,$$

we see that the canonical parameter corresponding to our choice of canonical statistic is

$$\eta(\sigma^2) = -\frac{1}{\sigma^2} ,$$

and so following (6), we have

$$\begin{aligned} \theta(\sigma^2) &= \eta(\sigma^2) - \eta(\sigma_0^2) \\ &= \frac{1}{\sigma_0^2} - \frac{1}{\sigma^2} . \end{aligned}$$

So F_θ corresponds to $\mathcal{N}(\mu_0, \sigma^2(\theta))$, where

$$\sigma^2(\theta) = \frac{\sigma_0^2}{1-\theta\sigma_0^2} ,$$

and the usual Gamma scale restriction that $\theta < 1/2$ corresponds to

$$\sigma^2 \quad < \quad \frac{2\sigma_0^2}{2 - \sigma_0^2} \ .$$

All other general restrictions apply as they do in the general Gamma scale mixtures case.

5.3.5 Beta

Let G_α be the distribution on $(0, 1)$ with density

$$\begin{aligned} g_\alpha(x) &= x^{\alpha-1}(1-x)^{\beta_0-1}/B(\alpha, \beta_0) \\ &= \exp\{(\alpha - 1)\log x + (\beta_0 - 1)\log(1-x) - \log B(\alpha, \beta_0)\} \ . \end{aligned}$$

where the Beta function is given by

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \ .$$

Suppose we are interested in mixtures of the α parameter, $\beta_0 > 0$ is fixed and the true distribution is G_{α_0}.

It is convenient to take the canonical statistic as just $\log x$, which under G_{α_0} has density

$$f_0(x) = \exp\{\alpha_0 x - (\beta_0 - 1)\log(1 - e^x) - \log B(\alpha_0, \beta_0)\}$$

for $x < 0$ and 0 otherwise.

$$\begin{aligned} \int_{-\infty}^{0} e^{\theta x} f_0(x)\, dx &= \frac{B(\alpha_0 + \theta, \beta_0)}{B(\alpha_0, \beta_0)} \int_{-\infty}^{0} e^{\theta x} e^{x(\alpha_0 + \theta)}\, (1 - e^x)^{\beta_0 - 1}\ dx/B(\alpha_0 + \theta, \beta_0) \\ &= \frac{B(\alpha_0 + \theta, \beta_0)}{B(\alpha_0, \beta_0)} = \frac{\Gamma(\alpha_0 + \theta)\Gamma(\beta_0)}{\Gamma(\alpha_0 + \beta_0 + \theta)} \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} \\ &= \frac{\Gamma(\alpha_0 + \theta)\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0 + \beta_0 + \theta)\Gamma(\alpha_0)} \ , \end{aligned}$$

so long as $\theta > -\alpha_0$. So $K(\theta) = \log B(\alpha_0 + \theta, \beta_0) - \log B(\alpha_0, \beta_0)$ and we define F_θ as the distribution on $(-\infty, 0)$ with density

$$\begin{aligned} f_\theta(x) &= e^{\theta x - K(\theta)} f_0(x) \\ &= e^{x(\theta + \alpha_0)}(1 - e^x)^{\beta_0 - 1}/B(\theta + \alpha_0, \beta_0) \ , \end{aligned}$$

and the full family is

$$\widetilde{\mathcal{F}} = \{ F_\theta \,|\, \theta > -\alpha_0 \}$$

so the restricted family is

$$\mathcal{F} = \{ F_\theta \,|\, \theta > -\alpha_0/2 \} \ .$$

We also have that

$$\begin{aligned}
K(\theta) &= \log\Gamma(\alpha_0+\theta) - \log\Gamma(\alpha_0+\beta_0+\theta) + \log\Gamma(\alpha_0+\beta_0) - \log\Gamma(\alpha_0) \ , \\
K'(\theta) &= \psi(\alpha_0+\theta) - \psi(\alpha_0+\beta_0+\theta) \ , \\
K''(\theta) &= \psi'(\alpha_0+\theta) - \psi'(\alpha_0+\beta_0+\theta) \ .
\end{aligned}$$

Utilising the approximations in (14), we can say that for large θ,

$$\begin{aligned}
\psi'(\alpha_0+\theta) - \psi'(\alpha_0+\beta_0+\theta) &\approx \frac{1}{\alpha_0+\theta} - \frac{1}{\alpha_0+\beta_0+\theta} \\
&= \frac{\beta_0}{(\alpha_0+\theta)(\alpha_0+\beta_0+\theta)} \\
&\sim \beta_0/\theta^2
\end{aligned}$$

as $\theta \to \infty$. Thus

$$\phi(\theta) \approx \beta_0^{1/2} \int_0^\theta \eta^{-1}\, d\eta \ = \ \beta_0^{1/2} \log\theta$$

for large θ. For θ near $-\alpha_0$,

$$\begin{aligned}
\psi'(\alpha_0+\theta) - \psi'(\alpha_0+\beta_0+\theta) &\approx (\alpha_0+\theta)^{-2} - \psi'(\beta_0) \\
&\sim (\alpha_0+\theta)^{-2}
\end{aligned}$$

as $(\alpha_0+\theta) \downarrow 0$. So then

$$\begin{aligned}
\phi(\theta) &\approx \int_0^\theta \frac{1}{\alpha_0+\eta}\, d\eta \\
&= \log(\alpha_0+\theta) - \log\alpha_0 = \log\left(1 - \frac{\theta}{\alpha_0}\right) \ .
\end{aligned}$$

for small $(\alpha_0 + \theta)$. So

$$\tau(\theta) = \phi(2\theta)/2 \begin{cases} \to \infty & \text{as } \theta \to \infty \\ \to -\infty & \text{as } \theta \to -\alpha_0/2 \ . \end{cases}$$

So any sequence of intervals $\{\Theta_n \mid n \in \mathbb{N}\}$ such that either $\inf \Theta_n \to -\alpha_0/2$ or $\sup \Theta_n \to \infty$ will yield sequences of families $\mathcal{F}_n = \{F_\theta \mid \theta \in \Theta_n\}$ satisfying (12).

5.3.6 Poisson

Let Pois(λ) denote the distribution on $0, 1, 2, \ldots$ with probability mass function

$$\begin{aligned} g_\lambda(x) &= \frac{e^{-\lambda}\lambda^x}{x!} \\ &= \exp\{x \log \lambda - \log x! - \lambda\} \ . \end{aligned}$$

Suppose $\widetilde{\mathcal{G}} = \{G_\lambda = \text{Pois}(\lambda) \mid \lambda > 0\}$, and we wish to test for homogeneity of the mean parameter λ against an alternative of a two-component mixture, and λ_0 is the true value.

If we let $F_0 = G_{\lambda_0}$, then

$$\begin{aligned} K(\theta) &= \log \int e^{\theta x} \, dF_0(x) \\ &= \log \sum_{x=0}^{\infty} \frac{e^{-\lambda_0} \left(\lambda_0 e^\theta\right)^x}{x!} \\ &= \log \left\{ e^{-\lambda_0} e^{\lambda_0 e^\theta} \right\} \\ &= \lambda_0 (e^\theta - 1) < \infty \end{aligned}$$

for all real θ. Thus $\widetilde{\Theta} = \Theta = \mathbb{R}$, that is the full and restricted families coincide:

$$\mathcal{F} = \left\{ F_\theta = \text{Pois}\left(\lambda_0 e^\theta\right) \mid \theta \in \mathbb{R} \right\} \ .$$

Also $K'(\theta) = K''(\theta) = \lambda_0 e^\theta$. So

$$\begin{aligned} \phi(\theta) = \lambda_0^{1/2} \int_0^\theta e^{\eta/2} \, d\eta &= \lambda_0^{1/2} \left[2e^{\eta/2}\right]_0^\theta \\ &= 2\lambda_0^{1/2} [e^{\theta/2} - 1] \end{aligned}$$

and

$$\tau(\theta) = \phi(2\theta)/2 = \lambda_0^{1/2}[e^\theta - 1] \begin{cases} \to \infty & \text{as } \theta \to \infty \\ \to -\lambda_0^{1/2} & \text{as } \theta \to -\infty \ . \end{cases}$$

So any sequence of intervals $\{\Theta_n \,|\, n \in \mathbb{N}\}$ with $\sup \Theta_n \to \infty$ will yield a sequence of families $\mathcal{F}_n = \{F_\theta \,|\, \theta \in \Theta_n\}$ satisfying (12).

5.3.7 Negative Binomial

Let $\mathrm{Neg}(m, q)$ denote the distribution on $0, 1, 2 \ldots$ with probability mass function

$$g_q(x) = \binom{m+x-1}{m-1} q^x (1-q)^m \quad \text{for } x = 0, 1, 2 \ldots \ .$$

This is the so-called negative binomial distribution; if X gives the number of failures before the m-th success in a series of independent Bernoulli trials with success probability $1-q$, then $X \sim \mathrm{Neg}(m, q)$. The case $m = 1$ yields the familiar Geometric distribution. Suppose $\widetilde{\mathcal{G}} = \{G_q = \mathrm{Neg}(m, q) \,|\, 0 \le q < 1\}$, and we wish to test for homogeneity of the failure probability q against the alternative of a two-component mixture. Suppose the true distribution is G_{q_0} for some fixed $0 < q_0 < 1$. We can write the density (with respect to counting measure) as

$$g_q(x) = \exp\left\{ x \log q + \log \binom{m+x-1}{m-1} + m \log(1-q) \right\} \ .$$

Let the canonical statistic be $t(x) = x$, and put $f_0(x) = g_{q_0}(x)$.

$$\begin{aligned} \int e^{\theta x} dF_0(x) &= \sum_{x=0}^{\infty} e^{\theta x} \binom{m+x-1}{m-1} q_0^x (1-q_0)^m \\ &= \left(\frac{1-q_0}{1-q_0 e^\theta} \right)^m \sum_{x=0}^{\infty} \binom{m+x-1}{m-1} \left(q_0 e^\theta\right)^x (1-q_0 e^\theta)^m \\ &= \left(\frac{1-q_0}{1-q_0 e^\theta} \right)^m \end{aligned}$$

if $q_0 e^\theta < 1$, or $\theta < -\log q_0$. Let

$$\begin{aligned} f_\theta(x) &= e^{\theta x - K(\theta)} f_0(x) \\ &= \binom{m+x-1}{m-1} (q_0 e^\theta)^x (1-q_0 e^\theta)^m \\ &= g_{q_0 e^\theta}(x) \end{aligned}$$

be the density of $F_\theta = G_{q_0 e^\theta} = \text{Neg}(m, q_0 e^\theta)$. So the full family is

$$\widetilde{\mathcal{F}} = \{F_\theta \,|\, \theta < -\log q_0\} = \{\text{Neg}(m,q) \,|\, 0 < q < 1\}$$

and the restricted family is

$$\mathcal{F} = \{F_\theta \,|\, \theta < -(\log q_0)/2\} = \left\{\text{Neg}(m,q) \,\middle|\, 0 < q < q_0^{1/2}\right\} .$$

We have

$$\begin{aligned} K(\theta) &= m[\log(1-q_0) - \log\left(1-q_0 e^\theta\right)] , \\ K'(\theta) &= \frac{m q_0 e^\theta}{1-q_0 e^\theta} , \\ K''(\theta) &= \frac{m q_0 e^\theta}{(1-q_0 e^\theta)^2} . \end{aligned}$$

So

$$\begin{aligned} \phi(\theta) &= m^{1/2} \int_0^\theta \frac{q_0^{1/2} e^{\eta/2}}{1-q_0 e^\theta}\, d\eta \\ &= m^{1/2} \int_{1-q_0 e^\theta}^{1-q_0} \frac{1}{z\sqrt{1-z}}\, dz \end{aligned}$$

using a change of variable $z = 1 - q_0 e^\eta$, $d\eta = -dz/(1-z)$. Now since

$$\frac{d}{dz} \log\left(\frac{1-\sqrt{1-z}}{1+\sqrt{1-z}}\right) = \frac{1}{z\sqrt{1-z}}$$

we have that

$$\phi(\theta) = \log\left(\frac{1-\sqrt{q_0}}{1+\sqrt{q_0}}\right) - \log\left(\frac{1-\sqrt{q_0 e^\theta}}{1+\sqrt{q_0 e^\theta}}\right)$$

and

$$\begin{aligned}\tau(\theta) &= \phi(2\theta)/2 = \tfrac{1}{2}\left[\log\left(\frac{1-\sqrt{q_0}}{1+\sqrt{q_0}}\right) - \log\left(\frac{1-\sqrt{q_0}e^{\theta}}{1+\sqrt{q_0}e^{\theta}}\right)\right] \\ & \begin{cases}\to \infty & \text{as } \theta \to -\tfrac{1}{2}\log q_0 \\ \to \tfrac{1}{2}\log\left(\frac{1-\sqrt{q_0}}{1+\sqrt{q_0}}\right) & \text{as } \theta \to -\infty\ .\end{cases}\end{aligned}$$

So any sequence of intervals $\{\Theta_n \mid n \in \mathbb{N}\}$ with $\sup \Theta_n \to -\frac{1}{2}\log q_0$ yields a sequence of families $\mathcal{F}_n = \{F_\theta \mid \theta \in \Theta_n\}$ satisfying (12).

5.3.8 Binomial

Let $\mathrm{Bin}(m,p)$ denote the distribution on $0,1,\ldots,m$ with probability mass function

$$\begin{aligned}\binom{m}{x}p^x(1-p)^{m-x} &= \exp\left\{x\log\left(\frac{p}{1-p}\right) + \log\binom{m}{x} + m\log(1-p)\right\} \\ &= \exp\left\{x\eta(p) + \log\binom{m}{x} - m\log(1+e^{\eta(p)})\right\}\ ,\end{aligned}$$

for $x = 0,1,\ldots,m$, where $\eta(p) = \log\left(\frac{p}{1-p}\right)$ is the canonical parameter. This is the well-known Binomial distribution, governing the number of successes in m independent Bernoulli trials with success probability p.

Suppose for the moment that m is fixed and known, and we are interested in mixtures of the success probability, and the true distribution is $F_0 = \mathrm{Bin}(m,p_0)$, taking x as the canonical statistic.

So

$$\begin{aligned}\int e^{\theta x}dF_0(x) &= \sum_{x=0}^{m}\exp\left\{x[\theta+\eta(p_0)] + \log\binom{m}{x} - m\log\left[1+e^{\eta(p_0)}\right]\right\} \\ &= \exp\left\{m\log\left[1+e^{\theta+\eta(p_0)}\right] - m\log\left[1+e^{\eta(p_0)}\right]\right\} \\ &\quad \sum_{x=0}^{m}\exp\left\{x[\theta+\eta(p_0)] + \log\binom{m}{x} - m\log\left[1+e^{\eta(\theta+p_0)}\right]\right\} \\ &= \left(\frac{1+e^{\theta+\eta(p_0)}}{1+e^{\eta(p_0)}}\right)^m\end{aligned}$$

for all real θ. So $K(\theta) = m\left\{\log\left[1 + e^{\theta+\eta(p_0)}\right] - \log\left[1 + e^{\eta(p_0)}\right]\right\}$. Then F_θ is the distribution with density

$$f_\theta(x) = e^{\theta x - K(\theta)} f_0(x)$$

so $F_\theta = \mathrm{Bin}(m, p(\theta))$, where

$$p(\theta) = \frac{e^{\theta+\eta(p_0)}}{1 + e^{\theta+\eta(p_0)}} \ .$$

So the full and restricted families are the same:

$$\widetilde{\mathcal{F}} = \mathcal{F} = \{F_\theta \,|\, \theta \in \mathbb{R}\} = \{\mathrm{Bin}(m,p) \,|\, 0 < p < 1\} \ .$$

However,

$$\begin{aligned} K'(\theta) &= mp(\theta) \ , \\ K''(\theta) &= mp(\theta)[1 - p(\theta)] \ , \end{aligned}$$

so

$$\begin{aligned} \phi(\theta) &= m^{1/2} \int_0^\theta \sqrt{p(\eta)[1-p(\eta)]}\, d\eta \\ &= m^{1/2} \int_{p_0}^{p(\theta)} \frac{1}{\sqrt{y(1-y)}}\, dy \\ &= m^{1/2}[\arcsin(\sqrt{p(\theta)}) - \arcsin(\sqrt{p_0})] \ , \end{aligned}$$

using a change of variable $y = p(\theta)$, $d\theta = \frac{dy}{y(1-y)}$. Since arcsin only takes values on $[-\pi/2, \pi/2]$, for fixed m (12) cannot be satisfied. Tests of homogeneity in this 'finite arc-length' Binomial example are in fact treated in Chernoff & Lander (1995) and Lindsay (1995).

However we can construct an admittedly artifical situation which is at least of theoretical interest; if we allow $m = m_n \to \infty$ as $n \to \infty$, then we can satisfy (12), although we need to be careful that all our subsequent methods can handle having a different base measure for each n. We do not pursue this any further here, but defer for later work.

5.4 Summary of our method

Define a third (empirical) process

$$S_n(\theta) = n^{-1/2}\sum_{i=1}^{n} [y_\theta(X_i) - F_0 y_\theta] \quad \text{where} \tag{15}$$

$$y_\theta(x) = [f_\theta(x)/f_0(x)] \left\{ \int [f_\theta/f_0]^2 \, dF_0 \right\}^{-1/2} . \tag{16}$$

and its maximum $M_n = \sup_{\theta\in\Theta_n} S_n(\theta)$. Our plan in brief is to derive the asymptotic distribution of M_n, and then to show that both $M_n^\dagger$ and M_n^* have the same asymptotic distribution as M_n.

When $f_\theta(x) = e^{\theta x - K(\theta)} f_0(x)$, $y_\theta(x) = e^{\theta x - \frac{1}{2}K(2\theta)}$ and

$$S_n(\theta) = n^{-1/2}\sum_{i=1}^{n} \left(e^{\theta X_i - K(\theta)} - 1\right) e^{K(\theta)-\frac{1}{2}K(2\theta)} .$$

The method we outline below borrows heavily from Bickel & Chernoff (1993) who treated the case $F_0 = \mathcal{N}(0,1)$, but there are significant differences and generalisations. There are essentially three stages:

1. We show that there exists a sequence of mean-zero Gaussian processes $\{Z_n(\theta) \,|\, \theta \in \Theta_n, n \in \mathbb{N}\}$ such that

   $$\sup_{\theta\in\Theta_n} |S_n(\theta) - Z_n(\theta)| = O_p(r_n)$$

   for some $r_n \to 0$ as $n \to \infty$.

2. Defining

   $$\widetilde{Z}_n(\theta) = Z_n(\theta) + c_n(\theta)\widetilde{Y} ,$$

   where $\widetilde{Y}$ is a standard normal random variable independent of $\{Z_n(\theta) \,|\, \theta \in \Theta_n\}$ and $\{c_n(\theta) \,|\, \theta \in \Theta_n, n \in \mathbb{N}\}$ is a sequence of real-valued functions gives us a sequence of Gaussian processes with a nice covariance structure.

   We then present methods for deriving the asymptotic distribution of $\widetilde{M}_n = \sup_{\theta\in\Theta_n} \widetilde{Z}_n(\theta)$. In the cases we consider (that is where (12) holds) for a certain $B_n \to \infty$ depending on F_0,

   $$P(B_n(\widetilde{M}_n - B_n) + \log 2\pi \le x) \to e^{-e^{-x}} \tag{17}$$

   for each x as $n \to \infty$. Typically $B_n = O\left(\log\log n\right)^{1/2}$.

3. We show that for relatively large θ all of the processes $\widetilde{Z}_n(\theta)$, $Z_n(\theta)$ $S_n(\theta)$, $S_n^\dagger(\theta)$ and $S_n^*(\theta)$ are very close, and that for relatively small θ the contributions to their suprema are (asymptotically) negligible. Thus the limit theorem for $\widetilde{M}_n$ is applicable to M_n, and so also to $M_n^\dagger$ and M_n^*.

6 Approximation Theorem

Our first step is to approximate M_n with $\sup_{\theta\in\Theta_n} Z_n(\theta)$, where $\{Z_n(\theta) \mid \theta \in \Theta_n\}$ is a certain sequence of mean-zero Gaussian processes. This is done by viewing M_n as a certain function of a sample $U_1, U_2, \ldots, U_n$ of independent and identically distributed $U(0,1)$ random variables as follows:

$$\begin{aligned} M_n &= \sup_{\theta\in\Theta_n} S_n(\theta) \\ &= \sup_{\theta\in\Theta_n} n^{-1/2}\sum_{i=1}^{n} [y_\theta(X_i) - F_0 y_\theta] \\ &= \sup_{\theta\in\Theta_n} n^{-1/2}\sum_{i=1}^{n} \left[y_\theta \circ F_0^{-1}(U_i) - \int_0^1 y_\theta \circ F_0^{-1}(u)\, du\right] \\ &= \sup_{\theta\in\Theta_n} \int_0^1 y_\theta \circ F_0^{-1}(u)\, d\alpha_n(u) \ , \end{aligned} \tag{18}$$

where $F_0^{-1}(u) = \inf\{x \mid F_0(x) \geq u\}$ is the inverse cdf of the X_i's,

$$\alpha_n(u) = n^{-1/2}\{\mathbb{F}_n(u) - u\} \tag{19}$$

is the empirical process based on

$$\mathbb{F}_n(u) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{U_i \leq u\}} \ ,$$

the empirical cdf of $U_1, U_2, \ldots, U_n$. Note that if F_0 is right-continuous then F_0^{-1} is left-continuous. We present below a strong approximation theorem from Csorgo *et al.* (1986) which permits the approximation of the empirical process α_n by a sequence of Brownian Bridges. We then present a slight refinement suited to our purposes.

Let $\mathcal{L}^*$ denote the collection of all left-continuous nondecreasing functions defined on $(0,1)$. We say that any class $\mathcal{L}$ of functions defined on $(0,1)$ is $\mathcal{L}^*$-*decomposable* if each $\ell \in \mathcal{L}$ can be written as $\ell = \ell_1 - \ell_2$ where $\ell_i \in \mathcal{L}^*$, $i = 1, 2$.

Let L be a function defined on $(0, \frac{1}{2}]$ slowly varying near zero. For any $\mathcal{L}^*$-decomposable class $\mathcal{L}$ define

$$N(\delta, \mathcal{L}, L) = \sup_{\ell\in\mathcal{L}} \sup_{0\leq u\leq\delta} \left\{(|\ell_1(u)| + |\ell_2(u)| + |\ell_1(1-u)| + |\ell_2(1-u)|)\, u^{1/2}/L(u)\right\} \ .$$

Proposition 6.1 (Csorgo *et al.* (1986) Theorem 3.2). *Suppose $\mathcal{L}_n$,$n = 1, 2\ldots$ is a sequence of $\mathcal{L}^*$-decomposable classes of functions and that there exists a slowly-varying-near-zero function L defined on $(0, \frac{1}{2}]$ such that, with $\delta_n = n^{-1/2}\log n$,*

$$N(\delta_n, \mathcal{L}_n, L) \to 0 \tag{20}$$

as $n \to \infty$. Then there exists a probability space $(\Omega, \mathcal{B}, P)$ with independent $U(0, 1)$ random variables $U_1, U_2, \ldots$ and a sequence of Brownian Bridges

$$\{B_i(u) \,|\, 0 \leq u \leq 1, i = 1, 2, \ldots\} \tag{21}$$

such that the empirical process given by (19) satisfies

$$\frac{\sup_{\ell\in\mathcal{L}_n}\left|\int_0^1 \ell(u)d\alpha_n(u) - \int_{1/n}^{1-1/n}\ell(u)dB_n(u)\right|}{L(1/n)} = o_P(1)\ . \tag{22}$$

This theorem is proved by breaking up the left-hand side of (22) into pieces and bounding each piece above by a quantity of the form

$$O_P(1)N_n(\delta_n, \mathcal{L}_n, L)\ .$$

The condition (20) then gives the result. If it is possible to quantify $N_n(\delta_n, \mathcal{L}_n, L)$, then it is possible to quantify the rate of convergence of the $o_P(1)$ term. The proof of the following proposition, where we consider the special case where $L \equiv 1$ (which is certainly slowly varying near zero), is immediate.

Proposition 6.2. *Using the notation of Proposition 6.1, suppose that $N(\delta_n, \mathcal{L}_n, 1) = O(r_n)$, for some rate $r_n \to 0$ as $n \to \infty$. Then*

$$\sup_{\ell\in\mathcal{L}_n}\left|\int_0^1 \ell(u)d\alpha_n(u) - \int_{1/n}^{1-1/n}\ell(u)dB_n(u)\right| = O_P(r_n)\ . \tag{23}$$

If the functions in each class $\mathcal{L}_n$ are either non-decreasing or non-increasing, then $N(\delta, \mathcal{L}_n, 1)$ can be expressed in terms of the envelope function (sequence)

$$\check{\ell}_n = \sup_{\ell\in\mathcal{L}_n}|\ell|\ . \tag{24}$$

Proposition 6.3. *If all elements of each class $\mathcal{L}_n$ are left-continuous and either non-decreasing or non-increasing, then*

$$N(\delta, \mathcal{L}_n, 1) \leq \sup_{0 \leq u \leq \delta} \left\{ \left[\check{\ell}_n(u) + \check{\ell}_n(1-u) \right] u^{1/2} \right\} \ .$$

Proof. If each $\ell \in \mathcal{L}_n$ is left-continuous and either non-increasing or non-decreasing, then it can be written as $\ell = \ell_1 - \ell_2$, where $\ell_i \in \mathcal{L}^*$ and either $\ell_1 = \ell, \ell_2 \equiv 0$ or $\ell_2 = -\ell$, $\ell_1 \equiv 0$. In either case, $|\ell_1| + |\ell_2| = |\ell|$. So then

$$\begin{aligned} N(\delta, \mathcal{L}_n, 1) &= \sup_{\ell \in \mathcal{L}_n} \sup_{0 \leq u \leq \delta} \left\{ \left[|\ell(u)| + |\ell(1-u)| \right] u^{1/2} \right\} \\ &\leq \sup_{0 \leq u \leq \delta} \left\{ \left[\sup_{\ell \in \mathcal{L}_n} |\ell(u)| + \sup_{\ell \in \mathcal{L}_n} |\ell(1-u)| \right] u^{1/2} \right\} \end{aligned}$$

□

We are now in a position to address our original problem. We are interested in $M_n = \sup_{\theta \in \Theta_n} S_n(\theta)$ where

$$S_n(\theta) \ = \ \int_0^1 y_\theta \circ F_0^{-1}(u) \, d\alpha_n(u)$$

and $U_1, U_2, \ldots, U_n$ are those $U(0,1)$ random variables referred to in Proposition 6.1. Define now

$$Z_n(\theta) = \int_{1/n}^{1-1/n} y_\theta \circ F_0^{-1}(u) \, dB_n(u) \ , \tag{25}$$

where $\{B_n(u) \,|\, 0 \leq u \leq 1, \, n \in \mathbb{N}\}$ are the Brownian Bridges referred to in Proposition 6.1. Then $\{Z_n(\theta) \,|\, \theta \in \Theta, n \in \mathbb{N}\}$ is a sequence of mean-zero Gaussian processes.

We state the following condition:

Condition 1 (Monotone likelihood ratio). *The family $\mathcal{F} = \{F_\theta \,|\, \theta \in \Theta \subset \mathbb{R}\}$ is said to have monotone likelihood ratio if for $\theta \neq \eta$ both in Θ, the distributions F_θ and F_η are distinct, and for $\eta < \theta$, the ratio of the densities f_θ / f_η is a nondecreasing function on the sample space.*

We note that this is a special case of the more general definition of monotone likelihood ratio given in, for example, Lehmann (1986), which is in terms of a statistic $T(x)$ which we have taken here as $T(x) = x$.

Define the envelope function of $\mathcal{Y}_n = \{y_\theta \,|\, \theta \in \Theta_n\}$ as

$$\check{y}_n(x) = \sup_{\theta\in\Theta_n} |y_\theta(x)| \ . \tag{26}$$

We now come to our first main theorem.

Theorem 6.4. *Suppose the family $\mathcal{F} = \{F_\theta \,|\, \theta \in \Theta \subset \mathbb{R}\}$ has monotone likelihood ratio. If in addition the envelope function defined in* (26) *satisfies (with $\delta_n = n^{-1/2}\log n$)*

$$\sup_{F_0(x)\le\delta_n} \check{y}_n(x) F_0(x)^{1/2} = O(r_n) \tag{27}$$

and

$$\sup_{F_0(x)\ge 1-\delta_n} \check{y}_n(x) \left[1 - F_0(x)\right]^{1/2} = O(r_n) \tag{28}$$

for some rate $r_n \to 0$ as $n \to \infty$, then $S_n(\theta)$ defined in (15) *and $Z_n(\theta)$ defined in* (25) *satisfy*

$$\sup_{\theta\in\Theta_n} |S_n(\theta) - Z_n(\theta)| = O_P(r_n) \ . \tag{29}$$

Proof. The monotone likelihood ratio condition implies that the functions y_θ, for $\theta > 0$ are nondecreasing and for $\theta < 0$ are nonincreasing. Thus proposition 6.3 gives that

$$\begin{aligned} &N(\delta_n, \mathcal{L}_n, 1) \\ &\quad\le \sup_{0\le u\le\delta_n} \left\{ \left[\check{y}_n \circ F_0^{-1}(u) + \check{y}_n \circ F_0^{-1}(1-u)\right] u^{1/2} \right\} \ , \\ &\quad\le \sup_{0\le u\le\delta_n} \left\{ \left[\check{y}_n \circ F_0^{-1}(u)\right] u^{1/2} \right\} + \sup_{0\le u\le\delta_n} \left\{ \left[\check{y}_n \circ F_0^{-1}(1-u)\right] u^{1/2} \right\} \end{aligned} \tag{30}$$

since $\check{\ell}_n = \check{y}_n \circ F_0^{-1}$. Next note that

$$F_0^{-1} \circ F_0(x) = \inf\{w \,|\, F(w) \ge F(x)\} \le x$$

since x is included in the set over which the infimum is taken. Then writing $u = F_0(x)$ in the first term in (30), and $u = 1 - F_0(x)$ in the second gives

$$\begin{aligned} N(\delta_n, \mathcal{L}_n, 1) &\le \sup_{F_0(x)\le\delta_n} \check{y}_n(x) F_0(x)^{1/2} + \sup_{F_0(x)\ge 1-\delta_n} \check{y}_n(x) \left[1 - F_0(x)\right]^{1/2} \\ &= O\left(r_n\right) \end{aligned}$$

by (27) and (28). Proposition 6.2 then implies (29). □

6.1 Linear exponential family examples

In this subsection we apply Theorem 6.4 to each of the examples from section 5.3. We need to show that they have monotone likelihood ratio and we need to compute the rate r_n from (29). For each the density is of the form

$$f_\theta(x) = e^{\theta x - K(\theta)} f_0(x) \ ,$$

so they all have monotone likelihood ratio; for $\eta < \theta$,

$$f_\theta(x)/f_\eta(x) = \exp\{x(\theta - \eta) - K(\theta) + K(\eta)\}$$

which is nondecreasing in x. So all that remains is to find r_n for each case. We outline below a general approach.

To use Theorem 6.4 we need to compute the envelope functions for the classes $\mathcal{Y}_n = \{y_\theta \,|\, \theta \in \Theta_n\}$ where

$$y_\theta(x) = e^{\theta x - \frac{1}{2}K(2\theta)} \ .$$

The functions y_θ are such that this is relatively easy:

$$\partial y_\theta(x)/\partial\theta \ \ = \ \ [x - K'(2\theta)]y_\theta(x) \ .$$

Recalling that $\hat{\theta}(x)$ is the solution of $\mu_\theta = K'(\theta) = x$, since K' is necessarily increasing (see Barndorff-Nielsen (1978, 1980)), as a function of θ

$$y_\theta(x) \ \text{ is } \begin{cases} \text{increasing in } \theta & \text{for } \theta < \hat{\theta}(x)/2 \ , \\ \text{decreasing in } \theta & \text{for } \theta > \hat{\theta}(x)/2 \end{cases}$$

for fixed x. So if the interval Θ_n is of the form $\Theta \cap [\theta_n^-, \theta_n^+]$, then the envelope function is given by

$$\check{y}_n(x) = \begin{cases} y_{\theta_n^-}(x) & \text{for } x \le \mu_{2\theta_n^-} \\ y_{\hat{\theta}(x)/2}(x) & \text{for } \mu_{2\theta_n^-} < x < \mu_{2\theta_n^+} \\ y_{\theta_n^+}(x) & \text{for } x \ge \mu_{2\theta_n^+} \ , \end{cases} \tag{31}$$

since $2\theta < \hat{\theta}(x) \Leftrightarrow \mu_{2\theta} < x$.

We also need to approximate tail probabilities and quantiles. Appendix C details an integration-by-parts approach for approximating tail probabilities, applicable in many of our examples. In particular conditions are given under which

$$\frac{f_0(x)}{-G'(x)}[1+r_1(x)] \le F_0(x) \le \frac{f_0(x)}{-G'(x)} ,$$
$$\frac{f_0(x)}{G'(x)}[1+r_2(x)] \le 1-F_0(x) \le \frac{f_0(x)}{G'(x)} ,$$

where $G = -\log f_0$, $r_1(x) \to 0$ as $F_0(x) \to 0$ and $r_2(x) \to 0$ as $F_0(x) \to 1$. In cases where $\Theta_n = \Theta$, or at least when $\mu_{2\theta_n^-} < x < \mu_{2\theta_n^+}$, we then have that, for suitably small x,

$$\begin{aligned}
\check{y}_n(x)F_0(x)^{1/2} &= y_{\hat{\theta}(x)/2}(x)F_0(x)^{1/2} \\
&\le \exp\Big\{\tfrac{1}{2}[\hat{\theta}(x)x - K(\hat{\theta}(x))]\Big\}\left(\frac{f_0(x)}{-G'(x)}\right)^{1/2} \\
&= \left(\frac{f_{\hat{\theta}(x)}(x)}{-G'(x)}\right)^{1/2} . \qquad (32)
\end{aligned}$$

Similarly for the upper tail

$$\check{y}_n(x)[1-F_0(x)]^{1/2} \le \left(\frac{f_{\hat{\theta}(x)}(x)}{G'(x)}\right)^{1/2} \qquad (33)$$

for suitably large x (see appendix C for details).

We also give in appendix C.3 an approach for approximating quantiles based on a convergent iterative scheme method, which applies in many cases where an approximation for a tail probability is dominated by an exponential factor.

6.1.1 Normal mean mixtures

For $F_0 = \mathcal{N}(0,1)$, $K(\theta) = \theta^2/2$, $\mu_\theta = K'(\theta) = \theta$, so $\hat{\theta}(x) = x$. Thus $f_{\hat{\theta}(x)}(x) \equiv (2\pi)^{-1/2}$. Also $G(x) = x^2/2$, so $G'(x) = x$. It is well-known that (32) and (33) hold in this case. So then for any $x < 0$,

$$\check{y}_n(x)F_0(x)^{1/2} \le \frac{1}{\sqrt{-x2\pi}} .$$

Similarly for any $x > 0$,

$$\check{y}_n(x)[1 - F_0(x)]^{1/2} \leq \frac{1}{\sqrt{x 2\pi}} \ .$$

Defining x_n via $1 - F_0(x_n) = \delta_n = n^{-1/2} \log n$, we have from (108) in appendix C.3 that

$$x_n^2 = \log n - 3 \log \log n - \log 2\pi + o(1) \ .$$

So then

$$\begin{aligned} \sup_{F_0(x) \geq 1 - \delta_n} \check{y}_n(x)[1 - F_0(x)]^{1/2} &\leq \sup_{F_0(x) \geq 1 - \delta_n} \frac{1}{\sqrt{x 2\pi}} \\ &= \frac{1}{\sqrt{x_n 2\pi}} \\ &= O(\log n)^{-1/4} \ , \end{aligned}$$

and by symmetry the same holds for the lower tail. So proposition 6.4 holds with $r_n = (\log n)^{-1/4}$ for $\Theta_n \equiv \Theta = \mathbb{R}$.

A faster rate of convergence to zero can be obtained by restricting the parameter space; if $\Theta_n = [-\theta_n, \theta_n]$ where $\theta_n = \hat{\theta}(x_n)/2 - \Delta_n = x_n/2 - \Delta_n$, for some $0 < \Delta_n = o\,(x_n) = o(\log n)^{1/2}$ tending to infinity, we have, for large x,

$$\check{y}_n(x) = y_{\theta_n}(x) = e^{x\theta_n - \theta^2} \ .$$

So then

$$\begin{aligned} \check{y}_n(x_n)[1 - F_0(x_n)]^{1/2} &\sim \frac{e^{x_n\theta_n - \theta^2 - x_n^2/4}}{x_n^{1/2}(2\pi)^{1/4}} \\ &= O\left(\exp\left\{-(x_n/2 - \theta_n)^2\right\} x_n^{-1/2}\right) \\ &= O\left(\exp\left\{-\Delta_n^2\right\} (\log n)^{-1/4}\right) \ . \end{aligned}$$

The same holds in the lower tail by symmetry.

6.1.2 Gamma shape mixtures

In what follows we use the following approximations for the Gamma function and its derivatives. We are only interested in the behaviour for very large and very small values. For details see appendix A.2 of Jensen (1995) which in turn refers to Section 6 of Abramowitz & Stegun (1964).

$$
\begin{aligned}
\log \Gamma(x) &\approx \begin{cases} -\log x & \text{for small } x\ . \\ (x-\frac{1}{2})\log x - x + \frac{1}{2}\log(2\pi) & \text{for large } x\ , \end{cases} \\
\psi(x) &= \frac{d}{dx}\log\Gamma(x) \approx \begin{cases} \psi(1) - \frac{1}{x} & \text{for small } x\ , \\ \log x & \text{for large } x\ , \end{cases} \\
\psi'(x) &= \frac{d^2}{dx^2}\log\Gamma(x) \approx \begin{cases} \frac{1}{x^2} & \text{for small } x\ , \\ \frac{1}{x} & \text{for large } x\ . \end{cases}
\end{aligned}
$$

where $-\psi(1) \approx 0.577$ is Euler's constant.

Let F_0 be the distribution on $\mathbb{R}$ with density

$$f_0(x) = e^{\alpha_0 x - e^x - \log\Gamma(\alpha_0)}\ .$$

The cumulant generating function is $K(\theta) = \log\Gamma(\alpha_0+\theta) - \log\Gamma(\alpha_0)$, for $\theta \in (-\alpha_0, \infty) = \widetilde{\Theta}$. So $\mu_\theta = K'(\theta) = \psi(\alpha_0+\theta)$. According to appendix C, the integration-by-parts method of approximating tail probabilities is valid for both tails, so we can use both (32) and (33).

We start with the upper tail. $x = \mu_{\hat\theta(x)} \approx \log(\alpha_0 + \hat\theta(x))$ for large x. So we approximate $\hat\theta(x) \approx e^x - \alpha_0$. Then

$$
\begin{aligned}
f_{\hat\theta(x)}(x) &= e^{x[\alpha_0+\hat\theta(x)] - e^x - \log\Gamma(\alpha_0+\hat\theta(x))} \\
&\approx \exp\big\{xe^x - e^x - [(e^x - \tfrac{1}{2})x - e^x + \tfrac{1}{2}\log(2\pi)]\big\} \\
&= \frac{e^{x/2}}{\sqrt{2\pi}}\ .
\end{aligned}
$$

Also $G'(x) = \frac{d}{dx}\log f_0(x) = e^x - \alpha_0$. So then

$$
\begin{aligned}
\left(\frac{f_{\hat\theta(x)}(x)}{G'(x)}\right)^{1/2} &= O\left(\frac{e^{x/2}}{e^x-\alpha_0}\right)^{1/2} \\
&= O\left(e^{-x/4}\right)
\end{aligned}
$$

as $x \to \infty$.

Writing $x_n = F_0^{-1}(1-\delta_n)$, where $\delta_n = n^{-1/2}\log n$, we have from (111) in appendix C.3 that x_n satisfies

$$e^{x_n} = \tfrac{1}{2}\log n + (2-\alpha_0)\log\log n + \log\Gamma(\alpha_0) + (\alpha_0 - 1)\log 2\ ,$$

so

$$\begin{aligned}
\sup_{F_0(x)\geq 1-\delta_n} \check{y}(x)[1-F_0(x)]^{1/2} &\leq \sup_{F_0(x)\geq 1-\delta_n} \left(\frac{f_{\hat\theta(x)}(x)}{G'(x)}\right)^{1/2} \\
&= \left(\frac{f_{\hat\theta(x_n)}(x_n)}{G'(x_n)}\right)^{1/2} \\
&= O(\log n)^{-1/4} \ .
\end{aligned}$$

Now for the lower tail. For small x, writing $\psi = \psi(1)$,

$$\begin{aligned}
x &= K'(\hat\theta(x)) \approx \psi - \frac{1}{\alpha_0 + \hat\theta(x)} \\
\hat\theta(x) &\approx \frac{1}{\psi - x} - \alpha_0 \ .
\end{aligned}$$

So for small x,

$$\begin{aligned}
f_{\hat\theta(x)}(x) &= \exp\Big\{x(\alpha_0 + \hat\theta(x)) - e^x - [\log\Gamma(\alpha_0 + \hat\theta(x))]\Big\} \\
&\approx \exp\left\{\frac{x}{\psi - x} - e^x - \log(\psi - x)\right\} \\
&= \frac{e^{\frac{x}{\psi - x}} e^{-e^x}}{\psi - x} \sim -\frac{e^{-1}}{x} \\
&= O(|x|)^{-1} \ \text{ as } x \to -\infty \ .
\end{aligned}$$

Also $G'(x) = \alpha_0 - e^x \to \alpha_0$ as $x \to -\infty$, and $y_n = F_0^{-1}(\delta_n)$ satisfies

$$y_n = \frac{1}{\alpha_0}\left[-\tfrac{1}{2}\log n + \log\log n + \log\Gamma(\alpha_0)\right] + o(1)$$

(see (112) in appendix C.3). So

$$\begin{aligned}
\sup_{F_0(x)\leq \delta_n} \check{y}(x) F_0(x)^{1/2} &\leq \sup_{F_0(x)\leq\delta_n} \left[\frac{f_{\hat\theta(x)}(x)}{-G'(x)}\right]^{1/2} \\
&= \left[\frac{f_{\hat\theta(x_n)}(x_n)}{-G'(x_n)}\right]^{1/2} \\
&= O(\log n)^{-1/2} \ .
\end{aligned}$$

6.1.3 Gamma scale mixtures

Let $F_0 = \text{Gamma}(\alpha_0, 1)$. Then

$$f_\theta(x) = \exp\{\theta x + [(\alpha_0 - 1)\log x - x - \log\Gamma(\alpha_0)] - [-\alpha_0 \log(1-\theta)]\} \ ,$$

$K(\theta) = -\alpha_0 \log(1-\theta)$, $K'(\theta) = \alpha_0/(1-\theta)$ and

$$\hat{\theta}(x) = \frac{x - \alpha_0}{x} \ .$$

So the unrestricted envelope function is

$$\begin{aligned} \check{y}(x) &= \exp\left\{\frac{x\hat{\theta}(x)}{2} - \frac{1}{2}K(\hat{\theta}(x))\right\} \\ &= e^{\frac{x-\alpha_0}{2}} \left(\frac{\alpha_0}{x}\right)^{\alpha_0/2} \ . \end{aligned} \tag{34}$$

From (101) we get that

$$1 - F_0(x) \sim \frac{x^{\alpha_0 - 1} e^{-x}}{\Gamma(\alpha_0)} \tag{35}$$

as $x \to \infty$. So

$$\begin{aligned} \check{y}(x)[1 - F_0(x)]^{-1/2} &\sim e^{\frac{x-\alpha_0}{2}} \left(\frac{\alpha_0}{x}\right)^{\alpha_0/2} \left[\frac{x^{\alpha_0-1}e^{-x}}{\Gamma(\alpha_0)}\right]^{1/2} \\ &= \frac{e^{-\alpha_0/2}\alpha_0^{\alpha_0/2}}{\sqrt{x\Gamma(\alpha_0)}} \\ &= O(x)^{-1/2} \end{aligned}$$

as $x \to \infty$.

Since (35) is true, (see appendix C), $x_n = F_0^{-1}(1 - \delta_n)$ where $\delta_n = n^{-1/2}\log n$ is of the form

$$x_n = \tfrac{1}{2}\log n + (\alpha - 2)\log\log n - \log\Gamma(\alpha_0) - (\alpha_0 - 1)\log 2 + o(1) \ .$$

So for the upper tail bound

$$\sup_{F_0(x) \geq 1 - \delta_n} \check{y}(x)[1 - F_0(x)]^{1/2} = O(x_n)^{-1/2} = O(\log n)^{-1/2} \ . \tag{36}$$

As for the lower tail bound, (103) or (102) give that

$$F_0(x) \sim \frac{x^{\alpha_0}}{\Gamma(\alpha_0+1)} \tag{37}$$

so from (34), as $x \to 0$ we get

$$\check{y}(x)F_0(x)^{1/2} \to \left(\frac{\alpha_0}{e}\right)^{\alpha_0/2} \Gamma(\alpha_0+1)^{-1/2} \ ,$$

not tending to zero.

So a further restriction on the parameter space is needed so that this lower tail bound goes to zero, moreover at a sufficiently fast rate. We describe below the effect of restricting the parameter space to $\Theta_n = \left(\theta_n, \frac{1}{2}\right)$ with $\theta_n \to -\infty$ sufficiently slowly.

Redefining $x_n = F_0^{-1}(\delta_n)$, (37) gives us that

$$\delta_n = F_0(x_n) \sim \frac{x_n^{\alpha_0}}{\Gamma(\alpha_0)} \ .$$

so that

$$x_n = O(\delta_n)^{1/\alpha_0} \ . \tag{38}$$

For any $\theta_n \to -\infty$, the envelope function evaluated at x_n (according to (31)) is

$$\check{y}_n(x_n) = y_{\theta_n}(x) = e^{\theta_n x_n + \frac{\alpha_0}{2}\log(1-2\theta_n)} \ ,$$

so long as $\theta_n \geq \hat{\theta}(x_n)/2$.

Let $\delta_n^* \to 0$ in such a way that $\delta_n/\delta_n^* \to 0$, that is $\delta_n = o(\delta_n^*)$. If we then define $x_n^* = F_0^{-1}(\delta_n^*)$ and $\theta_n = \hat{\theta}(x_n^*)/2$, note that $x_n^* = O(\delta_n^*)^{1/\alpha_0}$, and $x_n/x_n^* \to 0$. So then since $\hat{\theta}(x) = 1 - \alpha_0/x$,

$$\begin{aligned}
\check{y}_n(x_n)F_0(x_n)^{1/2} &= e^{x_n\hat{\theta}(x_n^*)/2}[1-\hat{\theta}(x_n^*)]^{\alpha_0/2}\delta_n^{1/2} \\
&= \exp\left\{\frac{x_n}{2} - \frac{\alpha_0 x_n}{2x_n^*}\right\}\left(\frac{\alpha_0}{x_n^*}\right)^{\alpha_0/2}\delta_n^{1/2} \\
&= O\left(\frac{\delta_n}{\delta_n^*}\right)^{1/2}
\end{aligned}$$

since both terms in the exponent go to zero and $x_n^* = O(\delta_n^*)^{1/\alpha_0}$.

So if we want to match the upper tail and have $\check{y}_n(x_n)F_0(x_n)^{1/2} = O(\log n)^{-1/2}$, choose any $\delta_n^* = C(\log n)^{3/2}n^{-1/2}$ for some constant C so that $\delta_n/\delta_n^* = O(\log n)^{-1/2}$, in which case

$$\begin{aligned} x_n^* &\sim \left\{C(\log n)^{3/2}n^{-1/2}\Gamma(\alpha_0)\right\}^{1/\alpha_0} \\ \text{and } \theta_n &= \hat{\theta}(x_n^*)/2 \\ &\approx \frac{1}{2} - \frac{\alpha_0}{x_n^*} \\ &= \frac{1}{2} - \frac{C^* n^{1/(2\alpha_0)}}{(\log n)^{3/(2\alpha_0)}} \end{aligned}$$

for another constant C^*. So it seems that any sequence $\theta_n \to -\infty$ of order $O\left(n^{1/2}(\log n)^{-3/2}\right)^{1/\alpha_0}$ yields a sequence of families

$$\mathcal{F}_n = \left\{F_\theta \,\middle|\, \theta_n < \theta < \tfrac{1}{2}\right\}$$

satisfying theorem 6.4 with the rate in (29) equal to $(\log n)^{-1/2}$.

The fact that this is only determined up to a multiplicative constant means that it is difficult to specify in practice, and so we note that this is of primarily theoretical interest, at least as an example of a sequence of families with a changing parameter space.

6.1.4 Poisson

For the case $F_0 = \text{Pois}(\lambda_0)$, $K(\theta) = \lambda_0(e^\theta - 1)$, $K'(\theta) = \lambda_0 e^\theta$, $\hat{\theta}(x) = \log(x/\lambda_0)$. So the envelope function is

$$\begin{aligned} \check{y}(x) &= y_{\hat{\theta}(x)/2}(x) \\ &= \exp\left\{\frac{x\hat{\theta}(x)}{2} - \frac{1}{2}K(\hat{\theta}(x))\right\} \\ &= \exp\left\{\frac{x}{2}\log\left(\frac{x}{\lambda_0}\right) - \frac{\lambda_0}{2}\left(e^{\log(x/\lambda_0)} - 1\right)\right\} \\ &= \left(\frac{x}{\lambda_0}\right)^{x/2} e^{-x/2}e^{\lambda_0/2}\ . \end{aligned}$$

Also from (107), as $x \to \infty$,

$$1 - F_0(x) \sim \frac{e^{-\lambda_0}\lambda_0^x}{x!} \sim e^{-\lambda_0}\left(\frac{\lambda_0 e}{x}\right)^x (2\pi x)^{-1/2}$$

using Stirlings formula. So

$$\begin{aligned}\check{y}(x)[1-F_0(x)]^{1/2} &\sim \left(\frac{x}{\lambda_0}\right)^{x/2} e^{-x/2}e^{\lambda_0/2}\left[e^{-\lambda_0}\left(\frac{\lambda_0 e}{x}\right)^{x}(2\pi x)^{-1/2}\right]^{1/2} \\ &= (2\pi x)^{-1/4}\end{aligned}$$

as $x \longrightarrow \infty$.

Defining $x_n = F_0^{-1}(1-\delta_n)$, where again $\delta_n = n^{-1/2}\log n$, we have from appendix C.3 that

$$\begin{aligned}x_n &= \frac{\frac{1}{2}\log n - \frac{3}{2}\log_2 n + \frac{1}{2}\log_3 n + \frac{1}{2}\log 4\pi - \lambda_0}{\log_2 n - \log_3 n - \log 2} + o(1) \\ &= O\left(\frac{\log n}{\log_2 n}\right) .\end{aligned}$$

So then

$$\check{y}(x_n)[1-F_0(x_n)]^{1/2} = O\left(\frac{\log_2 n}{\log n}\right)^{1/4} .$$

For the lower tail, for large enough n $F_0(0) > \delta_n$, so the set $\{x \mid F_0(x) \leq \delta_n\}$ is empty and only the upper tail is relevant.

6.1.5 Negative Binomial

Let $F_0 = \text{Neg}(m, q_0)$. Then $K(\theta) = m\left[\log(1-q_0) - \log\left(1-q_0e^{\theta}\right)\right]$ and $K'(\theta) = mq_0e^{\theta}/\left[1-q_0e^{\theta}\right]$. So

$$\hat{\theta}(x) = \log\left[\frac{x}{q_0(m+x)}\right] .$$

So

$$K(\hat{\theta}(x)) = m\log\left[\frac{(1-q)(m+x)}{m}\right]$$

and the unrestricted envelope function is

$$\begin{aligned}\check{y}(x) &= y_{\hat{\theta}(x)/2}(x)\\ &= \exp\left\{\frac{x\hat{\theta}(x)}{2} - \frac{K(\hat{\theta}(x))}{2}\right\}\\ &= \exp\left\{\frac{x}{2}\log\left[\frac{x}{q(x+m)}\right] + \frac{m}{2}\log\left[\frac{m}{(1-q)(x+m)}\right]\right\}\\ &= \left[\frac{x}{q(x+m)}\right]^{x/2}\left[\frac{m}{(1-q)(x+m)}\right]^{m/2} .\end{aligned}$$

From appendix C.2, we have that

$$\log[1-F_0(x)] = x\log q_0 + (m-1)\log x + C + o(1)$$

for some uniformly bounded (in $x \geq 0$) constant $C = C_x$ and that $1-F_0(x) = O\left(q_0^x x^{m-1}\right)$ as $x \to \infty$.

So we have that

$$\begin{aligned}\check{y}(x)[1-F_0(x)]^{1/2} &\sim C\left[\frac{x}{q_0(x+m)}\right]^{x/2}\left[\frac{m}{(1-q_0)(x+m)}\right]^{m/2} q_0^{x/2}x^{(m-1)/2}\\ &= C\left[\frac{m}{1-q}\right]^{m/2}\left[\frac{x}{x+m}\right]^{\frac{x+m}{2}} x^{-1/2}\\ &= C\left[\frac{m}{1-q}\right]^{m/2}\left[1-\frac{m}{x+m}\right]^{(x+m)/2} x^{-1/2}\\ &\sim C\left[\frac{m}{1-q}\right]^{m/2} e^{-m/2}x^{-1/2}\\ &= O\left(x^{-1/2}\right) .\end{aligned}$$

Defining $x_n = F_0^{-1}(1-\delta_n)$, where $\delta_n = n^{-1/2}\log n$ we have from appendix C.3 that

$$x_n = \frac{\frac{1}{2}\log n - m\log_2 n - C + (m-1)[\log\frac{1}{2} + \log(-\log q_0)]}{-\log q_0} + o(1)$$

as $n \to \infty$. In particular $x_n = O(\log n)$. So then

$$\check{y}(x_n)[1-F_0(x_n)]^{1/2} = O(\log n)^{-1/2} .$$

For the lower tail, for large enough n $F_0(0) > \delta_n$, so the set $\{x \mid F_0(x) \leq \delta_n\}$ is empty and only the upper tail is relevant.

7 Gaussian Processes

We now consider the second stage of our general procedure. Having shown that

$$\sup_{\theta\in\Theta_n} |S_n(\theta) - Z_n(\theta)| = O_p(r_n)$$

for some rate $r_n \to 0$, we now derive the limiting of distribution of $\sup_{\theta\in\Theta_n} Z_n(\theta)$, or more precisely $\sup_{\theta\in\Theta_n} \widetilde{Z}_n(\theta)$, where $\widetilde{Z}_n$ is similar to Z_n but has a simpler covariance function.

Define $a_n = F_0^{-1}(1/n)$ and $b_n = F_0^{-1}(1-1/n)$, and write $A_n = [a_n, b_n]$. The covariance of $Z_n(\theta)$ and $Z_n(\eta)$ is the same as that of the random variables $U_n(\theta)$ and $U_n(\eta)$ where

$$U_n(\theta) = e^{\theta X - \frac{1}{2}K(2\theta)} 1_{A_n}(X) \ , \quad \text{and } X \sim F_0 \ .$$

We have

$$\begin{aligned} EU_n(\theta) &= \int_{A_n} e^{\theta x - \frac{1}{2}K(2\theta)} f_0(x)\,dx \\ &= e^{K(\theta) - \frac{1}{2}K(2\theta)} \int_{A_n} e^{\theta x - K(\theta)} f_0(x)\,dx \\ &= e^{K(\theta) - \frac{1}{2}K(2\theta)} F_\theta A_n \end{aligned}$$

and

$$\begin{aligned} EU_n(\theta)U_n(\eta) &= \int_{A_n} e^{\theta x - \frac{1}{2}K(2\theta)} e^{\eta x - \frac{1}{2}K(2\eta)} f_0(x)\,dx \\ &= e^{K(\theta+\eta) - \frac{1}{2}K(2\theta) - \frac{1}{2}K(2\eta)} \int_{A_n} e^{(\theta+\eta)x - K(\theta+\eta)} f_0(x)\,dx \\ &=: \rho(\theta,\eta) F_{\theta+\eta} A_n \ . \end{aligned}$$

Writing

$$\rho_0(\theta) = \rho(0,\theta) = e^{K(\theta) - \frac{1}{2}K(2\theta)} \ ,$$

we can write

$$\mathrm{Cov}(U_n(\theta), U_n(\eta)) = \rho(\theta,\eta) F_{\theta+\eta} A_n - (\rho_0(\theta) F_\theta A_n)(\rho_0(\eta) F_\eta A_n) \ .$$

This is the covariance function of $Z_n(\theta)$. As in Bickel & Chernoff (1993) we obtain a new process with a less complicated covariance function by adding a relatively insignificant term:

$$\widetilde{Z}_n(\theta) = Z_n(\theta) + \tilde{X}\rho_0(\theta)F_\theta A_n \ ,$$

where $\widetilde{X} \sim \mathcal{N}(0,1)$ independently of $Z_n(\theta)$. The new process $\widetilde{Z}_n(\theta)$ now has covariance function

$$\rho(\theta,\eta)F_{\theta+\eta}A_n \ .$$

The idea is that the extra term only makes a difference for small θ. We show later that the contribution to the maximum for small θ can be ignored, and so it suffices to derive the limiting distribution of

$$\widetilde{M}_n \;=\; \sup_{\theta\in\Theta_n} \widetilde{Z}_n(\theta) \ .$$

Now $\widetilde{Z}_n(\theta)$ has mean zero but is not standardised; the variance of $\widetilde{Z}_n(\theta)$ is $F_{2\theta}A_n$, that is the amount of weight the distribution $F_{2\theta}$ assigns to the interval $A_n = [a_n, b_n]$. For $\theta = 0$ this is $1 - 2/n$. For θ close to 0 this should be close to 1, for θ far from 0 this should be close to 0. It turns out that for θ values in the bulk of the interval $[\hat{\theta}(a_n)/2, \hat{\theta}(b_n)/2]$, $Var\widetilde{Z}_n(\theta) \approx 1$, for most of the range outside this interval $Var\widetilde{Z}_n(\theta) \approx 0$, and there are relatively short transitional intervals containing $\hat{\theta}(a_n)/2$ and $\hat{\theta}(b_n)/2$ where $Var\widetilde{Z}_n(\theta)$ is somewhere between 0 and 1.

7.1 Extreme values of Gaussian processes

We now present an overview of results for the asymptotic distribution of the maxima of stationary and locally stationary Gaussian processes. Suppose $\{X(\theta) \,|\, \theta \geq 0\}$ is a standardised Gaussian process with correlation function

$$\rho(\theta,\eta) = EX(\theta)X(\eta) \ .$$

If $\rho(\theta,\eta) = r(|\theta-\eta|)$ for some symmetric function r then the process is *stationary*. If r can be written as

$$r(h) = 1 - \frac{\lambda h^2}{2} + o(h^2)$$

as $h \to 0$, then the process is *differentiable in quadratic mean* (or just differentiable). That is there is a process $\{X'(\theta) \mid \theta \geq 0\}$ such that as $h \to 0$,

$$Var\left(X'(\theta) - \frac{X(\theta+h) - X(\theta)}{h}\right) \to 0$$

and $X'(\theta) \sim \mathcal{N}(0, \lambda)$ independently of $X(\theta)$ for each θ. So we can interpret the quantity λ as the variance of the derivative of X.

Any differentiable stationary standardised Gaussian process can be transformed into one which has *standardised derivative* by a change of scale. The process

$$\left\{Y(t) = X(\lambda^{-1/2}t) \mid t \geq 0\right\} \tag{39}$$

satisfies

$$\begin{aligned} EY(t)Y(t+h) &= EX(\lambda^{-1/2}t)X(\lambda^{-1/2}(t+h)) \\ &= 1 - \frac{h^2}{2} + o(h^2) \end{aligned} \tag{40}$$

as $h \to 0$.

Let $\{Y(t) \mid t \geq 0\}$ be a standardised, stationary Gaussian process with standardised derivative, and define $M(T) = \sup\{Y(t) \mid 0 \leq t \leq T\}$. We now summarise two slightly different approaches to the derivation of the asymptotic distribution of $M(T)$ as $T \to \infty$.

If u_T satisfies

$$\frac{Te^{-u_T^2/2}}{2\pi} \to C \ , \tag{41}$$

then so long as r decays quickly enough, as $T \to \infty$,

$$P\{M(T) \leq u_T\} \to e^{-C} \ . \tag{42}$$

In particular, if we write $a_T = (2\log T)^{1/2}$, and define $u_T = a_T + \frac{x - \log 2\pi}{a_T}$, we have that

$$\frac{Te^{-u_T^2/2}}{2\pi} = e^{-x}[1 + o(1)] \ ,$$

so that

$$P\{a_T[M(T) - a_T] + \log 2\pi \leq x\} \to \exp\left\{-e^{-x}\right\} \ , \tag{43}$$

the cumulative distribution function of the so-called Gumbel distribution. See Leadbetter *et al.* (1983) for the details of the proof of this result under the decay condition

$$r(t)\log t \to 0 \tag{44}$$

as $t \to \infty$.

A refinement of this result is derived in Piterbarg (1996). Write

$$\ell_T = \sqrt{2\log\frac{T}{2\pi}} \ .$$

Suppose that for some $a > 0$ the apparently stronger decay condition

$$\int_0^\infty |r(t)|^a \, dt < \infty \tag{45}$$

holds. Then there exists a $q > 0$ such that

$$P\left\{\ell_T[M(T) - \ell_T] < x\right\} = \exp\left\{-e^{-x-x^2/2\ell_T^2}\right\} + O(T^{-q}) \ . \tag{46}$$

This is not a limit theorem as such, as the first term on the right hand side is not a distribution function, but an approximating function that converges to a distribution function. Noting that

$$\exp\left\{-e^{-x-x^2/2\ell_T^2}\right\} = \exp -e^{-x}\left(1 + e^{-x}\frac{x^2}{2\ell_T^2}\right) \ ,$$

we have that

$$P\left\{\ell_T[M(T) - \ell_T] < x\right\} = \exp -e^{-x} + O(\log T)^{-1} \ . \tag{47}$$

So the polynomial rate of convergence of the approximating functions in (46) is an improvement on the logarithmic rate of convergence of the limit theorem form (47).

A result of Hall (1991) is relevant here. It says that if the correlation function satisfies

$$\begin{aligned} r(t) &= 1 - \frac{\lambda_2 t^2}{2} + \frac{\lambda_4 t^4}{24} + o(t^4) \quad \text{as } t \to 0 \ , \\ r(t) &\to 0 \ \text{ as } t \to \infty \ \ , \ \int_0^\infty r(t)^2 \, dt < \infty \end{aligned}$$

then the rate of convergence to the Gumbel distribution is $O(\log T)^{-1}$, even if the convergence is to be uniform over as few as three points, for example the upper 0.1, 0.05 and 0.01 quantiles. We refer back to this point in a later section.

The method of proof of (43) does not depend heavily on the stationarity of the process; the important features are the local behaviour described by conditions like (39) and (40), and long range behaviour, described by decay conditions like (44) and (45). Analogous results for processes which satisfy these conditions but are not necessarily stationary are derived in Hüsler (1990, 1995). The processes are termed *locally stationary* following a definition given in Berman (1974).

A standardised Gaussian process $\{X(\theta) \mid \theta \geq 0\}$ is said to be *locally stationary* if the correlation function can be written as

$$EX(\theta)X(\theta+\Delta) \;=\; 1 - \lambda(\theta)\frac{\Delta^\alpha}{2} + o(\Delta^\alpha) \qquad (48)$$

for some $0 < \alpha \leq 2$, where $\lambda(\theta)$ is continuous in θ and $o(\Delta^2)$ is uniform in $\theta \geq 0$. If (48) holds with $\alpha = 2$, the process is also differentiable, and as before we can interpret $\lambda(\theta)$ as the variance of the derivative.

As in the stationary case, we can transform any differentiable, locally stationary standardised Gaussian process into one which has standardised derivative by applying a variance-stabilising transformation (that is one which stabilises the variance of the *derivative*). Define the transformation

$$\tau(\theta) = \int_0^\theta \sqrt{\lambda(\theta)}\, d\theta,$$

and define the inverse mapping $t \mapsto \theta_t$ implicitly via $\tau(\theta_t) = t$. Then the process $\{Y(t) = X(\theta_t) \mid t \geq 0\}$ is locally stationary with standardised derivative. In the case where $\lambda(\theta) \equiv \lambda$ is constant (or indeed the process is stationary), this reduces exactly to the change of scale in (39).

It is shown then that if Y is locally stationary with standardised derivative with correlation function $r(\cdot,\cdot)$, then (43) holds (with the same definition of $M(T)$) under the decay condition

$$\sup_{|s-t|\geq h} r(s,t)\log h \to 0$$

as $h \to \infty$.

An additional refinement contained in the results of Hüsler (1990, 1995) is that the probability of exceeding a growing level in (42) is generalised to the probability of crossing a general boundary:

$$P\left\{X(t) \le u_T(t), 0 \le t \le T\right\} \ .$$

The corresponding generalisation of the condition (41) is

$$\frac{1}{2\pi}\int_0^T e^{-u_T(t)^2/2}\,dt \to C \tag{49}$$

(note that when $u_T(t) \equiv u_T$ is constant this reduces to (41).)

If the boundary functions $\{u_T(t)\,|\,0 \le t \le T\}$ satisfy other conditions, including a smoothness condition and others which among other things imply that the minimum

$$\inf_{0\le t\le T} u_T(t) \to \infty$$

as $T \to \infty$, then if (49) holds,

$$P\left\{X(t) \le u_T(t), 0 \le \theta \le T\right\} \to e^{-C} \ .$$

This enables us to derive the limiting distribution of the maximum of a non-standardised process; if $\{W(\theta)\,|\,\theta \ge 0\}$ is a Gaussian process with $EW(\theta) = \mu(\theta)$ and $VarW(\theta) = \sigma^2(\theta)$, then $W^*(\theta) = [W(\theta)-\mu(\theta)]/\sigma(\theta)$ is standardised and

$$P\left\{W(\theta) \le u\right\} = P\left\{W^*(\theta) \le u^*(\theta)\right\} \ ,$$

where $u^*(\theta) = [u - \mu(\theta)]/\sigma(\theta)$, and so the methods of Hüsler (1990, 1995) can perhaps be applied.

In the next section we consider *sequences* of locally stationary Gaussian processes. We build upon the idea of expressing the probability of a boundary crossing in terms of an integral. In particular, if we have a sequence of locally stationary standardised Gaussian processes with standardised derivative

$$\{X_n(t)\,|\,t \in \mathbb{R}, n \in \mathbb{N}\} \ ,$$

and have a corresponding sequence of boundary functions(possibly infinite-valued outside an interval)

$$\{u_n(t)\,|\,t \in \mathbb{R}, n \in \mathbb{N}\} \ ,$$

then if the sequence of integrals

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-u_n(t)^2/2} \, dt \to C \tag{50}$$

as $n \to \infty$, we might expect that under certain conditions

$$P\left\{X_n(t) \le u_n(t), t \in \mathbb{R}\right\} \to e^{-C} \ .$$

In the next section we give sufficient conditions on the correlation functions of the processes and also on the boundary functions for this to hold.

These results permit us to derive the limiting distribution of $\widetilde{M}_n$ by comparing it to

$$M_n^0 = \sup\left\{ Z_0(\theta) \,\middle|\, \hat{\theta}(a_n)/2 \le \theta \le \hat{\theta}(b_n)/2 \right\} \ , \tag{51}$$

where $\{Z_0(\theta) \,|\, \theta \in \Theta\}$ is a single standardised Gaussian process with correlation function

$$\rho(\theta, \eta) = \exp\left\{K(\theta + \eta) - \tfrac{1}{2}K(2\theta) - \tfrac{1}{2}K(2\eta)\right\} \ .$$

It turns out that the corresponding sequences of integrals for both cases are equivalent in the sense of (50), so their limiting distributions are the same.

7.2 Theorem 7.1

We now outline the statement of and sufficient conditions for our second main result. The proof is lengthy and appears in a separate section.

Suppose that $\{X_n(t) \,|\, t \in \mathbb{R}, n \in \mathbb{N}\}$ is a sequence of standardised differentiable Gaussian processes with standardised derivative, continuous sample functions and correlation functions

$$EX_n(s)X_n(t) = r_n(s,t) \ .$$

Define $R_n(t,h)$ via

$$r_n(t, t+h) = 1 - \frac{h^2}{2} + R_n(t,h) \ .$$

We assume that the remainder

$$R_n(t,h) = o(h^2) \ , \ \textit{uniformly in } n \textit{ and } t. \tag{52}$$

This implies that for each fixed n, $\{X_n(t) \,|\, t \in \mathbb{R}\}$ is locally stationary. We also make an assumption about the rate of decay of the correlation.

$$\delta_n(t) \log(t) \to 0 \ , \quad \text{uniformly in } n \ , \tag{53}$$

where

$$\delta_n(t) = \sup \{ r_n(s, s') \colon |s - s'| \geq t \} \ . \tag{54}$$

Let $\{u_n \colon \mathbb{R} \to (0, \infty] \,|\, n \in \mathbb{N}\}$ be a sequence of (possibly infinite-valued) boundary functions.

Writing $\phi(u) = (2\pi)^{-1/2} e^{-u^2/2}$ for the standard normal density, we wish to approximate the sequence of integrals

$$\int_{-\infty}^{\infty} \phi[u_n(t)] \, dt \ , \ n \in \mathbb{N}$$

by sequences of upper and lower Riemann sums We do this by dividing the real line into intervals of width $h_n \to 0$ as $n \to \infty$. The rate at which h_n gets small needs to be carefully controlled. It needs to be fast enough that the upper and lower Riemann sums converge to the integral, but slow enough so that maxima of $X_n(t)$ on each interval have appropriate large deviation behaviour.

Define $I_j = I_j(n) = ((j-1)h_n, jh_n]$, for $j = 0, \pm 1, \pm 2, \ldots$. Define also u_n^- and u_n^+ to be step functions identically equal to (respectively) the minima and maxima of X_n over each interval. That is define $u_n^-(jh_n) = \inf \{ u_n(t) \,|\, t \in I_j \}$ and $u_n^+(jh_n) = \sup \{ u_n(t) \,|\, t \in I_j \}$; also define $u_n^-(t) = u_n^-(jh_n)$, $t \in I_j$, and define $u_n^+(t)$ in a similar way.

Writing $m_n = \inf_{t \in \mathbb{R}} u_n(t)$, we assume that there is a sequence $0 < h_n \to 0$ as $n \to \infty$ such that both

$$h_n m_n (\log m_n)^{-1/2} \to \infty \tag{55}$$

and

$$\sum_{j=-\infty}^{\infty} h_n \left\{ \phi[u_n^-(jh_n)] - \phi[u_n^+(jh_n)] \right\} \to 0 \ . \tag{56}$$

Note that (55) implies

$$m_n \to \infty \quad \text{as } n \to \infty \ . \tag{57}$$

Theorem 7.1. *Let $\{X_n(t)\,|\,t \in \mathbb{R}\}$ be a sequence of standardised Gaussian processes with standardised derivative satisfying the uniformity assumption* (52) *and the decay condition* (53)*. Let $\{u_n : \mathbb{R} \to (0, \infty]\,|\,n \in \mathbb{N}\}$ be a sequence of (possibly infinite-valued) boundary functions satisfying* (55) *and* (56)*. If*

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \phi[u_n(t)]\, dt =: C_n \to C \tag{58}$$

as $n \to \infty$ then

$$P\{X_n(t) \le u_n(t), t \in \mathbb{R}\} \to e^{-C} \ .$$

The following corollary applies theorem 7.1 to a sequence of locally stationary standardised Gaussian processes that are do not necessarily have standardised derivative.

Corollary 7.2. *Suppose*

$$\{Y_n(\theta)\,|\,\theta \in \mathbb{R}, n \in \mathbb{N}\} \ ,$$

is a sequence of locally stationary standardised Gaussian processes with correlation functions $EY_n(\theta)Y_n(\eta) = \rho_n(\theta, \eta)$, and variances of derivatives given by

$$\lambda_n(\theta) = -\left.\frac{\partial^2 \rho_n(\theta, \eta)}{\partial \eta^2}\right|_{\theta=\eta} \ .$$

Then if we define

$$\tau_n(\theta) = \int_0^{\theta} \sqrt{\lambda_n(\eta)}\, d\eta \ ,$$

and θ_t implicitly via $\tau_n(\theta_t) = t$, then the processes $\{X_n(t) = Y_n(\theta_t)\,|\,t \in \mathbb{R}\}$ have standardised derivative, with correlation functions $r_n(s,t) = \rho_n(\theta_s, \theta_t)$.

For a sequence of boundary functions $\{w_n(\theta)\,|\,\theta \in \mathbb{R}, n \in \mathbb{N}\}$, define another sequence of boundary functions via $\{w_n^*(t) = w_n(\theta_t)\,|\,t \in \mathbb{R}, n \in \mathbb{N}\}$.

Suppose that the correlation functions $\{r_n(\cdot,\cdot)\,|\,n \in \mathbb{N}\}$ satisfy the conditions (52) *and* (53)*, and that the boundary functions $\{w_n^*(\cdot)\,|\,n \in \mathbb{N}\}$ satisfy* (55) *and* (56)*. If the sequence of integrals*

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \sqrt{\lambda_n(\theta)} e^{-\frac{1}{2} w_n(\theta)^2}\, d\theta \to C \ , \tag{59}$$

then

$$P\left\{Y_n(\theta) \le w_n(\theta), \theta \in \mathbb{R}\right\} \to e^{-C} \ . \tag{60}$$

Proof. Equation (60) follows after an application of theorem 7.1 to the sequence of events

$$\{X_n(t) \le w_n^*(t), t \in \mathbb{R}\} = \{Y_n(\theta) \le w_n(\theta), \theta \in \mathbb{R}\} \ .$$

If (59) holds then by a change of variable, $t = \tau_n(\theta)$, $dt = \sqrt{\lambda_n(\theta)}\, d\theta$, we get

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \sqrt{\lambda_n(\theta)} e^{-\frac{1}{2}w_n(\theta_t)^2}\, d\theta = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \phi[w_n^*(t)]\, dt \to C,$$

So

$$P\left\{Y_n(\theta) \le w_n(\theta), \theta \in \mathbb{R}\right\} = P\left\{X_n(t) \le w_n^*(t), t \in \mathbb{R}\right\} \ \ \to \ \ e^{-C}$$

by theorem 7.1, and so (60) is proved. □

7.3 Applying theorem 7.1

In this subsection we present a corollary that shows how we use theorem 7.1 to derive the limiting distribution of $\widetilde{M}_n$. We firstly recall some definitions.

Recall that we have a sequence of linear exponential families $\mathcal{F} = \{F_\theta \,|\, \theta \in \Theta_n\}$ generated by a distribution F_0 with cumulant generating function $K(\cdot)$. We are interested in deriving the limiting distribution of $\widetilde{M}_n$, the supremum over θ of $\left\{\widetilde{Z}_n(\theta) \,|\, \theta \in \Theta_n,\ n \in \mathbb{N}\right\}$, a sequence of mean-zero Gaussian processes with covariance functions

$$\rho(\theta, \eta) F_{\theta+\eta} A_n \ ,$$

where

$$\rho(\theta, \eta) \ = \ \frac{C(\theta, \eta)}{\sqrt{C(\theta, \theta)C(\eta, \eta)}} \ ,$$

$C(\theta, \eta) = \exp\{K(\theta + \eta)\}$, $A_n = [a_n, b_n]$, $a_n = F_0^{-1}(1/n)$ and $b_n = F_0^{-1}(1 - 1/n)$.

Define

$$\overline{\Theta}_n = \Theta_n \cap \left[\frac{\hat{\theta}(a_n)}{2}, \frac{\hat{\theta}(b_n)}{2}\right] .$$

and

$$T_n \quad = \quad \int_{\overline{\Theta}_n} \sqrt{K''(2\theta)}\, d\theta \ .$$

Define $v_n(\theta) = Var\widetilde{Z}_n(\theta) = F_{2\theta}A_n$, and the sequence of *standardised* Gaussian processes

$$\left\{Y_n(\theta) = v_n(\theta)^{-1/2}\widetilde{Z}_n(\theta) \,|\, \theta \in \Theta_n, n \in \mathbb{N}\right\} ,$$

with covariance functions

$$\rho_n(\theta, \eta) \quad = \quad \frac{C_n(\theta, \eta)}{\sqrt{C_n(\theta, \theta)C_n(\eta, \eta)}}$$

and

$$C_n(\theta, \eta) \quad = \quad C(\theta, \eta)F_{\theta+\eta}A_n \ .$$

Using results from appendix B and differentiating under the integral sign we have that

$$\begin{aligned}\lambda_n(\theta) \quad &= \quad -\left.\partial^2\rho_n(\theta, \eta)/\partial\eta^2\right|_{\eta=\theta} \\ &= \quad \frac{\int_{a_n}^{b_n} x^2 dF_{2\theta}(x)}{\int_{a_n}^{b_n} dF_{2\theta}(x)} - \left\{\frac{\int_{a_n}^{b_n} x dF_{2\theta}(x)}{\int_{a_n}^{b_n} dF_{2\theta}(x)}\right\}^2 ,\end{aligned}$$

which coincides with the *conditional variance* of a random variable $X \sim F_{2\theta}$ given that $a_n \le X \le b_n$.

Corollary 7.3. *Let $C > 0$ be some constant and $\{u_n \,|\, n \in \mathbb{N}\}$ a sequence with $u_n \to \infty$ as $n \to \infty$. Consider the following three statements:*

$$\frac{T_n e^{-\frac{1}{2}u_n^2}}{2\pi} \to C \ , \tag{61}$$

$$\frac{1}{2\pi}\int_{\Theta_n}\sqrt{\lambda_n(\theta)}\,e^{-\frac{1}{2}u_n^2 v_n(\theta)^{-1}}\,d\theta \to C\ , \tag{62}$$

and

$$\begin{array}{c}\text{If a sequence } \{u_n \,|\, n\in\mathbb{N}\} \text{ satisfies (61) for some } C,\\ \text{then it also satisfies (62) with the same } C.\end{array} \tag{63}$$

If (63) *holds, then with* $B_n = (2\log T_n)^{1/2}$*, for all* $x \in \mathbb{R}$*,*

$$P\left\{B_n\left(\widetilde{M}_n - B_n\right) + \log 2\pi \le x\right\} \to \exp\{-e^{-x}\}\ . \tag{64}$$

Proof. Consider the sequence

$$u_n = B_n + \frac{x - \log 2\pi}{B_n}\ .$$

Now

$$\begin{array}{rcl}\frac{1}{2}u_n^2 & = & \frac{1}{2}B_n^2 + (x - \log 2\pi) + O(B_n)^{-2}\\ & = & \log T_n + x - \log 2\pi + o(1)\ .\end{array}$$

So

$$\frac{e^{-\frac{1}{2}u_n^2}T_n}{2\pi} = e^{-x}[1 + o(1)]\ .$$

If (63) holds, then this implies

$$\frac{1}{2\pi}\int_{\Theta_n}\sqrt{\lambda_n(\theta)}\,e^{-\frac{1}{2}u_n^2 v_n(\theta)^{-1}}\,d\theta \to e^{-x} \tag{65}$$

also. So then

$$\begin{array}{rcl}P\left\{\widetilde{M}_n \le u_n\right\} & = & P\left\{\widetilde{Z}_n(\theta) \le u_n, \theta\in\Theta_n\right\}\\ & = & P\left\{Y_n(\theta) \le u_n v_n(\theta)^{-1/2}\,\theta\in\Theta_n\right\}\\ & \to & \exp\{-e^{-x}\}\ ,\end{array}$$

using (65) and corollary 7.2. Hence (64) is proved.

□

7.4 Examples

We now show that the condition (63) holds in some of our examples; we do not show it for all, as the proofs for the simpler examples we have given are lengthy. We defer the proof for our other examples as later work.

Before we proceed we provide a lemma concerning conditional variances that is useful in what follows.

Lemma 7.4. *Suppose X is a random variable with $E(X) = \mu$, $E(X-\mu)^2 = \sigma^2$. Then*

$$Var\,(X|a \le X \le b) \le \sigma^2/p\ ,$$

where $p = P(a \le X \le b)$.

Proof. Write F for the cumulative distribution function of X. Then the conditional mean is $\mu_c = p^{-1}\int_a^b x\,dF(x)$. Then

$$\begin{aligned}
&Var\,(X|a \le X \le b) \\
&= p^{-1}\int_a^b (x-\mu_c)^2\,dF(x) \\
&= p^{-1}\left\{\int_a^b [(x-\mu)+(\mu-\mu_c)]^2\ dF(x)\right\} \\
&= p^{-1}\left\{\int_a^b (x-\mu)^2\,dF(x) + 2(\mu-\mu_c)\int_a^b (x-\mu)\,dF(x) + (\mu-\mu_c)^2 p\right\} \\
&= p^{-1}\left\{\int_a^b (x-\mu)^2\,dF(x) + 2(\mu-\mu_c)(\mu_c p - \mu p)\,dF(x) + (\mu-\mu_c)^2 p\right\} \\
&= p^{-1}\left\{\int_a^b (x-\mu)^2\,dF(x) - (\mu-\mu_c)^2 p\right\} \\
&\le p^{-1}\int_{-\infty}^{\infty} (x-\mu)^2\,dF(x) = \sigma^2/p\ .
\end{aligned}$$

□

7.4.1 Normal mean mixtures

Consider the case $F_0 = \mathcal{N}(0,1)$ and $\Theta_n \equiv \mathbb{R}$, that is

$$\mathcal{F}_n \equiv \mathcal{F} = \{\mathcal{N}(\theta,1)\,|\,\theta \in \mathbb{R}\}\ .$$

Let Φ and ϕ now denote the $\mathcal{N}(0,1)$ cumulative distribution function and density, respectively. Let a_n be the solution of $\Phi(a_n) = 1/n$, and let $b_n = -a_n$. We have from (109) from appendix C.3 that

$$b_n = (2\log n - \log_2 n - \log 4\pi)^{1/2} + o(1) \ .$$

Define

$$\begin{aligned}
\Delta_n &= (\log_2 n - 3\log_3 n)^{1/2} \ , \\
\theta^-_{n2} &= (a_n - \Delta_n)/2 \ , \\
\theta^-_{n1} &= (a_n + \Delta_n)/2 \ , \\
\theta^+_{n1} &= (b_n - \Delta_n)/2 \ , \\
\theta^+_{n2} &= (b_n + \Delta_n)/2 \ , \\
\Theta^-_{n3} &= (-\infty, \theta^-_{n2}) \ , \\
\Theta^-_{n2} &= [\theta^-_{n2}, \theta^-_{n1}) \ , \\
\Theta_{n1} &= [\theta^-_{n1}, \theta^+_{n1}] \ , \\
\Theta^+_{n2} &= (\theta^+_{n1}, \theta^+_{n2}] \ , \\
\Theta^+_{n3} &= (\theta^+_{n2}, \infty) \ .
\end{aligned}$$

Define also

$$\begin{aligned}
T_n &= \int_{\hat{\theta}(a_n)/2}^{\hat{\theta}(b_n)/2} \sqrt{\lambda(\theta)}\, d\theta \\
&= \int_{a_n/2}^{b_n/2} d\theta \\
&= b_n = (2\log n - \log_2 n - \log 4\pi)^{1/2} + o(1) \ .
\end{aligned}$$

Suppose u_n is a positive increasing sequence satisfying

$$\frac{T_n e^{-u_n^2/2}}{2\pi} = \frac{b_n e^{-u_n^2/2}}{2\pi} \to C$$

as $n \to \infty$. Define $v_n(\theta) = \Phi(b_n - 2\theta) - \Phi(a_n - 2\theta)$.

Theorem 7.5. *If a positive sequence u_n satisfies*

$$\frac{e^{-u_n^2/2} b_n}{2\pi} \to C$$

Then

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\sqrt{\lambda_n(\theta)}e^{-\frac{1}{2}u_n^2 v_n(\theta)^{-1}}\,d\theta \to C\ .$$

We prove this theorem by partitioning the range of integration into the five regions Θ_{n3}^-, Θ_{n2}^-, Θ_{n1}, Θ_{n2}^+ and Θ_{n3}^+. We show that the integral over Θ_{n1} tends to C, and the others tend to 0.

Proposition 7.6. *As* $n \to \infty$,

$$\sup_{\theta\in\Theta_{n1}} |\lambda_n(\theta) - 1| \to 0\ .$$

Proof. $\lambda_n(\theta)$ is the conditional variance of X given $a_n \le X \le b_n$, when unconditionally $X \sim \mathcal{N}(2\theta, 1)$. This is the same as the conditional variance of Y given $a_n - 2\theta \le Y \le b_n - 2\theta$ when unconditionally $Y \sim \mathcal{N}(0,1)$. So we can write

$$\lambda_n(\theta) \quad = \quad \frac{\int_{a_n-2\theta}^{b_n-2\theta} y^2\,d\Phi(y)}{\int_{a_n-2\theta}^{b_n-2\theta}\,d\Phi(y)} - \left\{\frac{\int_{a_n-2\theta}^{b_n-2\theta} y\,d\Phi(y)}{\int_{a_n-2\theta}^{b_n-2\theta}\,d\Phi(y)}\right\}^2\ .$$

Define for $j = 0, 1, 2$, $I_{nj}(\theta) = \int_{a_n-2\theta}^{b_n-2\theta} y^j\,d\Phi(y)$. It suffices to show that

$$\inf_{\theta\in\Theta_{n1}} I_{nj}(\theta) \to 1$$

for $j = 0, 2$ and

$$\sup_{\theta\in\Theta_{n1}} I_{n1}(\theta) \to 0\ .$$

Now for $\theta \in \Theta_{n1}$,

$$a_n - 2\theta \le a_n - 2\theta_{n1}^- = -\Delta_n < \Delta_n = b_n - 2\theta_{n1}^+ \le b_n - 2\theta$$

so

$$\inf_{\theta\in\Theta_{n1}} I_{nj}(\theta) \ge \int_{-\Delta_n}^{\Delta_n} x^j\,d\Phi(y) \to 1\ , \tag{66}$$

for $j = 0, 2$ as $n \to \infty$.

$$I_1(\theta) \quad = \quad \int_{-\infty}^{a_n-2\theta} |y|\,d\Phi(y) - \int_{b_n-2\theta} |y|\,d\Phi(y)$$

so

$$\begin{aligned}\sup_{\theta\in\Theta_{n1}} |I_1(\theta)| &\leq \sup_{\theta\in\Theta_{n1}} \max\left\{\int_{-\infty}^{a_n-2\theta} |y|\, d\Phi(y), \int_{b_n-2\theta}^{\infty} |y|\, d\Phi(y)\right\} \\ &\leq \int_{\Delta_n}^{\infty} y\, d\Phi(y) \to 0\end{aligned}$$

as $n \to \infty$. □

Proposition 7.7.

$$\frac{1}{2\pi}\int_{\Theta_{n1}} \sqrt{\lambda_n(\theta)} e^{-\frac{1}{2}u_n^2 v_n(\theta)^{-1}}\, d\theta \to C\ .$$

Proof. We have from the previous proposition that there exists a positive sequence $c_n \to 0$ such that for all $\theta \in \Theta_{n1}$,

$$1 - c_n \leq \lambda_n(\theta) \leq 1 + c_n$$

and

$$1 - c_n \leq v_n(\theta) \leq 1\ .$$

The first equation follows from the statement of the proposition itself; the second from (66). Thus

$$\begin{aligned}[1-c_n]^{1/2} e^{-\frac{1}{2}u_n^2[1-c_n]^{-1}} \frac{1}{2\pi}\int_{\Theta_{n1}} d\theta &\leq \frac{1}{2\pi}\int_{\Theta_{n1}} \sqrt{\lambda_n(\theta)}\, e^{-\frac{1}{2}u_n^2 v_n(\theta)^{-1}}\, d\theta \\ &\leq [1+c_n]^{1/2} e^{-\frac{1}{2}u_n^2} \frac{1}{2\pi}\int_{\Theta_{n1}} d\theta\end{aligned}$$

Thus,

$$\begin{aligned}\frac{1}{2\pi}\int_{\Theta_{n1}} \sqrt{\lambda_n(\theta)}\, e^{-\frac{1}{2}u_n^2 v_n(\theta)^{-1}}\, d\theta &= \frac{e^{-u_n^2/2}}{2\pi}\int_{\Theta_{n1}} d\theta[1+o(1)] \\ &= \frac{e^{-u_n^2/2}}{2\pi}\left(b_n - \Delta_n\right)[1+o(1)] \\ &= \frac{e^{-u_n^2/2}}{2\pi} b_n[1+o(1)] \\ &\to C.\end{aligned}$$

□

Proposition 7.8.

$$\int_{\Theta_{n2}^+} \sqrt{\lambda_n(\theta)} e^{-u_n^2 v_n(\theta)^{-1}} \, d\theta \quad \to \quad 0$$
$$\int_{\Theta_{n2}^-} \sqrt{\lambda_n(\theta)} e^{-u_n^2 v_n(\theta)^{-1}} \, d\theta \quad \to \quad 0$$

Proof. Now $\lambda_n(\theta)$ is the conditional variance of X given $a_n \le X \le b_n$ when $X \sim \mathcal{N}(2\theta, 1)$, and $v_n(\theta) = P(a_n \le X \le b_n)$. So lemma 7.4 gives that

$$\lambda_n(\theta) \quad \le \quad v_n(\theta)^{-1} \ , \tag{67}$$

since if $X \sim \mathcal{N}(2\theta, 1)$, $E(X - 2\theta)^2 \equiv 1$. For $\theta_{n1}^+ \le \theta \le \theta_{n2}^+$ we have

$$v_n(\theta_{n1}^+)^{-1} \le v_n(\theta)^{-1} \le v_n(\theta_{n2}^+)^{-1} \ ,$$

so

$$\begin{aligned} \int_{\theta_{n1}^+}^{\theta_{n2}^+} \sqrt{\lambda_n(\theta)} e^{-u_n^2 v_n(\theta)^{-1}/2} d\theta \quad &\le \quad v_n(\theta_{n2}^+)^{-1/2} e^{-u_n^2 v_n(\theta_{n1}^+)/2} \int_{\theta_{n1}^+}^{\theta_{n2}^+} d\theta \\ &\le \quad v_n(\theta_{n2}^+)^{-1/2} e^{-u_n^2/2} \Delta_n \ . \end{aligned}$$

Now

$$\begin{aligned} v_n(\theta_{n1}^+) \quad &= \quad v_n\left([b_n - \Delta_n]/2\right) \\ &= \quad \Phi(b_n - b_n + \Delta_n) - \Phi(-2b_n + \Delta_n) \ . \end{aligned}$$

The first term tends to 1 as $n \to \infty$, the second tends to zero. On the other hand

$$\begin{aligned} v_n(\theta_{n2}^+) \quad &= \quad v_n\left([b_n + \Delta_n]/2\right) \\ &= \quad \Phi(-\Delta_n) - \Phi(-2b_n - \Delta_n) \end{aligned}$$

and both terms tend to zero, the first one more slowly. So then

$$\begin{aligned} v_n(\theta_{n2}^+)^{-1/2} \quad &\sim \quad \Phi(-\Delta_n)^{-1/2} \\ &\sim \quad \left(\frac{\Delta_n}{\phi(\Delta_n)}\right)^{1/2} \\ &= \quad O\left(\Delta_n^{1/2} e^{\Delta_n^2/4}\right) \ . \end{aligned}$$

Now $e^{-u_n^2/2} = O(\log n)^{1/2}$, $\Delta_n = O(\log_2 n)^{1/2}$ and $e^{\Delta_n^2/4} = \exp\{\frac{1}{4}\log_2 n - \frac{3}{4}\log_3 n\} = (\log n)^{1/4}(\log_2 n)^{-3/4}$, so

$$\begin{aligned}\int_{\theta_{n1}^+}^{\theta_{n2}^+} \sqrt{\lambda_n(\theta)}e^{-u_n^2 v_n(\theta)^{-1}/2}d\theta &= O\left(e^{-u_n^2/2}\Delta_n^{3/2}e^{\Delta_n^2/4}\right) \\ &= O(1)(\log n)^{-1/2}(\log_2 n)^{3/4}(\log n)^{1/4}(\log_2 n)^{-3/4} \\ &= O\left(\log_n\right)^{-1/4} = o(1)\ .\end{aligned}$$

Thus the proposition is proved for the integral over Θ_{n2}^+. By symmetry the same holds for the integral over Θ_{n2}^-. □

Proposition 7.9.

$$\begin{aligned}\int_{\Theta_{n3}^+} \sqrt{\lambda_n(\theta)}e^{-u_n^2 v_n(\theta)^{-1}}\,d\theta &\rightarrow 0 \\ \int_{\Theta_{n3}^-} \sqrt{\lambda_n(\theta)}e^{-u_n^2 v_n(\theta)^{-1}}\,d\theta &\rightarrow 0\end{aligned}$$

Proof. Using (67), we have that

$$I_3 = \int_{\theta_{n2}^+}^{\infty} \sqrt{\lambda_n(\theta)}e^{-u_n^2 v_n(\theta)^{-1}/2}\,d\theta \le \int_{\theta_{n2}^+}^{\infty} v_n(\theta)^{-1/2}e^{-u_n^2 v_n(\theta)^{-1}/2}\,d\theta$$

For $\theta > b_n/2$, $v_n(\theta)^{-1}$ is monotone increasing, so has a well defined inverse h_n satisfying

$$h_n[v_n(\theta)^{-1}] = \theta \quad \text{and} \quad v_n[h_n(x)]^{-1} = x\ .$$

If we change variables $x = v_n(\theta)^{-1}$, then $\theta = h_n(x)$, $d\theta = h_n'(x)\,dx$ and the integral above becomes

$$I_3 \le \int_{v_n(\theta_{n2}^+)^{-1}}^{\infty} x^{1/2}e^{-\frac{1}{2}u_n^2 x}h_n'(x)\,dx\ .$$

For $\theta \ge b_n/2$, $v_n(\theta)^{-1}$ is also *strictly convex*, so the inverse $h_n(x)$ is strictly concave, and so the derivative h_n' is strictly decreasing. Moreover, the derivative is given by

$$h_n'(x) = \frac{-1}{x^2 v_n'(h_n(x))}\ ,$$

where $v_n'(\theta) = -2[\phi(b_n - 2\theta) - \phi(a_n - 2\theta)]$. So

$$\begin{aligned} I_3 &\leq h_n'\left[v_n(\theta_{n2}^{+})^{-1}\right]\int_{v_n(\theta_{n2}^{+})^{-1}}^{\infty} x^{1/2}e^{-u_n^2 x/x}\,dx \\ &\leq \frac{-1}{[v_n(\theta_{n2}^{+})^{-1}]^2 v_n'\left\{h_n[v_n(\theta_{n2}^{+})^{-1}]\right\}}\int_0^{\infty} x^{1/2}e^{-u_n^2 x/2}\,dx \\ &= \frac{-v_n(\theta_{n2}^{+})^2}{v_n'(\theta_{n2}^{+})}\frac{\Gamma(3/2)}{(u_n^2/2)^{3/2}} \;. \end{aligned}$$

Since $v_n(\theta) = \Phi(b_n - 2\theta) - \Phi(a_n - 2\theta) \leq \Phi(b_n - 2\theta) \sim \phi(b_n - 2\theta)/(2\theta - b_n)$,

$$\frac{-v_n(\theta)^2}{v_n'(\theta)} = O(1)\frac{\phi(b_n - 2\theta)}{2(2\theta - b_n)^2} \to 0 \quad \text{as} \quad \theta \to \infty \;,$$

so since $u_n \to \infty$, $I_3 \to 0$ as $n \to \infty$. The result holds also for the integral over Θ_{n3}^{-} by symmetry. □

Hence the theorem is proved.

7.4.2 Exponential (Gamma(1,1)) scale

Now let $F_0 = \text{Gamma}(1,1)$, then $F_\theta = \text{Gamma}(1, 1-\theta)$ (see section 5.3.3). We show in section 6.1.3 that we cannot deal with the (full) two sided problem, so we consider the (maximal) one-sided version, letting $\mathcal{F}_n \equiv \mathcal{F} = \{F_\theta \,|\, 0 \leq \theta < 1/2\}$ (we note that we could consider a 'method-of-seive' type approach, involving a slowly growing two sided set, but restrict attention to the one-sided case for simplicity).

In this case we have $F_0(x) = 1 - e^{-x}$, so $F_0^{-1}(u) = -\log(1-u)$. Thus $a_n = F_0^{-1}(1/n) = -\log(1 - 1/n) \sim 1/n$, and $b_n = F_0^{-1}(1 - 1/n) = \log n$.

Also $K(\theta) = -\log(1-\theta)$, $K^{(j)}(\theta) = (j-1)!(1-\theta)^{-j}$ for $j = 1, 2, \ldots$ and $\hat{\theta}(x) = 1 - 1/x$. Further $\tau(\theta) = -\frac{1}{2}\log(1 - 2\theta)$.

Define

$$\begin{aligned} v_n(\theta) &= F_{2\theta}(b_n) - F_{2\theta}(a_n) \\ &= \int_{a_n}^{b_n} (1-2\theta)e^{-x(1-2\theta)}\,dx \\ &= e^{-(1-2\theta)a_n} - e^{-(1-2\theta)b_n} \\ &= (1 - 1/n)^{1-2\theta} - e^{-(1-2\theta)\log n} \end{aligned} \qquad (68)$$

$$\lambda(\theta) = K''(2\theta) = (1-2\theta)^2$$

and

$$\begin{aligned} T_n &= \int_0^{\hat{\theta}(b_n)/2} \sqrt{\lambda(\theta)}\, d\theta \\ &= \left[-\tfrac{1}{2}\log(1-2\theta)\right]_0^{\hat{\theta}(b_n)/2} \\ &= -\tfrac{1}{2}\log[1-\hat{\theta}(b_n)] = -\tfrac{1}{2}\log(1/\log n) = \tfrac{1}{2}\log\log n \ . \end{aligned}$$

Define $\lambda_n(\theta)$ as the conditional variance of X given $a_n \leq X \leq b_n$, when unconditionally $X \sim F_{2\theta}$. Then

$$\lambda_n(\theta) = \frac{\int_{a_n}^{b_n} x^2(1-2\theta)e^{-x(1-2\theta)}\, dx}{\int_{a_n}^{b_n}(1-2\theta)e^{-x(1-2\theta)}\, dx} - \left\{\frac{\int_{a_n}^{b_n} x(1-2\theta)e^{-x(1-2\theta)}\, dx}{\int_{a_n}^{b_n}(1-2\theta)e^{-x(1-2\theta)}\, dx}\right\}^2 .$$

Theorem 7.10. *Suppose that u_n satisfies*

$$\frac{T_n e^{-u_n^2/2}}{2\pi} \to C$$

as $n \to \infty$. Then

$$\frac{1}{2\pi}\int_0^{1/2} \sqrt{\lambda_n(\theta)} e^{-\frac{1}{2}u_n^2 v_n(\theta)^{-1}}\, d\theta \to C$$

also.

We prove this theorem by partitioning the interval $[0, \frac{1}{2})$ into pieces. Write $\log_2 n = \log\log n$ and $\log_3 n = \log\log_2 n$. Note that $\log_2 n > 0$ for $n \geq 3 > e$. Define

$$\begin{aligned} \theta_{n1} &= \frac{1}{2} - \frac{\log_2 n}{2\log n \log_3 n} \ , \\ \theta_{n2} &= \frac{1}{2} - \frac{\log_3 n}{2\log n \log_2 n} \ . \end{aligned}$$

So then for $n \geq 16 > e^e$,

$$\theta_{n1} < \frac{\hat{\theta}(b_n)}{2} = \frac{1}{2} - \frac{1}{2\log n} < \theta_{n2} \ .$$

We then show that the integral over $[0, \theta_{n1})$ tends to C, and the integrals over $[\theta_{n1}, \theta_{n2})$ and $[\theta_{n2}, \frac{1}{2})$ tend to zero.

Proposition 7.11. *As $n \to \infty$, for $i = 0, 1, 2,$*

$$\inf_{0 \le \theta \le \theta_{n1}} \frac{\int_{a_n}^{b_n} x^i \, dF_{2\theta}(x)}{\int_0^\infty x^i \, dF_{2\theta}(x)} \to 1 \ .$$

Proof. For $i = 0, 1, 2,$

$$\frac{\int_{a_n}^{b_n} x^i \, dF_{2\theta}(x)}{\int_0^\infty x^i \, dF_{2\theta}(x)} = \frac{\int_{a_n}^{b_n} x^i(1-2\theta)e^{-(1-2\theta)x} \, dx}{\int_0^\infty x^i(1-2\theta)e^{-(1-2\theta)x} \, dx} = \frac{\int_{a_n(1-2\theta)}^{b_n(1-2\theta)} y^i e^{-y} \, dy}{\int_0^\infty y^i e^{-y} \, dy}$$

So it suffices to show that $\inf_{0 \le \theta \le \theta_{n1}} b_n(1-2\theta) \to \infty$ and $\sup_{0 \le \theta \le \theta_{n1}} a_n(1-2\theta) \to 0$. But

$$\begin{aligned} \inf_{0 \le \theta \le \theta_{n1}} b_n(1-2\theta) &= b_n(1-2\theta_{n1}) \\ &= \log n \left(\frac{\log_2 n}{\log n \log_3 n} \right) \\ &= \frac{\log_2 n}{\log_3 n} \to \infty \end{aligned}$$

and

$$\sup_{0 \le \theta \le \theta_{n1}} a_n(1-2\theta) = a_n \to 0 \ .$$

□

Proposition 7.12.

$$\frac{1}{2\pi} \int_0^{\theta_{n1}} \sqrt{\lambda_n(\theta)} e^{-\frac{1}{2} u_n^2 v_n(\theta)^{-1}} \, d\theta \to C$$

Proof. The previous proposition implies that both

$$\inf_{0 \le \theta \le \theta_{n1}} v_n(\theta) \to 1$$

and

$$\sup_{0 \le \theta \le \theta_{n1}} \left| \frac{\lambda_n(\theta)}{\lambda(\theta)} - 1 \right| \to 0 \ .$$

So

$$\frac{1}{2\pi}\int_0^{\theta_{n1}} \sqrt{\lambda_n(\theta)}e^{-\frac{1}{2}u_n^2 v_n(\theta)^{-1}}\,d\theta \;=\; [1+o(1)]\frac{e^{-\frac{1}{2}u_n^2}}{2\pi}\int_0^{\theta_{n1}}\sqrt{\lambda(\theta)}\,d\theta$$

Now

$$\begin{aligned}
\int_0^{\theta_{n1}} \sqrt{\lambda(\theta)}\,d\theta &= -\tfrac{1}{2}\left[\log(1-2\theta)\right]_0^{\theta_{n1}} \\
&= -\tfrac{1}{2}\log\left(\frac{\log_2 n}{\log n \log_3 n}\right) \\
&= \tfrac{1}{2}\left(\log_2 n - \log_3 n + \log_4 n\right) \\
&= T_n[1+o(1)]\ .
\end{aligned}$$

So

$$\frac{1}{2\pi}\int_0^{\theta_{n1}} \sqrt{\lambda_n(\theta)}e^{-\frac{1}{2}u_n^2 v_n(\theta)^{-1}}\,d\theta \;=\; \frac{e^{-\frac{1}{2}u_n^2}T_n}{2\pi}[1+o(1)] \to C\ .$$

□

Proposition 7.13. *As $n \to \infty$,*

$$\int_{\theta_{n1}}^{\theta_{n2}} \sqrt{\lambda_n(\theta)}e^{-\frac{1}{2}u_n^2 v_n(\theta)^{-1}}\,d\theta \to 0$$

Proof. Suppose $X \sim F_{2\theta}$. Then $\lambda_n(\theta)$ is the conditional variance of X given $a_n \le X \le b_n$. Since $K^{(j)}(2\theta)$ is the j-th cumulant of X we have

$$\begin{aligned}
E(X) &= \mu_{2\theta} = K'(2\theta) = (1-2\theta)^{-1}\ , \\
E(X-\mu_{2\theta})^2 &= \sigma_{2\theta}^2 = K''(2\theta) = (1-2\theta)^{-2}(=\lambda(\theta)\text{ also })\ ,
\end{aligned}$$

Lemma 7.4 then gives that

$$\lambda_n(\theta) \;\le\; \lambda(\theta)v_n(\theta)^{-1} \tag{69}$$

Now, $(1-2\theta_{n2})\log n = \log_3 n/\log_2 n \to 0$, so (68) gives

$$\begin{aligned}
v_n(\theta_{n2}) &= \left[1-\frac{1-2\theta_{n2}}{n}+o\left(\frac{1-2\theta_{n2}}{n}\right)\right] - e^{-(1-2\theta_{n2})\log n} \\
&= \left[1-e^{-\log_3 n/\log_2 n}\right] + O\left(\frac{\log_3 n}{n\log n\log_2 n}\right) \\
&= \frac{\log_3 n}{\log_2 n}\left[1+o(1)\right]\ .
\end{aligned} \tag{70}$$

Also

$$\begin{aligned}
\int_{\theta_{n1}}^{\theta_{n2}} \sqrt{\lambda(\theta)}\, d\theta &= \left[-\tfrac{1}{2}\log(1-2\theta)\right]_{\theta_{n1}}^{\theta_{n2}} \\
&= -\tfrac{1}{2}\left\{\log\left(\frac{\log_3 n}{\log n \log_2 n}\right) - \log\left(\frac{\log_2 n}{\log n \log_3 n}\right)\right\} \\
&= -\tfrac{1}{2}\log\left\{\left(\frac{\log_3 n}{\log_2 n}\right)^2\right\} \\
&= \log\left(\frac{\log_2 n}{\log_3 n}\right) \\
&= \log_3 n - \log_4 n \ ,
\end{aligned}$$

where $\log_4 n = \log\log_3 n$. Note also that $e^{-\frac{1}{2}u_n^2} = O(T_n)^{-1} = O(\log_2 n)^{-1}$. So then

$$\begin{aligned}
\int_{\theta_{n1}}^{\theta_{n2}} \sqrt{\lambda_n(\theta)} e^{-\frac{1}{2}u_n^2 v_n(\theta)^{-1}}\, d\theta &\leq v_n(\theta_{n2})^{-1/2} e^{-\frac{1}{2}u_n^2} \int_{\theta_{n1}}^{\theta_{n2}} \sqrt{\lambda(\theta)}\, d\theta \\
&= O(1)\left(\frac{\log_2 n}{\log_3 n}\right)^{1/2} (\log_2 n)^{-1} \log_3 n \\
&= O\left(\frac{\log_3 n}{\log_2 n}\right)^{1/2} = o(1)\ .
\end{aligned}$$

□

Proposition 7.14. *As $n \to \infty$,*

$$\int_{\theta_{n2}}^{1/2} \sqrt{\lambda_n(\theta)} e^{-\frac{1}{2}u_n^2 v_n(\theta)^{-1}}\, d\theta \to 0$$

Proof. (69) gives that

$$I = \int_{\theta_{n2}}^{1/2} \sqrt{\lambda_n(\theta)} e^{-\frac{1}{2}u_n^2 v_n(\theta)^{-1}}\, d\theta = O(1) \int_{\theta_{n2}}^{1/2} v_n(\theta)^{-1/2} \sqrt{\lambda(\theta)}\, e^{-\frac{1}{2}u_n^2 v_n(\theta)^{-1}}\, d\theta\ .$$

Change variable to $t = \tau(\theta)$, $dt = \sqrt{\lambda(\theta)}\, d\theta$. Then with $\tau(\theta_t) = t$,

$$I = O(1) \int_{\tau(\theta_{n2})}^{\infty} v_n(\theta_t)^{-1/2} e^{-\frac{1}{2}u_n^2 v_n(\theta_t)^{-1}}\, dt\ .$$

Changing again, via $x = v_n(\theta_t)^{-1}$. The inverse of this transformation is h_n, satisfying $h_n[v_n(\theta_t)^{-1}] = t$ for all $t \geq \tau(\theta_{n2})$ and $v_n\left(\theta_{h_n(x)}\right)^{-1} = x$ for all $x \geq v_n(\theta_{n2})^{-1}$. Implicit differentiation gives that

$$h_n'(x) = \frac{-1}{x^2 v_n'\left(\theta_{h_n(x)}\right)\theta'_{h_n(x)}} \ ,$$

where $\theta_t' = \tau(\theta_t)^{-1}$.

We assume that the function $v_n(\theta_t)^{-1}$ is strictly convex in $t \geq \tau(\theta_{n2})$. Then h_n' is decreasing and we get

$$\begin{aligned} I &= O(1)h_n'\left[v_n(\theta_{n2})^{-1}\right]\int_{v_n(\theta_{n2})^{-1}}^{\infty} x^{1/2}e^{-\frac{1}{2}u_n^2 x}\,dx \\ &= O(1)\frac{v_n(\theta_{n2})^2}{-v_n'(\theta_{n2})\theta'_{\tau(\theta_{n2})}}\int_0^{\infty} x^{1/2}e^{-\frac{1}{2}u_n^2 x}\,dx \\ &= O(1)\frac{v_n(\theta_{n2})^2\tau(\theta_{n2})}{-v_n'(\theta_{n2})(u_n^2)^{3/2}} \ . \end{aligned}$$

It is shown in (70) that $v_n(\theta_{n2}) = O(\log_3 n/\log_2 n)$. Also $u_n^2 = O(\log T_n) = O(\log_3 n)$, and

$$\begin{aligned} \tau(\theta_{n2}) &= -\tfrac{1}{2}\log(1-2\theta_{n2}) \\ &= -\tfrac{1}{2}\log\left(\frac{\log_3 n}{\log n \log_2 n}\right) \\ &= \tfrac{1}{2}\left(\log_2 n + \log_3 n - \log_4 n\right) = O\left(\log_2 n\right) \ . \end{aligned}$$

Next,

$$\begin{aligned} -v_n'(\theta_{n2}) &= 2b_n e^{-(1-2\theta_{n2})b_n} - 2a_n e^{-(1-2\theta_{n2})a_n} \\ &= 2(\log n)e^{-\log_3 n/\log_2 n} - O(1/n) \\ &= O(\log n) \ . \end{aligned}$$

Substituting these rates into (71) gives

$$I = O(1)\left(\frac{\log_3 n}{\log_2 n}\right)^2 \frac{\log_2 n}{\log n(\log_3 n)^{3/2}} = O(1)\frac{(\log_3 n)^{1/2}}{(\log_2 n)(\log n)} = o(1) \ .$$

□

7.5 Proof of theorem 7.1

The proof follows very closely the proof in Hüsler (1990). In essence we are replacing finite-term sums with infinite-term sums, but the general argument is mostly unchanged. In this subsection we write $\phi(u) = (2\pi)^{-1/2}e^{-u^2/2}$ for the standard normal density to facilitate comparison with Hüsler (1990). It does not denote the variance-stablising transformation of an exponential family as it does in other sections.

Lemma 7.15 (Small intervals-stationary case). *Suppose* $\{X(t) \mid t \in \mathbb{R}\}$ *is a stationary standardised Gaussian process with standardised derivative. If* $h \to 0$, $u \to \infty$ *such that* $hu \to \infty$, *then*

$$P\{X(t) > u, 0 \le t \le h\} / h\phi(u) \to \frac{1}{\sqrt{2\pi}} . \tag{71}$$

Proof. See Leadbetter *et al.* (1983). □

Lemma 7.16. *Suppose* $\{X_n(t) \mid t \in \mathbb{R},\ n \in \mathbb{N}\}$ *is a sequence of standardised Gaussian processes with standardised derivatives, satisfying the uniformity assumption* (52). *Then if* $h_n \to 0$, $u_n \to \infty$ *and* $h_n u_n \to \infty$, *then as* $n \to \infty$,

$$P\{X_n(s) > u_n, t \le s \le t + h_n\} / h_n\phi(u_n) \to \frac{1}{\sqrt{2\pi}} \tag{72}$$

uniformly in t.

Proof. Define $\mathcal{D}_n = \{t \mid u_n(t) < \infty\}$. Note that

$$\begin{aligned}
\overline{\delta}(h) &:= \sup_n \left[\sup_{t,t+h \in \mathcal{D}_n} r_n(t, t+h) \right] \\
&= \sup_n \left\{ \sup_{t,t+h \in \mathcal{D}_n} \left[1 - \frac{h^2}{2} + R_n(t, h) \right] \right\} \\
&= 1 - \frac{h^2}{2} + \sup_n \sup_{t,t+h \in \mathcal{D}_n} R_n(t, h) \\
&= 1 - \frac{h^2}{2} + o(h^2)
\end{aligned}$$

by (52). In a similar way,

$$\begin{aligned}
\underline{\delta}(h) &= \inf_n \left[\inf_{t,t+h \in \mathcal{D}_n} r_n(t, t+h) \right] \\
&= 1 - \frac{h^2}{2} + o(h^2) .
\end{aligned}$$

Thus if $\left\{\overline{Y}(t)\,|\,t\in\mathbb{R}\right\}$ and $\{\underline{Y}(t)\,|\,t\in\mathbb{R}\}$ are two stationary standardised Gaussian processes with respective correlation functions $\overline{\delta}(\cdot)$ and $\underline{\delta}(\cdot)$, they also *both* have standardised derivative. Now since

$$E\underline{Y}(s)\underline{Y}(t)\leq EX_n(s)X_n(t)\leq E\overline{Y}(s)\overline{Y}(t)$$

for all s and t, Slepian's Lemma (see Leadbetter *et al.* (1983), Theorem 7.4.2 and Piterbarg (1996), Theorem C.1) gives that for any u and any appropriate time set S,

$$P\left\{\sup_{s\in S}\underline{Y}(s)\leq u\right\}\leq P\left\{\sup_{s\in S}X_n(s)\leq u\right\}\leq P\left\{\sup_{s\in S}\overline{Y}(s)\leq u\right\}\ . \qquad (73)$$

So then (71) gives that

$$\begin{aligned}\frac{1}{\sqrt{\pi}} &\leq \liminf_n P\left\{\underline{Y}(s)\leq u_n, t\leq s\leq t+h_n\right\}\\ &\leq \liminf_n P\left\{X_n(s)\leq u_n, t\leq s\leq t+h_n\right\}\\ &\leq \limsup_n P\left\{X_n(s)\leq u_n, t\leq s\leq t+h_n\right\}\\ &\leq \limsup_n P\left\{\overline{Y}(s)\leq u_n, t\leq s\leq t+h_n\right\}\\ &\leq \frac{1}{\sqrt{\pi}}\end{aligned}$$

□

Lemma 7.17.

1.

$$0\ \leq\ P\left\{X_n(t)\leq u_n^+(t), t\in\mathbb{R}\right\}-P\left\{X_n(t)\leq u_n(t), t\in\mathbb{R}\right\}\to 0$$

2.

$$0\ \leq\ P\left\{X_n(t)\leq u_n(t), t\in\mathbb{R}\right\}-P\left\{X_n(t)\leq u_n^-(t), t\in\mathbb{R}\right\}\to 0$$

Proof. Following Hüsler (1990), both differences are proved in the same way.

$$
\begin{aligned}
0 &\le P\left\{X_n(t) \le u_n^+(t), t \in \mathbb{R}\right\} - P\left\{X_n(t) \le u_n(t), t \in \mathbb{R}\right\} \\
&\le \sum_{j=-\infty}^{\infty} P\left\{X_n(t) \le u_n^+(t), t \in I_j, \sup_{s \in I_j}[X_n(s) - u_n(s)] > 0\right\} \\
&\le \sum_{j=-\infty}^{\infty} P\left\{u_n^-(jh) \le M_n(I_j) \le u_n^+(jh)\right\} \\
&= \frac{1}{\sqrt{\pi}} \sum_{j=-\infty}^{\infty} h_n\left\{\phi[u_n^-(jh_n)][1+o(1)] - \phi[u_n^+(jh_n)][1+o(1)]\right\} && (74) \\
&= o(1), && (75)
\end{aligned}
$$

where in (74) the $o(1)$ is uniform in j (by (72)) and (75) follows from (56). □

Instead of writing statements concerning both $u_n^+(\cdot)$ and $u_n^-(\cdot)$, as in Hüsler (1990), if we say that a property holds for $u_n^*(\cdot)$, we mean that it holds for both $u_n^+(\cdot)$ and $u_n^-(\cdot)$.

Lemma 7.18. *Let $\{h_n \,|\, n \in \mathbb{N}\}$ satisfy (55) and (56). Let $\varepsilon = \varepsilon_n$ be such that $\varepsilon_n/h_n \to 0$ but $\varepsilon_n m_n \to \infty$, where $m_n = \inf_{t\in\mathbb{R}} u_n(t)$. Define $I_j^* = I_j \setminus (jh_n - \varepsilon_n, jh_n]$. Then as $n \to \infty$,*

$$
0 \le P\left\{X_n(t) \le u_n^*(t), t \in \bigcup_{j=-\infty}^{\infty} I_j^*\right\} - P\left\{X_n(t) \le u_n^*(t), t \in \mathbb{R}\right\} \to 0\ .
$$

Proof. The difference is bounded above by

$$
\begin{aligned}
&\sum_{j=-\infty}^{\infty} P\left\{\sup\left(X_n(t), t \in (jh_n - \varepsilon_n, jh_n]\right) > u_n^*(jh_n)\right\} \\
&= \frac{1}{\sqrt{\pi}} \sum_{j=-\infty}^{\infty} \varepsilon_n \phi[u_n^*(jh_n)][1+o(1)] && (76) \\
&= (\varepsilon_n/h_n)\, \frac{1}{\sqrt{\pi}} \sum_{j=-\infty}^{\infty} h_n \phi[u_n^*(jh_n)] + o(1) && (77) \\
&= (\varepsilon_n/h_n)\, O(1) = o(1)\ ,
\end{aligned}
$$

where in (76) $o(1)$ is uniform in j by (72) and (77) follows from (56). □

For a given $q_0 > 0$, define

$$q_j = q_j^*(n) = q_0 / u_n^*(jh_n) \ .$$

Lemma 7.19.

$$\begin{aligned} 0 \ \leq \ & \limsup_n \left(P\left\{ X_n(iq_j) \leq u_n^*(jh), iq_j \in I_j^*, j = 0, \pm 1, \pm 2, \ldots \right\} \right. \\ & \left. - \, P\left\{ X_n(t) \leq u_n^*(t), t \in \cup_{j=-\infty}^{\infty} I_j^* \right\} \right) \to 0 \end{aligned}$$

as $q_0 \to 0$.

Proof. Recal the proof of (72), the processes $\underline{Y}(t)$ and $\overline{Y}(t)$ and the inequalities in (73). Writing $M(I, X) = \sup\{X(t) \,|\, t \in I\}$, we have that for any I_j and any u,

$$\begin{aligned} & P\{X_n(iq_j) \leq u, iq_j \in I_j\} - P\{M(I_j, X_n) \leq u\} \\ \leq \ & P\left\{\overline{Y}(iq_j) \leq u, iq_j \in I_j\right\} - P\{M(I_j, \underline{Y}) \leq u\} \\ = \ & \left[P\left\{\overline{Y}(iq_j) \leq u, iq_j \in I_j\right\} - P\left\{M(I_j, \overline{Y}) \leq u\right\}\right] \\ & + \left[P\left\{M(I_j, \overline{Y}) \leq u\right\} - P\{M(I_j, \underline{Y}) \leq u\}\right] \end{aligned}$$

As in the proof for Lemma 3.2 in Hüsler (1990), the first difference is bounded by

$$\frac{1}{\sqrt{\pi}} h\phi(u)\rho(q_0) + \phi(u)/u$$

with $\rho(q_0) \to 0$ as $q_0 \to 0$ since

$$hu = h_n u_n^*(jh_n) \geq h_n m_n \to \infty$$

and $u_n^*(jh_n) = q_0$. Also $\phi(u_n^*)/u_n^* = o\left(h_n \phi(u_n^*)\right)$ using the same argument. Both terms in the second difference are $\{h_n \phi[u_n^*(jh_n)]/\sqrt{\pi}\}\,[1 + o(1)]$, with $o(1) \to 0$ as $n \to \infty$, not depending on j, by (71).

The sum over j is thus bounded by

$$[\rho(q_0) + o(1)] \frac{1}{\sqrt{\pi}} \sum_{j=-\infty}^{\infty} h_n \phi[u_n^*(jh_n)] \ .$$

Letting $n \to \infty$, (56) gives that the limsup is $O\left(\rho(q_0)\right)$. The result follows letting $q_0 \to 0$. □

Having reduced attention to a Gaussian sequence, we utilise the normal comparison lemma, the basic form of which appears below:

Lemma 7.20 (Normal comparison lemma). *Suppose* $(X_1, X_2, \ldots, X_n)$ *and* $(Y_1, Y_2, \ldots, Y_n)$ *are* $N(0,1)$ *random vectors with respective correlation matrices* $\{r^X\}$ *and* $\{r^Y\}$. *Let* $r_{ij} = \max\left(\left|r^X_{ij}\right|, \left|r^Y_{ij}\right|\right)$. *Let* $u_1, u_2, \ldots, u_n$ *be real numbers. Then*

$$\begin{aligned} & P\{X_i \le u_i, i = 1, 2, \ldots, n\} - P\{Y_i \le u_i, i = 1, 2, \ldots, n\} \\ \le \; & \frac{1}{2\pi} \sum_{1 \le i < j \le n} \left(r^X_{ij} - r^Y_{ij}\right)^+ \left(1 - r^2_{ij}\right)^{-1/2} \exp\left\{ - \frac{\frac{1}{2}(u_i^2 + u_j^2)}{1 + r_{ij}} \right\} , \end{aligned}$$

where $(x)^+ = \max(0, x)$.

Proof. See Leadbetter *et al.* (1983), Theorem 4.2.1. □

We intend to show that $\sup\left\{X_n(iq_j) \,\middle|\, iq_j \in I^*_j\right\}$ are asymptotically independent for different j. We use the normal comparison lemma, but following Hüsler (1990), we first show that we can ignore certain intervals. This is an adaptation of a method first introduced in Hüsler (1983).

Define, for each n,

$$a_1 = \min\{u^*_n(jh_n) \,|\, j = 0, \pm 1, \pm 2, \ldots\} \text{ and } J_1 = \{j \,|\, a_1 \le u^*_n(jh_n) \le 2a_1\} .$$

Then define

$$a_2 = \min\{u^*_n(jh_n) > 2a_1\} \text{ and } J_2 = \{j \,|\, a_2 \le u^*_n(jh_n) \le 2a_2\} ,$$

and continue defining a_m, J_m, for each positive integer m. A finite or possibly countably infinite number of J_m's are defined in this way, depending on whether $\sup\{u_n(t) \,|\, u_n(t) < \infty\}$ is finite or infinite. Thus $\{J_m \,|\, m \in \mathbb{N}\}$ forms a partition of the positive integers (possibly truncated at some $H_n < \infty$).

For each m, define

$$\begin{aligned} f_m &= \sum_{j \in J_m} \phi[u^*_n(jh_n)]h/q_j \\ &= \sum_{j \in J_m} \phi[u^*_n(jh_n)]u^*_n(jh_n)h_n/q_0 \end{aligned}$$

The size of f_m for each m determines whether the corresponding set J_m contributes significantly to the probability of an upcrossing. Define

$$G = \{m \,|\, f_m \geq \exp\{-a_m^2/4\}\} \ ,$$

and then define $J_0 = \{J_m \,|\, m \in G\}$. Note that since there is at least one j in each J_m,

$$f_m \geq \phi(a_m) a_m h_n / q_0 \ . \tag{78}$$

Lemma 7.21. *For each fixed q_0,*

$$\sum_{j \notin J_0} h_n \phi[u_n^*(jh_n)] \to 0 \ . \tag{79}$$

Moreover,

$$\sum_{j \notin J_0} h_n \phi[u_n^*(jh_n)]/q_0 \to 0 \sum_{j \notin J_0} h_n \phi[u_n^*(jh_n)] \to 0 \ . \tag{80}$$

not depending on q_0.

Proof. Since $j \in J_m \Rightarrow a_m \leq u_n^*(jh_n) \leq 2a_m$,

$$\begin{aligned}
\sum_{j \in J_m} h_n \phi[u_n^*(jh_n)] &\leq \sum_{j \in J_m} h_n \frac{u_n^*(jh_n)}{a_m} \phi[u_n^*(jh_n)] + o(1) \\
&= \frac{1}{a_m} f_m q_0 + o(1)
\end{aligned}$$

where $o(1)$ is uniform in m by (56). Now if $m \notin G$ we get

$$\frac{1}{a_m} f_m q_0 \leq \frac{q_0}{a_m} e^{-a_m^2/4} \leq \frac{q_0}{a_m} e^{-a_1^2/4} \ .$$

Note that since $a_{m+1} \geq 2a_m$, we have $a_m \geq 2^{m-1} a_1$. Adding over all $m \notin G$ gives

$$\begin{aligned}
\sum_{j \in J_m} h_n \phi[u_n^*(jh_n)] &\leq q_0 e^{-a_1^2/4} \sum_{m \notin G} \frac{1}{a_m} \\
&\leq q_0 e^{-a_1^2/4} \sum_{m=1}^{\infty} \frac{1}{a_m} \\
&\leq \frac{q_0 e^{-a_1^2/4}}{a_1} \sum_{m=1}^{\infty} \frac{1}{2^{m-1}} \\
&= q_0 o(1)
\end{aligned}$$

as $a_1 \to \infty$, for every $q_0 > 0$, proving the first (79). Dividing by q_0 proves (80). □

Lemma 7.22. *As $n \to \infty$,*

$$\begin{aligned} 0 \ \le \ & P\{X_n(iq_j) \le u_n^*(jh_n), iq_j \in I_j, j \in J_0\} \\ & -P\{X_n(iq_j) \le u_n^*(jh_n), iq_j \in I_j, j = 0, \pm 1, \pm 2, \ldots\} \to 0 \end{aligned}$$

Proof. The difference is bounded by

$$\begin{aligned} & \sum_{m \notin G} P\{X_n(iq_j) > u_n^*(jh_n) \ \text{ for some } \ iq_j \in I_j, j \in J_m\} \\ \le \ & \sum_{m \notin G} \sum_{j \in J_m} \sum_{iq_j \in I_j} \phi[u_n^*(jh_n)]/[u_n^*(jh_n)] \\ \le \ & \sum_{m \notin G} \sum_{j \in J_m} (h_n/q_j)\phi[u_n^*(jh_n)]/[u_n^*(jh_n)] \\ \le \ & \le \sum_{m \notin G} \sum_{j \in J_m} h_n \phi[u_n^*(jh_n)]/q_0 \\ \to \ & 0 \end{aligned}$$

by (80). □

Having shown that we can restrict attention to the time subset J_0, we need to consider the possibility that J_0 is made up of a countably infinite number of intervals. Intuition suggests this is not the case, but it is possible to prove the result without enforcing this.

As in the proof of Slepian's Lemma in Piterbarg (1996) (Theorem C.1), the comparison lemma can be extended to two infinite $N(0,1)$ sequences using a separability argument.

Then we need the following lemma:

Lemma 7.23. *Define*

$$S_n := \sum_{\substack{iq_j \in I_j^* \\ i'q_{j'} \in I_{j'}^*}} \sum_{j \ne j' \in J_0} |r_n(iq_j, i'q_{j'})| \exp\left\{ -\frac{\frac{1}{2}\left(u_n^*(jh_n)^2 + u_n^*(j'h_n)^2\right)}{1 + r_n(iq_j, i'q_{j'})} \right\}$$

If $S_n \to 0$ then

$$P\left\{X_n(iq_j) \le u_n^*(jh_n), iq_j \in I_j^*, j \in J_0\right\} - \prod_{j \in J_0} P\left\{X_n(iq_j) \le u_n^*(jh_n), iq_j \in I_j^*\right\} \to 0$$

Proof. Uses the normal comparison lemma, where the two correlation matrices r^X and r^Y are the same, except that off-diagonal blocks corresponding to correlations between points in different intervals in r^X are replaced by zero matrices in r^Y. □

Lemma 7.24. *If* (52),(53),(55) *and* (56) *hold, then* $S_n \to 0$ *as* $n \to \infty$.

Proof. Exactly as Lemma 3.4 in Hüsler (1990). The only difference is that here we allow the sum S_n to have an infinite number of terms. At every stage in that proof each finite-term sum may be replaced with an infinite-term sum without affecting the argument. The proof is quite lengthy so we omit it here. □

Lemma 7.25.

$$\limsup_{n\to\infty}\left(\prod_{j\in J_0} P\left\{X_n(iq_j) \le u_n^*(jh_n), iq_j \in I_j^*\right\} - \prod_{j\in J_0} P\left\{X_n(t) \le u_n^*(jh_n), t \in I_j^*\right\}\right) \to 0$$

as $q_0 \to 0$.

Proof. Uses same argument as Lemma 7.19. □

Lemma 7.26.

$$\prod_{j\in J_0} P\left\{X_n(t) \le u_n^*(jh_n), t \in I_j^*\right\} \to e^{-C} \ . \qquad (81)$$

Proof. We use Lemma (A.2), which says that if

1. $\sum_{j\in J_0} P\left\{X_n(t) > u_n^*(jh_n), t \in I_j^*\right\} \to C$ and
2. $\sup_{j\in J_0} P\left\{X_n(t) > u_n^*(jh_n), t \in I_j^*\right\} \to 0$

as $n \to \infty$ then (81) holds. Firstly we have from (72) that

$$\begin{aligned} P\left\{X_n(t) > u_n^*(jh_n), t \in I_j^*\right\} &= (h_n - \varepsilon_n)\phi[u_n^*(jh_n)][1+o(1)] \\ &= h_n\phi[u_n^*(jh_n)][1+o(1)] \qquad (82) \end{aligned}$$

where $o(1)$ is uniform in j.

This implies that

$$\limsup_j P\left\{X_n(t) > u_n^*(jh_n), t \in I_j^*\right\} \le h_n \sup_j \phi[u_n^*(jh_n)] \to 0$$

since $u_n^*(jh_n) \geq \min_{t\in\mathbb{R}} u_n(t) \to \infty$ by (57), thus proving 2. Finally (82) implies that

$$\begin{aligned}
\sum_{j\in J_0} P\left\{X_n(t) > u_n^*(jh_n), t \in I_j^*\right\} &= \left\{\sum_{j\in J_0} h_n\phi[u_n^*(jh_n)]\right\}[1+o(1)] \\
&= \sum_{j=-\infty}^{\infty} h_n\phi[u_n^*(jh_n)] - \sum_{j\notin J_0} h_n\phi[u_n^*(jh_n)] + o(1) \\
&= \int_{\mathbb{R}} \phi[u_n(t)]\,dt + o(1) \\
&= C + o(1)
\end{aligned}$$

proving 1.

□

8 Final Assembly

In subsection 7.3 we derive the limiting distribution of $\widetilde{M}_n$. We need to show that this has the same limiting distribution as M_n, and hence $M_n^\dagger$ and M_n^*.

Recall that

$$T_n = \int_{\overline{\Theta}_n} \sqrt{\lambda(\theta)}\, d\theta \ ,$$

where $\overline{\Theta}_n = \Theta_n \cap [\hat{\theta}(a_n)/2, \hat{\theta}(b_n)/2]$, where $a_n = F_0^{-1}(1/n)$ and $b_n = F_0^{-1}(1 - 1/n)$. So then we have that

$$\begin{aligned} T_n &\leq \int_{\hat{\theta}(a_n)/2}^{\hat{\theta}(b_n)/2} \sqrt{\lambda(\theta)}\, d\theta \\ &= \int_{\hat{\theta}(a_n)/2}^{\hat{\theta}(b_n)/2} \sqrt{K''(2\theta)}\, d\theta \\ &= \tau[\hat{\theta}(b_n)/2] - \tau[\hat{\theta}(a_n)/2] \\ &= [\phi(\hat{\theta}(b_n)) - \phi(\hat{\theta}(a_n))]/2 \\ &= [\hat{\phi}(b_n) - \hat{\phi}(a_n)]/2 \ , \end{aligned}$$

with the inequality becoming an equality if $\Theta_n \equiv \Theta$ (or indeed for any Θ_n such that $[\hat{\theta}(a_n)/2, \hat{\theta}(b_n)/2] \subset \Theta_n$).

Define

$$t_{n0} = \log\log T_n \ .$$

Recall also that $\mu_\theta = K'(\theta)$, that θ_t satisfies $\tau(\theta_t) = \frac{1}{2}\phi(2\theta_t) = t$, and

$$\rho_0(\theta) = \rho(0, \theta) = e^{K(\theta) - \frac{1}{2}K(2\theta)} \ .$$

The following condition is used to prove the results of this section:

$$\sup_{|t|>t_{n0}} (\mu_{\theta_t} - \mu_0)\, \rho_0(\theta_t) \quad = \quad o\,(\log T_n)^{-1/2} \ . \tag{83}$$

Note that since (12) holds, $\sup_{|t|>t_{n0}} (\mu_{\theta_t} - \mu_0)$ is bounded away from zero, so (83) implies

$$\sup_{|t|>t_{n0}} \rho_0(\theta_t) \quad = \quad o\,(\log T_n)^{-1/2} \ . \tag{84}$$

Lemma 8.1. *If* (84) *holds and*

$$\sup_{\theta\in\Theta_n} |S_n(\theta) - Z_n(\theta)| = o_p\left(\log T_n\right)^{-1/2} \ , \tag{85}$$

then $M_n = \sup_{\theta\in\Theta_n} S_n(\theta)$ *is distributed as*

$$(2\log T_n)^{1/2} + \frac{V_n - \log 2\pi}{(2\log T_n)^{1/2}} \ , \tag{86}$$

where for each $x \in \mathbb{R}$,

$$P\left\{V_n \le x\right\} \to \exp\left\{-e^{-x}\right\}$$

as $n \to \infty$. *Moreover, with high probability the supremum is attained in* $\{\theta_t \colon |t| \ge t_{n0}\}$, *that is with probability tending to 1*

$$\sup_{\theta_t\in\Theta_n} S_n(\theta_t) = \sup_{\theta_t\in\Theta_n, |t|>t_{n0}} S_n(\theta_t) \ . \tag{87}$$

Proof. We already have that $\widetilde{M}_n = \sup_{\theta\in\Theta_n} \widetilde{Z}_n(\theta)$ has the same asymptotic distribution as

$$\sup_{\frac{\hat\theta(a_n)}{2} \le \theta \le \frac{\hat\theta(b_n)}{2}} Z_0(\theta)$$

(see (51)), so $\widetilde{M}_n = (2\log T_n)^{1/2} + O_p\left(\log T_n\right)^{-1/2}$. A similar argument shows that

$$\sup_{|t|\le t_{n0}} \widetilde{Z}_n(\theta_t) \quad \text{and} \quad \sup_{|t|\le t_{n0}} Z_0(\theta_t)$$

have the same asymptotic distribution, so that

$$\begin{aligned} \sup_{|t|\le t_{n0}} \widetilde{Z}_n(\theta_t) &= (2\log t_{n0})^{1/2} + O_p\left(\log t_{n0}\right)^{-1/2} \\ &= (2\log_3 T_n)^{1/2} + O_p\left(\log_3 T_n\right)^{-1/2} \ . \end{aligned}$$

So then with high probability,

$$\sup_t \widetilde{Z}_n(\theta_t) = \sup_{|t|\ge t_{n0}} \widetilde{Z}_n(\theta_t) \ . \tag{88}$$

Since $\widetilde{Z}_n(\theta) = Z_n(\theta) + O_p(1)\,\rho_0(\theta)$, by (84),

$$Z_n(\theta) = \begin{cases} \widetilde{Z}_n(\theta) + O_p(1) & \text{for } |t| \le t_{n0} , \\ \widetilde{Z}_n(\theta) + o_p(\log T_n)^{-1/2} & \text{for } |t| > t_{n0} . \end{cases}$$

So then

$$\begin{aligned} \sup_{|t|\le t_{n0}} Z_n(\theta) &= \sup_{|t|\le t_{n0}} \widetilde{Z}_n(\theta) + O_p(1) \\ &= O_p(\log_3 T_n)^{1/2} \end{aligned}$$

and

$$\begin{aligned} \sup_{|t|> t_{n0}} Z_n(\theta) &= \sup_{|t|> t_{n0}} \widetilde{Z}_n(\theta) + o_p(\log T_n)^{-1/2} \\ &= O_p(\log T_n)^{1/2} \end{aligned}$$

So then with high probability,

$$\begin{aligned} \sup_t Z_n(\theta_t) &= \sup_{|t|> t_{n0}} Z_n(\theta) \qquad (89) \\ &= \sup_{|t|> t_{n0}} \widetilde{Z}_n(\theta) + o_p(\log T_n)^{-1/2} \\ &= \sup_t \widetilde{Z}_n(\theta) + o_p(\log T_n)^{-1/2} \end{aligned}$$

because of (88). Equation(86) follows from (85) and an application of theorem 7.1.

(85) also implies that

$$\begin{aligned} \sup_{\theta_t\in\Theta_n} S_n(\theta_t) &= \sup_{\theta_t\in\Theta_n} Z_n(\theta_t) + o_p(\log T_n)^{-1/2} \\ &= \sup_{\theta_t\in\Theta_n, |t|>t_{n0}} Z_n(\theta_t) + o_p(\log T_n)^{-1/2} \qquad (90) \\ &= \sup_{\theta_t\in\Theta_n, |t|>t_{n0}} S_n(\theta_t) + o_p(\log T_n)^{-1/2} , \end{aligned}$$

where (90) follows from (89). Thus (87) is proved.

□

We now proceed to show that $M_n^\dagger$ and M_n^* have the same asymptotic distribution as M_n. We firstly express $S_n^\dagger(\theta)$ and $S_n^*(\theta)$ as certain functions of $S_n(\theta)$. Firstly write

$$\begin{aligned} s_\theta(x) &= \left(e^{\theta x - K(\theta)} - 1\right) e^{K(\theta) - \frac{1}{2}K(2\theta)} \\ &= \left(e^{\theta x - K(\theta)} - 1\right) \rho_0(\theta) \ . \end{aligned}$$

Now write

$$\begin{aligned} s_\theta^\dagger(x) &= \left(e^{\theta x - K(\theta)} - 1\right) \left(e^{K(2\theta) - 2K(\theta)} - 1\right)^{-1/2} \\ &= \left(e^{\theta x - K(\theta)} - 1\right) \left(\rho_0(\theta)^{-2} - 1\right)^{-1/2} \\ &= \left(e^{\theta x - K(\theta)} - 1\right) \rho_0(\theta) \left(1 - \rho_0(\theta)^2\right)^{-1/2} \\ &=: s_\theta(x) h^\dagger(\theta) \ . \end{aligned}$$

Now

$$h^\dagger(\theta) = \left(1 - \rho_0(\theta)^2\right)^{-1/2} \ = \ 1 + O(\rho_0(\theta))^2 \ .$$

So

$$\text{for } |t| > t_{n0}, \ \ h^\dagger(\theta_t) = 1 + o(\log T_n)^{-1} \qquad (91)$$

by (84).

In a similar way we write

$$\begin{aligned} s_\theta^*(x) &= \frac{\left(e^{\theta x - K(\theta)} - 1 - (x - \mu_0)(\mu_\theta - \mu_0)/\sigma_0^2\right)}{\left(e^{K(2\theta) - 2K(\theta)} - 1 - [(\mu_\theta - \mu_0)/\sigma_0]^2\right)^{1/2}} \\ &= \frac{\left(e^{\theta x - K(\theta)} - 1 - (x - \mu_0)(\mu_\theta - \mu_0)/\sigma_0^2\right)}{\left(\rho_0(\theta)^{-2} - 1 - [(\mu_\theta - \mu_0)/\sigma_0]^2\right)^{1/2}} \\ &= \frac{\left(e^{\theta x - K(\theta)} - 1 - (x - \mu_0)(\mu_\theta - \mu_0)/\sigma_0^2\right) \rho_0(\theta)}{\left(1 - \rho_0(\theta)^2 \left\{1 + [(\mu_\theta - \mu_0)/\sigma_0]^2\right\}\right)^{1/2}} \\ &= \left(e^{\theta x - K(\theta)} - 1 - (x - \mu_0)(\mu_\theta - \mu_0)/\sigma_0^2\right) \rho_0(\theta) h^*(\theta) \\ &= s_\theta(x) h^*(\theta) - \rho_0(\theta) h^*(\theta)(x - \mu_0)(\mu_\theta - \mu_0)/\sigma_0^2 \end{aligned}$$

So, as in (91) above we have that

$$\begin{aligned} h^*(\theta) &= \left(1 - \rho_0(\theta)^2 \left\{1 + [(\mu_\theta - \mu_0)/\sigma_0]^2\right\}\right)^{-1/2} \\ &= 1 + O\left[(\mu_\theta - \mu_0)\rho_0(\theta)\right]^2 \ . \end{aligned}$$

So

$$\text{for } |t| > t_{n0},\ h^*(\theta_t) = 1 + o(\log T_n)^{-1} \tag{92}$$

by (83).

Writing $Z_i = (X_i - \mu_0)/\sigma_0$, and $\overline{Z}_n = n^{-1/2}\sum_{i=1}^n Z_i$, we have

$$\begin{aligned} S_n(\theta) &= n^{-1/2}\sum_{i=1}^n s_\theta(X_i) \ , \\ S_n^\dagger(\theta) &= n^{-1/2}\sum_{i=1}^n s_\theta^\dagger(X_i) \\ &= S_n(\theta)h^\dagger(\theta) \\ S_n^*(\theta) &= n^{-1/2}\sum_{i=1}^n s_\theta^*(X_i) \\ &= S_n(\theta)h^*(\theta) - \rho_0(\theta)h^*(\theta)\overline{Z}_n(\mu_\theta - \mu_0)/\sigma_0 \ . \end{aligned}$$

Now, assume that the derivatives of $h^\dagger$ and h^* tend to zero. Assume that there exists a constant c such that for all s, t,

$$2[1 - \rho(\theta_s, \theta_t)] \le c(s-t)^2 \ . \tag{93}$$

The following lemma is taken from Bickel & Chernoff (1993).

Lemma 8.2 (Kolmogorov Bound). *If $\{Z(t) \,|\, 0 \le t \le 1\}$ is a stochastic process satisfying*

$$E[Z(s) - Z(t)]^2 \le c(s-t)^2 \quad \textit{for all} \quad 0 \le s \le t \le 1 \ , \tag{94}$$

then

$$P\left\{\sup_{0\le t\le 1} |Z(t) - Z(0)| \ge z\right\} \le Kc/z^2 \ ,$$

where K is an absolute constant.

Note that for any given $\varepsilon > 0$, this implies that there is an $M_\varepsilon = \sqrt{Kc/\varepsilon}$ such that

$$\begin{aligned} P\left\{\sup_{0\le t\le 1} |Z(t) - Z(0)| \ge M_\varepsilon\right\} &= P\left\{c^{-1/2}\sup_{0\le t\le 1} |Z(t) - Z(0)| \ge \sqrt{K/\varepsilon}\right\} \\ &\le \varepsilon \ , \end{aligned}$$

So then (94) implies

$$\sup_{0\le t\le 1} |Z(t) - Z(0)| = c^{1/2} O_p(1) \ . \tag{95}$$

Now suppose we have a process $\{X(t)\,|\,0 \le t \le L\}$ satisfying

$$E[X(s) - X(t)]^2 \le c(s-t)^2 \ . \tag{96}$$

Then defining $\{Z(t) = X(tL)\,|\,0 \le t \le 1\}$ we have

$$E[Z(s) - Z(t)]^2 = E[X(sL) - X(tL)]^2 \le c(sL - tL)^2 = \left[cL^2\right](s-t)^2 \ ,$$

so this and (95) imply

$$\sup_{0\le t\le L} |X(t) - X(0)| = \sup_{0\le t\le 1} |Z(t) - Z(0)| = cL^{1/2} O_p(1)$$

(compare with Bickel & Chernoff (1993)).

Lemma 8.3. *If* (83) *and* (93) *hold then* $M_n^\dagger$ *and* M_n^* *have the same asymptotic distribution as* M_n*.*

Proof. Firstly we have that

$$S_n^\dagger(\theta) \;\; = \;\; S_n(\theta) + S_n(\theta)[h^\dagger(\theta) - 1]$$

so

$$\begin{aligned}\sup_{|t|>t_{n0}} S_n^\dagger(\theta_t) &= \sup_{|t|>t_{n0}} S_n(\theta_t) + O_p(\log T_n)^{-1/2}\, o(\log T_n)^{-1} \\ &= \sup_{\theta\in\Theta_n} S_n(\theta) + o_p(\log T_n)^{-1/2}\end{aligned}$$

by (87). In a similar way

$$S_n^*(\theta) \;\; = \;\; S_n(\theta) + S_n(\theta)[h^*(\theta) - 1] - \rho_0(\theta) h^*(\theta) \overline{Z}_n(\mu_\theta - \mu_0)/\sigma_0$$

so

$$\begin{aligned}\sup_{|t|>t_{n0}} S_n^*(\theta_t) &= \sup_{|t|>t_{n0}} S_n^*(\theta_t) + O_p(\log T_n)^{1/2}\, o(\log T_n)^{-1} + O_p(1)\, o(\log T_n)^{-1/2} \\ &= \sup_{\theta\in\Theta_n} S_n(\theta) + o_p(\log T_n)^{-1/2} \ ,\end{aligned}$$

using (92), (83) and (87).

Now, for $|s-t| \geq \varepsilon$, for some fixed $\varepsilon > 0$, we have

$$2[1-\rho^\dagger(\theta_s,\theta_t)] \leq 4 = \frac{4}{\varepsilon^2}\varepsilon^2 \leq C_\varepsilon (s-t)^2 \ .$$

where $C_\varepsilon = \frac{4}{\varepsilon^2}$. Now the transformation $t \mapsto \theta_t$ is chosen so that

$$-\left.\frac{\partial^2 \rho(\theta_s,\theta_t)}{\partial s^2}\right|_{s=t} \equiv 1 \ .$$

Now $\rho(\theta,\eta)$ and $\rho^\dagger(\theta,\eta)$ are very close. We can in fact show that for some constant C,

$$-\left.\frac{\partial^2 \rho^\dagger(\theta_s,\theta_t)}{\partial s^2}\right|_{s=t} \leq C$$

for all t, and that the corresponding first derivative is identically zero. So we can show that for some other constant C'_ε,

$$\rho^\dagger(\theta_t,\theta_{t+h}) \geq 1 - \frac{C'_\varepsilon h^2}{2} \ ,$$

for all t and $|h| < \varepsilon$. Thus there is a $c^\dagger$ so that

$$\begin{aligned} E\left\{\left[S_n^\dagger(\theta_s) - S_n^\dagger(\theta_t)\right]^2\right\} &= 2[1-\rho^\dagger(\theta_s,\theta_t)] \\ &\leq c^\dagger(s-t)^2 \end{aligned}$$

Similar (but more tedious) calculations show in a similar way that

$$\begin{aligned} E\left\{\left[S_n^*(\theta_s) - S_n^*(\theta_t)\right]^2\right\} &= 2[1-\rho^*(\theta_s,\theta_t)] \\ &\leq c^*(s-t)^2 \end{aligned}$$

for some constants $c^\dagger$ and c^*. Hence by the Kolmogorov Bound,

$$\sup_{|t|\leq t_{n0}} |S_n^\dagger(\theta_t) - S_n^\dagger(0)| = O_p\left(\log\log T_n\right) = o_p\left(\log T_n\right)^{1/2}$$

and

$$\sup_{|t|\leq t_{n0}} |S_n^*(\theta_t) - S_n^*(0)| = O_p\left(\log\log T_n\right) = o_p\left(\log T_n\right)^{1/2} \ .$$

Hence with probability tending to one,

$$\begin{aligned}
M_n^\dagger = \sup_{\theta\in\Theta_n} S_n^\dagger(\theta) &= \sup_{|t|>t_{n0},\theta_t\in\Theta_n} S_n^\dagger(\theta) \\
&= \sup_{\theta\in\Theta_n} S_n(\theta) + o_p\left(\log T_n\right)^{-1/2}
\end{aligned}$$

and

$$\begin{aligned}
M_n^* = \sup_{\theta\in\Theta_n} S_n^*(\theta) &= \sup_{|t|>t_{n0},\theta_t\in\Theta_n} S_n^*(\theta) \\
&= \sup_{\theta\in\Theta_n} S_n(\theta) + o_p\left(\log T_n\right)^{-1/2} \quad .
\end{aligned}$$

And so

$$\begin{aligned}
M_n^\dagger &= M_n + o_p\left(\log T_n\right)^{-1/2} \\
&= \left(2\log T_n\right)^{-1/2} + \frac{V_n - \log 2\pi}{\left(2\log T_n\right)^{-1/2}} + o_p\left(\log T_n\right)^{-1/2} \\
&= \left(2\log T_n\right)^{-1/2} + \frac{\left[V_n + o_p(1)\right] - \log 2\pi}{\left(2\log T_n\right)^{-1/2}} \\
&= \left(2\log T_n\right)^{-1/2} + \frac{V_n' - \log 2\pi}{\left(2\log T_n\right)^{-1/2}}
\end{aligned}$$

where

$$P\left\{V_n' \le x\right\} \to \exp\left\{-e^{-x}\right\}$$

as $n \to \infty$. Thus the result is proved for $M_n^\dagger$. It is proved for M_n^* in the same way. □

8.1 Examples

Here we show that conditions (83), (85) and (93) hold for certain examples.

8.1.1 Normal mean

Consider the case where $F_0 = \mathcal{N}(0,1)$ and $\Theta_n \equiv \mathbb{R}$. We show earlier that $T_n = b_n = F_0^{-1}(1-1/n) = (2\log n)^{1/2}[1+o(1)]$. We also show in subsection 6.1.1 that

$$\begin{aligned}
\sup_{\theta\in\mathbb{R}} \left|S_n(\theta) - Z_n(\theta)\right| &= O\left(\log n\right)^{-1/4} \\
&= o\left(\log_2 n\right)^{-1/2} \quad .
\end{aligned}$$

Since $(\log T_n)^{1/2} = O\left(\log_2 n\right)^{1/2}$, (85) holds in this case.

Also $t_{n0} = \log_2 T_n = \log_3 n + o(1)$, $\theta_t = \mu_{\theta_t} = t$ and $\rho_0(\theta_t) = e^{-t^2/2}$. So as $n \to \infty$,

$$\begin{aligned}
\sup_{|t|>t_{n0}} (\mu_{\theta_t} - \mu_0)\rho_0(\theta_t) &= O\left(\log_3 n \exp\left\{-\frac{1}{2}(\log_3 n)^2\right\}\right) \\
&= O\left(\log_3 n \exp\left\{-\frac{3}{2}\log_3 n\right\}\right) \\
&= O\left(\log_3 n \left(\log_2 n\right)^{-3/2}\right) \\
&= O\left(\frac{\log_3 n}{\log_2 n}\right)\left(\log_2 n\right)^{-1/2} \\
&= o\left(\log_2 n\right)^{-1/2} \ .
\end{aligned}$$

Again, since $(\log T_n)^{-1/2} = O(\log_2 n)^{-1/2}$, (83) holds.

Finally, since $K(\theta) = \theta^2/2$, $K''(\theta) \equiv 1$, $\phi(\theta) = \tau(\theta) = \theta$, so $\theta_t = t$. So $K(\theta_s + \theta_t) = (s+t)^2/2$, and

$$K(\theta_s + \theta_t) - \tfrac{1}{2}K(2\theta_t) - \tfrac{1}{2}K(2\theta_s) = \tfrac{1}{2}s^2 + st + \tfrac{1}{2}t^2 - s^2 - t^2 = -(s-t)^2/2 \ .$$

So

$$\begin{aligned}
EZ_0(\theta_s)Z_0(\theta_t) &= \rho(\theta_s, \theta_t) \\
&= \exp\left\{K(\theta_s + \theta_t) - \tfrac{1}{2}K(2\theta_s) - \tfrac{1}{2}K(2\theta_t)\right\} \\
&= \exp\left\{-(s-t)^2/2\right\} \ .
\end{aligned}$$

Since this is a function of $|s-t|$, the process $\{Z_0(\theta_t) \,|\, t \in \mathbb{R}\}$ is stationary. Since $1 - e^{-}x \leq x$, we have that

$$\begin{aligned}
2[1 - \rho(\theta_t, \theta_{t+h})] &= 2[1 - e^{-h^2/2}] \\
&\leq h^2 \ ,
\end{aligned}$$

for all t, so (93) holds with $c = 1$.

8.1.2 Gamma scale

Consider the one-sided Gamma-scale example where $F_0 = \Gamma(\alpha_0, 1)$, and $\Theta_n \equiv (0, \frac{1}{2})$.

We have that

$$1 - F_0(x) \sim x^{\alpha_0 - 1} e^{-x} / \Gamma(\alpha_0) \ ,$$

so if $b_n = F_0^{-1}(1 - 1/n)$, then using a convergent iterative scheme we find that

$$\begin{aligned} b_n &= \log n + (\alpha_0 - 1) \log_2 n - \log \Gamma(\alpha_0) + o(1) \\ &= \log n[1 + o(1)] \ . \end{aligned}$$

Also, $\hat{\theta}(x) = 1 - \alpha_0 / x$ and $\phi(\theta) = -\alpha_0^{1/2} \log(1-\theta)$, so $\hat{\phi}(x) = -\alpha_0^{1/2} \log(\alpha_0/x)$, and

$$\hat{\phi}(b_n) = \alpha_0^{1/2} [\log_2 n - \log(\alpha_0)] + o(1) \ .$$

$$\begin{aligned} T_n &= \int_0^{\hat{\theta}(b_n)/2} \sqrt{K''(\theta)}\, d\theta \\ &= \tau(\hat{\theta}(b_n)/2) \\ &= \hat{\phi}(b_n)/2 \\ &= O\left(\log_2 n\right) \ . \end{aligned}$$

So $(\log T_n)^{-1/2} = O(\log_3 n)^{-1/2}$. According to (36) we have that

$$\begin{aligned} \sup_{0<\theta<\frac{1}{2}} |S_n(\theta) - Z_n(\theta)| &= O\left(\log n\right)^{-1/2} \\ &= o\left(\log T_n\right)^{-1/2} \ , \end{aligned}$$

so (85) holds.

Let $F_0 = \text{Gamma}(\alpha_0, 1)$. Then $K(\theta) = -\alpha_0 \log(1 - \theta)$, and $\tau(\theta) = -\alpha_0^{-1/2} \log(1 - 2\theta)/2$ and

$$\theta_t = \tfrac{1}{2} \left[1 - \exp\left\{ -2t/\alpha_0^{1/2} \right\} \right] \ .$$

So

$$\begin{aligned} K(\theta_s + \theta_t) &= -\alpha_0 \log(1 - \theta_s - \theta_t) \\ &= -\alpha_0 \log \left[\exp\left\{ -2t/\alpha_0^{1/2} \right\} + \exp\left\{ -2s/\alpha_0^{1/2} \right\} \right] \ . \end{aligned}$$

So

$$\begin{aligned} r(s,t) = EZ_0(\theta_s)Z_0(\theta_t) &= \rho(\theta_s,\theta_t) \\ &= \exp\{K(\theta_s+\theta_t) - \tfrac{1}{2}K(2\theta_s) - \tfrac{1}{2}K(2\theta_t)\} \\ &= \left[\frac{\exp\left\{-2t/\alpha_0^{1/2}\right\} + \exp\left\{-2s/\alpha_0^{1/2}\right\}}{2\exp\left\{-(t+s)/\alpha_0^{-1/2}\right\}}\right]^{-\alpha_0} . \end{aligned}$$

In particular,

$$\begin{aligned} r(t,t+h) &= \left[\frac{\exp\left\{-2t/\alpha_0^{1/2}\right\} + \exp\left\{-2(t+h)/\alpha_0^{1/2}\right\}}{2\exp\left\{-(2t+h)/\alpha_0^{-1/2}\right\}}\right]^{-\alpha_0} \\ &= \left[\frac{\exp\left\{h/\alpha_0^{1/2}\right\} + \exp\left\{-h/\alpha_0^{1/2}\right\}}{2}\right]^{-\alpha_0} \\ &= \cosh(h/\alpha_0^{1/2})^{-\alpha_0} , \end{aligned}$$

so again the process $\{Z_0(\theta_t) \,|\, t \in \mathbb{R}\}$ is stationary (incidentally, $\cosh(h/\alpha_0^{1/2})^{-\alpha_0} \to e^{-h^2/2}$ as $\alpha_0 \to \infty$).

Since $\cosh(h/\alpha_0^{1/2})^{-\alpha_0} \geq 1 - h^2/2$ for all h and all α_0, (93) holds with $c = 1$. Also we have that as $t \to \infty$, $\theta_t \to \frac{1}{2}$ and $\mu_{\theta_t} \to 2\alpha_0$. Finally we have that

$$\rho_0(\theta_t) = r(0,t) = \cosh(t/\alpha_0)^{-\alpha} = O\left(e^{-\alpha^{1/2}t}\right)$$

as $t \to \infty$. So since $t_{n0} = \log_2 T_n = O(\log_4 n)$,

$$\sup_{|t|>t_{n0}} (\mu_{\theta_t} - \mu_0)\rho_0(\theta_t) = O\left(e^{-\alpha^{1/2}t_{n0}}\right) = O\left(\log_3 n\right)^{-\alpha^{1/2}} .$$

Since $\log T_n = O\left(\log_3 n\right)$, (83) holds for all $\alpha_0 > 1/4$ (which includes mixtures of normal variances).

8.1.3 Poisson

Consider the case where $F_0 = \text{Pois}(\lambda_0)$ and $\Theta_n \equiv \mathbb{R}$. Define $a_n = F_0^{-1}(1/n)$ and $b_n = F_0^{-1}(1 - 1/n)$. For large enough n, $a_n = 0$.

Now, $K'(\theta) = \lambda_0 e^{\theta}$, so $\hat{\theta}(x) = \log(x/\lambda_0)$. Also $\phi(\theta) = 2\lambda_0^{1/2}\left[e^{\theta/2} - 1\right]$. So

$$\hat{\phi}(x) \;=\; 2\left[\sqrt{x} - \sqrt{\lambda_0}\right] .$$

So for large enough n,

$$\begin{aligned} T_n &= \tau(\hat{\theta}(b_n)/2) - \tau(\hat{\theta}(a_n)/2) \\ &= [\hat{\phi}(b_n) - \hat{\phi}(a_n)]/2 \\ &= \sqrt{b_n} . \end{aligned}$$

From appendix C.3, we have that

$$b_n \;=\; O\left(\frac{\log n}{\log_2 n}\right) ,$$

so we have that

$$T_n = O\left(\frac{\log n}{\log_2 n}\right)^{1/2} .$$

We have from section 6.1.4 that

$$r_n = O\left(\frac{\log_2 n}{\log n}\right)^{1/4} = O(T_n^{-1/2}) = o(\log T_n)^{-1/2}$$

so (85) holds in this example.

Then $K(\theta) = \lambda_0(e^{\theta} - 1)$, $\tau(\theta) = \lambda_0^{1/2}\left[e^{\theta} - 1\right]$ and $\theta_t = \log(1 + t/\lambda_0^{1/2})$.

$$\begin{aligned} K(\theta_s + \theta_t) &= \lambda_0\left[\left(1 + \frac{s}{\lambda_0^{1/2}}\right)\left(1 + \frac{t}{\lambda_0^{1/2}}\right) - 1\right] \\ &= st + \lambda_0^{1/2}(s + t) . \end{aligned}$$

So

$$K(\theta_s + \theta_t) - \tfrac{1}{2}K(2\theta_s) - \tfrac{1}{2}K(2\theta_t) = -\tfrac{1}{2}(s - t)^2 .$$

Thus $\rho(\theta_s, \theta_t) = e^{-(s-t)^2/2}$ is the same as it is for the normal case; in particular (93) holds with $c = 1$.

Finally, we have $\mu_{\theta_t} = \lambda_0 e^{\theta_t} = \lambda_0 + \lambda_0^{1/2} t = O(t)$ as $t \to \infty$, and also

$$\begin{aligned} \log T_n &= \frac{1}{2}[\log_2 n - \log_3 n] = \frac{1}{2}\log_2 n[1 + o(1)] \ , \\ t_{n0} &= \log_2 T_n = \log_3 n - \log 2 + o(1) \ , \end{aligned}$$

and $\rho_0(\theta_t) = r(0,t) = e^{-t^2/2}$. So

$$\begin{aligned} \sup_{|t|>t_{n0}} (\mu_{\theta_t} - \mu_0)\rho_0(\theta_t) &= O\left(t_{n0}\, e^{-t_{n0}^2/2}\right) \\ &= O\left(\log_3 n\, e^{-\frac{1}{2}[\log_3 n]^2}\right) \\ &= O\left(\log_3 n\, e^{-\frac{3}{2}\log_3 n}\right) \ \text{(for large enough } n\text{)} \\ &= O\left(\frac{\log_3 n}{\log_2 n}\,(\log_2 n)^{-1/2}\right) \\ &= o\,(\log_2 n)^{-1/2} = o(\log T_n)^{-1/2} \ , \end{aligned}$$

so (83) holds.

8.1.4 Negative Binomial

Consider the case where $F_0 = \text{Neg}(m, q_0)$ and $\Theta_n \equiv (-\infty, -\frac{1}{2}\log q_0)$. Now as $n \to \infty$, $T_n = O\left(\hat{\phi}(b_n)/2\right)$, where $b_n = F_0^{-1}(1 - 1/n)$. Now

$$\phi(\theta) \ = \ C_0 - \log\left(\frac{1 - \sqrt{q_0 e^{\theta}}}{1 + \sqrt{q_0 e^{\theta}}}\right) \ ,$$

where C_0 is a constant depending only on q_0. Also

$$\hat{\theta}(x) = \log\left[\frac{x}{q_0(m + x)}\right] \ .$$

So

$$\begin{aligned}
\hat{\phi}(x) &= C_0 - \log\left(\frac{1-\left[1-\frac{m}{x+m}\right]^{1/2}}{1+\left[1-\frac{m}{x+m}\right]^{1/2}}\right) \\
&= C_0 - \log\left(\frac{\frac{m}{2(x+m)}[1+o(1)]}{2+o(1)}\right) \\
&= C_0 - \log\left(\frac{m}{4(x+m)}[1+o(1)]\right) \\
&= O\left(\log x\right)
\end{aligned}$$

as $x \to \infty$. We also have that

$$b_n \log q_0 + (m-1)\log b_n + C_n = -\log n + o(1)$$

for some uniformly bounded sequence $\{C_n \,|\, n \in \mathbb{N}\}$. A convergent iterative scheme based on this relation yields that

$$b_n = \frac{\log n + (m-1)\log -2n + C_n'}{-\log q_0} + o(1)$$

for some other uniformly bounded sequence $\{C_n' \,|\, n \in \mathbb{N}\}$. So

$$T_n = O\left(\log_2 n\right) \ .$$

Since already have that $r_n = (\log n)^{-1/2}$, we certainly have that $r_n = o(\log T_n)^{-1/2}$, so (85) holds.

Now

$$\tau(\theta) = \sqrt{m}\left[\log C_0 - \log\left(\frac{1+q_0^{1/2}e^{\theta}}{1-q_0^{1/2}e^{\theta}}\right)\right]$$

and

$$\theta_t = \log(e^{t/\sqrt{m}} - C_0) - \log(e^{t/\sqrt{m}} + C_0) - \tfrac{1}{2}\log q_0 \ ,$$

where

$$C_0 = \log\left(\frac{1-q_0^{1/2}}{1+q_0^{1/2}}\right) \ .$$

Then

$$q_0 e^{\theta_s+\theta_t} = \frac{\left(e^{s/\sqrt{m}} - C_0\right)}{\left(e^{s/\sqrt{m}} + C_0\right)} \frac{\left(e^{t/\sqrt{m}} - C_0\right)}{\left(e^{t/\sqrt{m}} + C_0\right)}$$

and

$$1 - q_0 e^{\theta_s+\theta_t} = \frac{2C_0(e^{s/\sqrt{m}} + e^{t/\sqrt{m}})}{\left(e^{s/\sqrt{m}} + C_0\right)\left(e^{t/\sqrt{m}} + C_0\right)} \ .$$

Finally,

$$\begin{aligned}
\rho(\theta_s, \theta_t) &= \frac{\exp\{K(\theta_s + \theta_t)\}}{\sqrt{\exp\{K(2\theta_s)\}\exp\{K(2\theta_t)\}}} \\
&= \left[\frac{\sqrt{(1 - q_0 e^{2\theta_s})(1 - q_0 e^{2\theta_t})}}{1 - q_0 e^{\theta_s+\theta_t}}\right]^m \\
&= \left[\frac{2\exp\{(s+t)/(2\sqrt{m})\}}{\exp\{s/\sqrt{m}\} + \exp\{t/\sqrt{m}\}}\right]^m \ .
\end{aligned}$$

In particular,

$$\begin{aligned}
\rho(\theta_t, \theta_{t+h}) &= \left[\frac{2\exp\{(2t+h)/(2\sqrt{m})\}}{\exp\{t/\sqrt{m}\} + \exp\{(t+h)/\sqrt{m}\}}\right]^m \\
&= \left[\frac{e^{-h/(2\sqrt{m})} + e^{h/(2\sqrt{m})}}{2}\right]^{-m} \\
&= \cosh[h/(2\sqrt{m})]^{-m} \ ,
\end{aligned}$$

so as in the Gamma scale case, (93) holds with $c = 1$.

Finally, $\rho_0(\theta_t) = \cosh(t/(2\sqrt{m}))^{-m} = O\left(e^{-t\sqrt{m}/2}\right)$ as $t \to \infty$. We have that as $\theta \to -\frac{1}{2}\log q_0$, $\tau(\theta) \to \infty$. So since θ_t satisfies $\tau(\theta_t) = t$, we have that as $t \to \infty$,

$$\mu_{\theta_t} = K'(\theta_t) \to K'\left(-[\log q_0]/2\right) = m q_0^{1/2}/(1 - q_0^{1/2}) < \infty \ .$$

So with $t_{n0} = \log\log T_n$,

$$\begin{aligned}
\sup_{|t|>t_{n0}} (\mu_{\theta_t} - \mu_0)\rho_0(\theta_t) &= O\left(e^{-t_{n0}\sqrt{m}/2}\right) \\
&= O\left(\log T_n\right)^{-\sqrt{m}/2} \ ,
\end{aligned}$$

so (83) holds for all $m > 1$.

9 Computer Simulation

In this section we use computer simulation to assess the accuracy of the limit theorem in approximating the finite-sample distribution of the maximum of the standardised score process. In brief, the approximation is poor, however this is not surprising, as we now explain.

The approximation can be thought of as being in two stages. Firstly the maximum of the standardised score process is approximated by the maximum of a Gaussian process. The second stage is then to approximate the distribution of the maximum of the Gaussian process by its limiting distribution. It is well known that the rate of convergence of the maxima of Gaussian processes to their limit distributions is extremely slow, see for instance Hall (1991), so we expect the error involved in the second stage of the approximation to be poor. However, the error involved in the first stage may not be so bad.

In practical applications we wish to approximate upper tail areas to obtain p-values. When testing a simple hypothesis, no matter how complicated the sampling distribution of the statistic is, we can always compute an arbitrarily accurate Monte-Carlo p-value, that is by generating a large number of pseudo-random samples from the null distribution and computing the statistic in each case. For a large enough number of such samples we can approximate the null distribution of the statistic arbitrarily accurately. This being the case, the only reason to consider some other method is that it provides comparable accuracy to the Monte-Carlo method while at the same time using considerably less computing resources.

Given that there are fast methods of simulating Gaussian processes available, then a possible alternative to the Monte-Carlo p-value is to instead simulate a large number of realisations of a Gaussian process, and then compute the maximum of each. If

1. the error involved in the first stage of our approximation is small, that is the distribution of the Gaussian process maximum provides a good approximation to the sampling distribution *and*

2. the computing resources involved are much less than generating a comparable number of pseudo-random samples from the null distribution,

then the alternate method would be "admissible" in some sense.

In our simulations we identify a particular Gaussian process that ought to satisfy these requirements, but unfortunately again we find that the approximation is poor, although considerably less computing resources are required. All is not in vain, however, because the distribution of the same Gaussian process maximum provides a very accurate approximation to the sampling distribution of the maximum of the *Studentised* score process $\sup_{\theta\in\Theta} U_n(\theta)$, where

$$U_n(\theta) = \frac{\sum_{i=1}^n u_\theta(X_i)}{\{\sum_{i=1}^n u_\theta^2(X_i)\}^{1/2}}$$

and $u_\theta(x) = e^{\theta x - K(\theta)} - 1$ (this is in fact the statistic studied in Hartigan (1985)). We present a heuristic discussion as to why this is to be expected after the results of our simulations.

9.1 Simulation method

Our method of simulation is as follows. We firstly, for $F_0 = \mathcal{N}(0,1)$ and Gamma$(1,1)$, generate N pseudo-random samples from F_0 of size n, and compute in each case the (simple-hypothesis) standardised score process

$$\begin{aligned} S_n^\dagger(\theta_t) &= n^{-1/2}\sum_{i=1}^n \left(e^{\theta_t X_i - K(\theta_t)} - 1\right)\left(e^{K(2\theta_t)-2K(\theta_t)} - 1\right)^{-1/2} \\ &= \int_0^1 y_{\theta_t}^\dagger \circ F_0^{-1}(u)\, d\alpha_n(u) \end{aligned}$$

and Studentised score process $U_n(\theta_t)$ for $m = 200$ equally spaced values of t over a certain interval $[t_{\min}, t_{\max}]$, for $(n, N) = (100, 10000)$, $(1000, 10000)$, $(10000, 1000)$ (we reduce N in the third case due to constraints on computing resources).

The second stage is to generate $M = 10000$ realisations of the Gaussian process

$$Z_n^\dagger(\theta_t) = \int_{1/n}^{1-1/n} y_{\theta_t}^\dagger \circ F_0^{-1}(u)\, dB_n(u)\ ,$$

where B_n is a $U(0,1)$ Brownian Bridge, evaluted over the same grid of t values as the two empirical score processes. The Gaussian processes are computed as follows. $Z_n^\dagger(\theta_s, \theta_t)$ has covariance function $r_n^\dagger(s,t) = \rho_n^\dagger(\theta_s, \theta_t)$, where

$$\rho_n^\dagger(\theta, \eta) \quad = \quad \frac{\int_{F_0^{-1}(1/n)}^{F_0^{-1}(1-1/n)} \left(e^{\theta x - K(\theta)} - 1\right)\left(e^{\eta x - K(\eta)} - 1\right) dF_0(x)}{\left\{\left(e^{K(2\theta)-2K(\theta)} - 1\right)\left(e^{K(2\eta)-2K(\eta)} - 1\right)\right\}^{1/2}} \ .$$

Denote the grid of t-values by $t_{\min} = t_1, t_2, t_3, \ldots, t_m = t_{\max}$. An m-by-m matrix C is computed, with (i,j)-th element $r_n^\dagger(t_i, t_j)$. A square-root matrix is then computed as $C^{1/2} = VDV^T$, where D is a (diagonal) matrix with the square-roots of the eigenvalues of C on the diagonal and zeroes elsewhere, and V is a matrix whose columns are the corresponding eigenvectors of C. Finally an M-by-m matrix Y of pseudo-random $\mathcal{N}(0,1)$ variates is generated, and then the matrix $YC^{1/2}$ is a pseudo-random realisation of a mean-zero multivariate Normal random m-vector with covariance matrix C. Our computations are performed using S-PLUS, using the function `eigen(...,symmetric=TRUE)` (technical note: in forming the matrix D, the positive part of each eigenvalue is used, as theoretically zero eigenvalues are sometimes computed as small negative values, due to numerical error).

9.2 Typical realisations

The phenomenon driving our whole analysis is that there is an interval of θ values within which the score process has more or less expected behaviour, but outside this interval the score process degrades rapidly to zero. Also, this 'interval of activity' slowly increases with the sample size n. The Gaussian process, defined via an integral with respect to a Brownian Bridge over the interval $[1/n, 1-1/n]$, also has this 'degrading' behaviour.

Below are three typical realisations of the standardised score process $S_n^\dagger(\theta_t)$ (with the corresponding Studentised score process $U_n(\theta_t)$ shown as a dotted line) followed by three typical realisations of the Gaussian process $Z_n^\dagger(\theta_t)$, for various choices of F_0 and n.

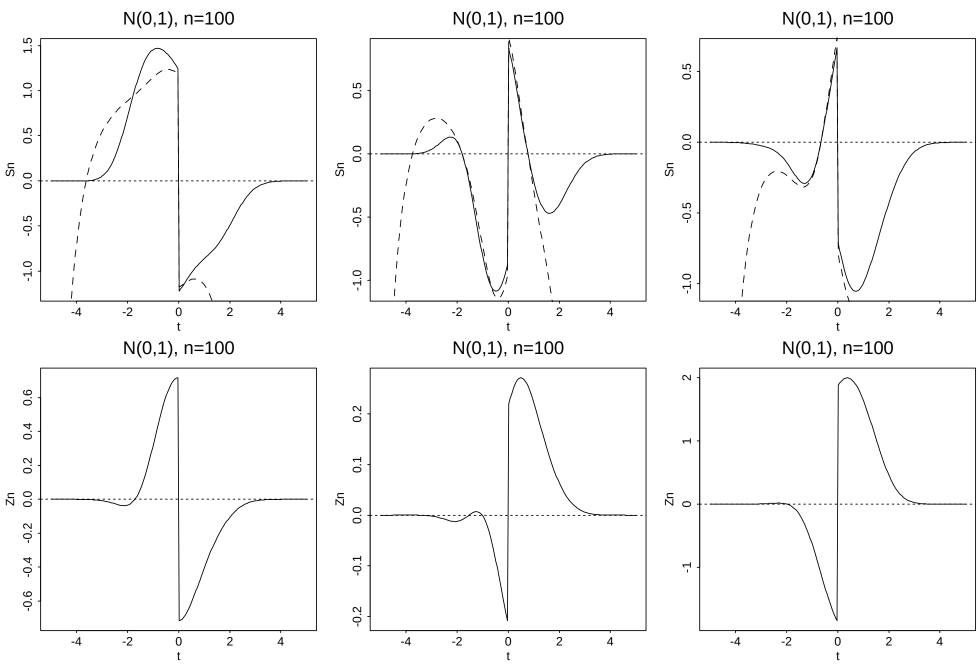
Note that there is a discontinuity at $t = 0$. This is because as $|\theta| \to 0$,

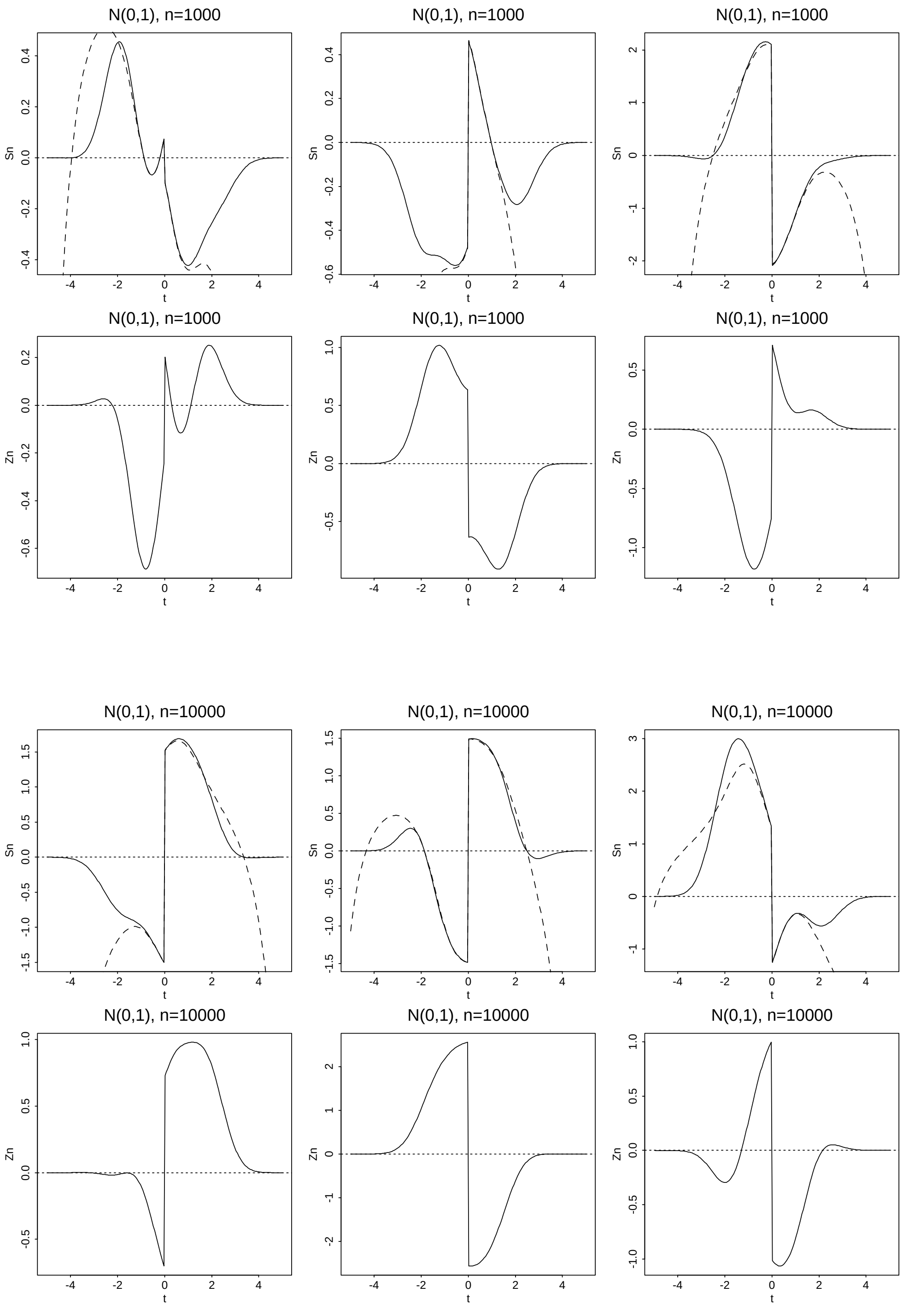
$$\begin{aligned} s_\theta^\dagger(x) &= \frac{e^{\theta x - K(\theta)} - 1}{\left\{e^{K(2\theta)-2K(\theta)} - 1\right\}^{1/2}} \\ &= \frac{\theta[x - K'(0)] + o(|\theta|)}{\left\{\theta^2 K''(0) + o(\theta^2)\right\}^{1/2}} \\ &= \operatorname{sign}(\theta)\frac{x - K'(0)}{K''(0)^{1/2}} + o(|\theta|) \ . \end{aligned}$$

So then as $\theta \downarrow 0$, $S_n(\theta) \to [nK''(0)]^{-1/2}[\bar{X} - K'(0)]$, and as $\theta \uparrow 0$, $S_n(\theta) \to -[nK''(0)]^{-1/2}[\bar{X} - K'(0)]$.

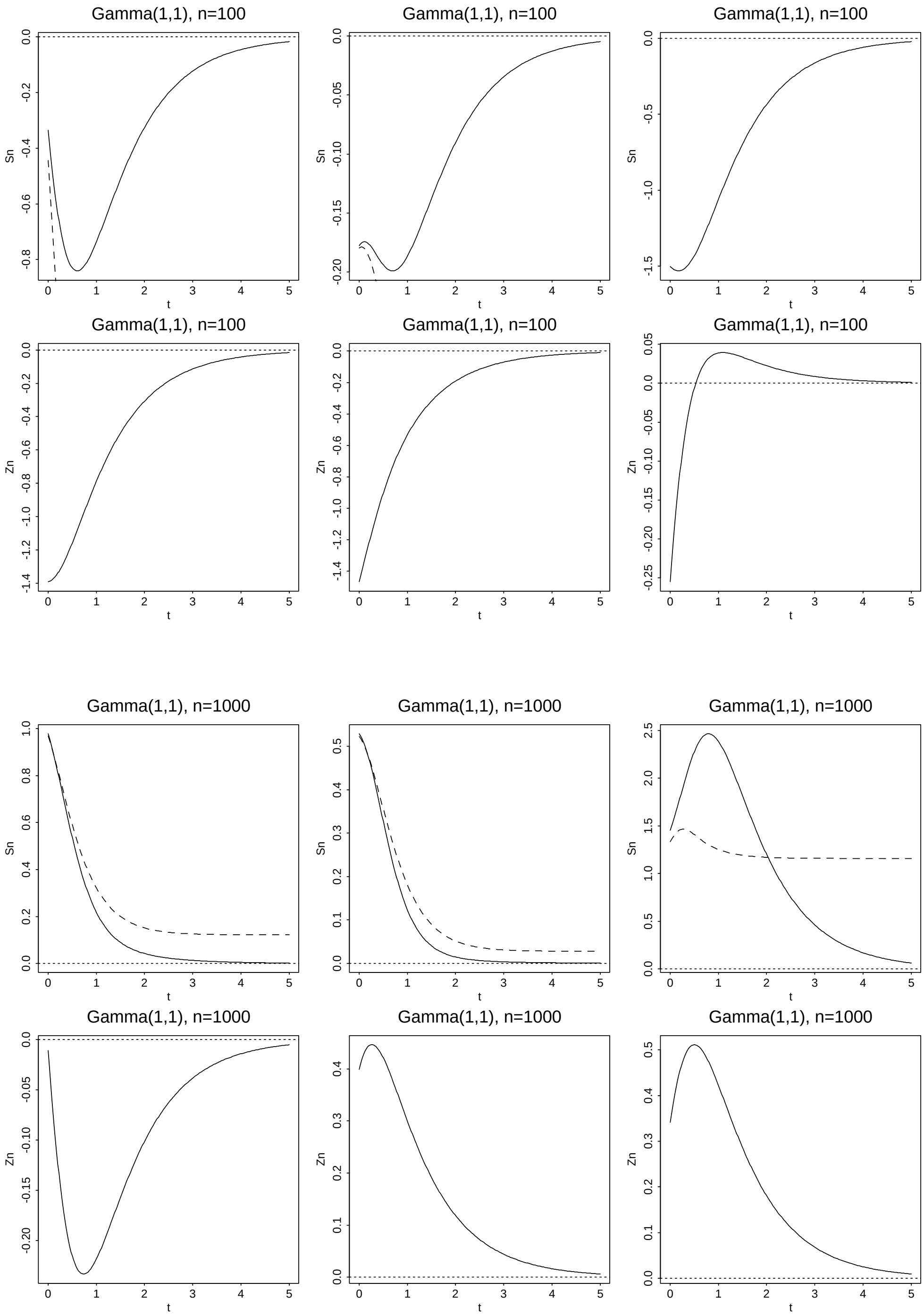
Note also that $U_n(\theta_t)$ has strange behaviour for large t. It is not completely clear why this is the case, but we remark that it has no apparent affect on the behaviour of the maximum.

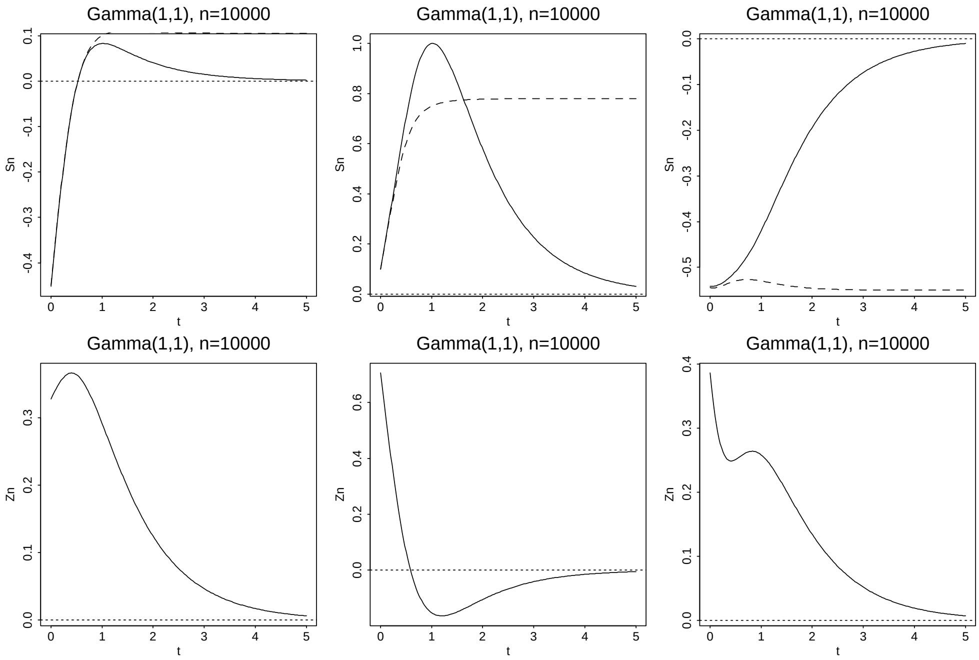
3 typical realisations of $S_n^\dagger(\theta_t)$, followed by
3 realisations of $Z_n^\dagger(\theta_t)$,
for $F_0 = \mathcal{N}(0,1)$ and $n = 100, 1000, 10000$.




N(0,1), n=1000
Sn
t
N(0,1), n=1000
Sn
t
N(0,1), n=1000
Sn
t
N(0,1), n=1000
Zn
t
N(0,1), n=1000
Zn
t
N(0,1), n=1000
Zn
t
N(0,1), n=10000
Sn
t
N(0,1), n=10000
Sn
t
N(0,1), n=10000
Sn
t
N(0,1), n=10000
Zn
t
N(0,1), n=10000
Zn
t
N(0,1), n=10000
Zn
t

Typical Realisations, $F_0 = \text{Gamma}(1,1)$.



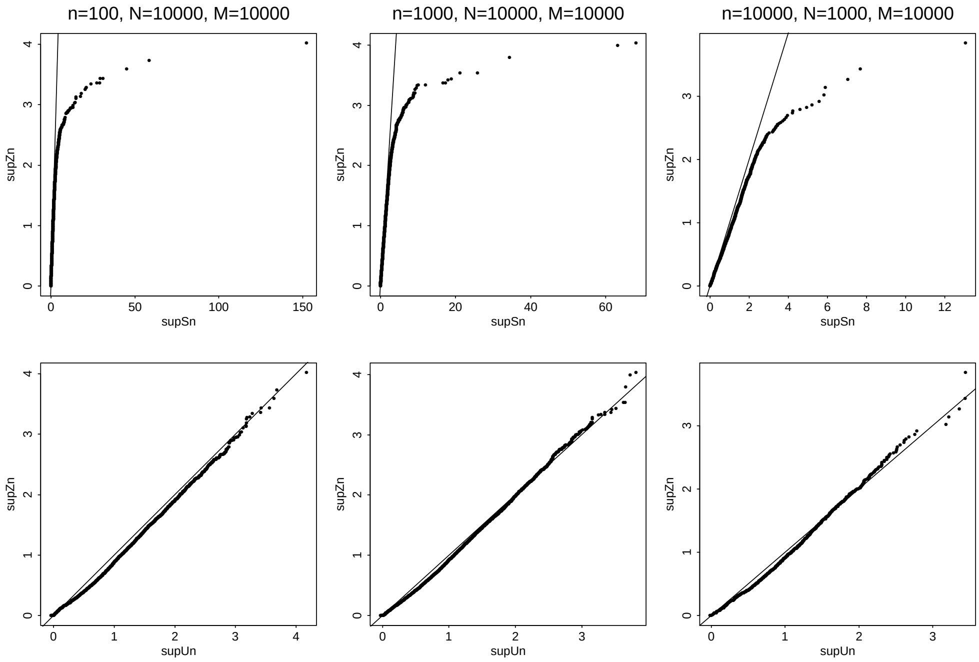


9.3 Simulation Results

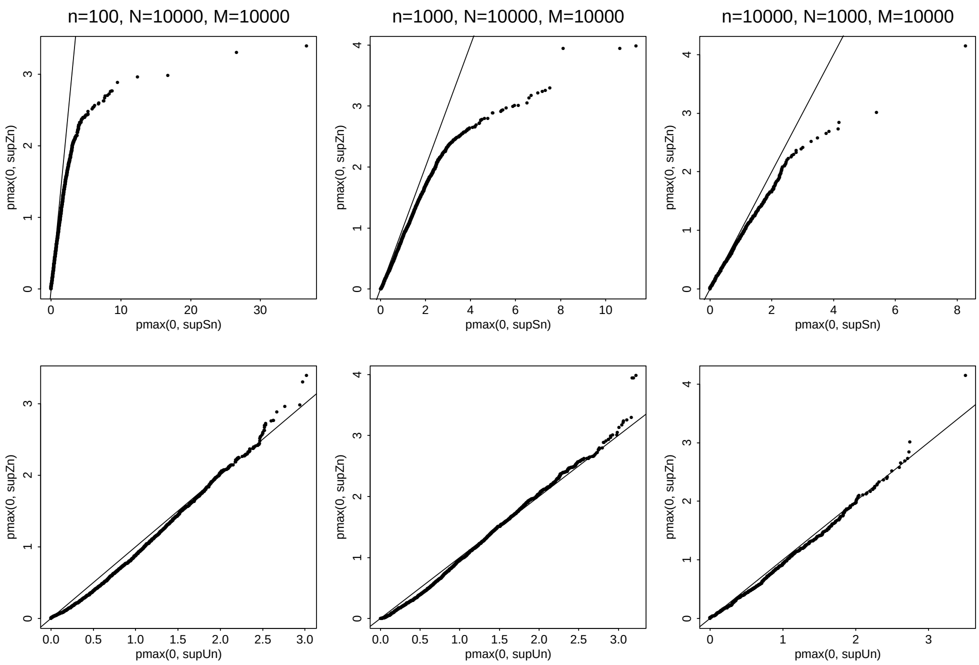
We now present the main simulation results, in the form of QQ-plots of simulated $\sup_t Z_n^\dagger(\theta_t)$ against both $\sup_t S_n^\dagger(\theta_t)$ and $\sup_t U_n(\theta_t)$, along with tables showing the actual level of accept-reject tests at certain nominal levels, using the distribution of $\sup_t Z_n^\dagger(\theta_t)$ to determine critical values.

Recall that for the case $F_0 = \mathcal{N}(0,1)$, the generating family is the full (two-sided) family $\mathcal{F} = \{F_\theta \,|\, \theta \in \mathbb{R}\}$, whereas when $F_0 = \text{Gamma}(1,1)$, we are only considering the one-sided family $\mathcal{F} = \left\{F_\theta \,\middle|\, 0 < \theta < \frac{1}{2}\right\}$. Consequently, in light of the discontinuity at $t = 0$, we always get a positive maximum $\sup_t S_n^\dagger(\theta_t)$ for the normal case, but for the Gamma scale (exponential) case we expect to get some instances where $\sup_t S_n^\dagger(\theta_t)$, and thus also $\sup_t Z_n^\dagger(\theta_t)$, is zero. Thus we have plotted $\left\{0 \vee \sup_t Z_n^\dagger(\theta_t)\right\}$ against $\left\{0 \vee \sup_t S_n^\dagger(\theta_t)\right\}$ and $\{0 \vee \sup_t U_n(\theta_t)\}$ in the Gamma$(1,1)$ case.

$F_0 = \mathcal{N}(0, 1)$



$F_0 = \text{Gamma}(1, 1)$



In the tables below, $M_n^\dagger = \sup_t S_n(\theta_t)$ and $T_n = \sup_t U_n(\theta_t)$.

Actual level of tests using simulated distribution of $\sup_t Z_n^\dagger(\theta_t)$ to determine critical values (based on Monte-Carlo simulations).

$F_0 = \mathcal{N}(0,1)$

		Nominal Level						
n	Stat.	0.1	0.05	0.025	0.01	0.005	0.0025	0.001
100	$M_n^\dagger$	0.2206	0.1481	0.1034	0.0646	0.0514	0.0398	0.0284
	T_n	0.1233	0.0643	0.0321	0.0124	0.0070	0.0029	0.0008
1000	$M_n^\dagger$	0.1906	0.1188	0.0811	0.0502	0.0387	0.0289	0.0222
	T_n	0.1081	0.0503	0.0242	0.0077	0.0040	0.0020	0.0010
10000	$M_n^\dagger$	0.155	0.093	0.057	0.039	0.028	0.024	0.019
	T_n	0.093	0.041	0.017	0.007	0.005	0.003	0.002

$F_0 = \text{Gamma}(1,1)$

		Nominal Level						
n	Stat.	0.1	0.05	0.025	0.01	0.005	0.0025	0.001
100	$M_n^\dagger$	0.1688	0.1088	0.0742	0.0454	0.0338	0.0271	0.0189
	T_n	0.1211	0.0580	0.0277	0.0088	0.0049	0.0027	0.0004
1000	$M_n^\dagger$	0.1406	0.0829	0.0484	0.0278	0.0192	0.0145	0.0102
	T_n	0.1034	0.0478	0.0230	0.0081	0.0038	0.0026	0.0008
10000	$M_n^\dagger$	0.133	0.085	0.054	0.022	0.014	0.012	0.008
	T_n	0.115	0.059	0.024	0.011	0.005	0.001	0.001

Note that in each case, $M = 10000$ replications of $\sup_t Z_n(\theta_t)$ are used to determine critical values. Note also that for $n = 100, 1000$, the number of Monte-Carlo simulations is $N = 10000$, but for $n = 10000$, computing resources limit us to $N = 1000$.

9.4 Discussion of simulation results

The simulations reveal several interesting things. Firstly, note that in all cases the distribution of $M_n^\dagger = \sup_t S_n^\dagger(\theta_t)$ is close to that of $\sup_t Z_n^\dagger(\theta_t)$ in the lower tail, but the upper tail is much larger. This leads to a high false rejection rate, which admittedly does decrease slightly across all levels as the sample size n increases, although even at $n = 10000$ the test is still grossly over-sensitive, rejecting up to 10-20 times more often than it should at the lower levels.

On the other hand, the distribution of $T_n = \sup_t U_n(\theta_t)$ closely follows the distribution of $\sup_t Z_n^\dagger(\theta_t)$ in all cases across all levels, and is particularly close in the upper tail. If anything the test is slightly conservative, at least for the larger sample sizes, so that in these cases the actual level of the test does not exceed its nominal level.

It is also of interest to compare the computing time required to complete the simulations. The table below shows the approximate time required to generate N pseudo-random samples of size n and evaluate $S_n^\dagger(\theta_t)$ at $m = 200$ values of t,

(n, N)	Approx. time
(100,10000)	30 mins.
(1000,10000)	4 hours
(10000,1000)	4 hours

whereas for all values of n, generating $M = 10000$ pseudo-realisations of the Gaussian process $Z_n^\dagger(\theta_t)$ at the same $m = 200$ values of t requires 25-30 seconds of computer time. Now the absolute times are not so important; on a sufficiently high-powered machine these times could be made arbitrarily small. The important point is that over the same grid of $m = 200$ t values, it takes about 30 seconds to generate $M = 10000$ $\sup_t Z_n^\dagger(\theta_t)$ values, but it takes about $30(n/2)$ seconds to generate the same number $N = 10000$ of $\sup_t S_n^\dagger(\theta_t)$ values based on samples of size n. So in this sense the computing time for the empirical process simulation is about $n/2$ times the computing time for the Gaussian process simulation.

9.5 The Studentised score process

It is quite remarkable that the maximum of the Studentised score process follows the distribution of the maximum of the Gaussian process so closely, whereas the maximum of the standardised score process follows it so poorly. We can offer a heuristic explanation for this.

The standardised score process is a normalised sum of independent and identically distributed standardised random variables which are *very skewed*.

In fact the skewness of $s_{\theta_t}(X)$, where $X \sim F_0$, increases as $|t|$ increases. Consequently for larger values of t, a larger n is needed before $S_n^\dagger(\theta_t)$ is approximately normal. So for any fixed n, for a range of small $|t|$ values $S_n^\dagger(\theta_t)$ is 'quite Normal', and so is well approximated by the Gaussian process $Z_n^\dagger(\theta_t)$. However for larger values of $|t|$, the Gaussian process provides a poor approximation to $S_n^\dagger(\theta_t)$, which for larger values of $|t|$ is very non-normal, in particular has a much fatter upper tail than a $\mathcal{N}(0,1)$ variable. Thus when $S_n^\dagger(\theta_t)$ is maximised for small values of $|t|$, the behaviour is much like the maximum of $Z_n^\dagger(\theta_t)$. However when $S_n^\dagger(\theta_t)$ is maximised for larger $|t|$, the maximum will tend to be larger than the maximum of $Z_n^\dagger(\theta_t)$. This explains why the distributions of the two maxima $\sup_t S_n^\dagger(\theta_t)$ $\sup_t Z_n^\dagger(\theta_t)$ disagree most in the upper tails.

However the *Studentised* score process

$$U_n(\theta) = \frac{\sum_{i=1}^n \left(e^{\theta X_i - K(\theta)} - 1\right)}{\left\{\sum_{i=1}^n \left(e^{\theta X_i - K(\theta)} - 1\right)^2\right\}^{1/2}}$$

can be thought of as a partly "skewness corrected" version of $S_n^\dagger(\theta_t)$, because when the numerator is large, the denominator is large too. In fact, technically the Studentised score statistic is *over*-corrected, as in fact it theoretically has negative skewness. However for each fixed t the upper tail of the distribution of $U_n(\theta_t)$ is still quite 'normal-like' by comparison to the fat upper tail of its non-Studentised counterpart, especially for large $|t|$. Also, examination of the typical realisation graphs earlier in this section reveal that as a process, $U_n(\theta_t)$ is very unlike the Gaussian process $Z_n^\dagger(\theta_t)$; however, as far as the *maximum* is concerned they are very alike.

A Infinite Products

We recall some results from real analysis regarding convergence of infinite products.

Lemma A.1. *If $\{x_i \mid i \in \mathbb{N}\}$ is a real sequence with each $x_i > -1$, then*

$$\begin{aligned}\sum_{i=1}^{\infty} x_i \text{ converges absolutely} \quad &\Rightarrow \quad \prod_{i=1}^{\infty}(1+x_i) \text{ converges} \\ &\Leftrightarrow \quad \sum_{i=1}^{\infty} \log(1+x_i) \text{ converges.}\end{aligned}$$

Proof. See for instance Bartle (1976), page 305. □

Next we apply these results to a sequence of such products.

Lemma A.2. *Let $\{p_{nj} \mid n \in \mathbb{N}, j \in \mathbb{N}\}$ be a triangular array of real numbers such that $0 \leq p_{nj} < 1$ for all $n, j \in \mathbb{N}$. Suppose that for each n,*

$$\sum_{j=1}^{\infty} p_{nj} = C_n < \infty \ , \tag{97}$$

and the maximum term $q_n = \sup_j p_{nj} \to 0$ as $n \to \infty$. Suppose also that $C_n \to C < \infty$ as $n \to \infty$. Then for each n the infinite product

$$\prod_{j=1}^{\infty}(1-p_{nj})$$

converges, and tends to e^{-C} as $n \to \infty$.

Proof. Since $0 \leq p_{nj} < 1$ for all n and j, for each n the sequence $\{-p_{nj} \mid j \in \mathbb{N}\}$ satisfies the conditions for Lemma A.1. Further, because each summand has the same sign, $\sum_j(-p_{nj})$ converges absolutely, and so both $\prod_j(1-p_{nj})$ and $\sum_j \log(1-p_{nj})$ converge for each n. In particular,

$$\prod_{j=1}^{\infty}(1-p_{nj}) \quad = \quad \exp\left\{\sum_{j=1}^{\infty}\log(1-p_{nj})\right\} \ .$$

Using Taylor's theorem with the integral form of the remainder we see that

$$\begin{aligned} |\log(1-p)+p| &\leq \int_{1-p}^{1} |1-p-t|/t^2\, dt \\ &\leq p^2/(1-p)^2 \ . \end{aligned}$$

So then

$$\begin{aligned} \left|\sum_{j=1}^{\infty} \log(1-p_{nj})+p_{nj}\right| &\leq \sum_{j=1}^{\infty} |\log(1-p_{nj})+p_{nj}| \\ &\leq \sum_{j=1}^{\infty} \frac{p_{nj}^2}{(1-p_{nj})^2} \\ &\leq \frac{q_n}{(1-q_n)^2} \sum_{j=1}^{\infty} p_{nj} \\ &\to 0 \end{aligned}$$

since by assumption $q_n = \sup_j p_{nj} \to 0$, and the sum $\sum_j p_{nj} \to C < \infty$. We have that

$$\begin{aligned} \sum_{j=1}^{\infty} \log(1-p_{nj}) &= -\sum_{j=1}^{\infty} p_{nj} + \sum_{j=1}^{\infty} \log(1-p_{nj})+p_{nj} \\ &= -\sum_{j=1}^{\infty} p_{nj} + o(1) \\ &\to -C \end{aligned}$$

so

$$\prod_{j=1}^{\infty}(1-p_{nj}) = \exp\left\{\sum_{j=1}^{\infty} \log(1-p_{nj})\right\} \to e^{-C} \ .$$

□

Note that since a finite sequence can be identified with an infinite sequence which has all terms beyond a certain point equal to zero, the above holds for a infinite sequence of *finite* sequences, that is triangular arrays of the form $\{p_{nj} \,|\, j = 1, 2, \ldots, m_n < \infty, n \in \mathbb{N}\}$ also.

B Differentiating correlation-like functions

Suppose $C\colon \Theta \times \Theta \to \mathbb{R}$ is symmetric in its arguments, for some $\Theta \subset \mathbb{R}$, and has continuous mixed partial derivatives up to order 2. Define

$$\rho(\theta_1, \theta_2) := \frac{C(\theta_1, \theta_2)}{\{C(\theta_1, \theta_1)C(\theta_2, \theta_2)\}^{1/2}}$$

Then, with

$$C^{(i,j)}(\theta) = \left.\frac{\partial^{i+j} C(\theta_1, \theta_2)}{\partial\theta_1^i \partial\theta_2^j}\right|_{\theta_1=\theta_2=\theta} ,$$

we have that

$$\rho(\theta, \theta + \varepsilon) = 1 - \lambda(\theta)\frac{\varepsilon^2}{2} + o(\varepsilon^2) ,$$

as $\varepsilon \to 0$ for each θ, where

$$\begin{aligned} \lambda(\theta) &= \frac{C^{(1,1)}(\theta)}{C(\theta, \theta)} - \left(\frac{C^{(1,0)}(\theta)}{C(\theta, \theta)}\right)^2 \\ &= \left.\frac{\partial^2 \log C(\theta_1, \theta_2)}{\partial\theta_1 \partial\theta_2}\right|_{\theta_1=\theta_2=\theta} \end{aligned}$$

C Approximating tail probabilities and quantiles

We collect together various results on asymptotic approximations of tail probabilities and quantiles for exponential families. We present three subsections:

1. a subsection on the general integration-by-parts method of approximating tail probabilities, pertaining to the Normal and Gamma distributions

2. a collection of elementary and established equalities and inequalities for tail probabilities for the Gamma, Beta, Binomial, Poisson and Negative Binomial distributions;

3. a subsection on asymptotic approximation of quantiles for all the examples considered in the first two subsections.

Although most of the material is well-known (the integration-by-parts subsection contains a few novel remarks), we collect it here in an appendix as a convenient reference. In particular although the convergent iterative scheme used in the quantile subsection is not new, there is apparently no easily accessible reference for the approximations obtained there.

C.1 The integration-by-parts method of approximating tail probabilities

Suppose F is a continuous cdf on the real line with density f, and write $L = \log f$. Formally integrating by parts once gives

$$\int_a^b e^{L(y)}\,dy = \int_a^b \left(\frac{1}{L'(y)}\right) L'(y) e^{L(y)}\,dy = \left[\frac{f(y)}{L'(y)}\right]_a^b + \int_a^b \frac{L''(y)}{L'(y)^2} f(y)\,dy \ .$$

In many cases of interest

$$-\infty < \int_a^b \frac{L''(y)}{L'(y)^2} f(y)\,dy \le 0 \ , \quad \text{that is} \quad 0 \le \int_a^b -\frac{L''(y)}{L'(y)^2} f(y)\,dy < \infty \ . \tag{98}$$

Integrating by parts once more gives

$$\begin{aligned}\int_a^b e^{L(y)}\,dy &= \left[\frac{f(y)}{L'(y)}\right]_a^b + \left[f(y)\frac{L''(y)}{L'(y)^3}\right]_a^b + \int_a^b \frac{3L''(y)^2 - L'(y)L^{(3)}(y)}{L'(y)^4}\,f(y)\,dy \\ &= \left[\frac{f(y)}{L'(y)}\left(1+\frac{L''(y)}{L'(y)^2}\right)\right]_a^b + \int_a^b \frac{3L''(y)^2 - L'(y)L^{(3)}(y)}{L'(y)^4}\,f(y)\,dy\ .\end{aligned}$$

In many cases of interest

$$0 \le \int_a^b \frac{3L''(y)^2 - L'(y)L^{(3)}(y)}{L'(y)^4}\,f(y)\,dy\ < \infty\ . \tag{99}$$

When both (98) and (99) hold, we have

$$\left[\frac{f(y)}{L'(y)}\left(1+\frac{L''(y)}{L'(y)^2}\right)\right]_a^b \le \int_a^b e^{L(y)} \le \left[\frac{f(y)}{L'(y)}\right]_a^b\ .$$

We now consider some examples where we verify (98) and (99). We begin with the well known results for Normal tail probabilities to illustrate our methods.

Normal Bounds for tail probabilities for the Normal distribution are well known, but for the sake of illustration we verify them using our general method.

$f(x) = \phi(x) = (2\pi)^{-1/2}e^{-x^2/2}$. Then $L'(x) = -x$, $L''(x) = -1$, $L^{(3)}(x) = 0$. Also we have

$$-\frac{L''(y)}{L'(y)^2} = \frac{1}{x^2}\ ,\quad \frac{3L''(y)^2 - L'(y)L^{(3)}(y)}{L'(y)^4} = \frac{3}{x^4}\ .$$

For any $x > 0$,

$$0 \le \int_x^\infty -\frac{L''(y)}{L'(y)^2}\,f(y)\,dy = \int_x^\infty \frac{1}{y^2}\phi(y)\,dy \le \frac{1}{x^2}\int_x^\infty \phi(y)\,dy \le \frac{1}{2x^2} < \infty$$

and

$$0 < \int_x^\infty \frac{3L''(y)^2 - L^{(3)}(y)L'(y)}{L'(y)^4}\,f(y)\,dy = \int_x^\infty \frac{3}{x^4}\phi(y)\,dy \le \frac{3}{2x^4} < \infty\ .$$

So (98) and (99) hold in this case, giving,

$$\frac{\phi(x)}{x}\left(1-\frac{1}{x^2}\right) \le 1-\Phi(x) \le \frac{\phi(x)}{x}\ .$$

Since ϕ is symmetric about 0 the corresponding result holds for the lower tail.

Gamma Consider the general gamma density which we write as

$$g(y) = y^{\alpha-1}e^{-\lambda y}\lambda^{\alpha}/\Gamma(\alpha)\ , \quad \text{for } y > 0\ .$$

We apply our general method to the upper tail. We get

$$\begin{aligned} L'(y) &= (\alpha-1)y^{-1} - \lambda\ , \\ L''(y) &= -(\alpha-1)y^{-2}\ , \\ L^{(3)}(y) &= 2(\alpha-1)y^{-3}. \end{aligned}$$

Write

$$\begin{aligned} I_1(y) &= -\frac{L''(y)}{L'(y)^2} = \frac{\alpha-1}{[(\alpha-1)-\lambda y]^2} \quad \text{and} \\ I_2(y) &= \frac{3L''(y)^2 - L'(y)L^{(3)}(y)}{L'(y)^4} \\ &= \frac{3(\alpha-1)^2 - (\alpha-1)[(\alpha-1)-\lambda y]}{[(\alpha-1)-\lambda y]^4} \\ &= \frac{2(\alpha-1)^2 + (\alpha-1)\lambda y}{[(\alpha-1)-\lambda y]^4} \end{aligned}$$

Now these two functions are, for $\alpha > 1$, postive everywhere except for vertical asymptotes at $y = (\alpha-1)/\lambda$. Also,

$$\begin{aligned} I_1'(y) &= \frac{2(\alpha-1)\lambda}{[(\alpha-1)-\lambda y]^3}\ , \\ I_2'(y) &= \frac{3(\alpha-1)\lambda[3(\alpha-1)+\lambda y]}{[(\alpha-1)-\lambda y]^5}\ . \end{aligned}$$

So for $\alpha > 1$, both I_1 and I_2 are increasing on $(0, (\alpha-1)/\lambda)$ and decreasing on $((\alpha-1)/\lambda), \infty)$. For $\alpha < 1$, the *opposite* is true for (for sufficiently large y in the case of I_2). However so long as $\alpha > 1$ we can verify (98) and (99) for upper (and lower) tail probabilities.

For any $y > (\alpha-1)/\lambda$,

$$\begin{aligned} 0 < \int_y^\infty I_1(x)g(x)\,dx &\leq I_1(y)\int_y^\infty g(x)\,dx \\ &\leq \frac{\alpha-1}{[(\alpha-1)-\lambda y]^2} < \infty \end{aligned}$$

and

$$0 < \int_y^\infty I_2(x) g(x)\, dx \;\leq\; I_2(y) \int_y^\infty g(x)\, dx$$
$$\leq\; \frac{2(\alpha-1)^2 + (\alpha-1)\lambda y}{[(\alpha-1)-\lambda y]^4} < \infty\ .$$

So for any $\alpha > 1$, we have

$$\frac{y^\alpha e^{-\lambda y} \lambda^\alpha}{\Gamma(\alpha)[\lambda y - (\alpha-1)]} \left[1 - \frac{\alpha-1}{[(\alpha-1)-\lambda y]^2}\right] \leq 1 - G(y) \leq \frac{y^\alpha e^{-\lambda y} \lambda^\alpha}{\Gamma(\alpha)[\lambda y - (\alpha-1)]} \tag{100}$$

We can in fact improve these bounds in the sense that we can find similar bounds that are valid for all $\alpha > 0$. The trick is to perform a monotone transformation to get a new distribution, apply the general method to obtain upper and lower bounds, then transform back. The bounds should be close to the original ones, but they may be valid for all $\alpha > 0$.

Suppose that Y has cdf G, and let $X = \log \lambda Y$. Then X has cdf $F(x) = G(e^x/\lambda)$, and density

$$f(x) = e^{\alpha x - e^x - \log \Gamma(\alpha)}\ , \quad \text{for } -\infty < x < \infty.$$

Then $L'(x) = \alpha - e^x$, $L''(x) = -e^x = L^{(3)}(x)$. Also

$$J_1(x) = -\frac{L''(x)}{L'(x)^2} \;=\; \frac{e^x}{(e^x-\alpha)^2}\ .$$
$$J_1'(x) \;=\; \frac{-(e^{2x} + \alpha e^x)}{(e^x-\alpha)^3}\ ,$$

so is increasing for $x < \log \alpha$ and decreasing for $x > \log \alpha$. Also

$$J_2(x) \;=\; \frac{3L''(x)^2 - L^{(3)}(x) L'(x)}{L'(x)^4} = \frac{3e^{2x} + e^x(\alpha - e^x)}{(e^x-\alpha)^4} = \frac{2e^{2x} + \alpha e^x}{(e^x-\alpha)^4} \text{ and}$$
$$J_2'(x) \;=\; \frac{-(4e^{3x} + 7\alpha e^{2x} + \alpha^2 e^x)}{(e^x-\alpha)^5}\ ,$$

so is also increasing for $x < \log \alpha$ and decreasing for $x > \log \alpha$. Both functions have vertical asymptotes at $x = \log \alpha$.

For the upper tail, for $i = 1, 2$, since J_i is decreasing on $(\log \alpha, \infty)$,

$$\begin{aligned} 0 &< \int_x^\infty J_i(y) f(y)\, dy \\ &\le J_i(x) \int_x^\infty f(y)\, dy \le J_i(x) < \infty \end{aligned}$$

for any $\alpha > 0$ and any $x > \log \alpha$. So (98) and (99) hold and we have

$$\frac{e^{\alpha x - e^x - \log \Gamma(\alpha)}}{e^x - \alpha} \left[1 - \frac{e^x}{(e^x - \alpha)^2} \right] \le 1 - F(x) \le \frac{e^{\alpha x - e^x - \log \Gamma(\alpha)}}{e^x - \alpha} \, .$$

Now, since $G(y) = F(\log \lambda y)$, we can transform back to get, for any $y > \alpha/\lambda$,

$$\frac{y^\alpha e^{-\lambda y} \lambda^\alpha}{\Gamma(\alpha)[\lambda y - \alpha)]} \left[1 - \frac{\lambda y}{[\lambda y - \alpha]^2} \right] \le 1 - G(y) \le \frac{y^\alpha e^{-\lambda y} \lambda^\alpha}{\Gamma(\alpha)[\lambda y - \alpha)]} \, . \qquad (101)$$

It is of interest to compare (101) with (100). It seems we have got something for nothing here; the bounds in (101) are valid for all $\alpha > 0$, whereas those in (100) are only valid for $\alpha > 1$. However, we have paid a price for this, since the relative error factor [in square brackets] is of the form $1 + O(y^{-1})$ as $y \to \infty$ in (101), whereas it is of the form $1 + O(y^{-2})$ in (100).

Since J_i is increasing on $(-\infty, \log \alpha)$ for $i = 1, 2$, we can perform identical calculations to verify (98) and (99) in the lower tail case. We thus get for any $\alpha > 0$ and $x < \log \alpha$,

$$\frac{e^{\alpha x - e^x - \log \Gamma(\alpha)}}{\alpha - e^x} \left[1 - \frac{e^x}{(\alpha - e^x)^2} \right] \le F(x) \le \frac{e^{\alpha x - e^x - \log \Gamma(\alpha)}}{\alpha - e^x}$$

and thus using $G(y) = F(\log \lambda y)$, for any $\alpha > 0$ and $y < \alpha/\lambda$,

$$\frac{y^\alpha e^{-\lambda y} \lambda^\alpha}{\Gamma(\alpha)(\alpha - \lambda y)} \left[1 - \frac{\lambda y}{(\alpha - \lambda y)^2} \right] \le G(y) \le \frac{y^\alpha e^{-\lambda y} \lambda^\alpha}{\Gamma(\alpha)(\alpha - \lambda y)} \, , \qquad (102)$$

which affords a slight refinement of the elementary bounds obtained below in (103).

C.2 Other equalities and inequalities

Gamma We can use a simple method for bounding the lower tail of the Gamma distribution as follows:

$$\frac{e^{-\lambda y} y^\alpha \lambda^\alpha}{\alpha \Gamma(\alpha)} = e^{-\lambda y} \int_0^y \frac{x^{\alpha-1} \lambda^x}{\Gamma(\alpha)}\, dy \le \int_0^y \frac{x^{\alpha-1} e^{-\lambda x} \lambda^\alpha}{\Gamma(\alpha)}\, dy \le \int_0^y \frac{x^{\alpha-1} \lambda^\alpha}{\Gamma(\alpha)}\, dy = \frac{y^\alpha \lambda^\alpha}{\alpha \Gamma(\alpha)} \qquad (103)$$

Negative Binomial, Binomial and Beta Tail probabilties for the Beta distribution are related to those of the Binomial and Negative Binomial distributions. Suppose that X represents the number of failures occuring before the m-th success in a series of independent Bernoulli trials with success probability $1-q$. We then say that X has a Negative Binomial distribution with parameters m and q, or just $X \sim \text{Neg}(m,q)$; we call m the required number of successes and q the failure probability.

If $X \geq x$, then the m-th success falls on or after the $(x+m)$-th trial. So in the first $(x+m-1)$ trials, no more than $(m-1)$ successes occur. So if we define Y as the number of successes in the first $(x+m-1)$ trials, $\{X \geq x\} \Leftrightarrow \{Y \leq m-1\}$. So then

$$P\{X \geq x\} = P\{Y \leq m-1\} \ .$$

where $Y \sim \text{Bin}(x+m-1, 1-q)$.

Now let $Z \sim \text{Beta}(x,m)$; that is, Z has pdf

$$f(z) = z^{x-1}(1-z)^{m-1}/B(x,m) \ ,$$

where

$$B(\alpha,\beta) = \int_0^1 z^{\alpha-1}(1-z)^{\beta-1}\,dz \quad = \quad \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

is the Beta function and

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$$

is the Gamma function. Note that for positive integer α, $\Gamma(\alpha) = (\alpha-1)!$. So $B(x,m) = [(x-1)!(m-1)!]/(x+m-1)!$.

By changing variable via $z = q(1-v)$ it can be shown (see Kendall *et al.* (1987), section 5.7) that

$$\begin{aligned} \int_0^q f(z)\,dz &= \sum_{j=0}^{x+m-1} (1-q)^j q^{x-j} \frac{x!}{j!(x-j)!} \\ &= P(Y \leq m-1) \ , \end{aligned}$$

thus connecting the Negative Binomial, Binomial and Beta distribution functions:

$$P\{\text{Neg}(m,q) \geq x\} = P\{\text{Bin}(x-m+1, 1-q) \leq m-1\} = P\{\text{Beta}(x,m) \leq q\} \ . \tag{104}$$

Thus by approximating tail probabilities and quantiles for the Beta distribution we also do the same for the Negative Binomial and Binomial. We are interested in approximating the upper tail of the Negative Binomial distribution and also both tails of the Binomial (m_n, p_0), with $m_n \to \infty$.

Upper and Lower bounds for Beta tail probabilities are obtained from the simple relation

$$\frac{x^\alpha(1-x)^{\beta-1}}{\alpha} = (1-x)^{\beta-1}\int_0^x z^{\alpha-1}\,dz \le \int_0^x z^{\alpha-1}(1-z)^{\beta-1}\,dz \le \int_0^x z^{\alpha-1}\,dz = \frac{x^\alpha}{\alpha} \tag{105}$$

for $\beta \ge 1$ and

$$\frac{x^\alpha}{\alpha} \le \int_0^x z^{\alpha-1}(1-z)^{\beta-1}\,dz \le \frac{x^\alpha(1-x)^{\beta-1}}{\alpha}$$

for $\beta < 1$.

So (104) gives that

$$\frac{q^x(1-q)^{m-1}(x+m-1)!}{x!(m-1)!} \le P\left\{\text{Neg}(m,q) \ge x\right\} \le \frac{q^x(x+m-1)!}{x!(m-1)!}\ .$$

Now,

$$\begin{aligned}(x+m-1)! &= (x+m-1)(x+m-2)! \\ &= (x+m-1)(x+m-2)(x+m-3)! \\ &= (x+m-1)(x+m-2)\cdots(x+1)x!\ .\end{aligned}$$

So

$$\begin{aligned}\frac{(x+m-1)!}{x!} &= (x+m-1)(x+m-2)\cdots(x+1) \\ &= x^{m-1}\{(1+[m-1]/x)(1+[m-2]/x)\cdots(1+1/x)\} \\ &= x^{m-1}[1+o(1)]\end{aligned}$$

as $x \to \infty$. So $\log P\left\{\text{Neg}(m,q) \ge x\right\}$ differs from

$$x\log q + (m-1)\log x - \log(m-1)!$$

by a constant C satisfying $(m-1)\log(1-q) \le C \le 0$ plus a term that goes to zero as $x \to \infty$. In particular,

$$P\left\{\text{Neg}(m,q) \ge x\right\} = O\left(q^x x^{m-1}\right)\ .$$

Poisson We can use the previous section to establish approximations for the upper tail of the Poisson distribution. Suppose $X \sim \text{Pois}(\lambda)$ and $Y \sim \text{Gamma}(x, 1)$, for some integer x. Since $\Gamma(x) = (x-1)!$ for integer x, repeated integration by parts gives

$$\begin{aligned}
P(Y \leq \lambda) &= \frac{1}{(x-1)!} \int_0^\lambda y^{x-1} e^{-y}\, dx \\
&= \frac{1}{(x-1)!} \int_0^\lambda e^{-y} d\left(\frac{y^x}{x}\right) \\
&= \frac{1}{(x-1)!} \left\{ \left[\frac{e^{-y} y^x}{x}\right]_0^\lambda - \int_0^\lambda \frac{y^x}{x} d\left(e^{-y}\right) \right\} \\
&= \frac{e^{-\lambda}\lambda^x}{x!} + \frac{1}{x!} \int_0^\lambda y^x e^{-y}\, dy \\
&= \frac{e^{-\lambda}\lambda^x}{x!} + \frac{1}{x!} \int_0^\lambda e^{-y} d\left(\frac{y^{x+1}}{x+1}\right) \\
&= \frac{e^{-\lambda}\lambda^x}{x!} + \frac{e^{-\lambda}\lambda^{x+1}}{(x+1)!} + \frac{1}{(x+1)!} \int_0^\lambda y^{x+1} e^{-y}\, dy \\
&= \frac{e^{-\lambda}\lambda^x}{x!} + \frac{e^{-\lambda}\lambda^{x+1}}{(x+1)!} + \frac{e^{-\lambda}\lambda^{x+2}}{(x+2)!} + \ldots \qquad (106) \\
&= P(X \geq x) \ .
\end{aligned}$$

So we can use our *lower* tail inequalities for the Gamma distribution to obtain bounds on the upper tail of the Poisson distribution. Using (102) we have

$$\frac{\lambda^x e^{-\lambda}}{(x-1)![x-\lambda]} \left[1 - \frac{\lambda}{(x-\lambda)^2}\right] \leq P(X \geq x) \leq \frac{\lambda^x e^{-\lambda}}{(x-1)![x-\lambda]} \ .$$

In particular, as $x \to \infty$,

$$P(X \geq x) \sim \frac{\lambda^x e^{-\lambda}}{x!} \ , \qquad (107)$$

confirming the well known result that the first term in (106) dominates the sum as $x \to \infty$.

C.3 Quantiles

In the course of our analysis we need to be able to approximate quantiles for various distributions corresponding to tails of $n^{-1/2} \log n$ and n^{-1}. We now provide these approximations all in one place for each of the examples we consider.

Having established methods of approximating tail probabilities, we have a complementary and quite general procedure for approximating quantiles as the corresponding tail probability gets small at a certain rate. The method is described as a *convergent iterative scheme*, and is outlined in Barndorff-Nielsen & Cox (1989), Section 3.5, Example 3.13. The method is generally applicable wherever the corresponding tail probability is approximated by a function including an exponential term that dominates as the tail probability gets small.

Normal If $X \sim \mathcal{N}(0,1)$, the well-known approximation for the upper tail is

$$P(X \geq x) \quad = \quad \frac{e^{-x^2/2}}{\sqrt{2\pi}x}[1+o(1)]$$

as $x \to \infty$. Suppose x_n satisfies $P(X \geq x_n) = n^{-1/2}\log n$. Then as $n \to \infty$,

$$\begin{aligned}
\frac{e^{-x_n^2/2}}{\sqrt{2\pi}x_n} &= n^{-1/2}\log n[1+o(1)] \\
-\frac{x_n^2}{2} - \tfrac{1}{2}\log 2\pi - \log x_n &= -\tfrac{1}{2}\log n + \log_2 n + o(1) \\
x_n^2 &= \log n - 2\log_2 n - \log 2\pi - 2\log x_n + o(1)\ .
\end{aligned}$$

after taking logs and writing $\log_2 n = \log\log n$.

The plan is to define a sequence $x_{(1)}, x_{(2)}, \ldots$ where $x_{(k+1)}$ is defined in terms of $x_{(k)}$ via

$$x_{(k+1)}^2 \quad = \quad \log n - 2\log_2 n - \log 2\pi - 2\log x_{(k)}\ .$$

If all goes well, we choose a sensible $x_{(1)}$, and iterate until $x_{(k+1)}^2 = x_{(k)}^2 + o(1)$. Such a $x_{(k+1)}^2$ will agree with x_n^2 up to $o(1)$ as $n \to \infty$.

A sensible starting point would be $x_{(1)} = (\log n)^{1/2}$, but we start with $x_{(1)} = 1$ to illustrate our method. Then we get

$$\begin{aligned}
x_{(2)}^2 &= \log n - 2\log_2 n - \log 2\pi - 0 \\
&= \log n[1+o(1)]\ .
\end{aligned}$$

Taking logs gives

$$\begin{aligned} 2\log x_{(2)} &= \log_2 n + o(1)\ , \quad \text{so then} \\ x_{(3)}^2 &= \log n - 2\log_2 n - \log 2\pi - 2\log x_{(2)} \\ &= \log n - 3\log_2 n - \log 2\pi + o(1)\ . \\ &= \log n[1 + o(1)]\ . \end{aligned}$$

Taking logs again gives

$$\begin{aligned} 2\log x_{(3)} &= \log_2 n + o(1)\ , \text{ so} \\ x_{(4)}^2 &= \log n - 2\log_2 n - \log 2\pi - 2\log x_{(3)} \\ &= \log n - 3\log_2 n - \log 2\pi + o(1) \\ &= x_{(3)}^2 + o(1)\ . \end{aligned}$$

Thus x_n satisfies

$$x_n^2 = \log n - 3\log_2 n - \log 2\pi + o(1)\ .$$

Now write $a_n = F_0^{-1}(1/n)$ and $b_n = F_0^{-1}(1 - 1/n)$. The upper tail approximation of $1 - F_0(x)$ gives that, as $n \to \infty$,

$$\begin{aligned} 1 - F_0(b_n) = 1/n &= \frac{e^{-b_n^2/2}}{\sqrt{2\pi} b_n}[1 + o(1)]\ , \\ \text{so} \quad -\log n &= -\frac{b_n^2}{2} - \log b_n - \tfrac{1}{2}\log 2\pi + o(1) \\ b_n^2 &= 2\log n - 2\log b_n - \log 2\pi + o(1)\ . \end{aligned}$$

Defining $b_{(1)}^2 = 2\log n$ and iterating via

$$b_{(k+1)}^2 = 2\log n - 2\log b_{(k)} - \log 2\pi$$

we construct a convergent sequence iteratively; after a few iterations we find that

$$\lim_{k\to\infty} b_{(k)}^2 = 2\log n - \log_2 n - \log 4\pi + o(1)\ .$$

In summary, if $F_0 = \mathcal{N}(0, 1)$,

$$\begin{aligned} F_0^{-1}(1 - n^{-1/2}\log n) &= (\log n - 3\log_2 n - \log 2\pi)^{1/2} + o(1)\ , && (108) \\ F_0^{-1}(n^{-1/2}\log n) &= -F_0^{-1}(1 - n^{-1/2}\log n) \\ F_0^{-1}(1 - 1/n) &= (2\log n - \log_2 n - \log 4\pi)^{1/2} + o(1) && (109) \\ F_0^{-1}(1/n) &= -F_0^{-1}(1 - 1/n)\ . && (110) \end{aligned}$$

Gamma shape When considering mixtures of the Gamma shape parameter we consider the generating distribution F_0 with density

$$f_0(x) = \exp\{\alpha_0 x - e^x - \log\Gamma(\alpha_0)\} \ .$$

From the previous subsections we have that as $x \to \infty$,

$$1 - F_0(x) \sim \frac{e^{(\alpha_0-1)x - e^x}}{\Gamma(\alpha)} \ .$$

So writing $x_n = F_0^{-1}(1-\delta_n)$, where $\delta_n = n^{-1/2}\log n$, taking logs and rearranging gives

$$e^{x_n} = \tfrac{1}{2}\log n - \log_2 n + (\alpha_0 - 1)x_n - \log\Gamma(\alpha_0) + o(1)$$

We use this to construct a convergent sequence $x_{(1)}, x_{(2)}, \ldots$ constructed iteratively via

$$e^{x_{(k+1)}} = \tfrac{1}{2}\log n - \log_2 n + (\alpha_0 - 1)x_{(k)} - \log\Gamma(\alpha_0) \ .$$

Start with $x_{(1)} = 1$. Then

$$\begin{aligned} e^{x_{(2)}} &= \tfrac{1}{2}\log n - \log_2 n + (\alpha_0 - 1) - \log\Gamma(\alpha_0) \\ &= \tfrac{1}{2}\log n[1 + o(1)] \ . \end{aligned}$$

So $x_{(2)} = \log\frac{1}{2} + \log_2 n + o(1)$ and

$$\begin{aligned} e^{x_{(3)}} &= \tfrac{1}{2}\log n - \log_2 n + (\alpha_0 - 1)\left[\log\tfrac{1}{2} + \log_2 n\right] - \log\Gamma(\alpha_0) + o(1) \\ &= \tfrac{1}{2}\log n + (\alpha_0 - 2)\log_2 n + (\alpha_0 - 1)\log\tfrac{1}{2} - \log\Gamma(\alpha_0) + o(1) \\ &= \tfrac{1}{2}\log n[1 + o(1)] \ . \end{aligned}$$

So $x_{(3)} = \log\frac{1}{2} + \log_2 n + o(1) = x_{(2)} + o(1)$, so we need iterate no further. So we conclude that x_n satisfies

$$e^{x_n} = \tfrac{1}{2}\log n + (\alpha_0 - 2)\log_2 n + (\alpha_0 - 1)\log\tfrac{1}{2} - \log\Gamma(\alpha_0) + o(1) \tag{111}$$

Things are easier for the lower tail. From earlier sections we have that Since

$$F_0(x) \sim e^{\alpha_0 x - e^x}\Gamma(\alpha_0 + 1) \ ,$$

as $x \to -\infty$ (see appendix C), defining $y_n = F_0^{-1}(\delta_n)$ we have that

$$y_n = \frac{1}{\alpha_0}\left[-\tfrac{1}{2}\log n + \log\log n + \log\Gamma(\alpha_0)\right] + o(1) \ . \tag{112}$$

Poisson We can also use (107) and the convergent iterative procedure outlined in appendix C.3 to approximate $x_n = F_0^{-1}(1-\delta_n)$, where again $\delta_n = n^{-1/2}\log n$. We have

$$\frac{e^{-\lambda_0}\lambda_0^{x_n}}{x_n!} \sim \delta_n$$

so

$$-\lambda_0 + x(\log\lambda_0 + 1 - \log x) - \frac{1}{2}\log 2\pi - \frac{1}{2}\log x = \log_2 n - \frac{1}{2}\log n + o(1)\ .$$

We can use this to define the iterative scheme via

$$x_{(k+1)} = \frac{\frac{1}{2}\log n - \log_2 n - \frac{1}{2}\log x_{(k)} - \frac{1}{2}\log 2\pi - \lambda_0}{\log x_{(k)} - \log\lambda_0 - 1}\ .$$

Starting with $x_{(1)} = \frac{1}{2}\log n$, and iterating three times gives that

$$\begin{aligned} x_n &= \frac{\frac{1}{2}\log n - \frac{3}{2}\log_2 n + \frac{1}{2}\log_3 n + \frac{1}{2}\log 4\pi - \lambda_0}{\log_2 n - \log_3 n - \log 2} + o(1) \\ &= O\left(\frac{\log n}{\log_2 n}\right)\ . \end{aligned}$$

Now suppose that $Y \sim \text{Pois}(\lambda)$. Using (107), we have, as $y \to \infty$,

$$P(Y \geq y) = \frac{e^{-\lambda}\lambda^y}{y!}[1 + o(1)]\ .$$

Define $y_n = \inf\{y \mid P(Y \geq y) \geq 1/n\}$. Then

$$P(Y \geq y_n) = \frac{1}{n}[1+o(1)] = \frac{e^{-\lambda}\lambda^{y_n}}{y_n!}[1+o(1)]\ .$$

Using $\log x! = x[\log x - 1] + \frac{1}{2}\log x + \frac{1}{2}\log 2\pi + o(1)$,

$$\begin{aligned} -\log n + o(1) &= -\lambda + y_n\log\lambda - \left\{y_n[\log y_n - 1] + \tfrac{1}{2}\log y_n + \tfrac{1}{2}\log 2\pi\right\} \\ y_n &= \frac{\log n - \frac{1}{2}\log y_n - \frac{1}{2}\log 2\pi - \lambda}{\log y_n - \log\lambda - 1} + o(1)\ . \end{aligned}$$

We can use this last relation to construct a convergent interative scheme to approximate y_n. Use a sensible starting value $y_{(1)}$, define $y_{(k+1)}$ in terms of $y_{(k)}$ via

$$y_{(k+1)} = \frac{\log n - \frac{1}{2}\log y_{(k)} - \frac{1}{2}\log 2\pi - \lambda}{\log y_{(k)} - \log\lambda - 1}$$

and continue until $y_{(k+1)} = y_{(k)} + o(1)$. Such $y_{(k+1)}$ agrees with y_n up to $o(1)$.

As in the previous example, we use a constant (if unusual), starting value of $y_{(1)} = \lambda e^2$ to see how the method performs. Then $\log y_{(1)} = 2 + \log \lambda$ and

$$\begin{aligned} y_{(2)} &= \frac{\log n - \frac{1}{2}\log y_{(1)} - \frac{1}{2}\log 2\pi - \lambda}{\log y_{(1)} - \log\lambda - 1} \\ &= \frac{\log n - 1 - \frac{1}{2}\log\lambda - \frac{1}{2}\log 2\pi - \lambda}{2 + \log\lambda - \log\lambda - 1} \\ &= \frac{\log n[1+o(1)]}{1} \quad \text{(hence the strange } y_{(1)}\text{)} \\ \log y_{(2)} &= \log_2 n + o(1)\ . \end{aligned}$$

We now iterate until the increment is $o(1)$ (note that $\log_3 n = \log\log\log n$):

$$\begin{aligned} y_{(3)} &= \frac{\log n - \frac{1}{2}\log y_{(2)} - \frac{1}{2}\log 2\pi - \lambda}{\log y_{(2)} - \log\lambda - 1} \\ &= \frac{\log n - \frac{1}{2}\log_2 n - \frac{1}{2}\log 2\pi - \lambda}{\log_2 n - \log\lambda - 1} + o(1) \\ &= \frac{\log n[1+o(1)]}{\log_2 n[1+o(1)]} \\ \log y_{(3)} &= \log_2 n - \log_3 n + o(1) \\ y_{(4)} &= \frac{\log n - \frac{1}{2}\log y_{(3)} - \frac{1}{2}\log 2\pi - \lambda}{\log y_{(3)} - \log\lambda - 1} \\ &= \frac{\log n - \frac{1}{2}\log_2 n + \frac{1}{2}\log_3 n - \frac{1}{2}\log 2\pi - \lambda}{\log_2 n - \log_3 n - \log\lambda - 1} + o(1) \\ &= \frac{\log n[1+o(1)]}{\log_2 n[1+o(1)]} \\ \log y_{(4)} &= \log_2 n - \log_3 n + o(1)\ . \end{aligned}$$

Since this is the same as $\log y_{(3)}$ up to $o(1)$, no more iterations are necessary, and we find that as $n \to \infty$,

$$y_n = \frac{\log n - \frac{1}{2}\log_2 n + \frac{1}{2}\log_3 n - \frac{1}{2}\log 2\pi - \lambda}{\log_2 n - \log_3 n - \log\lambda - 1} + o(1)\ .$$

So in summary, if $F_0 = \text{Pois}(\lambda)$,

$$F_0^{-1}(1 - 1/n) = \frac{\log n - \frac{1}{2}\log_2 n + \frac{1}{2}\log_3 n - \frac{1}{2}\log 2\pi - \lambda}{\log_2 n - \log_3 n - \log\lambda - 1} + o(1) \tag{113}$$

Negative Binomial Defining $x_n = F_0^{-1}(1-\delta_n)$, where $\delta_n = n^{-1/2}\log n$ gives us that

$$x_n \quad = \quad \frac{\frac{1}{2}\log n - \log_2 n - (m-1)\log x_n - C}{-\log q_0} \quad + o(1) \ .$$

We use this to construct a convergent iterative scheme. Starting with $x_{(1)} = \left(\frac{1}{2}\log n - \log_2 n\right)/(-\log q_0)$, after two interations we find that

$$x_n = \frac{\frac{1}{2}\log n - m\log_2 n - C + (m-1)[\log\frac{1}{2} + \log(-\log q_0)]}{-\log q_0} + o(1)$$

as $n \to \infty$. In particular $x_n = O(\log n)$. So then

$$\breve{y}(x_n)[1-F_0(x_n)]^{1/2} = O(\log n)^{-1/2} \ .$$

References

Abramowitz, M., & Stegun, I. A. 1964. *Handbook of Mathematical Functions.* Dover, New York.

Barndorff-Nielsen, O. 1978. *Information and Exponential Families in Statistical Theory.* Wiley series in probability and mathematical statistics. John Wiley and Sons.

Barndorff-Nielsen, O. 1980. Exponential Families. *Memoirs, Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus.*

Barndorff-Nielsen, O. E., & Cox, D. R. 1989. *Asymptotic Techniques for Use in Statistics.* Monographs on Statistics and Applied Probability. Chapman and Hall.

Bartle, R. G. 1976. *The Elements of Real Analysis.* 2nd edn. John Wiley & Sons., New York.

Berman, S. M. 1974. Sojourns and extremes of Gaussian processes. *Annals of Probability*, **2**, 999–1026.

BICKEL, P., & CHERNOFF, H. 1993. Asymptotic distribution of the likelihood ratio statistic in a prototypical non regular problem. *Pages 83–96 of:* GHOSH, J. K., MITRA, S. K., PARTHASARARTHY, K. R., & PRAKASA RAO, B. L. S. (eds), *Statistics and Probability: A Raghu Raj Bahadur Festschrift*. Wiley Eastern Limited.

CHERNOFF, H., & LANDER, E. 1995. Asymptotic distribution of the likelihood ratio test that a mixture of two binomials is a single binomial. *Journal of Statistical Planning and Inference*, 19–40.

CHOI, S., HALL, W.J., & SHICK, A. 1996. Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Annals of Statistics*, **24**(2), 841–861.

COLQUHOUN, D., & HAWKES, A. G. 1981. On the stochastic properties of ion channels. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **211**(Mar.), 205–235.

CSORGO, M., CSORGO, S., HORVATH, L., & MASON, D. M. 1986. Weighted empirical and quantile processes. *Annals of Probability*, **14**(1), 31–85.

DACUNHA-CASTELLE, D., & GASSIAT, É. 1997. Testing in Locally Conic Models, and Application to Mixture Models. *ESAIM: Probability and Statistics*, **1**(July), 285–317.

DAVIES, R. B. 1977. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, **64**(2), 247–254.

DAVIES, R. B. 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, **74**(1), 33–43.

GHOSH, J., & SEN, P. 1985. On the asymptotic performance of the log-likelihood ratio statistic for the mixture model and related results. *In: Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*.

HALL, P. 1991. On convergence rates of suprema. *Probability Theory and Related Fields*, 447–455.

HALL, W. J., & MATHIASON, D. 1990. On large-sample estimation and testing in parametric models. *International Statistical Review*, **58**, 77–97.

HARTIGAN, J. A. 1985. A failure of likelihood ratio asymptotics for normal mixtures. *In: Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer.*

HÜSLER, J. 1983. Asymptotic approximation of crossing probabilities of random sequences. *Z. Wahrsch. Verw. Gebiete*, **63**, 257–270.

HÜSLER, J. 1990. Extreme values and high boundary crossings of locally stationary gaussian processes. *Annals of Probability*, **18**(3), 1141–1158.

HÜSLER, J. 1995. A note on extreme values of locally stationary Gaussian processes. *Journal of Statistical Planning and Inference*, 203–213.

JENSEN, J. L. 1995. *Saddlepoint Approximations.* Oxford University Press.

KENDALL, SIR M., STUART, A., & ORD, J.K. 1987. *Kendall's Advanced Theory of Statistics.* Charles Griffin & company limited.

KOMLÓS, J., MAJOR, P., & TUSNÁDY, G. 1975. An Approximation of Partial Sums of Independent RV'-s, and the Sample DF.I. *Z.Wahrsch. verw. Geb.*, 111–131.

LEADBETTER, M.R., LINDGREN, G., & ROOTZÉN, H. 1983. *Extremes and Related Properties of Random Sequences and Processes.* Springer-Verlag.

LEHMANN, E. L. 1986. *Testing Statistical Hypotheses.* John Wiley Sons, Inc.

LINDSAY, B. G. 1995. *Mixture Models: Theory, Geometry and Applications.* NSF-CBMS Regional Conference Series in Probability and Statistics, vol. 5. Institute of Mathematical Statistics and American Statistical Association.

LIU, X., & SHAO, Y. 2001. *Asymptotic Distribution of the Likelihood Ratio Test in a two-component normal mixture model.* Unpublished at time of writing. Contact author for preprint. See also `http://bruce.unibw-hamburg.de/mix01/abstracts/Liu,_Xin/Liu,_Xin.html`.

PITERBARG, V. I. 1996. *Asymptotic Methods in the Theory of Gaussian Processes and Fields.* American Mathematical Society.

REDNER, R. 1981. Note on the consistency of the maximum likelihood estimate for non-identfiable distributions. *Annals of Statistics*, **9**, 225–228.