ASYMPTOTIC PROPERTIES OF POSTERIOR DISTRIBUTIONS AND STUDY OF SOME NONREGULAR CASES

SUBHASHIS GHOSAL

Thesis submitted to the Indian Statistical Institute

in partial fulfilment of the requirements

for the award of the degree of

Doctor of Philosophy

CALCUTTA

1994

Contents

Acknowledgements

The present work has been done under the guidance of Professor J. K. Ghosh. It is extremely difficult for me to express my sense of gratitude towards him in words. He has played a very significant role in my academic work during the last few years. I learnt a lot from his way of thinking and approach to a problem. In spite of his heavy burden of work, I had the privilege of occupying much of his precious time, even in his busiest moments. The great pleasure of working under him will always be remembered. Also, I would like to thank him for allowing me to include our joint works in my thesis.

Dr. T. Samanta was a co-worker throughout the present work. I must thank him for allowing me to include our joint works in this thesis. But I owe him much more for sharing the ups and downs as a friend and co-researcher.

At various stages, I received encouragement and co-operation from most of the faculty members in the Stat-Math Unit. I have benefitted substantially from the discussions with Professor B. V. Rao. Professor T. K. Chandra always showed interest in my work. It is a pleasure to record my thanks to all of them.

I have been fortunate to have a number of good friends whose constant encouragement inspired me to continue my work. My friend Mr. M. Mitra provided many useful suggestions regarding the presentation of the thesis. Also, I am indebted to my fellow research scholars for sharing our common problems.

I had the opportunity to work in the pleasant atmosphere of the Indian Statistical Institute. My work has been funded by the National Board of Higher Mathematics, Department of Atomic Energy of the Government of India. I record my sincere thanks to these organizations.

I would be failing in my duties if I do not mention the role of non-scientific workers of the institute. It would have been a difficult job to continue the work without the co-operation from the staff of the Stat-Math Unit, library, reprography unit, Dean's office and other parts of the institute.

I have benefitted much from the advantages of the TEX-group software typing systems. I thank the Stat-Math Unit once more for making these available.

Finally, the role of my family in my academic career can never be denied. Even though I spent a considerable amount of family-time in my work, my parents and brothers were always patient with me. Without their encouragement, this work would never have been materialized.

Calcutta, April 1994 Subhashis Ghosal

Basic Notations

$\mathbb{R}$	real line
$\mathbb{R}^d$	d-dimensional Euclidean space
$\mathcal{B}$	Borel σ-field on $\mathbb{R}$
$\mathcal{B}^d$	Borel σ-field on $\mathbb{R}^d$
χ	indicator function
A^c	complement of a set A
$\|\cdot\|$	Euclidean norm
$'$	transpose of a vector, derivative of a function
$\xrightarrow{p}$	convergence in probability
$\xrightarrow{d}$	convergence in distribution
$\stackrel{d}{=}$	equality in distribution
$a_n = o(b_n)$	$a_n/b_n \to 0$
$a_n = O(b_n)$	a_n/b_n is bounded
$a_n \sim b_n$	$a_n/b_n \to 1$
$X_n = o_p(a_n)$	$X_n/a_n \xrightarrow{p} 0$
$X_n = O_p(a_n)$	X_n/a_n is stochastically bounded
$\mathcal{L}(X)$	distribution of X
$X \sim F$	X has distribution F
$\square$	end of a proof
$:=$	equality in definition

Result $x.y.z$ (which could be a theorem, lemma, etc.) means Result z of Section y of Chapter x. Result $y.z$ means Result z of Section y of the same chapter. Result A.x means Result x of Appendix A.

Introduction

The asymptotic approach to statistical estimation is frequently adopted because of its general applicability and relative simplicity. The modern study of asymptotic theory, initiated in Le Cam (1953), has undergone a vigorous development through the classic works of Le Cam, Hájek, Bahadur, Ibragimov and Has'minskii (Khas'minskii), Bickel, Pfanzagl, Millar and many other scholars; see Le Cam (1986), Le Cam and Yang (1990), Ibragimov and Has'minskii (1981) and the review article by Ghosh (1985) for an account of this development.

Most of the results in asymptotic theory of estimation are obtained under the classical Cramér-Rao type regularity conditions or their substantial generalizations like LAN, LAMN etc. While undoubtably these are the most important cases, they are by no means the only cases of interest. It is well known that quite different and interesting phenomena occur when regularity conditions are violated. For example, in the case of a family of discontinuous densities, the best rate of convergence is n^{-1} instead of $n^{-1/2}$ in the regular cases. Also in contrast to the regular cases, the MLE is not efficient (but Bayes estimates are) and actually the notion of efficiency depends on the loss function involved. These "nonregular" cases attracted the attention of researchers from an early period and often were used to produce counterexamples. Convergence rates of estimates (particularly the MLE) were considered in several papaers, see Polfeldt (1970) and Woodroofe (1972, 1974) for instance. In Polfeldt (1970a,b), the questions related to the order of variance of minimum variance unbiased estimators were investigated. A general theory of nonregular (as well as regular) cases was first attempted in Weiss and Wolfowitz (1974) who showed that the MLE may not be efficient, but maximum probabilty estimators (which are basically Bayes estimators depending on the chosen loss function) are always efficient. In fact, in general, the posterior distribution and Bayes procedures may behave well even if the MLE in not well behaved, see Schwartz (1965) in this context. A modern treatment of both regular and nonregular cases appeared in the works of Ibragimov and Has'minskii (1970, 1971, 1972a, b, 1973a, b, 1974, 1975a, b, 1976, 1977) which are now collectively available in Ibragimov and Has'minskii (1981). Since this will be the main reference throughout the present work, it is convenient to abbreviate the above as IH. The general formulation of IH views the (normalized) likelihood ratios as a stochastic process in the following way:

Fix a point θ in the parameter space and define the "likelihood ratio process" (LRP) parametrized by u by

$$Z_n(u) = p^n(x^n; \theta + \varphi_n u)/p^n(x^n; \theta),$$

where $p^n(\cdot)$ stands for the joint density and φ_n for appropriate normalizing constants. It is assumed in IH that the LRP satisfies the following three conditions (these are stated more formally in Section 1 of Chapter 1):

(1) The map $u \mapsto Z_n^{1/2}(u)$ into $L^2(P_\theta)$ is Lipschitz continuous and the growth of the Lipschitz constant is at most like a polynomial.

(2) For some sequence $g_n(\cdot)$ of nonnegative functions which increase to infinity for every $n \geq 1$ and satisfy $\lim_{\substack{y\to\infty \\ n\to\infty}} y^N \exp[-g_n(y)] = 0$ for every N, $EZ_n^{1/2}(u) \leq \exp[-g_n(\|u\|)]$.

(3) The finite dimensional distributions of $Z_n(u)$ converge to those of a stochastic process $Z(u)$.

Under this general set up, the asymptotic properties of the MLE and Bayes estimates are derived in IH. This formulation is suitable for handling both the regular and nonregular cases. The likelihood ratio process was used earlier by Rubin (1961) and Prakasa Rao (1968), and also occured in Hoeffding's lecture notes at Chapel Hill. To be more specific, in Prakasa Rao (1968), the problem of finding the asymptotic distribution of the MLE was reduced to the problem of finding the distribution of maxima of a Gaussian process in the case of densities with singularities of the second type (see Section 1.2). *In a part of the present work, we show that these three properties of the LRP provide useful informations on the asymptotic properties of the posterior distributions as well.*

The present work can be divided into two major portions. In Chapter 1, the asymptotic properties of the posterior distributions (and some related results) are treated in the general set up of IH. The conditions of IH stated above are more formally described in Section 1 of Chapter 1. It can be noted that the version of the conditions are somewhat *weaker than those that actually appeared in* IH; *this relaxation is crucially used in Chapters* 2 *and* 3. Several examples of different kinds fall in the set up of IH; they are briefly reviewed in Section 1.2 for convenience in later developments. For ready references, the results of IH on Bayes estimates and the MLE are presented in Appendix B. The main results of Chapter 1 consist of the results on the asymptotic properties of posterior distributions, which are discussed in the following paragraphs. Some new results on Bayes estimates and the MLE are also presented. In particular, it is shown that *under a very general "regular" situation, the MLE and Bayes estimates are asymptotically equivalent* (Corollary 1.5.1). While this result is quite well known in many particular cases or under conditions much stronger than ours, it seems to be *new in this general form.* In the last section of this chapter (Section 1.6), relations with convergence of statistical experiments, efficiency, regularity property of Bayes estimates and MLE are discussed. A notion of asymptotic independence of estimation problems of components of the parameter is also introduced here. This means that the limit of posterior distributions, the limiting experiment associated with the decision problem and the asymptotic distribution of Bayes estimates and the MLE are all of product type. Such a phenomenon occurs whenever the limiting LRP can be factorized into two independent processes parametrized by different components. It will be seen later that the estimation problems for the "regular" and "nonregular" components for the cases in Chapters 2 and 3 are asymptotically independent.

In Bayesian analysis, one starts with a prior and the resultant analysis is based on the posterior, given data. Since this involves a prior, naturally we are interested to know to what extent the Bayesian analysis is sensitive to the choice of a prior. In Theorem 1.3.1, we show that *posterior consistency implies that the total variation distance between the posterior distributions corresponding to any two "reasonable" priors converges to zero* whereas Proposition 3.1 shows that *posterior consistency always holds under the conditions of IH.* Thus under mild conditions, the prior has little or no effect if the sample size is large, so almost the same conclusion will follow from any reasonable prior, i.e., we have almost complete prior robustness for large sample sizes.

The next natural question is whether the posterior converges to a limit as the sample size increases indefinitely. In this case, the inference stabilizes and the Bayesian analysis can be remarkably simplified. The approximate computations based on the much simpler limiting form is often quite accurate, even for moderately large sample sizes. (For an example, see Berger (1985, p. 225).)

We observe that *a weak limit of posterior probability that the normalized parameter u belongs to a given Borel set always exists* under the general set up of IH (Theorem 1.3.2). Indeed, it is shown in Theorem 1.3.3 that the *posterior densities of the normalized parameter u, as random densities, converge weakly* and the limit is also identified. Although theoretically interesting (these results are used as important theoretical tools in later developments), the weak limit itself is not quite useful, and so we want to improve the mode of convergence by making a data-dependent transformation of the parameter. In the "regular" cases, it is well known that for a wide variety of priors, the posterior normalized and centered at the MLE, converges to a normal distribution. This fact is referred to as the *Bernstein-von Mises theorem* or *Bayesian central limit theorem.* A rigorous proof of this fact in case of i.i.d. observations first appeared in Le Cam (1953). Various modifications and extensions of this result have been made by several authors; see Section 3 of Chapter 1 for a more detailed discussion.

So far, the asymptotic behaviour of posterior distributions has not been studied in nonregular cases except in Samanta (1988), where a non-normal limit of a posterior has been obtained for a particular type of discontinuous densities. (A refined version of this result including asymptotic expansions will appear in a subsequent work.) This suggests the possibility of finding a posterior limit (not necessarily a normal distribution) in a stronger sense (either almost surely or in probability) by a suitable centering (not necessarily at the MLE). Under the general set up of IH, we try to find out the necessary and sufficient condition for this phenomenon to occur. The reason for considering the possibilty of a constant limit can be justifed in the i.i.d. case; we show in Proposition 1.3.2 that *a limit of normalized and centered posterior probabilities, if it exists, must be constant almost surely.* The main theorem in this context (Theorem 1.3.4) can be stated as follows:

If the limit of suitably centered and normalized posterior distributions exists, then the limiting LRP $Z(u)$ must satisfy the relation

$$Z(u)/\int Z(u')du' = g(u+W)$$

for some random variable W and a fixed probability density $g(\cdot)$.

Using this criterion, we investigate in Section 1.4 the possibility of the existence of a limiting posterior in many cases that appeared in the literature. We conclude that in most of the nonregular cases, there cannot be a posterior convergence. Also it is of importance to see that the necessary condition described in Theorem 1.3.4 is also sufficient to imply a posterior convergence. In fact, we show in Theorem 1.3.6 that *if $Z(u)$ is of the stated form, then, centering with a Bayes estimate, the posterior distribution converges in probability to a constant limit.* This result appears to be a very *general posterior limit theorem.* For example, it *implies a version of the Bernstein-von Mises theorem* (here we interpret the convergence in probability sense) in *a very general regular situation.* Under a little more restriction, the Bayes estimate can be replaced by the MLE in the above result; see the discussion following Corollary 1.5.1 in this context.

In Chapter 2, we treat a very important multiparameter nonregular case. We consider i.i.d. observations with a common density on $\mathbb{R}$ such that the densities are discontinuous in one component of the parameter (say, θ) but the problem is "smooth" with respect to the other components (say, φ). We state the set up and assumptions formally in Section 2.2 where several possible examples from the literature are also described. Such examples were treated earlier in Smith (1985) and Cheng and Iles (1987) who considered the problem of finding the asymptotic distribution of the MLE or its alternatives using the methods of extreme value theory. In contrast, we study the whole estimation problem using methods similar to that of IH (Ch. V) who treated the case of one-parameter discontinuous densities. More specifically, in Section 2.3, we show that the likelihood ratios satisfy certain properties similar to the conditions of IH described in Chapter 1 and consequently derive asymptotic properties of Bayes estimates and posterior distributions (Section 2.4). A version of the first two conditions of IH are satisfied because the squared Hellinger distance $r_2^2((\theta,\varphi),(\theta+u,\varphi+v))$ between the densities corresponding to two paramter points (θ,φ) and $(\theta+u,\varphi+v)$ is shown to be bounded (see Section I.5 of IH):

$$c(|u|+\|v\|^2) \leq r_2^2((\theta,\varphi),(\theta+u,\varphi+v)) \leq C(|u|+\|v\|^2)$$

for some constants $C > c > 0$, and the dependence of C on (θ,φ) satisfies certain growth conditions. The key idea of proving these inequalities is to show that the densities satisfy a property very similar to quadratic mean differentiablity if a suitable neighbourhood of the points of discontinuities is excluded (see Lemma 2.3.1). To show the convergence of finite dimensional distributions of the LRP $Z_n(u,v)$, we first derive an approximation $\widetilde{Z}_n(u,v)$ of $Z_n(u,v)$ which is much simpler to deal with than the original LRP (see Theorem 2.3.1). Here we use techniques very similar to those in Chapter V of IH, but we need a judicious combination of the techniques of regular and nonregular cases. (For example, Lemma 2.3.1 has been crucially used.) Now using a simple weak convergence result (Lemma 2.3.5), we derive in Theorem 2.3.2 the following result:

The finite dimensional distributions of $Z_n(u,v)$ converge to those of the process $Z(u,v) = Z^{(1)}(u)Z^{(2)}(v)$ where $Z^{(1)}(u)$ is an explicitly stated function of at most $2r$ independent homogeneous Poisson processes, $Z^{(2)}(v) = \exp[v'\Delta - (1/2)v'Iv]$ where Δ is a $N(0,I)$ random vector, I is a positive definite matrix *and the processes $Z^{(1)}(u)$ and $Z^{(2)}(v)$ are independent.* (Here, r is the number of points of discontinuity of the density as a function of θ.)

Using the above facts, the asymptotic distribution of the Bayes estimates is (almost) immediately obtained from the results of IH in terms of the limiting LRP. The results of Chapter 1 on posterior distributions can now also be utilized to show posterior consistency, asymptotic prior robustness and to investigate the possible existence of a posterior limit. When the likelihood ratios have a particular expansion, which is, roughly speaking, locally asymptotically normal in v and locally asymptotically exponential (see Section V.5 of IH) in u, we derive a convolution theorem in Theorem 2.5.1 along the lines of Millar (1983). This result can be stated as follows:

If the likelihood ratios satisfy

$$Z_n(u,v) = \exp[uc + v'\Delta_n - (1/2)v'Iv]\chi\{u < \sigma_n\} + o_p(1)$$

where $c > 0$, I is positive definite, $\mathcal{L}((\Delta_n, \sigma_n)) \Rightarrow \mathcal{L}((\Delta, \sigma))$, $\Delta \sim N(0, I)$, σ has an exponential distribution with mean $1/c$ and Δ and σ are independent, then the limiting distribution of any regular estimator of (θ, φ) can be written as a convolution of the product of an exponential distribution with mean $1/c$ and an $N(0, I)$ with some probability measure.

As mentioned in the context of Theorem 2.3.2, $Z(u,v)$ can be factored into two independent processes $Z^{(1)}(u)$ and $Z^{(2)}(v)$. It is also worthnoting that $Z^{(1)}(u)$ would have been the limiting LRP had φ been known and $Z^{(2)}(v)$ would have been the limiting LRP had θ been known. The fact that $Z(u,v)$ is factored into two independent processes parametrized by u and v respectively implies the asymptotic independence of the estimation problems of θ and φ in the sense of Section 1.6.

In Chapter 3, we study another important multiparameter nonregular case where the problem is smooth for φ but has singularity (instead of discontinuity as in Chapter 2) in θ. The idea of the solution of this problem is very similar to that of Chapter 2, but there are certain differences as well. The case is more complex than that of Chapter 2 and we derive results for the particular case when θ is a location parameter and there is only one point of singularity; a more general case can be treated following similar lines but does not lead to substantially new phenomena. Here also, we establish certain properties of the LRP similar to those in the conditions of IH. The method of proof, however, is somewhat different from that of Chapter 2. We first split $\log Z_n(u,v)$ into two terms as follows:

$$\log Z_n(u,v) = \log \frac{p^n(x^n;\theta+k_n u,\varphi)}{p^n(x^n;\theta,\varphi)} + \log \frac{p^n(x^n;\theta+k_n u,\varphi+n^{-1/2}v)}{p^n(x^n;\theta+k_n u,\varphi)}$$

where $k_n = n^{-1/(1+\alpha)}$ and α is the order of the singularity. An approximation of the first term is immediately obtained from the results of Chapter VI of IH on the corresponding one-parameter case. We then show that the second term has an appropriate expansion under certain assumptions on the pointwise behaviour of the density function and its derivatives. (It can be noted that the approach in Chapter 2 needs less assumptions, but here it does not apply because an analogoue of Lemma 2.3.1 fails to be true.) This implies (Theorem 3.2.1) that $\log Z_n(u,v)$ is approximately the sum of two processes, one parametrized by u and the other by v. These two terms are asymptotically independent and thus we conclude in Theorem 3.2.2 that the following result holds:

The limiting LRP $Z(u,v)$ is given by

$$\log Z(u,v) = Y(u) + v'\Delta - (1/2)v'Iv,$$

where $Y(u)$ is an explicit function two independent nonhomogeneous Poisson processes (or a Gaussian process if the singularity is of the second type) and Δ is an $N(0,I)$ random vector independent of $Y(u)$.

It can be noted that $Y(u)$ would have been the limiting log-LRP if φ were known whereas $v'\Delta-(1/2)v'Iv$ would have been the limiting log-LRP if θ were known. The asymptotic properties of Bayes estimates and posterior distributions are obtained in terms of the limiting LRP $Z(u,v)$ in view of the results of IH and Chapter 1. As in Chapter 2, the estimation problems for θ and φ are asymptotically independent.

Results of Chapter 4 are of somewhat different spirit and are not derived as a consequence of the results on Chatper 1, although occasionally the facts in Chapter 1 and Chapter V of IH are used. Here we derive the asymptotic expansion of the expected Kullback-Leibler distance between the posterior and the prior for a class of one-parameter discontinuous densities. We consider a class of one-parameter discontinuous densities which are supported on an interval (finite or infinite) which increases or decreases with θ. The main result of this section (Theorem 4.2.1) can be stated as follows:

Under certain assumptions, for any prior $\pi(\theta)$ positive, continuous and supported on a compact set $K \subset \mathbb{R}$, we have as $n \to \infty$,

$$I(\pi;X^n) = \log(n/e) + \int_K \pi(\theta)\log(|c(\theta)|/\pi(\theta))d\theta + o(1),$$

where $I(\pi; X^n)$ *is the expected Kullback-Leibler distance between the posterior and the prior and* $c(\theta) = E_\theta((\partial/\partial\theta)\log f(X_1;\theta))$.

This expansion is then used to define and find out the reference prior in the sense of Bernardo (1979). We thus reach the following conclusion:

The reference prior for the problem is given by $\pi(\theta) \propto |c(\theta)|$.

Importance of the study of reference priors in nonregular cases was mentioned in Bernardo (1979), but the present results seem to be the first rigorous treatment of reference priors in nonregular cases. It will also be of interest to get similar results for the multiparameter case treated in Chapter 2. This is currently under investigation.

For many reasons, it is also important to study the higher order asymptotic properties which we have not attempted here. For an account of higher order asymptotic properties under regularity conditions, one can see the recent monograph by Ghosh (1993). It is of importance to study the higher order asymptotic properties of posterior distributions, Bayes estimates and the MLE under a suitable modification of the conditions of IH.

Chapter 1

General Theory

1 The Set Up

We shall investigate asymptotic properties in finite dimensional parametric problems. The underlying set up is same as that considered by IH which we shall describe first. Let $\{\mathfrak{X}^n, \mathcal{A}^n, P_\theta^n : \theta \in \Theta\}$ be a sequence of statistical experiments generated by observations $X^n \in \mathfrak{X}^n$, where $\Theta \subset \mathbb{R}^d$ is a nonempty open set. (Actually IH considered a more general real indexing variable than the integer variable n considered here. The difference however is very nominal, and all the claims made here will be valid for real indexing variable also whenever it is meaningful.) For example, in the very important case of i.i.d. observations $X_1, X_2, \ldots$, the sample size plays the role of indexing variable and $X^n = (X_1, \ldots, X_n)$. We shall assume that for each $n \geq 1$, the family of probability measures $\{P_\theta^n : \theta \in \Theta\}$ is dominated, i.e., there exists a σ-finite measure ν^n on $(\mathfrak{X}^n, \mathcal{A}^n)$ such that for each $\theta \in \Theta$, P_θ^n is absolutely continuous with respect to ν^n. We denote the Radon-Nikodym derivative by $p^n(x^n; \theta)$, i.e.,

$$p^n(x^n; \theta) = \frac{dP_\theta^n}{d\nu^n}(x^n); \quad x^n \in \mathfrak{X}^n,\ n \geq 1.$$

We fix a point $\theta_0 \in \Theta$ which we regard as the "true" parameter; all the asymptotic properties will be established under the model $\theta = \theta_0$. Let $\{\varphi_n\}$ be a sequence of $d \times d$ positive definite matrices which converges to zero as $n \to \infty$ (entrywise). The sequence $\{\varphi_n\}$ is regarded as the sequence of normalizing constants and the asymptotic properties are understood after scaling by this factor. The choice of this sequence depends on the particular problem concerned. For example, in the familiar i.i.d. case of "smooth" densities, it is well known that $\varphi_n = n^{-1/2} I_d$ is the appropriate normalizing factor, some further examples will soon appear. The (local) *likelihood ratio process* (LRP) $Z_{n,\theta_0}(u) = Z_n(u)$ is defined by

$$Z_n(u) = \frac{p^n(x^n; \theta + \varphi_n u)}{p^n(x^n; \theta_0)}, \quad u \in U_n,$$

where $U_n = U_{n,\theta_0}$ stands for the set $\varphi_n^{-1}(\Theta - \theta_0)$. Since φ_n converges to zero, $\{U_n\}$ finally fills up the whole of $\mathbb{R}^d$.

The general theory of IH has been developed under certain assumptions on the LRP $\{Z_n(u) : u \in U_n\}$ described below. In what follows, all the probability statements refer to the true parameter θ_0 unless otherwise explicitly mentioned. Also we use the following notations:

By Pol(x), we mean any function of the form $B(1+|x|^b)$ where $B, b > 0$. Any function of the form $B\exp[b|x|]$ will be denoted by Exp(x). If $f(x)$ is a function, by $f(x) \le \text{Pol}(x)$, we mean that $f(x) \le B(1+|x|^b)$ for some $B, b > 0$. When we write a limiting relation involving Pol(x) (for example, $\lim_{x\to\infty} e^{-x}\text{Pol}(x) = 0$), we mean that the relation is satisfied by every function of the form $B(1+|x|^b)$ with $B, b > 0$. A similar convention is adopted for Exp(x) also.

Conditions.

(IH 1) For some $\alpha_1, \ldots, \alpha_d > 0$, with $u = (u_1, \ldots, u_d)$, $v = (v_1, \ldots, v_d)$,

$$E|Z_n^{1/2}(u) - Z_n^{1/2}(v)|^2 \le \prod_{i=1}^{d} \text{Pol}(R_i) \sum_{i=1}^{d} |u_i - v_i|^{\alpha_i}$$

for all $u, v \in U_n$ with $|u_i| \le R_i$, $|v_i| \le R_i$, $i = 1, \ldots, d$.

(IH 2) For every $u = (u_1, \ldots, u_d) \in U_n$,

$$EZ_n^{1/2}(u) \le \exp[-g_n(|u_1|, \ldots, |u_d|)],$$

where $\{g_n(\cdot)\}$ is a sequence of nonnegative real valued functions on $[0,\infty)^d$ satisfying the following conditions:

(a) For any $n \ge 1$, $g_n(\cdot)$ is increasing to infinity in each of its arguments.

(b) With $y = (y_1, \ldots, y_d)$,

$$\lim_{\substack{\|y\|\to\infty \\ n\to\infty}} \prod_{i=1}^{d} \text{Pol}(y_i) \exp[-g_n(y_1, \ldots, y_d)] = 0.$$

(IH 3) The finite dimensional distributions of the stochastic process $\{Z_n(u) : u \in U_n\}$ converge to those of a stochastic process $\{Z(u) : u \in \mathbb{R}^d\}$.

Conditions (IH 1), (IH 2) and (IH 3) together will be referred to as Conditions (IH). However, we must say that there is a little difference between the conditions assumed by IH and those stated above. Firstly, IH assumed that the conditions are satisfied uniformly over compact subsets of Θ, and as a consequence, their conclusions are valid uniformly over compact subsets also. *Secondly, we have relaxed the first two conditions to allow different dependence on different components in contrast to* IH. *This generalization, although not a significant improvement, gives us additional flexibility so that the proofs can be easily* *adapted to some similar situations. For example, one can replce in* (IH 1) *the factor* $\mathrm{Pol}(R_i)$ *by* $\mathrm{Exp}(R_i)$, *provided* $\mathrm{Pol}(y_i)$ *in* (IH 2) *is also replaced by* $\mathrm{Exp}(y_i)$, $i = 1, \ldots, d$. *It is easily verified that all the asymptotic results then go through in an exactly same manner. This observation, although simple, is crucially used in Chapters 2 and 3.*

A lot of results regarding the asymptotic behaviour of posterior distribution and Bayes estimates can be derived from the above conditions. Also the general theory based on these conditions has a wide applicabilty because, under a variety of situations, Conditions (IH) are satisfied. (See Section 2 for a brief review.) In the i.i.d. case, easy sufficient conditions are available for Conditions (IH 1) and (IH 2) in terms of the Hellinger distance between the true and neighbouring densities. For two probabilities P and Q on some measurable space, recall that the Hellinger distance $r_2(P, Q)$ between P and Q is defined by

$$\begin{aligned} r_2^2(P, Q) &= \int (f^{1/2}(x) - g^{1/2}(x))^2 \nu(dx) \\ &= 2(1 - \int f^{1/2}(x) g^{1/2}(x) \nu(dx)), \end{aligned}$$

where $f = dP/d\nu$ and $g = dQ/d\nu$ are the Radon-Nikodym derivatives of P and Q respectively with respect to a common dominating σ-finite measure ν. Abbreviating P_θ as θ, $r_2(\theta, \theta + h)$ denotes the Hellinger distance between P_θ and $P_{\theta+h}$. Assume that for all $\theta \in \Theta$ and for all h with $\theta + h \in \Theta$, we have

(i) $r_2^2(\theta, \theta + h) \leq A(\theta) \|h\|^\alpha$,

where $A(\theta)$ has a polynomial growth,

(ii) $r_2^2(\theta, \theta + h) \geq a(\theta) \|h\|^\alpha / (1 + \|h\|^\alpha)$,

where $a(\theta) > 0$ (If uniformity is required, we assume that the left hand side is positive uniformly in compact sets),

(iii) For some $\gamma > 0$,

$$\int f^{1/2}(x; \theta) f^{1/2}(x; \theta + h) \nu(dx) \leq C(\theta) \|h\|^{-\gamma} \tag{1.1}$$

for some constant $C(\theta)$. (This condition is vacuously satified if Θ is bounded. If uniformity is reqired, we need boundedness of $C(\theta)$ on compact subsets of Θ.)

Then Conditions (IH 1) and (IH 2) are satisfied with $\varphi_n = n^{-1/\alpha} I_d$. (For a proof and more details, see Sec. I.5 of IH, pp. 51-57.)

In case of location families $f(x; \theta) = f(x - \theta)$ (densities are with respect to Lebesgue measure), Condition (iii) above holds if the mild moment condition

$$\int \|x\|^\delta f(x) dx < \infty \quad \text{for some } \delta > 0 \tag{1.2}$$

is satisfied (see, e.g., p. 189 of IH). However, outside the location and scale families, Condition (iii) may not be always a mild requirement, see Example 3 in Chapter 2 in this context.

2 Examples

In this section, we present some examples for which Conditions (IH) are satisfied; further details are available from IH.

EXAMPLE 1. INDEPENDENT HOMOGENEOUS OBSERVATIONS WITH A SMOOTH DENSITY. Let $X_1, X_2, \ldots$ be i.i.d. observations with a distribution P_θ, $\theta \in \Theta$. Assume that each P_θ has a density $f(\cdot;\theta)$ with respect to some σ-finite measure ν. Suppose $\{f(\cdot;\theta) : \theta \in \Theta\}$ is quadratic mean differentiable (QMD) with Fisher's information $I(\theta)$ which is positive definite, continuous and bounded away from zero and infinity in the sense of (III.3.1) of p. 185 of IH. Suppose further that the identifiability type condition (III.3.2) of p. 185 of IH and (1.1) are satisfied. Then the sequence of experiments generated by $\{X_1, \ldots, X_n\}$ satisfies Conditions (IH) with $\varphi_n = n^{-1/2} I_d$. Also here the local asymptotic normality (LAN) condition is satisfied, i.e.,

$$Z(u) = \exp[u'\Delta - (1/2)u'I(\theta_0)u], \tag{2.1}$$

where $\Delta \sim N_d(0, I(\theta_0))$. Particularly, for location families, the conditions reduces to a much simpler form. For sufficient conditions of QMD, see, e.g., Theorem II.2.1 of IH.

However, one may have to reparametrize θ to satisfy Conditions (IH). For example, consider the estimation of the scale parameter in a regular family

$$f(x;\theta) = \theta^{-1} f(x/\theta), \quad 0 < \theta < \infty,$$

on the basis of i.i.d. observations $X_1, X_2, \ldots, X_n$ where f is sufficiently smooth. Clearly Condition (3.1) in Theorem III.3.1 of IH does not hold since the Fisher information, being proportional to θ^{-1}, is unbounded. Indeed, for such examples, Condition (IH 1) also fails. To see this, take the simple example of normal distribution with mean zero and unknown variance θ^2. For $\theta_0 = 1$ and $-n^{1/2} < u < \infty$

$$Z_n(u) = (1 + n^{-1/2}u)^{-n} \exp[-(1/2)\{(1 + n^{-1/2}u)^{-2} - 1\} \sum_{i=1}^{n} X_i^2]. \tag{2.2}$$

So we have

$$E|Z_n^{1/2}(u_1) - Z_n^{1/2}(u_2)|^2 = 2\{1 - [\frac{2(1+n^{-1/2}u_1)(1+n^{-1/2}u_2)}{(1+n^{-1/2}u_1)^2 + (1+n^{-1/2}u_2)^2}]^{n/2}\}$$

whenever $-n^{1/2} < u_1, u_2 < \infty$. Using the fact that $1 - r^n \geq 1 - r$ if $0 < r < 1$, we have

$$\begin{aligned} E|Z_n^{1/2}(u_1) - Z_n^{1/2}(u_2)|^2 &\geq 2\{1 - [\frac{2(1+n^{-1/2}u_1)(1+n^{-1/2}u_2)}{(1+n^{-1/2}u_1)^2 + (1+n^{-1/2}u_2)^2}]\} \\ &= \frac{(2/n)(u_1-u_2)^2}{(1+n^{-1/2}u_1)^2 + (1+n^{-1/2}u_2)^2}. \end{aligned}$$

Now taking $u_1 = -n^{1/2}(1-\varepsilon), u_2 = -n^{1/2}(1-2\varepsilon), \varepsilon > 0$, we have for any $\alpha > 0$,

$$|u_1 - u_2|^{-\alpha} E|Z_n^{1/2}(u_1) - Z_n^{1/2}(u_2)|^2 \geq (2/(5n^{\alpha/2}))\varepsilon^{-\alpha}. \tag{2.3}$$

The right hand side of (2.3) can be made arbitrarily large by choosing $\varepsilon > 0$ sufficiently small. Consequently Condition (IH 1) cannot be satisfied. Similar calculations show that (IH 1) fails also for the exponential distribution with scale parameter θ.

This difficulty, however, is not so serious since a reparametrization $\sigma = \log\theta$ overcomes this problem. More generally, let $f(\cdot;\theta)$ be a QMD family of densities with a finite, continuous and positive definite Fisher's information $I(\theta)$, consider a reparametrization σ of θ which satisfies the differential equation

$$d\sigma = I^{1/2}(\theta)d\theta, \tag{2.4}$$

where $I^{1/2}(\theta)$ is the positive definite square root of $I(\theta)$. By results of Section III.3 of IH, Conditions (IH) are satisfied with σ as the parameter.

EXAMPLE 2. INDEPENDENT NONHOMOGENEOUS OBSERVATIONS (SMOOTH DENSITIES). Let $X_1, X_2, \ldots$ be independent observations with X_j having a density $f_j(\cdot;\theta)$ (with respect to some σ-finite measure ν_j). Assume that $\forall j$, $\{f_j(\cdot;\theta) : \theta \in \Theta\}$ is QMD with Fisher's information $I_j(\theta)$ positive definite and continuous. Let $\Psi^2(n;\theta) = \sum_{j=1}^n I_j(\theta)$. Under certain conditions (see Theorems II.3.1, II.6.1 and III.4.1 of IH), Conditions (IH) alongwith the LAN condition are satisfied with $\varphi_n = (\Psi^2(n;\theta))^{-1/2}$. A very important example of this kind apart from Example 1 is the "signal plus noise model" or a (nonlinear) regression model (see Sec. II.4 of IH). One can also consider a triangular array version of this; see Theorem II.3.1′ of IH in this context.

There are many other "regular" situations outside the i.i.d. set up. Among them there are certain Markov processes, Gaussian white noise (Sec. II.7 and III.5 of IH) and a planar Gibbsian point process model (Mase, 1992).

EXAMPLE 3. ALMOST SMOOTH DENSITY. In this case, the classical smoothness conditions are marginally violated and the Fisher information is infinite. However LAN condition is satisfied with a different normalizer. This interseting case has been treated in Sections II.5 and III.3 of IH; see also Woodroofe (1972) and Weiss and Wolfowitz (1974a). To describe briefly, consider i.i.d. observations from a location family $f(x-\theta)$ in $\mathbb{R}$. Suppose $f(x)$ is absolutely continuous and $(f'(x))^2/f(x)$ is integrable everywhere except in a neighbourhood of finitely many points $x_1,\ldots,x_l$ and in a neighbourhood of a point x_k, $k=1,\ldots,l$ the following representation is assumed to be valid:

$$f(x)=\begin{cases} a_k|x-x_k|+\psi_k(x), & \text{if } x\le x_k,\\ b_k|x-x_k|+\psi_k(x), & \text{if } x>x_k,\end{cases} \tag{2.5}$$

where ψ_k is twice continuously differentiable with $\psi_k(x_k)=\psi_k'(x_k)=0$, $a_k, b_k\ge 0$ and $B=\sum_{k=1}^{n}(a_k+b_k)>0$. Then LAN condition holds with $\varphi_n=(n\log n)^{-1/2}$. Also, Conditions (IH 1) and (IH 2) are satisfied; one needs to assume the moment condition (1.2) if Θ is unbounded.

Although the results of this chapter are valid in a wide generality, we are more interested in applying them to the non-smooth cases, or more familiarly known as the nonregular cases (because most of the results are already known in the "regular cases"). So far in nonregular cases, the examples of practical importance are only restricted to the i.i.d. case.

EXAMPLE 4. DENSITIES WITH JUMPS–ONE-PARAMETER CASE. The case of one-parameter family of discontinuous densities occupy a very distingiushed position since a long time because simple examples of this kind showed that a quite different phenomenon happens if the classical smoothness conditions are not satisfied (see Cramér (1946)). Many authors investigated this case; see, for example, Chernoff and Rubin (1956). A complete treatment is presented in IH (ch. V). Let $X_1, X_2, \ldots$ be i.i.d. with a density $f(x;\theta)$ in $\mathbb{R}$ which has only r discontinuities $a_1(\theta),\ldots,a_r(\theta)$ and at $a_k(\theta)$, $k=1,\ldots,r$, the right limit $p_k(\theta)$ and the left limit $q_k(\theta)$ exist. Then under certain assumptions (see p. 242 of IH), it is shown that Conditions (IH) are satisfied with $\varphi_n = n^{-1}$. Familiar examples of this kind are $U(0,\theta)$, $\theta > 0$ and $f(x;\theta) = f(x-\theta)$ with $f(x) = e^{-x}\chi\{x \geq 0\}$.

EXAMPLE 5. DENSITIES WITH SINGULARITIES. Let $X_1, X_2, \ldots$ be i.i.d. real valued random variables having a density $f(x;\theta)$ with respect to the Lebesgue measure and assume for simplicity that $f(x;\theta) = f(x-\theta)$, i.e., a location family. Sometimes, the lack of smoothness happens because of a singularity. Such cases are treated in Chapter VI of IH, and we briefly state the set up.

A point z is called a *singularity of order α of the first type* $(0 < \alpha < 1)$ for the density $f(x)$ if , in a neighbourhood of z, $f(x)$ admits a representation

$$f(x) = \begin{cases} p(x)|x-z|^{\alpha}, & \text{if } x > z, \\ q(x)|x-z|^{\alpha}, & \text{if } x < z, \end{cases} \tag{2.6}$$

where the functions $p(x)$ and $q(x)$ are continuous, $p(z) + q(z) > 0$ and there exist a number $\lambda > 1 + \alpha$ such that as $\eta \to 0$, $\eta > 0$

$$\begin{aligned} &\int_z |p^{1/\lambda}(x+\eta) - p^{1/\lambda}(x)|^{\lambda}|x-z|^{\alpha}dx \\ &+ \int^z |q^{1/\lambda}(x-\eta) - q^{1/\lambda}(x)|^{\lambda}|x-z|^{\alpha}dx = O(\eta^{\lambda}). \end{aligned} \tag{2.7}$$

Here and below $\int_z$ (or $\int^z$) denotes integration over an interval located to the right (left) of z where the representation (2.6) is valid.

A point z is called a *singularity of order α of the second type* $(0 < \alpha < 1)$ for the density $f(x)$ if , in a neighbourhood of z, $f(x)$ admits a representation

$$f(x) = \begin{cases} h(x)\exp[a(x)|x-z|^{\alpha/2}], & \text{if } x < z, \\ h(x)\exp[b(x)|x-z|^{\alpha/2}], & \text{if } x > z, \end{cases} \tag{2.8}$$

where the functions $a(x)$ and $b(x)$ are continuous with at least one of $a(z)$ and $b(z)$ is nonzero,

$$\int^z |a(x-\eta) - a(x)|^2 dx + \int_z |b(x+\eta) - b(x)|^2 dx = o(|\eta|^{1+\alpha}) \quad \text{as } \eta \to 0 \tag{2.9}$$

and there exist a number $\lambda > 1 + \alpha$ such that as $\eta \to 0$, we have

$$\int |h^{1/\lambda}(x-\eta) - h^{1/\lambda}(x)|^{\lambda} dx = O(|\eta|^{\lambda}). \tag{2.10}$$

A *singularity of the third type* is almost same as the first type of singularity with the only difference being $-1 < \alpha < 0$. The case $\alpha = 0$ is a transition phase and corresponds to the case of discontinuous densities.

The second type singularity appeared in the works of Prakasa Rao (1968) where incidentally the method of likelihood ratio process was first successfully applied (in the context of finding asymptotic distribution of the MLE). Densities with singularities have also been considered in Woodroofe (1974) and Polfeldt (1970). The special case of the location shift of a lognormal family was considerd by Hill (1963) who also used this model to analyze a point source epidemic data. The term singularity and its classification are due to IH.

Suppose f has a finite number of singularities and (1.2) is satisfied. It is shown in IH that the asymptotic properies of the LRP depends only on the highest order singularity. If α is the order of the highest order singularity, then an appropriate normalizing constant is $\varphi_n = n^{-1/(1+\alpha)}$. Familiar examples correspond to

(i) $f(x) = (1/\Gamma(\alpha))e^{-x}x^{\alpha-1}\chi\{x \geq 0\}$, $1 < \alpha < 2$ (first type),

(ii) $f(x) = C\exp[-a|x|^{\alpha/2}]$, $0 < \alpha < 1$ (second type),

(iii) $f(x) = (1/\Gamma(\alpha))e^{-x}x^{\alpha-1}\chi\{x \geq 0\}$, $0 < \alpha < 1$ (third type).

EXAMPLE 6. MULTIPARAMETER DISCONTINUOUS DENSITIES. Let $X_1, X_2, \ldots$ be i.i.d. $\mathbb{R}^k$-valued random variables having a density $f(x;\theta)$ with respect to the Lebesgue measure where $\theta \in \Theta \subset \mathbb{R}^d$ and the densities are discontinuous in θ. This is a generalization of Example 4 and treated in Rubin (1961) and Ermakov (1977). Also see Pflug (1982b). Under certain assumptions, it follows from Ermakov (1977) that Conditions (IH) are satisfied. Examples of this kind are

(i) $U(\theta - \varphi, \theta + \varphi)$, $\theta \in \mathbb{R}$, $\varphi > 0$,

(ii) For $\Omega \subset \mathbb{R}^d$, $x \in \mathbb{R}^d$, $\theta \in \mathbb{R}^d$,

$$f(x;\theta) = \begin{cases} (\mathrm{vol}(\Omega))^{-1}, & \text{if } x - \theta \in \Omega, \\ 0, & \text{otherwise,} \end{cases}$$

(iii) $f(x;\theta) = \begin{cases} C\exp[-\sum_{i=1}^d a_i(x_i - \theta_i)], & \text{if } x_i > \theta_i \quad \forall i \\ 0, & \text{otherwise.} \end{cases}$

There is another important type of family of multiparameter discontinuous densities very much different from Example 6. In this case, the density is discontinuous at some points depending on the first component of the parameter but the family of densities is smooth in the other components for every fixed value of the first component. The whole of Chapter 2 is devoted to the study of such cases. The case with a singularity in the density at a point depending on the first component and smoothness of the family of densities as the other components vary is dealt in Chapter 3.

3 Behaviour of Posterior Distributions

Conditions (IH) have several implications regarding the asymptotic properties of Bayes estimates and posterior distributions. The properties of Bayes estimates were studied in IH which are stated in Appendix B for ready references. In this section, we study the asymptotic properties of the posterior distributions under Conditions (IH).

Let Π be the class of (possibly improper) prior densities on Θ which are continuous and positive at θ_0 and have a polynomial majorant. For example, Jeffrys' prior in Example 1 is an element of Π. Let $\mathfrak{L}$ be the class of continuous "loss functions" $l : \mathbb{R}^d \to [0, \infty)$ satisfying the following conditions:

(i) $l(0) = 0$, $l(x) = l(-x)$ for all $x \in \mathbb{R}^d$.

(ii) The sets $\{x : l(x) < c\}$ are convex for all $c > 0$ and bounded if c is sufficiently small.

(iii) $l(x) \leq B_0(1 + \|x\|^b)$, $x \in \mathbb{R}^d$ for some $B_0, b > 0$.

(iv) There exist numbers $H_0, \eta > 0$ such that for all $H \geq H_0$,

$$\sup\{l(x) : x \leq H^\eta\} - \inf\{l(x) : x \geq H\} \leq 0.$$

The class $\mathfrak{L}$ of loss functions is sufficiently general to include all loss of the form $l(x) = \|x\|^p$, $p > 0$. Whenever we talk about posterior distribution and Bayes estimates, it is implicitly understood that the prior $\pi \in \Pi$ and loss $l \in \mathfrak{L}$. (Here, by loss function l, we actually mean that at stage n, the loss function is $l(\varphi_n^{-1}(a - \theta_0))$.)

It can be seen from the proofs of Lemma I.5.2 and Theorem I.5.2 of IH that the posterior is proper and the set of Bayes estimates is nonempty for all sufficiently large n almost surely. We shall, for simplicity, assume that the posterior is always proper and the set of Bayes estimates is nonempty. It then follows by the celebrated von Neumann selection theorem that a measurable choice of Bayes estimate is possible (provided the underlying probability is assumed to be complete). However, this is impractical to assume that the Bayes estimate is measurable since while computing from the data, we cannot assure that we are following a measurable choice. Therefore, we will not work with any measurability restriction and allow a completely arbitrary choice. Since the resulting random variables are not necessarily measurable, the weak convergence results are interpreted in the sense of Hoffman-Jørgessen-Dudley and others (see, e.g., Pollard (1990), Ch. 9). More specifically, for a sequence of (possibly nonmeasurable) random variables $\{X_n\}$, we say that $\{X_n\}$ has a limiting distribution P (a probability) if

$$\liminf_{n\to\infty} \Pr_*\{X_n \in A\} = \limsup_{n\to\infty} \Pr^*\{X_n \in A\} = P(A)$$

for all P-continuity sets A, where $\Pr_*$ and $\Pr^*$ stand for the inner and outer probabilities respectively. A similar remark also applies to the MLE.

Set

$$\xi_n(u) = \frac{Z_n(u)\pi(\theta_0 + \varphi_n u)}{\int_{U_n} Z_n(v)\pi(\theta_0 + \varphi_n v)dv}, \quad u \in U_n,$$

the posterior of the normalized parameter $u = \varphi_n^{-1}(\theta - \theta_0)$. We also denote the posterior probability of a set A by $\pi_n(A|X^n)$.

One important question is the consistency of posterior distributions, i.e., whether the posterior concentrates in the neighbourhoods of the true parameter point as more and more data is collected. More precisely, we have the following definition.

DEFINITION 3.1. We say that the posterior is (*strongly*) *consistent* if for any neighbourhood V of θ_0,

$$\lim_{n\to\infty} \pi_n(\theta \notin V|X^n) = 0 \ a.s. \tag{3.1}$$

The posterior is called *weakly consistent* if

$$\pi_n(\theta \notin V|X^n) \xrightarrow{P} 0. \tag{3.2}$$

REMARK 3.1. Strong and weak consistency of posterior can be viewed as Bayesian analogues of the strong and weak laws of large numbers respectively.

One nice consequence of the consistency is that the posterior is asymptotically free of the prior. Note that Conditions (IH) play no role here.

THEOREM 3.1. *Let θ_0 be an interior point of Θ and π_1 and π_2 be two prior densities which are positive and continuous at θ_0. If the posterior $\pi_{1n}(\theta|X^n)$ and $\pi_{2n}(\theta|X^n)$ are consistent, then*

$$\lim_{n\to\infty} \int |\pi_{1n}(\theta|X^n) - \pi_{2n}(\theta|X^n)|d\theta = 0 \quad a.s. \tag{3.3}$$

If the posteriors are weakly consistent, then

$$\int |\pi_{1n}(\theta|X^n) - \pi_{2n}(\theta|X^n)|d\theta \xrightarrow{P} 0. \tag{3.4}$$

PROOF. We shall prove only the first assertion; the other being exactly the same. Let V be any neighbourhood of θ_0. Given $\eta > 0$, outside a null set N, we can get $n_0 \geq 1$ (depending on the sample sequence) such that

$$\int |\pi_{1n}(\theta|X^n) - \pi_{2n}(\theta|X^n)| d\theta \leq \int_V \pi_{1n}(\theta|X^n) \left|1 - \frac{\pi_{2n}(\theta|X^n)}{\pi_{1n}(\theta|X^n)}\right| d\theta + 2\eta; \tag{3.5}$$

this follows from posterior consistency. Let $\delta > 0$. Now choose the neighbourhood V so that for all $\theta \in V$ and $i = 1, 2$,

$$\pi_i(\theta_0)(1-\delta) \leq \pi_i(\theta) \leq \pi_i(\theta_0)(1+\delta); \tag{3.6}$$

this is possible by continuity and positivity of π_1 and π_2 at θ_0. Thus

$$(1-\delta)\pi_i(\theta_0)C_n \leq \int_V \pi_i(\theta) p^n(X^n;\theta) d\theta \leq (1+\delta)\pi(\theta_0)C_n, \quad i = 1, 2, \tag{3.7}$$

where $C_n = \int_V p^n(X^n;\theta)d\theta$.

Clearly, outside the null set N, for $i = 1, 2$,

$$\begin{aligned}\int_V \pi_i(\theta)p^n(X^n;\theta)d\theta &\leq \int \pi_i(\theta)p^n(X^n;\theta)d\theta \\ &\leq (1-\eta)^{-1} \int_V \pi_i(\theta)p^n(X^n;\theta)d\theta.\end{aligned}$$

Using (3.6), for $\theta \in V$ and $i = 1, 2$, we have

$$\frac{(1-\eta)(1-\delta)}{(1+\delta)C_n} p^n(X^n;\theta) \leq \pi_{in}(\theta|X^n) \leq \frac{(1+\delta)}{(1-\delta)C_n} p^n(X^n;\theta). \tag{3.8}$$

Thus outside the null set N and for $\theta \in V$, we have

$$(1-\eta)\left(\frac{1-\delta}{1+\delta}\right)^2 \leq \frac{\pi_{1n}(\theta|X^n)}{\pi_{2n}(\theta|X^n)} \leq (1-\eta)^{-1}\left(\frac{1+\delta}{1-\delta}\right)^2. \tag{3.9}$$

Substituting in (3.5), we get the result since δ and η are arbitrarily small positive numbers. □

REMARK 3.2. If $\mathfrak{P}$ is a family of prior densities such that $\pi(\theta_0) > 0$ for all $\pi \in \mathfrak{P}$, $\mathfrak{P}$ is equicontinuous at θ_0 and the posterior consistency is uniform for priors $\pi \in \mathfrak{P}$, then

$$\lim_{n\to\infty} \sup\{\int |\pi_1(\theta|X^n) - \pi_2(\theta|X^n)| d\theta : \pi_1, \pi_2 \in \mathfrak{P}\} = 0 \quad \text{a.s.} \tag{3.10}$$

REMARK 3.3. Let $\widehat{\theta}_n = \widehat{\theta}_n(X^n)$ and $T_n = T_n(X^n)$ be any two sequence of statistics and let $v = T_n^{-1}(\theta - \widehat{\theta}_n)$ be the normalized and centered parameter. Denote the posterior densities of v for priors π_1 and π_2 by $\pi^*_{1n}(v|X^n)$ and $\pi^*_{2n}(v|X^n)$ respectively. Since L^1-distance is invariant under a location and scale change, the conclusion (3.3) can be restated as

$$\lim_{n\to\infty} \int |\pi^*_{1n}(v|X^n) - \pi^*_{2n}(v|X^n)| dv = 0 \quad \text{a.s.} \tag{3.11}$$

Similarly (3.4) can also be restated.

Posterior consistency is generally true in the finite dimensional situations. See, in this connection, Diaconis and Freedman (1986) and the references therein, where a subjective interpretation of posterior consistency is also discussed. Below, we show that posterior consistency is always guaranteed under Conditions (IH 1) and (IH 2).

PROPOSITION 3.1. *Assume Conditions* (IH 1) *and* (IH 2) *and let* $\pi \in \Pi$. *Then the posterior is weakly consistent. If further* $\sum_{n=1}^{\infty} \|\varphi_n\|^s < \infty$ *for some* $s > 0$, *then the posterior is strongly consistent also (provided it is meaningful).*

PROOF. Let V be a neighbourhood of θ_0 and let $r > 0$ be such that the open ball of radius r around θ_0 is contained in V. By Lemma I.5.2 of IH, for any $N > 0$, there is a constant C_N such that

$$E[\int_{\|u\|>H} \xi_n(u)du] \le C_N\, r^{-N} \|\varphi_n\|^N \quad \forall H > 0. \tag{3.12}$$

Consequently, we have

$$E[\pi_n(\theta \notin V|X^n)] \le E[\int_{\{\|u\|>r/\|\varphi_n\|\}} \xi_n(u)du] \le C_N\, r^{-N} \|\varphi_n\|^N. \tag{3.13}$$

The result is now immediate. □

Proposition 3.1 and Theorem 3.1 show that under Conditions (IH 1) and (IH 2), the posterior distributions stabilize. We now investigate the convergence properties of the normalized posterior distributions. Define

$$\xi(u) = Z(u) / \int_{\mathbb{R}^d} Z(v)dv.$$

In the following, for any set $A \in \mathcal{B}^d$, we shall abbreviate $\int_A \xi_n(u)du$ as $\xi_n(A)$ and $\int_A \xi(u)du$ as $\xi(A)$. Note that $\pi_n(u \in A|X^n) = \xi_n(A)$. The following theorem, actually implicit in Theorem I.10.2 of IH, shows weak convergence.

THEOREM 3.2. *Assume Conditions* (IH) *and let* $\pi \in \Pi$. *Then for any* $A \in \mathcal{B}^d$, *we have*

$$\xi_n(A) \xrightarrow{d} \xi(A). \tag{3.14}$$

PROOF. Fix $M > 0$ and note, by Theorem A.2 that

$$\begin{aligned}(\textstyle\int_{\|u\|\le M} \chi\{u \in A\}\pi(\theta_0 + \varphi_n u)Z_n(u)du, \int_{\|u\|\le M} \pi(\theta_0 + \varphi_n u)Z_n(u)du) \\ \xrightarrow{d} (\textstyle\int_{\|u\|\le M} \chi\{u \in A\}Z(u)du, \int_{\|u\|\le M} Z(u)du),\end{aligned} \tag{3.15}$$

and hence

$$\frac{\int_{\|u\|\le M} \chi\{u \in A\}\pi(\theta_0 + \varphi_n u)Z_n(u)du}{\int_{\|u\|\le M} \pi(\theta_0 + \varphi_n u)Z_n(u)du} \xrightarrow{d} \frac{\int_{\|u\|\le M} \chi\{u \in A\}Z(u)du}{\int_{\|u\|\le M} Z(u)du}. \tag{3.16}$$

But

$$\xi_n(A) = \frac{\int_{\|u\|\le M} \chi\{u \in A\}\pi(\theta_0 + \varphi_n u)Z_n(u)du}{\int_{\|u\|\le M} \pi(\theta_0 + \varphi_n u)Z_n(u)du}(1 + \gamma_n(M)), \tag{3.17}$$

where $\lim_{\substack{M\to\infty \\ n\to\infty}} \gamma_n(M) = 0$ (in probability) and

$$\xi(A) = \frac{\int_{\|u\|\le M} \chi\{u \in A\}Z(u)du}{\int_{\|u\|\le M} Z(u)du}(1 + \gamma(M)), \tag{3.18}$$

where $\gamma(M) \xrightarrow{P} 0$ by arguments very similar to those given in IH. An application of Theorem I.4.2 of Billingsley (1968, p. 25) now proves the validity of (3.14). □

REMARK 3.4. Let $\mathcal{A}$ be a countable subcollection of $\mathcal{B}^d$. Then by a slight modification of the proof of Theorem 3.2, it follows that for any $A_1, \ldots, A_r \in \mathcal{A}$, we have

$$(\xi_n(A_1), \ldots, \xi_n(A_r)) \xrightarrow{d} (\xi(A_1), \ldots, \xi(A_r)). \tag{3.19}$$

Hence by the weak convergence theory in $\mathbb{R}^\infty$, the $\mathbb{R}^\infty$-valued process $(\xi_n(A) : A \in \mathcal{A})$ converges weakly to the $\mathbb{R}^\infty$-valued process $(\xi(A) : A \in \mathcal{A})$.

Let $\mathcal{P}$ donote the space of all absolutely continuous probabilities on $\mathbb{R}^d$ equipped with the total variation distance. Then $\mathcal{P}$ is isometrically identified with the space of all probability densities on $\mathbb{R}^d$ with the L^1-distance. Note that by Lemma A.2, the processes $\xi_n(\cdot)$ and $\xi(\cdot)$ can be viwed as L^1-valued random variables. The conclusion in Remark 3.4 can be substantially strengthened as shown by the following result.

THEOREM 3.3. *The process* $\xi_n(\cdot)$ *converges weakly to* $\xi(\cdot)$ *in* $L^1(\mathbb{R}^d)$.

PROOF. In view of Remark 3.4, the finite dimensionals of $\xi_n(\cdot)$ converge to those of $\xi(\cdot)$ and thus we have to verify tightness. In view of Corollary A.1, we have to verify for some $n_0 \geq 1$, the quantities given by

(a) $\int_{\|u\| \leq M} |\xi_n(u+x) - \xi_n(u)| du$

and

(b) $\int_{\|u\| > M} \xi_n(u) du$

are arbitrarily small with probability arbitrarily close to one for all $\|x\| < \delta$ and $n \geq n_0$ by choosing δ sufficiently small and M sufficiently large. The second assertion follows from Lemma I.5.2 of IH and we now prove the first. For simplicity, assume that $\pi \equiv 1$ and note that

$$\begin{aligned}
&\int_{\|u\| \leq M} |\xi_n(u+x) - \xi_n(u)| du \\
= \;& \left(\int Z_n(u) du\right)^{-1} \int_{\|u\| \leq M} |Z_n^{1/2}(u+x) - Z_n^{1/2}(u)||Z_n^{1/2}(u+x) + Z_n^{1/2}(u)| du \\
\leq \;& 2\left(\int Z_n(u) du\right)^{-1} \left(\int_{\|u\| \leq M} |Z_n^{1/2}(u+x) - Z_n^{1/2}(u)|^2\right)^{1/2} \\
& \times \left(\int_{\|u\| \leq M+\delta} Z_n(u) du\right)^{1/2}; \qquad (3.20)
\end{aligned}$$

above we have used Cauchy-Schwartz inequality and the elementarary inequality $(a+b)^2 \leq 2(a^2+b^2)$. The expression in (3.20) is in turn less than or equal to

$$\leq 2\left(\int_{\|u\| \leq M+\delta} Z_n(u) du\right)^{-1} \left(\int_{\|u\| \leq M} |Z_n^{1/2}(u+x) - Z_n^{1/2}(u)|^2\right)^{1/2}. \qquad (3.21)$$

The first factor in (3.21) is clearly stochastically bounded. Now

$$\begin{aligned}
& E\left(\int_{\|u\| \leq M} |Z_n^{1/2}(u+x) - Z_n^{1/2}(u)|^2 du\right)^{1/2} \\
\leq \;& \left(\int_{\|u\| \leq M} E|Z_n^{1/2}(u+x) - Z_n^{1/2}(u)|^2 du\right)^{1/2} \\
\leq \;& A(K(1+(M+\delta)^{m_1})^{1/2} M^{d/2} \delta^{\alpha/2} \qquad (3.22)
\end{aligned}$$

by Condition (IH 1) where A is an absolute constant. The last expression can be made arbitrarily small, for any fixed M, by choosing $\delta > 0$ sufficiently small. Hence Assertion (a) follows and the proof is complete. □

The above results, although theoretically interesting, give only weak limit of posterior probabilities which has not much relevance for approximating posterior in practical situations. It is well known that in the regular cases, the posterior centered at the MLE converges to a normal distribution almost surely. This fact was first observed by Laplace (1774) and more recently by Bernstein (1917) and von Mises (1931) and subsequently referred to as the *Bernstein-von Mises theorem* or the *Bayesian central limit theorem.* A complete proof of this fact in case of i.i.d. observations first appeared in Le Cam (1953, 1958). Various modifications and extensions have been made by several authors including Bickel and Yahav (1969), Walker (1969), Chao (1970), Dawid (1970), Borwanker, Kallianpur and Prakasa Rao (1971), Heyde and Johnstone (1979), Chen (1985), Sweeting and Adekola (1987) and Clarke and Barron (1990b). A detailed discussion on the various conditions underlying the Bernstein-von Mises theorem can be found in Le Cam (1970). Refinements of posterior normality are considered in Johnson (1967, 1970), Ghosh, Sinha and Joshi (1982) and Woodroofe (1992). On the other hand, for a class of discontinuous densities, Samanta (1988) obtained a non-normal limit of the posterior distribution after a suitable centering. This suggests the possibility that the posterior, after a suitable centering (not necessarily at the MLE) may converge to a limit (not necessarily a normal distribution) almost surely, or at least in probability. In the remaining portion of this section, we find necessary and sufficient conditions for the above convergence. We shall introduce the notion of a "suitable" centering. But before that, we observe a very simple but interesting fact. The result is valid only in the i.i.d. case, but Conditions (IH) play no role here.

PROPOSITION 3.2. *Let $X_1, X_2, \ldots$ be i.i.d. random variables taking values in a standard Borel space $(\mathfrak{X}, \mathcal{A})$ and has a density $f(x;\theta)$, $\theta \in \Theta \subset \mathbb{R}^d$ with respect to a σ-finite measure ν on $(\mathfrak{X}, \mathcal{A})$. Let $\widehat{\theta}_n = \widehat{\theta}_n(X_1, \ldots, X_n)$ and $T_n = T_n(X_1, \ldots, X_n)$ be symmetric functions of $X_1, \ldots, X_n$ (which may or may not involve the true parameter θ_0). Let π be a prior density for θ, $v = T_n^{-1}(\theta - \widehat{\theta}_n)$ and $A \in \mathcal{B}^d$ be fixed. Let $c(X)$ be a random variable on $(\mathfrak{X}^\infty, \mathcal{A}^\infty)$ such that*

$$\pi_n(v \in A | X^n) \xrightarrow{P} c(X), \tag{3.23}$$

where X stands for the infinite sequence $(X_1, X_2, \ldots)$. Then $c(X)$ is almost surely constant.

PROOF. The posterior density of v is

$$\pi_n^*(v|X^n) = \frac{\pi(\widehat{\theta}_n + T_n v) \prod_{i=1}^n f(X_i; \widehat{\theta}_n + T_n v)}{\int \pi(\widehat{\theta}_n + T_n w) \prod_{i=1}^n f(X_i; \widehat{\theta}_n + T_n w) dw}.$$

This is a symmetric function of $X_1, \ldots, X_n$ and hence so is $\pi_n(v \in A|X^n) = \int_A \pi_n^*(v|X^n) dv$. Thus $c(X)$ is measurable with respect to the symmetric σ-field and hence the result follows by an application of the Hewitt-Savage zero-one law. □

Often it is true that $\varphi_n T_n \xrightarrow{P} \Sigma$ where Σ is a positive definite matrix. In that case, one may assume that $T_n = \varphi_n^{-1}$.

DEFINITION 3.2. An $\mathbb{R}^d$-valued random variable $\widehat{\theta}_n = \widehat{\theta}_n(X^n)$, is called a *proper centering* if for all $A \in \mathcal{B}^d$, there exists a non-random quantity $Q(A)$ such that

$$\sup\{|\pi_n(\varphi_n^{-1}(\theta - \widehat{\theta}_n) \in A|X^n) - Q(A)| : A \in \mathcal{B}^d\} \xrightarrow{P} 0. \tag{3.24}$$

If for each $A \in \mathcal{B}^d$, there is a constant $Q(A)$ such that

$$\pi_n(\varphi_n^{-1}(\theta - \widehat{\theta}_n) \in A)|X^n) \xrightarrow{P} Q(A), \tag{3.25}$$

we say that $\widehat{\theta}_n$ is a *semiproper centering*.

A statistic $\widehat{\theta}_n$ is called *compatible* (with the posterior) if the random element $((\varphi_n^{-1}(\widehat{\theta}_n - \theta_0), \xi_n(\cdot))$ in $\mathbb{R}^d \times L^1(\mathbb{R}^d)$ is weakly convergent.

REMARK 3.5. If (3.24) is satisfied, it automatically follows that Q is an absolutely continuous countably additive probability, and if (3.25) holds, then Q is a finitely additive probability. Although the motivation for a constant limit comes from Proposition 3.2, the concepts and the following results are certainly not resticted to the i.i.d. case.

REMARK 3.6. In view of Remark 3.3, if any of the above statements in Definition 3.2 holds for some prior $\pi \in \Pi$, then under the assumption of posterior consistency, it holds for any other prior in Π.

PROPOSITION 3.3. *Let $\widehat{\theta}_n$ be a proper centering such that $W_n := \varphi_n^{-1}(\widehat{\theta}_n - \theta_0)$ converges to a random variable W. Then for any countable subcollection $\mathcal{A}$ of $\mathcal{B}^d$, the $\mathbb{R}^\infty$-valued process $\{\pi_n(\varphi_n^{-1}(\theta - \theta_0) \in A|X^n) : A \in \mathcal{A}\}$ converges weakly to the process $\{Q(A - W) : A \in \mathcal{A}\}$.*

PROOF. By the weak convergence theory in $\mathbb{R}^\infty$, it is enough to prove the result for a finite collection $\{A_1, \dots, A_r\}$. Since Q is absolutely continuous, the mapping

$$x \mapsto (Q(A_1 - x), \dots, Q(A_r - x))$$

is continuous by Lemma A.1, and hence

$$(Q(A_1 - W_n), \dots, Q(A_r - W_n)) \xrightarrow{d} (Q(A_1 - W), \dots, Q(A_r - W)). \tag{3.26}$$

Errata

Equation (3.27), Chapter 1, page 25, should be corrected as

$$\begin{aligned}&\max_{1\le i\le r}|\pi_n(\varphi_n^{-1}(\theta-\theta_0)\in A_i|X^n)-Q(A_i-W_n)|\\ \le\ &\sup\{|\pi_n(\varphi_n^{-1}(\theta-\widehat{\theta}_n)\in A.|X^n)-Q(A)|\},\end{aligned}$$

with the same numbering as before (i.e., (3.27)).

In the proofs of Proposition 3.6 and Theorem 5.1 of Chapter 1, pages 28 and 37 respectively, the definition of $\psi_n(\cdot)$ is missing. It should be defined as

$$\psi_n(s)=\int_{\mathbf{R}^d} l(s-u)\xi_n(u)du.$$

Also,

$$\begin{aligned}
&\max_{1\le i\le r} |\pi_n(\varphi_n^{-1}(\theta-\theta_0)\in A_i|X^n) - Q(A_i)| \\
\le\ & \sup\{|\varphi_n^{-1}(\theta-\theta_0)\in A|X^n) - Q(A)|\},
\end{aligned} \tag{3.27}$$

which goes to zero in probability. By Slutsky's theorem, the result now follows. □

PROPOSITION 3.4. *Assume Conditions* (IH) *and let $\widehat{\theta}_n$ be a proper centering. Then $\varphi_n^{-1}(\widehat{\theta}_n - \theta_0)$ is weakly convergent.*

PROOF. We first show that $W_n := \varphi_n^{-1}(\widehat{\theta}_n - \theta_0)$ is stochastically bounded. If not, there exist $\varepsilon > 0$ such that for any $\lambda > 0$, there is a subsequence $\{m\}$ of $\{n\}$ for which

$$P\{\|W_m\| > \lambda\} > \varepsilon \quad \forall m. \tag{3.28}$$

Put $u = \varphi_n^{-1}(\theta - \theta_0)$ and $v = \varphi_n^{-1}(\theta - \widehat{\theta}_n)$. Then

$$\pi_n(v \in A|X^n) = \int_{A+W_n} \xi_n(u)du. \tag{3.29}$$

Fix a bounded set A and positive numbers ε and δ. Using Lemma I.5.2 of IH, find M large enough so that $\int_{\|u\|>M}\xi_n(u)du$, uniformly in $n \ge n_0$ (say), is less than ε with probability greater than $1-\delta$. Choose $\lambda > 0$ large enough so that $\|x\| > \lambda$ implies

$$A + x \subset \{u : \|u\| > M\}. \tag{3.30}$$

Combining (3.28) to (3.30) and using the definition of proper centering, it follows that we must have $Q(A) = 0$. Clearly, this cannot be true for every bounded set and hence $\{W_n\}$ is tight.

If W and W' are two subsequential limits of $\{W_n\}$, then by Proposition 3.3, we have

$$Q(A - W) \stackrel{d}{=} Q(A - W') \quad \forall A \in \mathcal{B}^d. \tag{3.31}$$

An application of Lemma A.3 makes the proof complete. □

Now we are in a position to prove one of our main results.

THEOREM 3.4. *Assume Conditions* (IH). *If a proper centering $\widehat{\theta}_n$ exists, then there exists a random variable W such that*

$$\varphi_n^{-1}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} W \tag{3.32}$$

and for almost all $u \in \mathbb{R}^d$,

$$\xi(u - W) \text{ is non-random.} \tag{3.33}$$

PROOF. In view of Proposition 3.4, such a W exists. Fix a countable field $\mathcal{A}$ which generates $\mathcal{B}^d$. By Lemma A.2, $\zeta(\cdot) := \{Q(A-W) : A \in \mathcal{B}^d\}$ is an $\mathfrak{M}(Q)$-valued random variable where $\mathfrak{M}(Q)$ stands for the set of all shifts of Q. Using Remark 3.4 and Proposition 3.3,

$$(\xi(A) : A \in \mathcal{A}) \stackrel{d}{=} (Q(A-W) : A \in \mathcal{A}) \tag{3.34}$$

and hence $\xi(\cdot) \stackrel{d}{=} \zeta(\cdot)$.

Since $P\{\zeta(\cdot) \in \mathfrak{M}(Q)\} = 1$, we also have $P\{\xi(\cdot) \in \mathfrak{M}(Q)\} = 1$. Define $\psi : \mathfrak{M}(Q) \to \mathbb{R}^d$ by $\psi(Q_x) = x$ where $Q_x(A) = Q(A+x) \ \ \forall A \in \mathcal{B}^d$. By Lemma A.3, we thus have

$$(\xi, \psi(\xi)) \stackrel{d}{=} (\zeta, \psi(\zeta)) = (\zeta, W) \tag{3.35}$$

since $\psi(\zeta) = W$ by definition. Put $W^* = \psi(\xi)$ and note that $W^* \stackrel{d}{=} W$. By Lemma A.1, we have

$$\int_{A+W^*} \xi(u)du \stackrel{d}{=} \int_{A+W} \zeta(u)du = Q(A)$$

and so

$$\int_A \xi(u-W^*)du = Q(A) \quad \text{a.s.} \tag{3.36}$$

The conclusion is now immediate. □

The next result shows that a proper centering, if exists, is essentially unique.

PROPOSITION 3.5. *Assume Conditions* (IH) *and let* $\widehat{\theta}_n$ *and* $\widetilde{\theta}_n$ *be two proper centerings. Then the associated probabilities and weak limits are shifts of each other.*

PROOF. Let Q_1, Q_2 denote the associated probabilities and W_1, W_2 denote the weak limits of $\{\varphi_n^{-1}(\widehat{\theta}_n - \theta_0)\}$ and $\{\varphi_n^{-1}(\widetilde{\theta} - \theta_0)\}$ respectively (which exist by Proposition 3.4). By Proposition 3.3, it follows that the $\mathcal{P}$-valued random process $\{Q_1(A-W_1) : A \in \mathcal{B}^d\}$ has the same distribution as $\{Q_2(A-W_2) : A \in \mathcal{B}^d\}$. Hence it folows that Q_2 is a shift of Q_1, say $Q_2(A) = Q_1(A+c) \ \forall A \in \mathcal{B}^d$. Using arguments similar to those in Theorem 3.4, it follows that $W_2 \stackrel{d}{=} W_1 + c$. □

We now give a partial answer to the question whether a semiproper centering exists.

THEOREM 3.5. *Assume Conditions* (IH) *and suppose that* $\widehat{\theta}_n$ *is a compatible and semiproper centering with associated finitely additive probability* Q*. Then* Q *is countably additive and there exist a random variable* W *satisfying*

(a) $\varphi_n^{-1}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} W$

and

(b) $\xi(u - W)$ *is non-random for almost all* $u \in \mathbb{R}^d$.

Further, if $\widetilde{\theta}_n$ *is another compatible semiproper centering with associated probability* Q' *and weak limit* W'*, then* Q' *and* W' *are shifts of* Q *and* W *respectively.*

PROOF. Let $(\varphi_n^{-1}(\widehat{\theta}_n - \theta_0), \xi(\cdot)) \xrightarrow{d} (W, \xi(\cdot))$. By Lemma A.1, we obtain

$$(\pi_n(\varphi_n^{-1}(\theta - \widehat{\theta}_n) \in A | X^n) : A \in \mathcal{B}^d) \xrightarrow{d} (\int_{A+W} \xi(u) du : A \in \mathcal{B}^d). \tag{3.37}$$

By given condition, for any $A \in \mathcal{B}^d$

$$\int_{A+W} \xi(u) du = Q(A) \tag{3.38}$$

and consequently

$$(\pi_n(\varphi_n^{-1}(\theta - \hat{\theta}_n) \in A | X^n) : A \in \mathcal{B}^d) \xrightarrow{d} (Q(A) : A \in \mathcal{B}^d). \tag{3.39}$$

Since the right hand side of (3.39) is non-random, we have a convergence in probability and so $\widehat{\theta}_n$ *is in fact a proper centering.* Thus the result follws from Theorem 3.4 or by direct arguments as before. □

REMARK 3.7. When Condition (3.33) in Theorem 3.4 is fulfilled, it is normally easy to exhibit a random variable W satisfying the requirement. But when (3.33) fails, it is relatively more difficult to show non-existence of such a W. A direct application of Theorem 3.4 thus may not be convenient. We present two alternative methods of checking non-existence. Let $Y(u) = \log Z(u)$. Then (3.33) is equivalent to saying that for almost all $u_1, u_2 \in \mathbb{R}^d$,

$$Y(u_1 - W) - Y(u_2 - W) \text{ is non-random.} \tag{3.40}$$

Also, if $\widehat{\xi}(t)$ denotes the Fourier transform of the (random) L^1-function $\xi(u)$, then by Theorem 3.3, a (further) necessary condition for the existence of posterior limit is

$$|\widehat{\xi}(t)|^2 \text{ is non-random.} \tag{3.41}$$

To show the non-existence of posterior limit, (3.40) and (3.41) may be easier to apply. Particularly, since the the last condition does not involve any unknown random variable, it is sometimes easier to check whether already (3.41) fails. It is to be expected that (3.33) is more likely to fail if the limiting LRP is more complicated. We however should admit that we have not been able to conclude the existence or non-existence in every possible nonregular cases. Fortunately for most of the important cases arising in practice, we have an answer. (See the next section for this.) Finally, even if a posterior limit does not exist, useful approximations may still be obtained. For example, Theorem 3.1 says that we can do our computations with respect to any convenient choice of prior if sample size is large enough.

Now it is natural to ask whether (3.33) is also sufficient to imply the posterior convergence. We see that this is indeed so, as shown by the following theorem.

THEOREM 3.6. *Assume Conditions* (IH) *and suppose that there exists a random variable W such that for almost all $u \in \mathbb{R}^d$, $\xi(u-W)$ is non-random. Let $l \in \mathfrak{L}$ and suppose that the random function*

$$\psi(s) := \int_{\mathbb{R}^d} l(s-u)\xi(u)du$$

attains its absolute minimum at a unique point τ. Then the Bayes estimate with respect to loss l works as a compatible proper centering, i.e., $(\varphi_n^{-1}(\widetilde{\theta}_n - \theta_0), \xi_n(\cdot))$ converges weakly (to $(\tau, \xi(\cdot))$ in $\mathbb{R}^d \times L^1(\mathbb{R}^d)$) and

$$\sup\{|\pi_n(\varphi_n^{-1}(\theta - \widetilde{\theta}_n) \in A|X^n) - Q(A)| : A \in \mathcal{A}\} \xrightarrow{p} 0$$

for some probability measure Q on $\mathbb{R}^d$.

Let $\widetilde{\theta}_n$ be a Bayes estimate with respect to a loss $\mathfrak{L}$ and prior $\pi \in \Pi$. Because of Remark 3.6, it is sufficient to consider the posterior with respect to the same prior π. We first establish the following result which also of some independent interest.

PROPOSITION 3.6. *Any Bayes estimate is compatible.*

PROOF. We have to show that

$$(\varphi_n^{-1}(\widetilde{\theta}_n - \theta_0), \xi_n(\cdot)) \xrightarrow{d} (\tau, \xi(\cdot)), \tag{3.42}$$

where τ is as in Theorem 3.6. By the arguments for the proof of Theorem I.10.2 of IH, it suffices to show that for all $M > 0$, in the space $C[-M, M]^d \times L^1(\mathbb{R}^d)$

$$(\psi_n(\cdot|M), \xi_n(\cdot)) \xrightarrow{d} (\psi(\cdot|M), \xi(\cdot)) \tag{3.43}$$

where $\psi_n(\cdot|M)$ and $\psi(\cdot|M)$ stand for the restrictions of $\psi_n(\cdot)$ and $\psi(\cdot)$ respectively on $[-M, M]^d$. From Theorems I.10.2 of IH and Theorem 3.3 respectively, we know that $\{\psi_n(\cdot|M)\}$ and $\{\xi_n(\cdot)\}$ are tight; hence it suffices to verify the finite dimensional convergences. Let $s_1, \ldots, s_m \in \mathbb{R}^d$ and $A_1, \ldots, A_k \in \mathcal{B}^d$. We have to show that

$$\begin{aligned} & (\psi_n(s_1), \ldots, \psi_n(s_m), \xi_n(A_1), \ldots, \xi_n(A_k)) \\ \xrightarrow{d} \;& (\psi(s_1), \ldots, \psi(s_m), \xi(A_1), \ldots, \xi(A_k)). \end{aligned} \tag{3.44}$$

This is an easy consequence of Theorem A.2 by the arguments used in Theorem I.10.2 of IH and Theorem 3.2. □

PROOF OF THEOREM 3.6. By the given condition,

$$\xi(u) = g(u + W), \tag{3.45}$$

where g is a fixed probability density. Let c be the unique minimizer of $\int l(s - u)g(u)du$. Then $\tau = W + c$ and hence, without loss of generality, we can assume that $W = \tau$. The posterior density of $v := \varphi_n^{-1}(\theta - \widetilde{\theta}_n)$ given by

$$\pi_n^*(v|X^n) = \xi_n(v - \tau_n), \tag{3.46}$$

where $\tau_n = \varphi_n^{-1}(\widetilde{\theta}_n - \theta_0)$. By (3.42) and Lemma A.1, in the space $L^1(\mathbb{R}^d)$, we have

$$(\pi_n^*(v|X^n) : v \in \mathbb{R}^d) \xrightarrow{d} (g(v) : v \in \mathbb{R}^d). \tag{3.47}$$

The result is now immediate since $g(\cdot)$ is a non-random element. □

We illustrate the usefulness of Theorem 3.6 by means of a simple example. Let $X_1, X_2, \ldots$ be a sequence of independent random variables with $X_n \sim N(\theta, \sigma_n^2)$, $n = 1, 2, \ldots$ with σ_n known and $\sum_{n=1}^{\infty} \sigma_n^{-2} = \infty$. Then by results of Section II.3.1 and III.4.1 of IH (or by straightforward computations), Conditions (IH) are satisfied with $\varphi_n = (\sum_{i=1}^{n} \sigma_i^{-2})^{-1/2}$ and $Z(u) = \exp[u\Delta - (1/2)u^2]$ where $\Delta \sim N(0, 1)$. Let $\pi \in \Pi$ and $\widetilde{\theta}_n$ be a Bayes estimate. Thus the condition given by (3.33) is satisfied with $W = \Delta$. Hence by Theorem 3.6, it follows that with $v = \varphi_n^{-1}(\theta - \widetilde{\theta}_n)$

$$\sup\{|\pi_n(v \in A|X^n) - Q(A)| : A \in \mathcal{B}\} \xrightarrow{P} 0,$$

where Q is the $N(0, 1)$ probability. In other words, we get a version of the Bernstein-von Mises theorem in this independent nonhomogeneous case. For a slightly differnt version, see Section 5.

The above obtained posterior limit theorem may not be fully satisfactory to a Bayesian because the convergence is in probability sense, while a Bayesian will like to know whether the posterior convergence holds for a given sequence of sample obsevations. In this context, let us observe the following simple fact, which is an easy consequence of Hewitt-Savage zero-one law.

ZERO-ONE LAW FOR POSTERIOR LIMIT. *Let $X_1, X_2, \ldots$ be i.i.d. observations and $\widehat{\theta}_n = \widehat{\theta}_n(X_1, \ldots, X_n)$ be symmetric in its arguments. Then the set of all samples for which*

$$\sup\{|\pi_n(\varphi_n^{-1}(\theta - \widehat{\theta}_n) \in A) - Q(A)| : A \in \mathcal{B}^d\} \to 0$$

has probability zero or one.

If (3.24) holds, there is no a priori reason that the corresponding almost sure convergence will happen. However, still we do not know of any example where (3.24) holds but the corresponding almost sure result fails.

4 Investigation of Posterior Convergence in Examples

In this section, we apply the criterion (3.33) to investigate whether a posterior convergence holds for the examples mentioned in Section 2.

(1) "REGULAR" CASES. In all the cases of Examples 1, 2, 3 and the examples mentioned after Example 2, the limiting LRP $Z(u)$ is of the form

$$Z(u) = \exp[u'\Delta - (1/2)u'\Sigma u], \tag{4.1}$$

where Σ is a constant positive definite matrix and $\Delta \sim N_d(0, \Sigma)$. It is clear that (3.33) is satisfied with $W = \Sigma^{-1}\Delta$. Consequently, Theorem 3.6 implies a posterior convergence with Bayes estimate as a proper centering and $N_d(0, \Sigma^{-1})$ as the limit. This fact is quite well known in the i.i.d. case (Example 1) and known as (a *in probability* version of) the Bernstein-von Mises theorem. It has also been established for for many stochastic processes (see Section 3 for the relevant references). For a different version, see also Section 5.

(2) ONE-PARAMETER DISCONTINUOUS DENSITIES. In the set up of Example 4, let

$$\Gamma = \{1 \le k \le r : p_k > 0, q_k > 0\},$$

$$\begin{aligned}\Gamma^+ &= \{1 \le k \le r : q_k = 0, a'_k > 0\} \cup \{1 \le k \le r : p_k = 0, a'_k < 0\},\\ \Gamma^- &= \{1 \le k \le r : p_k = 0, a'_k > 0\} \cup \{1 \le k \le r : q_k = 0, a'_k < 0\}\end{aligned}$$

and $c = \sum_{k=1}^{r}(p_k - q_k)a'_k$. The limiting LRP and hence the possibility of the existence of a posterior limit depends on the nature of the jumps. Let $\nu_1, \ldots, \nu_r, \widetilde{\nu}_1, \ldots, \widetilde{\nu}_r$ be independent unit rate homogeneous Poisson processes and for $k = 1, \ldots, r$, define

$$\nu_k^+ = \begin{cases} \nu_k(p_k a'_k u), & \text{if } a'_k > 0,\\ \nu_k(-q_k a'_k u), & \text{if } a'_k < 0,\end{cases}$$

if $u \ge 0$, $\nu_k^+(u) = 0$ if $u < 0$ and

$$\nu_k^- = \begin{cases} \widetilde{\nu}_k(-q_k a'_k u), & \text{if } a'_k > 0,\\ \widetilde{\nu}_k(p_k a'_k u), & \text{if } a'_k < 0,\end{cases}$$

if $u \le 0$, $\nu_k^+(u) = 0$ if $u > 0$.

The limiting LRP $Z(u)$ is given by

$$Z(u) = \exp\{uc + \sum_{k=1}^{r} \operatorname{sign}(ua'_k) \log(q_k/p_k)(\nu^+(u) + \nu_k^-(u))\} \tag{4.2}$$

(see Theorem V.2.1 of IH). Below, we consider several important cases.

Case 1. Assume that Γ is empty, i.e., there is no positive to positive jump.

Subcase 1(a). Suppose both Γ^+ and Γ^- are nonempty. In this case, $Z(u)$ can be written more simply as

$$Z(u) = \begin{cases} \exp[cu], & \text{if } -\tau^- < u < \tau^+,\\ 0, & \text{otherwise},\end{cases} \tag{4.3}$$

where τ^+ and τ^- are independent exponential random variables with parameters $\alpha := \sum_{k \in \Gamma^-}(q_k - p_k)a'_k$ and $\beta := \sum_{k \in \Gamma^+}(p_k - q_k)a'_k$ respectively. Note that $c = \beta - \alpha$.

If $c = 0$, we have

$$\xi(u) = \begin{cases} (\tau^+ + \tau^-)^{-1}, & \text{if } -\tau^- < u < \tau^+,\\ 0, & \text{otherwise}.\end{cases} \tag{4.4}$$

Had $\xi(\cdot)$ been of the form (3.33), the length of the support (i.e., where $\xi(u) > 0$) would have been constant. This is clearly not so for (4.4), and hence there canot be any posterior convergence. An example of this kind is $U(\theta, \theta + 1)$, $\theta \in \mathbb{R}$.

If $c \neq 0$, we have

$$\xi(u) = \begin{cases} c\exp[cu]/(\exp[c\tau^+] - \exp[c\tau^-]), & \text{if } -\tau^- < u < \tau^+, \\ 0, & \text{otherwise.} \end{cases} \tag{4.5}$$

Again, by the same argument, (3.33) is not satisfied and consequently there is no posterior limit. An example of this kind is $U(\theta, 2\theta)$, $\theta > 0$.

SUBCASE 1(b). Suppose one of Γ^- and Γ^+ is empty, say $\Gamma^- = \emptyset$. Here $Z(u)$ is simplified to

$$Z(u) = \begin{cases} \exp[cu], & \text{if } u < \tau^+, \\ 0, & \text{otherwise,} \end{cases} \tag{4.6}$$

and so

$$\xi(u) = \begin{cases} c\exp[c(u - \tau^+)], & \text{if } u < \tau^+, \\ 0, & \text{otherwise,} \end{cases} \tag{4.7}$$

where τ^+ is exponential with parameter c. (Note that in this case, $c = \beta > 0$.) It is clear that (3.33) is satisfied with $W = \tau^+$ and consequently we have a posterior limit.

Indeed, a limit has been obtained in Samanta (1988) for a special case of Subcase 1(b), where the support of the density is an interval which is either increasing or decreasing in θ. Samanta (1988) assumed conditions similar to those in Weiss and Wolfowitz (1974, Ch. 5) and a uniform integrability type condition on $\log f$ and obtained an exponential limit in an *almost sure* sense. In this situations, there exists a statistic T_n such that the set $\{(x_1, \ldots, x_n) : f(x_i; \theta) > 0 \text{ for all } i\}$ can be expressed as $\{T_n(x_1, \ldots, x_n) > \theta\}$ or $\{T_n(x_1, \ldots, x_n) < \theta\}$ according as the support is increasing or decreasing in θ. This T_n works as a proper centering. Important examples of this kind are shifts of exponential density, $U(0, \theta)$, $\theta > 0$ etc.

CASE 2. The set Γ is nonempty but both Γ^+ and Γ^- are empty.

We consider only the case $r = 1$ and $a_1' > 0$. In this case, we have $c = (p_1 - q_1)a_1'$ and

$$Z(u) = \begin{cases} \exp[cu + \nu(u)\log(q_1/p_1)], & \text{if } u \geq 0, \\ \exp[cu - \tilde{\nu}(-u)\log(q_1/p_1)], & \text{if } u < 0, \end{cases} \tag{4.8}$$

where $\nu(u)$ and $\tilde{\nu}(u)$ are homogeneous Poisson processes with rates $p_1 a_1'$ and $q_1 a_1'$ respectively. We shall show that posterior convergence does not hold by showing that (3.41) is violated. Let $\delta = \log(q_1/p_1)$, $0 = \tau_0 \leq \tau_1 \leq \tau_2 \leq \ldots$ be the occurence times of $\nu(\cdot)$ and $0 = \widetilde{\tau}_0 \leq \widetilde{\tau}_1 \leq \widetilde{\tau}_2 \leq \ldots$ be the occurence times of $\widetilde{\nu}(\cdot)$. Thus we have

$$Z(u) = \begin{cases} \exp[cu + \delta j], & \text{if } \tau_j \leq u < \tau_{j+1}, \\ \exp[cu - \delta j], & \text{if } -\widetilde{\tau}_{j+1} < u \leq -\widetilde{\tau}_j. \end{cases} \tag{4.9}$$

Now

$$\begin{aligned} \int Z(u)du &= \int_{\{u>0\}} Z(u)du + \int_{\{u<0\}} Z(u)du \\ &= \sum_{j=0}^{\infty} e^{\delta j} \int_{\tau_j}^{\tau_{j+1}} e^{cu} du + \sum_{j=0}^{\infty} e^{-\delta j} \int_{-\widetilde{\tau}_{j+1}}^{-\widetilde{\tau}_j} e^{cu} du \\ &= c^{-1}\{\sum_{j=0}^{\infty} e^{\delta j}(e^{c\tau_{j+1}} - e^{c\tau_j}) + \sum_{j=0}^{\infty} e^{-\delta j}(e^{-c\widetilde{\tau}_j} - e^{-c\widetilde{\tau}_{j+1}})\}. \end{aligned} \tag{4.10}$$

Similarly

$$\begin{aligned} &\int e^{itu} Z(u)du \\ = \;& (c+it)^{-1}\{\sum_{j=0}^{\infty} e^{\delta j}(\exp[(c+it)\tau_{j+1}] - \exp[(c+it)\tau_j]) \\ & \quad + \sum_{j=0}^{\infty} e^{-\delta j}(\exp[-(c+it)\widetilde{\tau}_j] - \exp[-(c+it)\widetilde{\tau}_{j+1}])\}. \\ = \;& (c+it)^{-1}\{\sum_{j=0}^{\infty} e^{\delta j}(e^{c\tau_{j+1}}(\cos t\tau_{j+1} + i\sin t\tau_{j+1}) \\ & \qquad - e^{c\tau_j}(\cos t\tau_j + i\sin t\tau_j)) \\ & \quad + \sum_{j=0}^{\infty} e^{-\delta j}(e^{-c\widetilde{\tau}_j}(\cos t\widetilde{\tau}_j - i\sin t\widetilde{\tau}_j) \\ & \qquad - e^{-c\widetilde{\tau}_{j+1}}(\cos t\widetilde{\tau}_j - i\sin t\widetilde{\tau}_j))\}. \end{aligned} \tag{4.11}$$

Hence we have

$$\widehat{\xi}(t) = (c/(c+it))N(t)/D(t),$$

where $N(t)$ equals

$$\begin{aligned} & \sum_{j=0}^{\infty} e^{\delta j}(e^{c\tau_{j+1}}(\cos t\tau_{j+1} + i\sin t\tau_{j+1}) - e^{c\tau_j}(\cos t\tau_j + i\sin t\tau_j)) \\ & + \sum_{j=0}^{\infty} e^{-\delta j}(e^{-c\widetilde{\tau}_j}(\cos t\widetilde{\tau}_j - i\sin t\widetilde{\tau}_j) - e^{-c\widetilde{\tau}_{j+1}}(\cos t\widetilde{\tau}_{j+1} - i\sin t\widetilde{\tau}_{j+1})) \\ = \;& \{\sum_{j=0}^{\infty} e^{\delta j}(e^{c\tau_{j+1}} \cos t\tau_{j+1} - e^{c\tau_j} \cos t\tau_j) \end{aligned}$$

$$
\begin{aligned}
&+\sum_{j=0}^{\infty} e^{-\delta j}(e^{-c\widetilde{\tau}_j}\cos t\widetilde{\tau}_j - e^{-c\widetilde{\tau}_{j+1}}\cos t\widetilde{\tau}_{j+1})\} \\
&+i\{\sum_{j=0}^{\infty} e^{\delta j}(e^{c\tau_{j+1}}\sin t\tau_{j+1} - e^{c\tau_j}\sin t\tau_j) \\
&+\sum_{j=0}^{\infty} e^{-\delta j}(e^{-c\widetilde{\tau}_{j+1}}\sin t\widetilde{\tau}_{j+1} - e^{-c\widetilde{\tau}_j}\sin t\widetilde{\tau}_j)\} \\
= {} & R(t) + iI(t) \quad \text{(say)}
\end{aligned}
\tag{4.12}
$$

and

$$
D(t) = \sum_{j=0}^{\infty} e^{\delta j}(e^{c\tau_{j+1}} - e^{c\tau_j}) + \sum_{j=0}^{\infty} e^{-\delta j}(e^{-c\widetilde{\tau}_j} - e^{-c\widetilde{\tau}_{j+1}}). \tag{4.13}
$$

To show that the function $|\widehat{\xi}(t)|^2$ is not non-random, it is enough to show that the random variable $|\widehat{\xi}(1)|^2$ is nondegenerate, i.e.,

$$
(R^2(1) + I^2(1))/D^2(1) \ \text{ is nondegenerate.} \tag{4.14}
$$

If (4.14) is not true, any conditional distribution is also degenerate. We condition on every random variable except τ_1, which we denote by x. Letting English letters (with or without subscript) to denote constants with respect to the conditional distribution, we have

$$
\begin{aligned}
& (R^2(1) + I^2(1))/D^2(1) \\
= {} & \frac{(e^{cx}\cos x - e^{\delta+cx}\cos x + a_1)^2 + (e^{cx}\sin x - e^{\delta+cx}\sin x + a_2)^2}{(e^{cx} - e^{\delta+cx}x + a)^2} \\
= {} & \frac{(e^{cx}\cos x + b_1)^2 + (e^{cx}\sin x + b_2)^2}{(e^{cx} + b)^2} \\
= {} & \frac{e^{2cx} + 2e^{cx}(b_1\cos x + b_2\sin x) + (b_1^2 + b_2^2)}{e^{2cx} + 2be^{cx} + b^2}.
\end{aligned}
\tag{4.15}
$$

It is easy to see that x is uniformly distributed over $(0, \tau_2)$ with respect to the conditional distribution. If (4.15) is degenerate at A (say), then for all $x \in (0, \tau_2)$,

$$
e^{2cx} + 2e^{cx}(b_1\cos x + b_2\sin x) + (b_1^2 + b_2^2) = A(e^{2cx} + 2be^{cx} + b^2). \tag{4.16}
$$

The identity in (4.16) can be written as

$$
A_1e^{2cx} + A_2e^{cx} + A_3e^{cx}\sin(x + \alpha) + A_4 = 0 \tag{4.17}
$$

for all $x \in (0, \tau_2)$. This forces $A_1 = A_2 = A_3 = A_4 = 0$ by the linear independence of the involved functions. But $A_1 = 1 - A$ and $A_2 = -Ab$ forces $b = 0$, and so $a = 0$.

It is obvious that $a > 0$ unless $\tau_2 = \tau_3 = \cdots$ and $0 = \tilde{\tau}_1 = \tilde{\tau}_2 = \cdots$ both happen. The last event has probability zero and so (4.14) is proved.

An important example of this kind is the change point problem

$$f(x;\theta) = \begin{cases} a\exp[-ax], & \text{if } 0 < x < \theta, \\ b\exp[-a\theta - b(x-\theta)], & \text{if } x > \theta, \end{cases}$$

where $a > b > 0$ are known constants and $0 < \theta < B$ (a known bound) is the parameter of interest. See, in this connection, Basu, Ghosh and Joshi (1988) and Ghosh, Joshi and Mukhopadhyay (1992a, b).

(3) DENSITY WITH SINGULARITIES. We consider the question of existence of a posterior limit for the set up given in Example 5. We consider a density with only one first or third type singularity at the point $x = 0$ satisfying $q = q(0) = 0$ and $p = p(0) > 0$. (The case $p = 0$ and $q > 0$ is exactly similar. The case treated here, althogh very special, is perhaps the most important one from a practical point of view.) In this case, we have $Z(u) = \exp[Y(u)]$,

$$Y(u) = \begin{cases} \alpha \int_0^\infty \log|1-u/x|(\nu(dx) - E\nu(dx)) & \\ -p\int_0^\infty(|1-u/x|^\alpha - 1 - \alpha\log|1-u/x|)x^\alpha dx & \\ +(p/(1+\alpha))u^{1+\alpha}, & \text{if } u \geq \tau, \\ -\infty, & \text{if } u < \tau, \end{cases} \tag{4.18}$$

where ν is a nonhomogeneous Poisson process with rate function $\lambda(x) = px^\alpha$ and τ is the first occurence time of ν (see Theorems VI.2.1 and VI.2.2 of IH). We shall show that a posterior limit does not exist by showing that (3.40) is violated. Suppose, if possible, (3.40) holds with a random variable W. Fix any $u_2 > u_1$ and note that $Y(u_1 - W) - Y(u_2 - W)$ is well defined and different from $-\infty$ on the set $\{W + \tau < u_1\}$. If (3.40) holds, it must be true that the set $\{W + \tau < u_1\}$ is a non-random set. By considering different u_1, it is easy to see that then $W + \tau$ must be constant, say $W = c - \tau$. Choose $u_2 > u_1 > c$ and put $u_1' = u_1 - c$, $u_2' = u_2 - c$. Thus

$$\begin{aligned} & Y(u_1 - W) - Y(u_2 - W) \\ = \; & Y(u_1' + \tau) - Y(u_2' + \tau) \\ = \; & -\alpha p \int_0^\tau [\log|1-(u_1'+\tau)/x| - \log|1-(u_2'+\tau)/x|]x^\alpha dx \\ & -p\int_0^\infty [g(u_1'+\tau, x) - g(u_2'+\tau, x)]x^\alpha dx \end{aligned}$$

$$+(p/(1+\alpha))[(u_1'+\tau)^\alpha-(u_2'+\tau)^\alpha]$$
$$+\alpha\int_\tau^\infty[\log|1-(u_1'+\tau)/x|-\log|1-(u_2'+\tau)/x|]\bar{\nu}(dx), \tag{4.19}$$

where

$$g(u,x)=|1-u/x|^\alpha-1-\alpha\log|1-u/x|$$

and

$$\bar{\nu}(dx)=\nu(dx)-E\nu(dx).$$

Clearly the first three terms are functions of τ only. If $Y(u_1-W)-Y(u_2-W)$ is non-random, then the conditional distribution of $Y(u_1-W)-Y(u_2-W)$ given τ is also degenerate. Therefore,

$$\int_\tau^\infty[\log|1-(u_1'+\tau)/x|-\log|1-(u_2'+\tau)/x|]\bar{\nu}(dx)$$
$$=\int_0^\infty\log|(x-u_1')/(x-u_2')|\bar{\mu}(dx)$$

has a degenerate conditional distribution given τ; here

$$\bar{\mu}(dx)=\mu(dx)-E\mu(dx)$$

and

$$\mu(dx)=\nu(\tau+dx).$$

However, given τ, μ is again a nonhomogeneous Poisson process with rate function $p(x+\tau)^\alpha$. Thus the conditional variance of $\int_0^\infty\log|(x-u_1')/(x-u_2')|\bar{\mu}(dx)$ is

$$p\int_0^\infty\log^2|(x-u_1')/(x-u_2')|(x+\tau)^\alpha dx>0.$$

This contradiction implies the nonexistence of a posterior limit.

Important examples of this kind are gamma density

$$f(x)=\begin{cases}(1/\Gamma(\alpha))e^{-x}x^{\alpha-1}, & \text{if } x\ge 0,\\ 0, & \text{otherwise,}\end{cases}$$

and the Weibull density

$$f(x)=\begin{cases}\alpha x^{\alpha-1}\exp[-x^\alpha], & \text{if } x\ge 0,\\ 0, & \text{otherwise,}\end{cases}$$

where $0<\alpha<2$, $\alpha\neq 1$.

Another example is provided by the beta density

$$f(x)=\begin{cases}(1/B(a,b))x^{a-1}(1-x)^{b-1}, & \text{if } 0\le x\le 1,\\ 0, & \text{otherwise,}\end{cases}$$

where $0<b<a<2$.

5 Some Further Properties of the MLE

The asymptotic properties of the MLE are treated in Sections I.5 and I.10 of IH and are also stated in Appendix B for convenience. In this section, we prove some further related properties. Let Condition (IH 1) be replaced by the following condition:

(IH 1)$'$ There exist numbers $m \geq \alpha > d$ such that

$$E|Z_n^{1/m}(u) - Z_n^{1/m}(v)|^m \leq \prod_{i=1}^{d} \mathrm{Pol}(R_i)\|u - v\|^\alpha$$

for all $u, v \in U_n$ with $|u_i|, |v_i| \leq R_i$, $i = 1, \ldots, d$.

Henceforth Conditions (IH 1)$'$, (IH 2) and (IH 3) together will be referred to as Conditions (IH)$'$. We use the notation $\widehat{\theta}_n$ for the MLE. For verification of Conditions (IH)$'$ in the i.i.d. case, we can take the convenient choice $m = \alpha = 2$ if the dimension $d = 1$. However for $d \geq 2$, this choice is not permissible. In the case when densities are smooth, a sufficient condition is given in Theorem III.3.2 of IH. Also, the restriction that the sample paths of the LRP and limiting LRP are continuous somewhat reduces the applicability of the above two theorems. For example, in case of Example 4, a different method was adopted in IH (Sec. V.3) for the treatment of the MLE.

If Conditions (IH)$'$ hold instead of Conditions (IH), it is not difficult to see that all the results on Bayes estimates and posterior distribution go through. But we can actually claim a stronger fact.

THEOREM 5.1. *Assume Conditions* (IH)$'$. *Let* $Z_n(\cdot)$ *and* $Z(\cdot)$ *have continuous sample paths and* $Z(\cdot)$ *attains its maximum at a unique point* $\widehat{u}$. *Suppose further the random function* $\psi(s) = \int l(s-u)\xi(u)du$ *attains its absolute minimum at a unique point* τ. *Then as* $n \to \infty$,

$$(\varphi_n^{-1}(\widehat{\theta}_n - \theta_0), \varphi_n^{-1}(\widetilde{\theta} - \theta_0)) \xrightarrow{d} (\widehat{u}, \tau). \tag{5.1}$$

PROOF. Let $C[-M, M]^d$ be the space of all continuous functions on $[-M, M]^d$ and for any function $f(\cdot)$ on $\mathbb{R}^d$, let $f(\cdot|M)$ denote the restriction of $f(\cdot)$ on $[-M, M]^d$. By arguments of Theorems I.10.1 and I.10.2 of IH, it is enough to show that as random elements in $C[-M, M]^d \times C[-M, M]^d$,

$$(Z_n(\cdot|M), \psi_n(\cdot|M)) \xrightarrow{d} (Z(\cdot|M), \psi(\cdot|M)). \tag{5.2}$$

Since tightness has already been verified, it only remains to prove the convergence of finite dimensional distributions. But this clearly follows from Theorem A.2. □

Theorem 5.1 has a very useful consequence which occurs in the LAN, LAMN and the LAQ situations (see Le Cam (1986) and Le Cam and Yang (1990) for definitions).

COROLLARY 5.1. *Assume conditions of Theorem* 5.1 *and suppose further the limiting LRP is of the form*

$$Z(u) = \exp[u'\Delta - (1/2)u'\Sigma u], \tag{5.3}$$

where Δ *is a random vector and* Σ *is an almost surely positive definite random matrix. Then*

$$\varphi_n^{-1}(\widetilde{\theta}_n - \widehat{\theta}_n) \xrightarrow{p} 0. \tag{5.4}$$

In other words, MLE and Bayes estimates are asymptotically equivalent.

PROOF. By Theorem 5.1,

$$\varphi_n^{-1}(\widetilde{\theta}_n - \widehat{\theta}_n) \xrightarrow{d} \tau - \widehat{u}. \tag{5.5}$$

If (5.3) is satisfied, Anderson's Lemma (see, e.g., Section II.10 of IH) implies that $\tau = \widehat{u} = \Sigma^{-1}\Delta$, and hence (5.4) is satisfied. □

Let us consider now the special situation where both hypotheses of Theorem 5.1 and the LAN condition are satisfied. In this case, Σ is a non-random matrix and Δ is distributed as $N(0, \Sigma)$. Then (3.33) holds and hence we can have a limit of the posterior with Bayes estimate as a proper centering. By Corollary 5.1, one can now replace the Bayes estimate by the MLE. Moreover, the limit of the posterior is $N_d(0, \Sigma^{-1})$. The same conclusion can be reached by a more direct route. By arguments similar to those in Theorem 5.1, one can show that *MLE is compatible* and then one gets the result by following the proof of Theorem 3.6. Thus we obtain an *in probability* version of the well known Bernstein-von Mises Theorem in a *much more general setting.*

6 Convergence of Experiments, Efficiency and Asymptotic Independence

The set up of IH has some connection with the theory of convergence of statistical experiments. Let

$$\Lambda = \{u \in \mathbb{R}^d : EZ(u) = 1\}. \tag{6.1}$$

Then by Le Cam's First Lemma (see, e.g., Hájek and Šidák (1967, p. 202)), $\{P^n_{\theta_0+\varphi_n u}\}$ is contiguous to $\{P^n_{\theta_0}\}$ whenever $u \in \Lambda$. By Proposition II.2.3 of Millar (1983), the experiment $\{P^n_{\theta_0+\varphi_n u} : u \in \Lambda\}$ converges to the experiment $\{Q_u : u \in \Lambda\}$, where

$$Q_u(A) = \int_A Z(u,\omega)Q_0(d\omega) \tag{6.2}$$

and $(\mathfrak{X}, \mathcal{A}, Q_0)$ is the probability space where the limiting LRP $\{Z(u) : u \in \mathbb{R}^d\}$ is defined. Note that Λ is always nonempty as $0 \in \Lambda$. In those cases where Λ is of a "nice" form, one may obtain some further asymptotic results. For example, in the regular case, $\Lambda = \mathbb{R}^d$ and a convolution theorem and a lower bound to the local asymptotic minimax risk can be obtained. (See Le Cam (1986) or Millar (1983)). If we have a class of discontinuous densities with increasing (or decreasing) support, then $\Lambda = [0, \infty)$ (or $(-\infty, 0]$). In this case, similar results has been obtained in IH (ch. V) using a different approach, and also by Samanta (1988a) using the limiting experiment approach.

A desirable property of estimates is regularity in the sense of Hájek (1970). Under conditions of respective theorems, both *MLE and Bayes estimates are regular* if some further mild conditions are satisfied. Indeed, if the mapping $\theta \longmapsto \mathcal{L}(\widehat{u}(\theta)|\theta)$ is continuous, using uniform convergence on compacts (of course, we have to assume uniform version of Conditions (IH)), for any $u \in \mathbb{R}^d$,

$$\lim_{n\to\infty} \mathcal{L}(\varphi_n^{-1}(\widehat{\theta}_n - \theta_0 - \varphi_n u)|P^n_{\theta_0+\varphi_n u}) = \mathcal{L}(\widehat{u}(\theta_0)|\theta_0). \tag{6.3}$$

Similarly, if the mapping $\theta \longmapsto \mathcal{L}(\tau(\theta)|P_\theta)$ is continuous, then the Bayes estimate is regular. In some situations, the regularity property can be used to derive a convolution theorem in the lines of Theorem III.2.10 of Millar (1983).

Asymptotic efficiency of a sequence of estimators $\widehat{\theta}_n$ was defined in IH as follows.

DEFINITION 6.1. Let l be a loss function. A sequence of estimators $\widehat{\theta}_n$ is said to be *asymptotically efficient in* $K \subset \Theta$ if for any nonempty open set $U \subset K$, the following relation holds:

$$\lim_{n\to\infty}[\inf_{T_n}\sup_{\theta\in U}E_\theta l(\varphi_n^{-1}(T_n-\theta))-\sup_{\theta\in U}E_\theta l(\varphi_n^{-1}(\widehat{\theta}_n-\theta))]=0. \tag{6.4}$$

The estimator $\widehat{\theta}_n$ is called *asymptotically efficient at* $\theta_0\in\Theta$ if

$$\lim_{\delta\to 0}\liminf_{n\to\infty}[\inf_{T_n}\sup_{|\theta-\theta_0|<\delta}E_\theta l(\varphi_n^{-1}(T_n-\theta))-\sup_{|\theta-\theta_0|<\delta}E_\theta l(\varphi_n^{-1}(\widehat{\theta}_n-\theta))]=0. \tag{6.5}$$

Here, the infima in (6.4) and (6.5) are taken over all estimators T_n of θ.

It is to be noted that asymptotic efficiency in K implies asymptotic efficiency at any interior point $\theta_0\in K$. Asymptotic efficiency at all points is also known as *local asymptotic minimaxity*.

As mentioned in IH, in order to prove asymptotic efficiency of an estimator $\widehat{\theta}_n$, it is sufficient to show that uniformly in $U\subset K$, the limit

$$\lim_{n\to\infty}E_\theta l(\varphi_n^{-1}(\widehat{\theta}_n-\theta))=L(\theta) \tag{6.6}$$

exists and then to prove that for any estimator T_n and any nonempty open set $U\subset K$,

$$\liminf_{n\to\infty}\sup_{\theta\in U}E_\theta l(\varphi_n^{-1}(T_n-\theta))\geq\sup_{\theta\in U}L(\theta). \tag{6.7}$$

In view of Theorem I.9.1 and Theorem I.10.2 of IH, a Bayes estimate is asymptotically efficient under a quite general situation stated below.

THEOREM 6.1. (IBRAGIMOV AND HAS'MINSKII). *Assume Conditions* (IH) *hold uniformly in compact subsets of* Θ *and the random function* $\psi_\theta(s)=\int l(s-u)\xi_\theta(u)du$ *attains its minimum at a unique point* $\tau(\theta)$. *Let* $\widetilde{\theta}_n$ *be a Bayes estimate with respect to the loss function* $l\in\mathfrak{L}$ *and prior density* $\pi\in\Pi$. *Then* $\widetilde{\theta}_n$ *is asymptotically efficient in any compact subset* K *of* Θ *provided the mapping* $\theta\mapsto El(\tau(\theta))$ *is continuous.*

Thus in nonregular cases, Bayes estimates deserve special attention since they are always efficient while the MLE is not so.

In some important cases, the parameter θ can be naturally split into two parts, say $\theta=(\vartheta,\sigma)$ and the limiting likelihood ratio at the normalized parameter (u,v) is of the form

$$Z(u,v)=Z^{(1)}(u)Z^{(2)}(v), \tag{6.8}$$

where the processes $Z^{(1)}(u)$ and $Z^{(2)}(v)$ are independent. In this case, we have several important phenomena.

(a) The weak limit of the posterior probability density of (u, v) is of the product form with probability one. Moreover, the two (random) factors are stochastically independent.

(b) The LRP $Z(\cdot)$ satisfies the criterion for posterior limit if and only if both $Z^{(1)}(\cdot)$ and $Z^{(2)}(\cdot)$, viewed as the limiting LRP for u and v respectively, satisfies the criterion. In this case, the (non-random) limit of the posterior probabilities of the normalized and centered paramters are of the product form.

(c) If the loss function l is of the form $l(x, y) = l_1(x) + l_2(y)$ and conditions of Theorem I.10.2 of IH are satisfied, then the asymptotic distribution of the nomalized Bayes estimates of ϑ and σ are of the product type.

(d) If the conditions of Theorem I.10.1 of IH are satisfied, then the asymptotic distribution of the normalized MLE of ϑ and σ are of product type.

(e) If Λ defined in (6.1) is of the form $\Lambda = \Lambda_1 \times \Lambda_2$, then the limiting experiment defined by (6.2) is product of two experiments $\{Q_u^{(1)} : u \in \Lambda_1\}$ and $\{Q_v^{(2)} : v \in \Lambda_2\}$, where $Q_u^{(1)}$ and $Q_v^{(2)}$ are obtained from $Z^{(1)}(\cdot)$ and $Z^{(2)}(\cdot)$ respectively by (6.2).

In view of (a), (b), (c), (d) and (e), in such a situation, we may say that the estimation problems of ϑ and σ are *asymptotically independent.* The cases dealt in Chapters 2 and 3 are of this type.

Chapter 2
Multiparameter Densities with Discontinuities

1 Introduction

We consider i.i.d. observations with a common density on $\mathbb{R}$. The estimation problem for a family of nonregular cases with discontinuous densities involving a single parameter (θ) was considered in Chapter V of IH. This case was also considered by Pflug (1982a) and earlier by Chernoff and Rubin (1956). IH studied the properties of the likelihood ratio process (LRP) and using these obtained the asymptotic properties of the Bayes estimates using their general results. They also studied the properties of the maximum likelihood estimate (MLE) in this context. Such a method of investigation was also used earlier by Rubin (1961) and Prakasa Rao (1968). Rubin (1961), Ermakov (1977) and Pflug (1982b) considered the problem of estimation for multiparameter family of discontinuous densities exhibiting nonregularity in each of the parameters.

There are, however, many important examples of multiparameter family of densities where in addition to the "nonregular"parameter θ, there is also a vector of parameters φ with respect to which the problem is regular (with θ being fixed). For important examples of this kind see, for example, Smith (1985) and Cheng and Iles (1987). Several classes of examples are also presented in Section 2 of this chapter. Smith (1985) and Cheng and Iles (1987) were concerned mainly with the problem of obtaining the asymptotic distribution of MLE or its alternatives. In this chapter, we study the behaviour of Bayes estimates and posterior distributions using a method similar to that in IH.

We first obtain properties of the likelihood ratios. It is shown that the likelihood ratios satisfy certain conditions similar to Conditions (IH). The general results of IH are then used to obtain asymptotic behaviour of the Bayes estimates. The results of Chapter 1 on the asymptotic behaviour of posterior distributions are also used to study the convergence of posterior in this case. As mentioned in Section 1.6, the estimation problems of θ and φ when considered together are asymptotically independent, since in this case, the limiting LRP is a product of two independent processes – one involving the normalized parameter corresponding to θ (i.e., u) only and the other involving only that corresponding to φ (i.e., v).

We derive asymptotic properties of Bayes estimates and posterior distributions using a modified version of Theorem I.10.2 of IH and results of Chapter 1. As explained in Remark 2.1 of Section 2, such a modification is necessary since the conditions of IH in the original form are not satisfied for the usual examples of multiparameter families considered in this chapter. Here the bounding constant appearing in Condition (IH 1) will have exponential growth in the v-components instead of having a polynomial growth. But here we can prove a stronger version of Condition (IH 2) and so can apply the general results in view of the discussion in Section 1.1. Since we consider two different kinds of parameter θ and φ simultaneously, the corresponding normalizing factors being different, we also need a judicious combination of the techniques for the "regular" and "nonregular"problems. Indeed, there are certain difficulties in considering the $(1+d)$-dimensional vector (θ, φ) and proving conditions of IH for the normalized parameter $(u, v) \in \mathbb{R}^{d+1}$. One has to treat u and v in a different manner as in Assertions (I) and (II) of Section 3.

The set up and assumptions have been stated in Section 2. We also present some natural and interesting examples in this section. In Section 3, we study the properties of the LRP. The results of Section 3 are then used in Section 4 to study the asymptotic behaviour of the Bayes estimates and to investigate the existence of a limit of a suitably centered posterior. In Section 5, we consider a specific family of distributions admitting certain local asymptotic expansions of the likelihood ratio and obtain, among other things, a convolution theorem characterizing the possible limiting distributions of a sequence of regular estimators.

2 Set up and Assumptions

In this section, we formally state the assumptions needed for the later developments. We consider a sequence of i.i.d. observations $X_1, X_2, \ldots$ with values in $\mathbb{R}$ and a common distribution $P_{\theta,\varphi}$ depending on unknown parameters $\theta \in \Theta$ and $\varphi \in \Phi$, where Θ and Φ are non-empty open subsets of $\mathbb{R}$ and $\mathbb{R}^d$ $(d \geq 1)$ respectively. We assume that $P_{\theta,\varphi}$ possesses a density $f(x; \theta, \varphi)$ with respect to the Lebesgue measure. We make the following assumptions:

(A1) For any $(\theta, \varphi) \in \Theta \times \Phi$,

$$\inf_{\|(u,v)\|>\varepsilon} \int |f(x; \theta + u, \varphi + v) - f(x; \theta, \varphi)| dx > 0, \quad \varepsilon > 0.$$

(A2) The density $f(\cdot; \theta, \varphi)$ possesses r jumps at $a_1(\theta), \ldots, a_r(\theta)$ depending on θ only. (There may be some other points of discontinuity free from the parameters.)

The functions $a_k(\theta)$, $k=1,\dots,r$, are strictly monotone, continuously differentiable with derivative $a'_k(\theta)$ and $a_i(\theta) < a_j(\theta)$ if $i<j$, $i,j=1,\dots,r$.

(A3) For every $(\theta,\varphi)\in\Theta\times\Phi$, the following limits exist:

$$p_k(\theta,\varphi) := \lim_{x\downarrow a_k(\theta)} f(x;\theta,\varphi), \qquad q_k(\theta,\varphi) := \lim_{x\uparrow a_k(\theta)} f(x;\theta,\varphi)$$

and the convergences are uniform over compact subsets of $\Theta\times\Phi$.

Also the functions $p_k(\theta,\varphi)$, $q_k(\theta,\varphi)$, $k=1,\dots,r$, are continuous and $p_k(\theta,\varphi)\neq q_k(\theta,\varphi)$, $k=1,\dots,r$.

(A4) The function $g(x;\theta,\varphi) := f^{1/2}(x;\theta,\varphi)$ is continuously differentiable in (θ,φ) for a fixed x in each of the regions

$$x<a_1(\theta), \quad a_1(\theta)<x<a_2(\theta),\dots,x>a_r(\theta).$$

Set $g_\theta(x;\theta,\varphi) = (\partial/\partial\theta)g(x;\theta,\varphi)$, $g_{\varphi i}(x;\theta,\varphi) = (\partial/\partial\varphi_i)g(x;\theta,\varphi)$, $i=1,\dots,d$ and $g_\varphi = (g_{\varphi 1},\dots,g_{\varphi d})$.

(A5) The functions defined by

$$\begin{aligned} J(\theta,\varphi) &= \int (g_\theta(x;\theta,\varphi))^2 dx, \\ J_i(\theta,\varphi) &= \int g_\theta(x;\theta,\varphi) g_{\varphi i}(x;\theta,\varphi)dx, \quad i=1,\dots,d, \\ J_{ij}(\theta,\varphi) &= \int g_{\varphi i}(x;\theta,\varphi) g_{\varphi j}(x;\theta,\varphi)dx, \quad i,j=1,\dots,d, \end{aligned}$$

are (finite and) continuous in (θ,φ). Moreover, the matrix $((J_{ij}(\theta,\varphi)))$ is postive definite.

(A6) The functions $a_k(\theta)$, $a'_k(\theta)$, $k=1,\dots,r$, $J(\theta,\varphi)$, $J_i(\theta,\varphi)$, $i=1,\dots,d$, $J_{ij}(\theta,\varphi)$, $i,j=1,\dots,d$, have polynomial majorants in θ and exponential majorants in φ, i.e., the majorants are products of a polynomial in $|\theta|$ and an exponential function in $\|\varphi\|$.

Also, for $k=1,\dots,r$, the function $F_k(\theta,\varphi,R) := \sup\{f(x;\theta,\varphi) : |x-a_k(\theta)|\le R\}$ possesses a polynomial majorant in θ and R and an exponential majorant in φ.

Let $h(x;\theta,\varphi) = \log g(x;\theta,\varphi)$ and h_θ, $h_{\varphi i}$ and h_φ stand for the derivatives of h.

(A7) For any fixed (θ,φ), the functions $h_{\varphi i}(x;\theta,\varphi)$ are bounded in x lying in a deleted neighbourhood of the discontinuity points $a_1(\theta),\dots,a_r(\theta)$.

If $\Theta\times\Phi$ is unbounded, we impose the following additional condition:

(A8) For some $\gamma>0$ and $C>0$, we have

$$\int f^{1/2}(x;\theta,\varphi) f^{1/2}(x;\theta+u,\varphi+v)dx \le C|u|^{-\gamma}\exp[-\gamma\|v\|] \tag{2.1}$$

as $\max\{|u|,\|v\|\}\to\infty$.

REMARK 2.1. In our asymptotic analysis, following the approach of IH, we first obtain certain properties of the LRP which are analogous to Conditions (IH). As noted in Example 2.1 of Chapter 1, one needs to apply a logarithmic transfomation on a scale paramter. In the multiparameter case, when there is a scale (or scale like) parameter σ in addition to the "nonregular" parameter θ, we may again use the transformation $\varphi = \log\sigma$. The functions mentioned in (A6) will then have majorants which are polynomial in θ, but not in φ. However, the majorants will usually be exponential in φ (as in case of $\varphi = \log\sigma$) and one can prove a variant of the conditions of IH as stated in Section 3 (Assertions (I), (II) and (III)). Assertion (I) of Section 3 is a weaker version of the first condition of IH; the polynomial majorant in the condition of IH corresponding to the normalization v of the paramter φ is replaced by an exponential majorant. This is compensated by the strength of Assertion (II). As mentioned in the discussion in Section 1 of Chapter 1, all the asymptotic results of Sections I.5 and I.10 of IH can be proved using Assertions (I), (II) and (III) proceeding along the same lines of IH.

REMARK 2.2. If the functions $a_k(\theta)$, $k = 1, \ldots, r$, are uniformly continuous, the second part of (A6) can be relaxed by replacing it by the assumption that for some $\varepsilon > 0$, the functions $F_k(\theta, \varphi, \varepsilon)$, $k = 1, \ldots, r$, have polynomial majorants in θ and exponential majorants in φ.

Also, if in (A6), all the functions have poynomial majorants in φ as well, we can relax (2.1) to

$$\int f^{1/2}(x;\theta,\varphi)f^{1/2}(x;\theta+u,\varphi+v)dx \le C|u|^{-\gamma}\|v\|^{-\gamma} \tag{2.2}$$

as $\max\{|u|, \|v\|\} \to \infty$ where $\gamma > 0$. In this case, we can prove the conditions of IH in its original form and use the results of IH without any modification. However, such cases are likely to be rare.

The above assumptions will be used to establish asymptotic properties in the subsequent sections. These are the analogoues of the assumptions made by IH (Ch. V) for discontinuous densities with a single parameter θ.

Below we present four important classes of examples which fall in our set up.

EXAMPLE 1. LOCATION-SCALE FAMILY. Let $f(x)$ be a probability density on $\mathbb{R}$ which has a discontinuity at *only one point*, differentiable at all the other points and the right and left limits exist at the point of discontinuity. Without loss of generality, we shall assume that the point of discontinuity is zero. Consider the family of densities

$$f(x;\theta,\varphi) = e^{\varphi} f(e^{\varphi}(x-\theta)), \qquad x,\theta,\varphi \in \mathbb{R}. \tag{2.3}$$

This class is of the type considered in this chapter. It is to be noted that if $f(x)$ has more than one discontinuity, then the discontinuity points depend on φ also and hence f does not fall in our set up. Clearly, the verification of (A1) is equivalent to verifying

$$\inf_{\|(u,v)\|>\varepsilon} \int |e^{v} f(e^{v}(x-u)) - f(x)|dx > 0, \quad \varepsilon > 0. \tag{2.4}$$

If (2.4) is not true, there is an $\varepsilon > 0$ and a sequence $\{(u_n, v_n)\}$ such that $\|(u_n, v_n)\| > \varepsilon$ and

$$\int |e^{v_n} f(e^{v_n}(x-u_n)) - f(x)|dx \to 0 \quad \text{as } n \to \infty. \tag{2.5}$$

If zero is the pth quantile of $f(x)$, then so is u_n for $f(x; u_n, v_n)$. Hence we must have $u_n \to 0$. Now, get a $\lambda \neq 0$ such that $f(\lambda) > 0$. Then by (2.5),

$$\begin{aligned} &\int_{-\infty}^{u_n+\lambda e^{-v_n}} f(x)dx - \int_{-\infty}^{u_n+\lambda e^{-v_n}} e^{v_n} f(e^{v_n}(x-u_n))dx \\ = &\int_{-\infty}^{u_n+\lambda e^{-v_n}} f(x)dx - \int_{-\infty}^{\lambda} f(x)dx \end{aligned}$$

which converges to zero as $n \to \infty$. Clearly, we must have $u_n + \lambda e^{-v_n} \to \lambda$, which together with $u_n \to 0$ implies that $v_n \to 0$. This contradicts the fact that $\|(u_n, v_n)\| > \varepsilon$ and hence (A1) is verified.

In this example, $r = 1$, $a_1(\theta) = \theta$, $a_1'(\theta) = 1$, $p_1(\theta,\varphi) = e^{\varphi} f(0+)$, $q_1(\theta,\varphi) = e^{\varphi} f(0-)$; $p_1(\theta,\varphi)$ and $q_1(\theta,\varphi)$ are continuous, do not depend on θ and have exponential majorants in φ.

It is clear that (A2), (A3) and (A4) are satisfied. To verify (A5), observe that

$$\begin{aligned} J(\theta,\varphi) &= (e^{2\varphi}/4) \int ((f'(x))^2/f(x))dx, \\ J_1(\theta,\varphi) &= (e^{\varphi}/4) \int f'(x)dx - (e^{\varphi}/4) \int x((f'(x))^2/f(x))dx, \\ J_{11}(\theta,\varphi) &= (1/4) \int (1 + xf'(x)/f(x))^2 f(x)dx \end{aligned}$$

and hence it is enough to assume that the integrals in the right side are finite. (In the above expressions, integrals are over all $x \neq 0$ such that $f(x) > 0$.) It may be noted that $\int f'(x)dx$ may not be equal to zero. Finally, (A7) is satisfied if $xf'(x)/f(x)$ is bounded in a deleted neighbourhood of zero. In case $\Theta \times \Phi$ is unbounded, to prove

(A8), we assume that for some $\delta > 0$, (a) $\int|x|^\delta f(x)dx < \infty$ and (b) $f(x) = O(|x|^{-\delta})$ as $|x| \to \infty$. For example, let $\varphi > 0$ and observe that

$$\begin{aligned}
& |\theta|^\gamma e^{\gamma\varphi} \int f^{1/2}(x) f^{1/2}(e^\varphi(x-\theta))dx \\
\leq\ & 2^\gamma \int f^{1/2}(x)(e^\varphi|x-\theta|)^\gamma f^{1/2}(e^\varphi(x-\theta))dx \\
& +2^\gamma \int (\chi\{|x-\theta|>1\} + \chi\{|x-\theta| \leq 1\})(|x|/|x-\theta|)^\gamma f^{1/2}(x) \\
& \qquad \times (e^\varphi|x-\theta|)^\gamma f^{1/2}(e^\varphi(x-\theta))dx \\
\leq\ & 2^\gamma e^{-\varphi/2}[(\int|x|^{2\gamma}f(x)dx)^{1/2} + \int|x|^{2\gamma}f(x)dx + M(\int|x|^{2\gamma}f(x)dx)^{1/2}];
\end{aligned}$$

the last expression is dominated by a constant by (a) if we choose $\gamma = \delta/2$. Here, the second inequality is obtained by applying Cauchy-Schwartz inequality to each term and using (b) for the third term. The case for $\varphi < 0$ is similar.

An important example corresponds to

$$f(x) = \begin{cases} e^{-x}, & \text{if } x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Example 2. Truncation Models. Let $g(x;\varphi)$ be a smooth density, positive on a finite or infinite interval $(a,b) \subset \mathbb{R}$, $\varphi \in \Phi \subset \mathbb{R}^d$. Let

$$f(x;\theta,\varphi) = \begin{cases} g(x;\varphi)/(\int_{a_1(\theta)}^{a_2(\theta)} g(y;\varphi)dy), & \text{if } a_1(\theta) < x < a_2(\theta), \\ 0, & \text{otherwise,} \end{cases} \tag{2.6}$$

where $\theta \in \Theta$, a finite or infinite open interval in $\mathbb{R}$ and $a \leq a_1(\theta) < a_2(\theta) \leq b$. It is possible that $a_1(\theta) \equiv a$ or $a_2(\theta) \equiv b$ (but not both). In this case, $r = 1$ or 2, $q_1(\theta,\varphi) = 0 = p_2(\theta,\varphi)$,

$$\begin{aligned}
p_1(\theta,\varphi) &= g(a_1(\theta);\varphi)/A(\theta,\varphi), \\
q_2(\theta,\varphi) &= g(a_2(\theta);\varphi)/A(\theta,\varphi),
\end{aligned}$$

where $A(\theta,\varphi) = \int_{a_1(\theta)}^{a_2(\theta)} g(x;\varphi)dx$. We assume that $a_1(\theta)$ and $a_2(\theta)$ satisfy (A2). One can find suitable conditions on the densities $g(x;\varphi)$ under which (A1) is saisfied. By continuity of $g(x;\varphi)$ in φ, $p_1(\theta,\varphi)$, $p_2(\theta,\varphi)$, $q_1(\theta,\varphi)$ and $q_2(\theta,\varphi)$ are continuous. Further,

$$\begin{aligned}
J(\theta,\varphi) &= (a_1'(\theta)g(a_1(\theta),\varphi) - a_2'(\theta)g(a_2(\theta),\varphi))^2/4A^2(\theta,\varphi), \\
J_i(\theta,\varphi) &= 0,
\end{aligned}$$

$$\begin{aligned} J_{ij}(\theta,\varphi) &= [A(\theta,\varphi)\int_{a_1(\theta)}^{a_2(\theta)}(g_{\varphi i}(x;\varphi)g_{\varphi j}(x;\varphi)/g(x;\theta,\varphi))dx \\ &\quad -\int_{a_1(\theta)}^{a_2(\theta)} g_{\varphi i}(x;\varphi)dx\int_{a_1(\theta)}^{a_2(\theta)} g_{\varphi j}(x;\varphi)dx]/4A^2(\theta,\varphi) \end{aligned}$$

for $i,j = 1,\ldots,d$. Under appropriate conditions on the density $g(x;\varphi)$ and the truncation points $a_1(\theta)$ and $a_2(\theta)$, (A6) is satisfied. Further, (A7) will be satisfied if $g_{\varphi i}(x;\varphi)/g(x;\varphi)$ are bounded in x. For a particular important case, consider the density

$$g(x;\varphi) = \begin{cases} (e^{\alpha\varphi}/\Gamma(\alpha))\exp[-xe^{\varphi}]x^{\alpha-1}, & \text{if } x \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

where $-\infty < \varphi < \infty$ is a parameter and $\alpha > 0$ is a known constant. Let $a_1(\theta) = \theta$ and $a_2(\theta) = \infty$, $\theta > 0$. Then

$$\begin{aligned} A(\theta,\varphi) &= \int_{\theta}^{\infty}(e^{\alpha\varphi}/\Gamma(\alpha))\exp[-xe^{\varphi}]x^{\alpha-1}dx \\ &= \int_{\theta e^{\varphi}}^{\infty}(1/\Gamma(\alpha))\exp[-x]x^{\alpha-1}dx \\ &\sim (1/\Gamma(\alpha))\exp[\alpha\varphi-\theta e^{\varphi}]\theta^{\alpha-1} \end{aligned}$$

if $\theta e^{\varphi} \to \infty$. (see, e.g., Johnson and Kotz (1970), p. 179). Using this fact, it can be seen that (A6) is satisfied. It is obvious that (A7) holds. A little calculation using the same fact mentioned above guarantees the validity of (A8).

One can also verify the validity of the assumptions in case α is also unknown, i.e., the parameter φ is replaced by (φ,α), if α belongs to a compact set. The case with a Weibull density can also be treated similarly.

EXAMPLE 3. CHANGE POINT MODELS. Let $g^{(1)}(x;\varphi)$ and $g^{(2)}(x;\varphi)$ be two smooth families of densities on a finite or infinite open interval $(a,b) \subset \mathbb{R}$ where $\varphi \in \Phi$, an open subset of $\mathbb{R}^d$. Further, for any $\theta \in (a,b)$ and $\varphi \in \Phi$,

$$g^{(1)}(\theta;\varphi) \neq \lambda(\theta,\varphi)g^{(2)}(\theta;\varphi) \tag{2.7}$$

where $\lambda(\theta,\varphi) = \int_\theta^b g^{(1)}(x;\varphi)dx/\int_\theta^b g^{(2)}(x;\varphi)dx$. Let

$$f(x;\theta,\varphi) = \begin{cases} g^{(1)}(x;\varphi), & \text{if } a < x < \theta, \\ \lambda(\theta,\varphi)g^{(2)}(x;\varphi), & \text{if } \theta \leq x < b. \end{cases} \tag{2.8}$$

The parameter θ is called a change point. This example falls in our set up with $r = 1$, $a_1(\theta) = \theta$, $q_1(\theta,\varphi) = g^{(1)}(\theta,\varphi)$ and $p_1(\theta,\varphi) = \lambda(\theta,\varphi)g^{(2)}(\theta;\varphi)$. The quantities in (A5) can also be computed, but they are a little messy. We consider a very important special case with

$$f(x;\theta,\varphi)=\begin{cases}\exp[\alpha-e^{\alpha}x], & \text{if } 0<x<\theta,\\ \exp[\beta-\theta(e^{\alpha}-e^{\beta})-e^{\beta}x], & \text{if } \theta\le x<\infty,\end{cases}$$

where $0<\theta<\infty$, $-\infty<\beta<\alpha<\infty$ are unknown parameters. This model has been traditionally used in reliability problems using "burn-in" technique; see Nguyen, Rogers and Walker (1984) and Basu, Ghosh and Joshi (1988). Also $q_1(\theta,\alpha,\beta)=\exp[\alpha-e^{\alpha}\theta]$, $p_1(\theta,\alpha,\beta)=\exp[\beta-(e^{\alpha}-e^{\beta})-e^{\beta}\theta]$ are continuous in (θ,α,β) and have polynomial majorants in θ and exponential majorants in (α,β). Further,

$$\begin{aligned}
J(\theta,\alpha,\beta) &= (e^{\alpha}-e^{\beta})^2\exp[-\theta e^{\alpha}]/4,\\
J_1(\theta,\alpha,\beta) &= \theta e^{\alpha}(e^{\alpha}-e^{\beta})\exp[-\theta e^{\alpha}]/4,\\
J_2(\theta,\alpha,\beta) &= -(e^{\alpha}-e^{\beta})\{1+(\theta e^{\beta}-1)\exp[-\theta e^{\beta}]\}\exp[-\theta(e^{\alpha}-e^{\beta})]/4,\\
J_{11}(\theta,\alpha,\beta) &= 1-\exp[-\theta e^{\alpha}](1+2\theta e^{\alpha})/4,\\
J_{12}(\theta,\alpha,\beta) &= -\theta e^{\alpha}\{1+(\theta e^{\beta}-1)\exp[-\theta e^{\beta}]\}\exp[-\theta(e^{\alpha}-e^{\beta})]/4,\\
J_{22}(\theta,\alpha,\beta) &= (1+\theta^2e^{2\beta})\exp[-\theta e^{\beta}-\theta^2(e^{a}-e^{\beta})^2]/4,
\end{aligned}$$

which are continuous and satisfy the required growth rate conditions. Also, it is not difficult to observe that (A1) and (A7) are satisfied. However, in this case, (A8) fails to hold. For if α,β remain fixed and $u\to\infty$, then $f(x;\theta,\alpha,\beta)$ and $f(x;\theta+u,\alpha,\beta)$ have a common portion; consequently their affinity cannot go to zero. So we have to restrict our attention to the case where Θ is bounded in which case (A8) is satisfied.

EXAMPLE 4. LOCATION SHIFT OF REGULAR FAMILY. Suppose $\{g(x;\varphi):\varphi\in\Phi\}$ is a family of probability densities on $\mathbb{R}$ which is smooth with respect to φ, but has at least one discontinuity as a function of x and the jump points are free from φ. Consider the family of densities defined by

$$f(x;\theta,\varphi)=g(x-\theta;\varphi),\qquad \theta\in\mathbb{R},\ \varphi\in\Phi. \tag{2.9}$$

This family falls within the scope of our set up and it is not difficult to give sufficient conditions on $\{g(x;\varphi):\varphi\in\Phi\}$ so that (A1)–(A8) are satisfied, but we omit this for brevity. Example 1 is a special case of this when the regular family is also a scale family. For an example which is not included in any of the previous ones, consider

$$g(x;\alpha,\beta)=\begin{cases}(1/2)e^{\alpha}\exp[-e^{\alpha}x], & \text{if } x\ge 0,\\ (1/2)e^{\beta}\exp[e^{\beta}x], & \text{if } x\le 0,\end{cases} \tag{2.10}$$

where $-\infty<\beta<\alpha<\infty$.

3 Properties of Likelihood Ratio Process

The asymptotic properties of the usual estimators depend on the behaviour of the local likelihood ratios. For a fixed $(\theta_0, \varphi_0) \in \Theta \times \Phi$, the (local) *likelihood ratio process* is defined by

$$Z_{n,\theta_0,\varphi_0}(u,v) = Z_n(u,v) = \prod_{i=1}^{n} \frac{f(X_i;\theta_0 + u/n, \varphi_0 + n^{-1/2}v)}{f(X_i, \theta_0, \varphi_0)},$$

where $u \in U_n := n(\Theta - \theta_0)$ and $v \in V_n := n^{1/2}(\Phi - \varphi_0)$. The appropriate normalizing constant is $n^{-1/2}$ when θ is known and n^{-1} when φ is known. Hence it is natural to consider the above normalizer. Below, we shall omit the argument (θ_0, φ_0) whenever there is no source of confusion. Also, all the probability statements made below refer to the parameter point (θ_0, φ_0).

All the asymptotic results will follow from the following three properties of the LRP:

(I) There exist $a_1, a_2, B > 0$ such that

$$\begin{aligned} &E|Z_n^{1/2}(u_1,v_1) - Z_n^{1/2}(u_2,v_2)|^2 \\ \leq\ & B(1 + R_1^{a_1})\exp(a_2 R_2)(|u_1 - u_2| + \|v_1 - v_2\|^2) \end{aligned}$$

for all $u_1, u_2 \in [-R_1, R_1]$, $v_1, v_2 \in \{v : \|v\| \leq R_2\}$.

(II) For all $u \in U_n$ and $v \in V_n$,

$$EZ_n^{1/2}(u,v) \leq \exp[-g_n(|u|, \|v\|)]$$

where $\{g_n(\cdot,\cdot)\}$ is a sequence of functions from $[0,\infty) \times [0,\infty)$ into $[0,\infty)$ satisfying the following properties:

(a) For each $n \geq 1$, $g_n(x,y)$ is increasing to infinity in each component.

(b) For any $N_1, N_2 \geq 0$, we have

$$\lim_{\substack{n\to\infty \\ \max(x,y)\to\infty}} x^{N_1} e^{N_2 y} \exp[-g_n(x,y)] = 0.$$

(III) The finite dimensional distributions of the process $Z_n(u,v)$ converge to those of a process $Z(u,v)$. (The expression for $Z(u,v)$ is given in Theorem 2.2.)

REMARK 3.1. Assertions (I), (II) and (III) above can also be shown to hold uniformly in (θ_0, φ_0) belonging to compact subsets of $\Theta \times \Phi$ under appropriate uniform versions of the Assumptions (A1)–(A8) of Section 2. Consequently, one can obtain the uniform version of Theorem 4.1.

In this section, we prove the properties (I), (II) and (III). We first prove a lemma which will be used below. Set $G_\theta(u) = \bigcup_{k=1}^{r}[a_k(\theta), a_k(\theta+u)]$ and abbreviate $G_{\theta_0}(u)$ as $G(u)$.

LEMMA 3.1. *Under Assumptions* (A2), (A4) *and* (A5), *we have as* $|u|^2+\|v\|^2 \to 0$,

$$\int |g(x;\theta_0+u,\varphi_0+v) - g(x;\theta_0,\varphi_0) - ug_\theta(x;\theta_0,\varphi_0) - v' g_\varphi(x;\theta_0,\varphi_0)|^2 \times \chi(x \notin G(u))dx = o(|u|^2+\|v\|^2).$$

PROOF. It is enough to prove the result along a sequence $\{(u_m, v_m)\}$ such that $|u_m|^2 + \|v_m\|^2 \to 0$. Without loss of generality, we assume that

$$\begin{aligned} &u_m/(|u_m|^2+\|v_m\|^2)^{1/2} \to \lambda, \\ &v_{mi}/(|u_m|^2+\|v_m\|^2)^{1/2} \to \lambda_i, \quad i=1,\dots,d, \end{aligned} \tag{3.1}$$

where $\lambda^2+\lambda_1^2+\cdots+\lambda_d^2 = 1$. Hence it is enough to show that as $m \to \infty$

$$\int \Big| \frac{g(x;\theta_0+u_m,\varphi_0+v_m) - g(x;\theta_0,\varphi_0)}{(|u_m|^2+\|v_m\|^2)^{1/2}} - \lambda g_\theta(x;\theta_0,\varphi_0) - \sum_{i=1}^{d} \lambda_i g_{\varphi i}(x;\theta_0,\varphi_0)\Big|^2 \chi(x \notin G(u_m))dx = o(1). \tag{3.2}$$

Now

$$\begin{aligned} &(|u_m|^2+\|v_m\|^2)^{-1} \int |g(x;\theta_0+u_m,\varphi_0+v_m) - g(x;\theta_0,\varphi_0)|^2 \chi(x \notin G(u_m))dx \\ = \;& (|u_m|^2+\|v_m\|^2)^{-1} \int \Big| \int_0^1 u_m g_\theta(x;\theta_0+u_m t,\varphi_0+v_m t) \\ &\quad + v_m' g_\varphi(x;\theta_0+u_m t,\varphi_0+v_m t)dt\Big|^2 \chi(x \notin G(u_m))dx \\ \le \;& (|u_m|^2+\|v_m\|^2)^{-1} \int_0^1 \int |u_m g_\theta(x;\theta_0+u_m t,\varphi_0+v_m t) \\ &\quad + v_m' g_\varphi(x;\theta_0+u_m t,\varphi_0+v_m t)|^2 dx dt \\ \le \;& (|u_m|^2+\|v_m\|^2)^{-1} \int_0^1 [|u_m|^2 J(\theta_0+u_m t,\varphi_0+v_m t) \\ &\quad + 2u_m \sum_{i=1}^{d} v_{mi} J_i(\theta_0+u_m t,\varphi_0+v_m t) \\ &\quad + \sum_{i=1}^{d}\sum_{j=1}^{d} v_{mi}v_{mj} J_{ij}(\theta_0+u_m t,\varphi_0+v_m t)]dt \end{aligned} \tag{3.3}$$

which, by (A5), converges to

$$\lambda^2 J(\theta_0,\varphi_0)+2\lambda\sum_{i=1}^{d}\lambda_i J_i(\theta_0,\varphi_0)+\sum_{i=1}^{d}\sum_{j=1}^{d}\lambda_i\lambda_j J_{ij}(\theta_0,\varphi_0). \tag{3.4}$$

Now by an application of a well known convergence theorem (see, e.g., Hájek and Šidák (1967), p. 154), the proof is complete. □

We now prove Assertion (I). Assume Conditions (A2)–(A6) and denote the Hellinger distance between $P_{\theta,\varphi}$ and $P_{\theta+u,\varphi+v}$ by $r_2((\theta,\varphi),(\theta+u,\varphi+v))$. As in IH (p. 54), we have

$$\begin{aligned} & E|Z_n^{1/2}(u_1,v_1)-Z_n^{1/2}(u_2,v_2)|^2 \\ \leq\ & nr_2^2((\theta_0+u_1/n,\varphi_0+n^{-1/2}v_1),(\theta_0+u_2/n,\varphi_0+n^{-1/2}v_2)). \end{aligned} \tag{3.5}$$

Now

$$\begin{aligned} & r_2^2((\theta,\varphi),(\theta+u,\varphi+v)) \\ =\ & \int|g(x;\theta+u,\varphi+v)-g(x;\theta,\varphi)|^2dx \\ \leq\ & \sum_{k=1}^{r}\left|\int_{a_k(\theta)}^{a_k(\theta+u)}|g(x;\theta+u,\varphi+v)-g(x;\theta,\varphi)|^2dx\right| \\ & +\int|g(x;\theta+u,\varphi+v)-g(x;\theta,\varphi)|^2\chi(x\notin G_\theta(u))dx. \end{aligned} \tag{3.6}$$

The first term is bounded above by

$$\begin{aligned} & \sum_{k=1}^{r}\left|\int_{a_k(\theta)}^{a_k(\theta+u)}|f(x;\theta+u,\varphi+v)-f(x;\theta,\varphi)|dx\right| \\ \leq\ & |u|\sum_{k=1}^{r}\sup_{\xi\in(\theta,\theta+u)}|a_k'(\xi)|(F_k(\theta+u,\varphi+v,|a_k(\theta+u)-a_k(\theta)|) \\ & +F_k(\theta,\varphi,|a_k(\theta+u)-a_k(\theta)|)), \end{aligned} \tag{3.7}$$

where F_k's are as in (A6). To get a bound for the second term, note that, as argued in (3.3), we have

$$\begin{aligned} & \int|g(x;\theta+u,\varphi+v)-g(x;\theta,\varphi)|^2\chi(x\notin G_\theta(u))dx \\ \leq\ & 2[|u|^2\int_0^1 J(\theta+ut,\varphi+vt)+\sum_{i=1}^{d}\sum_{j=1}^{d}v_iv_j\int_0^1 J_{ij}(\theta+ut,\varphi+vt)dt] \\ \leq\ & 2[|u|^2\sup_{(\theta^*,\varphi^*)\in L}J(\theta^*,\varphi^*)+d\|v\|^2\max_{i,j}\sup_{(\theta^*,\varphi^*)\in L}J_{ij}(\theta^*,\varphi^*)] \end{aligned} \tag{3.8}$$

where L stands for the line segment joining (θ, φ) and $(\theta+u, \varphi+v)$. Now relations (3.5)-(3.8) in conjugation with (A6) imply Assertion (I).

Let us now prove Assertion (II). We note that, as in IH (p. 53), we have

$$EZ_n^{1/2}(u,v) \le \exp[-(n/2)r_2^2((\theta_0,\varphi_0),(\theta_0+u/n,\varphi_0+n^{-1/2}v))]. \tag{3.9}$$

We shall show that for all (u,v) satisfying $(\theta_0+u, \varphi_0+v) \in \Theta \times \Phi$,

$$r_2^2((\theta_0,\varphi_0),(\theta_0+u,\varphi_0+v)) \ge a(|u|+\|v\|^2)/(1+|u|+\|v\|^2) \tag{3.10}$$

where $a = a(\theta_0, \varphi_0) > 0$. Now

$$\begin{aligned}
& r_2^2((\theta_0,\varphi_0),(\theta_0+u,\varphi_0+v)) \\
= & \sum_{k=1}^{r} \int |g(x;\theta_0+u,\varphi_0+v) - g(x;\theta_0,\varphi_0)|^2 \chi(x \in [a_k(\theta_0), a_k(\theta_0+u)])dx \\
& + \int |g(x;\theta_0+u,\varphi_0+v) - g(x;\theta_0,\varphi_0)|^2 \chi(x \notin G(u))dx.
\end{aligned} \tag{3.11}$$

For $|u|$ sufficiently small, we have by (A2) and (A3),

$$\begin{aligned}
& \sum_{k=1}^{r} \int |g(x;\theta_0+u,\varphi_0+v) - g(x;\theta_0,\varphi_0)|^2 \chi(x \in [a_k(\theta_0), a_k(\theta_0+u)])dx \\
\ge & (|u|/2)\sum_{k=1}^{r}(q_k^{1/2}(\theta_0,\varphi_0) - p_k^{1/2}(\theta_0,\varphi_0))^2 |a_k'(\theta_0)| \\
= & c_1|u| \quad \text{(say)}.
\end{aligned} \tag{3.12}$$

For the second term in (3.11), we use Lemma 3.1 and note that

$$\begin{aligned}
& \int |g(x;\theta_0+u,\varphi_0+v) - g(x;\theta_0,\varphi_0)|^2 \chi(x \notin G(u))dx \\
= & u^2 J(\theta_0,\varphi_0) + 2u\sum_{i=1}^{d} v_i J_i(\theta_0,\varphi_0) + \sum_{i=1}^{d}\sum_{j=1}^{d} v_i v_j J_{ij}(\theta_0,\varphi_0) + o(|u|^2+\|v\|^2) \\
\ge & 2u\sum_{i=1}^{d} v_i J_i(\theta_0,\varphi_0) + \sum_{i=1}^{d}\sum_{j=1}^{d} v_i v_j J_{ij}(\theta_0,\varphi_0) + o(|u|^2+\|v\|^2) \\
\ge & -c_1|u|/2 + \eta\|v\|^2
\end{aligned} \tag{3.13}$$

for sufficiently small $|u|$ and $\|v\|$ and some $\eta > 0$. Now by (A1), there exists $a > 0$ such that (3.10) holds for all u, v. From (3.9), we thus have

$$EZ_n^{1/2}(u,v) \le \exp[-g_n(|u|, \|v\|)],$$

where $g_n(x,y) = (a/2)(x+y^2)/(1+(x+y^2)/n)$. Then the sequence of functions $\{g_n(x,y)\}$ satisfies conditions (a) and (b) of Assertion (II) if $\Theta \times \Phi$ is bounded (see Lemma I.5.3 of IH). When $\Theta \times \Phi$ is unbounded, we further assume (A8) and the result follows as in Lemma I.5.4 of IH.

Before proving Assertion (III), we introduce the following notations:

$$\begin{aligned}
\Gamma = \quad & \Gamma(\theta_0) = & & \{k : p_k(\theta_0,\varphi_0) > 0, q_k(\theta_0,\varphi_0) > 0\},\\
\Gamma^+ = \quad & \Gamma^+(\theta_0) = & & \{k : q_k(\theta_0,\varphi_0) = 0, a_k'(\theta_0) > 0\}\\
& & & \bigcup\{k : p_k(\theta_0,\varphi_0) = 0, a_k'(\theta_0) < 0\},\\
\Gamma^- = \quad & \Gamma^-(\theta_0) = & & \{k : p_k(\theta_0,\varphi_0) = 0, a_k'(\theta_0) > 0\}\\
& & & \bigcup\{k : q_k(\theta_0,\varphi_0) = 0, a_k'(\theta_0) < 0\},\\
\sigma_n^+ = \quad & \sigma_n^+(\theta_0) = & & \inf\{u > 0 : X_i \in \bigcup_{k\in\Gamma^+} [a_k(\theta_0), a_k(\theta_0 + u/n)]\\
& & & \qquad \text{for some } i = 1,\ldots,n\},\\
\sigma_n^- = \quad & \sigma_n^-(\theta_0) = & & \sup\{u < 0 : X_i \in \bigcup_{k\in\Gamma^-} [a_k(\theta_0), a_k(\theta_0 + u/n)]\\
& & & \qquad \text{for some } i = 1,\ldots,n\}.
\end{aligned}$$

We define $\sigma_n^+ = +\infty$ if $\Gamma^+ = \emptyset$ and $\sigma_n^- = -\infty$ if $\Gamma^- = \emptyset$. We also set

$$\widetilde{Z}_n(u,v) = \begin{cases} 0, & \text{if } u \le \sigma_n^- \text{ or } u \ge \sigma_n^+,\\ \exp\{uc + \sum_{k\in\Gamma} \operatorname{sign}(ua_k') \log(q_k/p_k) & \\ \quad \times \sum_{i=1}^n \chi(X_i \in [a_k, a_k + (u/n)a_k']) & \\ \quad + v'\Delta_n - (1/2)v'Iv\}, & \text{if } \sigma_n^- < u < \sigma_n^+, \end{cases}$$

where $a_k = a_k(\theta_0)$, $a_k' = a_k'(\theta_0)$, $p_k = p_k(\theta_0,\varphi_0)$, $q_k = q_k(\theta_0,\varphi_0)$, $k = 1,\ldots,r$, $c = c(\theta_0,\varphi_0) = \sum_{k=1}^r (p_k - q_k)a_k'$, $\Delta_n = \Delta_n(\theta_0,\varphi_0) = 2n^{-1/2}\sum_{i=1}^n h_\varphi(X_i;\theta_0,\varphi_0)$ and $I = I(\theta_0,\varphi_0) = ((4J_{ij}(\theta_0,\varphi_0)))_{d\times d}$. The following theorem gives an approximation to the LRP $Z_n(u,v)$:

THEOREM 3.1. *Under Assumptions* (A1)-(A5), *the following relations hold uniformly in* (u,v) *belonging to compact subsets of* $\mathbb{R}^{1+d}$:

$$\begin{aligned}
\text{(a)} \quad Z_n(u,v) \quad &= \quad \exp\{uc + \sum_{k=1}^r \operatorname{sign}(ua_k')\log(q_k/p_k)\\
&\qquad \times \sum_{i=1}^n \chi(X_i \in [a_k, a_k + (u/n)a_k'])\\
&\qquad + v'\Delta_n - (1/2)v'Iv + o_p(1)\} + o_p(1).\\
\text{(b)} \quad Z_n(u,v) \quad &= \quad \widetilde{Z}_n(u,v) + o_p(1).
\end{aligned}$$

Decompose the likelihood ratio as

$$\log Z_n(u,v) = I_{1n}(u,v) + I_{2n}(u,v)$$

where

$$\begin{aligned} I_{1n}(u,v) &= \sum_{i=1}^{n} \log(\frac{f(X_i;\theta_0+u/n,\varphi_0+n^{-1/2}v)}{f(X_i;\theta_0,\varphi_0)})\chi(X_i \notin G(u/n)) \\ I_{2n}(u,v) &= \sum_{i=1}^{n} \log(\frac{f(X_i;\theta_0+u/n,\varphi_0+n^{-1/2}v)}{f(X_i;\theta_0,\varphi_0)})\chi(X_i \in G(u/n)). \end{aligned}$$

Let

$$\begin{aligned} Z_{1n}(u,v) &= \exp[I_{1n}(u,v)], \\ Z_{2n}(u,v) &= \exp[I_{2n}(u,v)], \\ \widetilde{Z}_{2n}(u,v) &= \begin{cases} \exp\{\sum_{k\in\Gamma} \operatorname{sign}(ua_k') \log(q_k/p_k) \\ \qquad \times \sum_{i=1}^{n} \chi(X_i \in [a_k, a_k+(u/n)a_k'])\} & \text{if } \sigma_n^- < u < \sigma_n^+, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

The following lemmas will be used to prove Theorem 3.1.

LEMMA 3.2. *As $n \to \infty$, uniformly on $\{(u,v) : |u| \le H, \|v\| \le H\}$, $H < \infty$, we have*

$$Z_{2n}(u,v) - \widetilde{Z}_{2n}(u,v) = o_p(1).$$

LEMMA 3.3. *As $n \to \infty$, uniformly on $\{(u,v) : |u| \le H, \|v\| \le H\}$, $H < \infty$, we have*

$$I_{1n}(u,v) = uc + v'\Delta_n - (1/2)v'Iv + o_p(1).$$

Proof of Lemma 3.2 is similar to that of Lemma V.2.1 of IH and hence omitted. To prove Lemma 3.3, we need the following auxiliary result:

For $i = 1, \ldots, n$, set

$$\eta_{in} = \begin{cases} (\frac{g(X_i;\theta_0+u/n,\varphi_0+n^{-1/2}v)}{g(X_i;\theta_0,\varphi_0)} - 1)\chi(X_i \notin G(u/n)), & \text{if } g(X_i;\theta_0,\varphi_0) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

LEMMA 3.4. *As $n \to \infty$, we have*

(a) $E\eta_{1n}^2 - E((u/n)h_\theta(X_1;\theta_0,\varphi_0) + n^{-1/2}v'h_\varphi(X_1;\theta_0,\varphi_0))^2 = o(1/n)$,

(b) $E|\eta_{1n}^2 - ((u/n)h_\theta(X_1;\theta_0,\varphi_0) + n^{-1/2}v'h_\varphi(X_1;\theta_0,\varphi_0))^2 \chi(X_1 \notin G(u/n))| = o(1/n)$,

(c) $P\{|\eta_{1n}| > \varepsilon\} = o(1/n)$, $\varepsilon > 0$,

(d) $E\eta_{1n} - cu/(2n) + (1/8n)v'Iv = o(1/n)$

and for any $H < \infty$, *the convergences are uniform on* $\{(u,v) : |u| \leq H,\ \|v\| \leq H\}$.

PROOF. For simplicity, we write down the proof for a fixed (u, v); the uniformity will be clear from the arguments. In view of (A5), (a) follows from (b). Also from Lemma 3.1, we have

$$E(|\eta_{1n} - ((u/n)h_\theta(X_1;\theta_0,\varphi_0) + n^{-1/2}v'h_\varphi(X_1;\theta_0,\varphi_0))|^2 \times\chi(X_1 \notin G(u/n))) = o(1/n). \tag{3.14}$$

Therefore, by (A5) and Cauchy-Schwartz inequality, (b) follows. Now

$$\begin{aligned} P\{|\eta_{1n}| > \varepsilon\} &\leq P\{|\eta_{1n} - ((u/n)h_\theta(X_1;\theta_0,\varphi_0) + n^{-1/2}v'h_\varphi(X_1;\theta_0,\varphi_0))| \\ &\qquad \times\chi(X_1 \notin G(u/n)) > \varepsilon/2\} \\ &\quad + P\{|(u/n)h_\theta(X_1;\theta_0,\varphi_0) + n^{-1/2}v'h_\varphi(X_1;\theta_0,\varphi_0))| > \varepsilon/2\}. \end{aligned}$$

By an application of Chebyshev's inequality and using (3.14), the first term is $o(1/n)$. Using the elementary inequality

$$P\{|Z| > a\} \leq a^{-1}E[|Z|\chi(|Z| > a)]$$

for a random variable Z, the second term can be shown to be $o(1/n)$ by virtue of (A5). Hence (c) is proved.

It remains to prove (d). We note that

$$\int(f(x;\theta_0+u/n,\varphi_0+n^{-1/2}v)-f(x;\theta_0,\varphi_0))\chi(x \notin G(u/n))dx = cu/n+o(1/n) \tag{3.15}$$

by (A2) and (A3). Thus by Lemma 3.1 and (A5),

$$\begin{aligned} E\eta_{1n} &= \int(g(x;\theta_0 + u/n, \varphi_0 + n^{-1/2}v) - g(x;\theta_0,\varphi_0))g(x;\theta_0,\varphi_0) \\ &\qquad \chi(x \notin G(u/n))dx \\ &= -(1/2)\int(g(x;\theta_0 + u/n, \varphi_0 + n^{-1/2}v) - g(x;\theta_0,\varphi_0))^2 \\ &\qquad \chi(x \notin G(u/n))dx + cu/(2n) + o(1/n) \\ &= -(1/2)\int((u/n)g_\theta(x;\theta_0,\varphi_0) + n^{-1/2}g_\varphi(x;\theta_0,\varphi_0))^2\chi(x \notin G(u/n))dx \\ &\quad +cu/(2n) + o(1/n) \\ &= cu/(2n) - (1/2n)\int(v'g_\varphi(x;\theta_0,\varphi_0))^2dx + o(1/n) \\ &= cu/(2n) - (1/8n)v'Iv + o(1/n) \end{aligned} \tag{3.16}$$

proving (d). $\square$

Proof of Lemma 3.3. From the definition of η_{in}, $i = 1, \ldots, n$, we have $I_{1n} = 2\sum_{i=1}^n \log(1+\eta_{in})$. Expanding the logarithm in a Taylor's series on the set $A_n := \{\max_{1\le i\le n} |\eta_{in}| \le \varepsilon_0\}$ for suitably chosen ε_0, we have on A_n,

$$I_{1n} = 2\sum_{i=1}^n \eta_{in} - \sum_{i=1}^n \eta_{in}^2 + 2\sum_{i=1}^n \alpha_{in}\eta_{in}^3,$$

where α_{in}, $i = 1, \ldots, n$ are random variables satisfying $|\alpha_{in}| < 1$, $i = 1, \ldots, n$. The result now follows from the following assertions:

(1) $\max_{1\le i\le n} |\eta_{in}| = o_p(1)$,
(2) $\sum_{i=1}^n \eta_{in}^2 - (1/4)v'Iv = o_p(1)$,
(3) $\sum_{i=1}^n \eta_{in}^3 = o_p(1)$,
(4) $\sum_{i=1}^n \eta_{in} - (1/2)uc - (1/2)v'\Delta_n + (1/8)v'Iv = o_p(1)$.

Assertion (1) immediately follows from part (c) of Lemma 3.4.

From part (b) of Lemma 3.4, we have

$$\begin{aligned}\textstyle\sum_{i=1}^n \eta_{in}^2 - \sum_{i=1}^n ((u/n)h_\theta(X_i;\theta_0,\varphi_0) + n^{-1/2}v'h_\varphi(X_i;\theta_0,\varphi_0))^2 \\ \times\chi(X_i \notin G(u/n)) = o_p(1).\end{aligned} \tag{3.17}$$

Also by repeated appplication of (A5),

$$\begin{aligned}&\sum_{i=1}^n ((u/n)h_\theta(X_i;\theta_0,\varphi_0) + n^{-1/2}v'h_\varphi(X_i;\theta_0,\varphi_0))^2\chi(X_i \notin G(u/n)) \\ = \;& (1/n)\sum_{i=1}^n (v'h_\varphi(X_i;\theta_0,\varphi_0))^2\chi(X_i \notin G(u/n)) + o_p(1) \\ = \;& (1/n)\sum_{i=1}^n (v'h_\varphi(X_i;\theta_0,\varphi_0))^2 + o_p(1).\end{aligned} \tag{3.18}$$

Assertion (2) now follows from (3.17), (3.18) and the Law of Large Numbers.

Assertion (3) is immediate from (1) and (2) by virtue of the inequality

$$\sum_{i=1}^n |\eta_{in}|^3 \le (\max_{1\le i\le n} |\eta_{in}|)\sum_{i=1}^n \eta_{in}^2.$$

We now prove (4). Put

$$T_n = \sum_{i=1}^n \eta_{in} - \sum_{i=1}^n ((u/n)h_\theta(X_i;\theta_0,\varphi_0) + n^{-1/2}v'h_\varphi(X_i;\theta_0,\varphi_0))\chi(X_i \notin G(u/n)).$$

We observe that

$$T_n - ET_n \xrightarrow{P} 0. \tag{3.19}$$

Indeed, for $\varepsilon > 0$,

$$\begin{aligned}
& P\{|T_n - ET_n| > \varepsilon\} \\
\leq\ & \varepsilon^{-2}\mathrm{var}(T_n) \\
\leq\ & n\varepsilon^{-2}E(|\eta_{1n} - ((u/n)h_\theta(X_i;\theta_0,\varphi_0) + n^{-1/2}v'h_\varphi(X_i;\theta_0,\varphi_0))|^2 \\
& \qquad \times\chi(X_1 \notin G(u/n)))
\end{aligned}$$

which converges to zero by (3.14). We next show that

$$E[(u/n)\sum_{i=1}^{n} h_\theta(X_i;\theta_0,\varphi_0)\chi(X_i \in G(u/n))]^2 = o(1) \tag{3.20}$$

and

$$E[n^{-1/2}v'\sum_{i=1}^{n} h_\varphi(X_i;\theta_0,\varphi_0)\chi(X_i \in G(u/n))]^2 = o(1). \tag{3.21}$$

Equation (3.20) follows from (A5) by virtue of the Dominated Convergence Theorem (DCT) and Minkowski's inequality. Now,

$$\begin{aligned}
& E[n^{-1/2}v'\sum_{i=1}^{n} h_\varphi(X_i;\theta_0,\varphi_0)\chi(X_i \in G(u/n))]^2 \\
=\ & E[(v'h_\varphi(X_1;\theta_0,\varphi_0)^2\chi(X_1 \in G(u/n))] \\
& +(n-1)(E[v'h_\varphi(X_1;\theta_0,\varphi_0)\chi(X_1 \in G(u/n))])^2.
\end{aligned}$$

The first term converges to zero by DCT and (A5), while the second term, by idempotency of indicator function and the Cauchy-Schwartz inequality, is dominated by

$$(n-1)E[(v'h_\varphi(X_1;\theta_0,\varphi_0))^2\chi(X_1 \in G(u/n))])P\{X_1 \in G(u/n)\}.$$

Since $P\{X_1 \in G(u/n)\} = O(1/n)$, (3.21) follows.

By a well known fact on regular experiments, (see, e.g., equation II.1.13 of IH)

$$Eh_\varphi(X_1;\theta_0,\varphi_0) = 0. \tag{3.22}$$

Also, by Lemma V.2.5 of IH, we have

$$Eh_\theta(X_1;\theta_0,\varphi_0) = c/2. \tag{3.23}$$

Therefore, from (3.20)-(3.23) and part (d) of Lemma 3.4, we have

$$ET_n = -(1/8)v'Iv + o(1). \tag{3.24}$$

Now from (3.20) and (3.21), we have

$$\begin{aligned}\sum_{i=1}^{n} \eta_{in} &= T_n + (u/n) \sum_{i=1}^{n} h_\theta(X_i;\theta_0,\varphi_0) \\ &\quad + n^{-1/2} v' \sum_{i=1}^{n} h_\varphi(X_i;\theta_0,\varphi_0) + o_p(1),\end{aligned}$$

and therefore part (4) follows from (3.19), (3.22), (3.23) and the Law of Large Numbers. □

PROOF OF THEOREM 3.1. By results of IH (Sec. V.2), $\widetilde{Z}_{2n}(u,v)$ is stochastically bounded and so is Δ_n by the Central Limit Theorem (CLT). Therefore, from Lemma 3.2 and 3.3, part (b) of Theorem 3.1 follows. Part (a) is an easy consequence of (b) and in fact, is equivalent to (b). □

We now obtain the limit of the LRP $Z_n(u,v)$. For this, we introduce the following notations:

Let $\nu_1(u), \ldots, \nu_r(u), \widetilde{\nu}_1(u), \ldots, \widetilde{\nu}_r(u)$ be $2r$ independent copies of a homogeneous Poisson process with rate one. For $u \geq 0$, define

$$\nu_k^+(u) = \begin{cases} \nu_k(p_k a_k' u), & \text{if } a_k' > 0, \\ \nu_k(-q_k a_k' u), & \text{if } a_k' < 0, \end{cases}$$

$k = 1, \ldots, r$ and for $u \leq 0$, define

$$\nu_k^-(u) = \begin{cases} \widetilde{\nu}_k(-q_k a_k' u), & \text{if } a_k' > 0, \\ \widetilde{\nu}_k(p_k a_k' u), & \text{if } a_k' < 0, \end{cases}$$

$k = 1, \ldots, r$. Also set $\nu_k^+(u) = 0$ if $u < 0$ and $\nu_k^-(u) = 0$ if $u > 0$. Let Δ be a $N_d(0, I)$ random variable independent of $\nu_1, \ldots, \nu_r, \widetilde{\nu}_1, \ldots, \widetilde{\nu}_r$.

THEOREM 3.2. *Under Assumptions* (A2)-(A5) *and* (A7), *the finite dimensional distributions of the LRP $Z_n(u,v)$ converge to those of the process*

$$Z(u,v) = Z^{(1)}(u) Z^{(2)}(v),$$

where

$$Z^{(1)}(u) = \exp\{uc + \sum_{k=1}^{r} \operatorname{sign}(u a_k') \log(q_k/p_k)(\nu_k^+(u) + \nu_k^-(u))\}, \quad u \in \mathbb{R}$$

and

$$Z^{(2)}(v) = \exp\{v'\Delta - (1/2) v' I v\}, \quad v \in \mathbb{R}^d.$$

PROOF. For $u_1, \ldots, u_m \in \mathbb{R}$, we find the joint asymptotic distribution of

$$(\sum_{i=1}^{n} \chi(X_i \in [a_k, a_k + (u_1/n)a_k']), \ldots, \sum_{i=1}^{n} \chi(X_i \in [a_k, a_k + (u_m/n)a_k']), k = 1, \ldots, r)$$

and $n^{-1/2} \sum_{i=1}^{n} h_\varphi(X_i; \theta_0, \varphi_0)$.

For simplicity, let us assume $0 < u_1 < \cdots < u_m$; the treatment for other cases are similar. Thus it is enough to find out the joint asymptotic distribution of

$$(\sum_{i=1}^{n} \chi(X_i \in [a_k, a_k + (u_1/n)a_k']), \sum_{i=1}^{n} \chi(X_i \in [a_k + (u_1/n)a_k', a_k + (u_2/n)a_k']),$$
$$\ldots, \sum_{i=1}^{n} \chi(X_i \in [a_k + (u_{m-1}/n)a_k', a_k + (u_m/n)a_k']), k = 1, \ldots, r)$$

and $n^{-1/2} \sum_{i=1}^{n} h_\varphi(X_i; \theta_0, \varphi_0)$.

In view of Theorem 3.1, the result follows from the following lemma.

LEMMA 3.5. *Let $X_1, X_2 \ldots$ be i.i.d. random variables having a density $f(\cdot)$ and $h : \mathbb{R} \to \mathbb{R}^d$ be a function satisfying $Eh(X_1) = 0$ and $\mathrm{COV}(h(X_1)) = \Sigma$, a positive definite matrix. Let $I_{1n}, I_{2n}, \ldots, I_{ln}$ be disjoint intervals such that for all $j = 1, \ldots, l$, the following hold:*

(a) $\mathrm{mes}(I_{jn}) = c_j/n + o(1/n)$ *for some* $c_j > 0$,

(b) $\sup\{|f(x) - b_j|; x \in I_{jn}\} = o(1)$ *as* $n \to \infty$ *for some* $b_j > 0$,

(c) $\sup_{n \geq 1}\{\|h(x)\| : x \in I_{jn}\} < \infty$.

Set $N_{jn} = \sum_{i=1}^{n} \chi\{X_i \in I_{jn}\}$, $j = 1, \ldots, l$, $n \geq 1$ and $\Delta_n = n^{-1/2} \sum_{i=1}^{n} h(X_i)$. Then

$$(N_{1n}, \ldots, N_{ln}, \Delta_n) \xrightarrow{d} (N_1, \ldots, N_l, \Delta),$$

where $N_1, \ldots, N_l$ and Δ are independent with $N_j \sim \mathrm{Poisson}(b_j c_j)$, $j = 1, \ldots, l$ and $\Delta \sim N_d(0, \Sigma)$.

PROOF. Let $\psi(s)$, $s \in \mathbb{R}^d$ be the characteristic function (c.f.) of $h(X_1)$. By Assumption (c), $\exp[in^{-1/2}s'h(x)] = 1 + o(1)$ uniformly in $x \in I_{jn}$, $j = 1, \ldots, l$. Also, by continuity of c.f., $\psi(n^{-1/2}s) = 1 + o(1)$. By CLT,

$$\psi^n(n^{-1/2}s) \to \exp[-(1/2)s'\Sigma s].$$

Now,

$$E[\exp\{it_1 N_{1n} + \cdots + it_l N_{ln} + is'\Delta_n\}]$$

$$
\begin{aligned}
&= \{\sum_{j=1}^{l}(\exp(it_j)-1)\int_{I_{jn}}\exp[in^{-1/2}s'h(x)]f(x)dx \\
&\quad + \int \exp[in^{-1/2}s'h(x)]f(x)dx\}^n \\
&= \psi^n(n^{-1/2}s)\{1+\sum_{j=1}^{l}(\exp(it_j)-1)b_jc_j(1+o(1))/n\}^n \\
&\to \exp[-(1/2)s'\Sigma s]\exp(\sum_{j=1}^{l}(\exp(it_j)-1)b_jc_j). \qquad (3.25)
\end{aligned}
$$

This completes the proof by virtue of the Levy Continuity Theorem. □

We end this section with a refinement of Theoerem 3.1 (in some particular situations). This will not be used anywhere in this chapter, but is of independent interest, and is analogous to Theorem V.2.4 of IH.

THEOREM 3.3. *Assume Conditions* (A1)–(A7) *and suppose that* $h_\varphi(X_1;\theta_0,\varphi_0)$ *has a finite moment generating function in a* [redacted] *neighbourhood of zero. Then for any* $H<\infty$,

(a) $\sup\{E|Z_n(u,v)-\widetilde{Z}_n(u,v)| : 0\le u\le H, \|v\|\le H\}\to 0$ *as* $n\to\infty$ *if* $\Gamma^-(\theta_0)=\emptyset$,

(b) $\sup\{E|Z_n(u,v)-\widetilde{Z}_n(u,v)| : -H\le u\le 0, \|v\|\le H\}\to 0$ *as* $n\to\infty$ *if* $\Gamma^+(\theta_0)=\emptyset$,

(c) $\sup\{E|Z_n(u,v)-\widetilde{Z}_n(u,v)| : |u|\le H, \|v\|\le H\}\to 0$ *as* $n\to\infty$ *if both* $\Gamma^+(\theta_0)=\emptyset$ *and* $\Gamma^-(\theta_0)=\emptyset$.

PROOF. We shall prove only (a); proof of (b) is similar and (c) is an easy consequence of (a) and (b). Observe that the convergence in Theorem 3.2 is uniform in (u,v) belonging to compacts. As argued in IH (Theorem V.2.4), it now suffices to verify that

$$\lim_{n\to\infty}EZ_n(u,v)=\lim_{n\to\infty}E\widetilde{Z}_n(u,v)=EZ(u,v)=1 \qquad (3.26)$$

where the convergences are uniform in the arguments. Now by Theorem 3.2,

$$EZ(u,v)=EZ^{(1)}(u)EZ^{(2)}(v)=1 \qquad (3.27)$$

by arguments of IH and the trivial equality $EZ^{(2)}(v)=1$. For convenience, assume that $a'_k>0\ \forall k=1,\dots,r$. Observe that

$$EZ_n(u,v)=(E[f(X_1;\theta_0+u/n,\varphi_0+n^{-1/2}v)/f(X_1;\theta_0,\varphi_0)])^n \qquad (3.28)$$

while

$$
\begin{aligned}
& E(f(X_1;\theta_0+u/n,\varphi_0+n^{-1/2}v)/f(X_1;\theta_0,\varphi_0)) \\
= \; & E(1+\eta_{1n})^2 + \int_{G(u/n)} f(x;\theta_0+u/n,\varphi_0+n^{-1/2}v)\chi\{f(x;\theta_0,\varphi_0)>0\}dx \\
& - \int_{G(u/n)} f(x;\theta_0,\varphi_0)dx \\
= \; & 1+2E\eta_{1n}+E\eta_{1n}^2+(u/n)\sum_{k=1}^{r}(q_k-p_k)a_k'+o(1/n) \\
= \; & 1+2[cu/(2n)-(1/8n)v'Iv]+(1/4n)v'Iv \\
& -(u/n)\sum_{k=1}^{r}(q_k-p_k)a_k'+o(1/n) \\
= \; & 1+o(1/n)
\end{aligned} \tag{3.29}
$$

by arguments similar to that of IH and an application of Lemma 3.4. Also, the convergence in (3.29) is uniform in the arguments. Consequently, (3.28) and (3.29) together imply one part of (3.26).

Now

$$
\begin{aligned}
& E\widetilde{Z}_n(u,v) \\
= \; & \exp[uc-(1/2)v'Iv]E\{\chi(\sigma_n^+>u)\exp[\sum_{i=1}^{n}\sum_{k\in\Gamma}\log(q_k/p_k) \\
& \quad \times\chi(X_i\in[a_k,a_k+(u/n)a_k'])+2n^{-1/2}\sum_{i=1}^{n}h_\varphi(X_i;\theta_0,\varphi_0)]\} \\
= \; & \exp[uc-(1/2)v'Iv](E\{\chi(X_1\notin\bigcup_{k\in\Gamma^+}[a_k,a_k(\theta_0+u/n)]) \\
& \quad \times\exp[\sum_{k\in\Gamma}\log(q_k/p_k)\chi(X_1\in[a_k,a_k+(u/n)a_k']) \\
& \quad +2n^{-1/2}h_\varphi(X_1;\theta_0,\varphi_0)]\})^n
\end{aligned} \tag{3.30}
$$

and

$$
\begin{aligned}
& E\{\chi(X_1\notin\bigcup_{k\in\Gamma^+}[a_k,a_k(\theta_0+u/n)])\exp[\sum_{k\in\Gamma}\log(q_k/p_k) \\
& \quad \times\chi(X_1\in[a_k,a_k+(u/n)a_k'])+2n^{-1/2}h_\varphi(X_1;\theta_0,\varphi_0]\} \\
= \; & E\{\exp[\sum_{k\in\Gamma}\log(q_k/p_k)\chi(X_1\in[a_k,a_k+(u/n)a_k']) \\
& \quad +2n^{-1/2}h_\varphi(X_1;\theta_0,\varphi_0)]\}-E\{\exp[\sum_{k\in\Gamma}\log(q_k/p_k) \\
& \quad \times\chi(X_1\in[a_k,a_k+(u/n)a_k'])+2n^{-1/2}h_\varphi(X_1;\theta_0,\varphi_0)]
\end{aligned}
$$

$$
\begin{aligned}
& \times \sum_{k\in\Gamma^+} \chi(X_1 \in [a_k, a_k(\theta_0 + u/n)])\} \\
= \; & \sum_{k\in\Gamma}(q_k/p_k)\int_{a_k}^{a_k+(u/n)a_k'} \exp[2n^{-1/2}h_\varphi(X_1;\theta_0,\varphi_0)]f(x;\theta_0,\varphi_0)dx \\
& -\sum_{k\in\Gamma}\int_{a_k}^{a_k+(u/n)a_k'} \exp[2n^{-1/2}h_\varphi(X_1;\theta_0,\varphi_0)]f(x;\theta_0,\varphi_0)dx \\
& +\int \exp[2n^{-1/2}h_\varphi(X_1;\theta_0,\varphi_0)]f(x;\theta_0,\varphi_0)dx \\
& -\sum_{k\in\Gamma^+}\int_{a_k}^{a_k+(u/n)a_k'} \exp[2n^{-1/2}h_\varphi(X_1;\theta_0,\varphi_0)]f(x;\theta_0,\varphi_0)dx \\
= \; & (u/n)\sum_{k\in\Gamma}(q_k/p_k)a_k'p_k - (u/n)\sum_{k\in\Gamma}a_k'p_k + 1 \\
& +(1/2n)v'Iv - (u/n)\sum_{k\in\Gamma^+}a_k'q_k + o(1/n) \\
= \; & -(u/n)\sum_{k=1}^{r}(p_k - q_k)a_k' + (1/2n)v'Iv + o(1/n) \\
= \; & -uc/n + (1/2n)v'Iv + o(1/n) \qquad (3.31)
\end{aligned}
$$

if n is large enough. In the above derivation, we have used (A7), a standard expansion of the moment generating function in terms of the moments and equation (3.22). It is obvious that (3.31) implies the other part of (3.26), and so the proof is complete. □

In the way of proving Theorem 3.3, we have obtained a very important result:

COROLLARY 3.1. *Under the conditions of Theorem 3.2, $\{P^n_{\theta_0+u/n,\varphi_0+n^{-1/2}v}\}$ is contiguous to $\{P^n_{\theta_0,\varphi_0}\}$ if any one of the following holds:*

(a) $u \geq 0$ *and* $\Gamma^- = \emptyset$,

(b) $u \leq 0$ *and* $\Gamma^+ = \emptyset$,

(c) $u \in \mathbb{R}$ *and* $\Gamma^+ \cup \Gamma^- = \emptyset$.

This follows from (3.27) and Le Cam's first lemma (see, Hájek and Šidák (1967, p. 202)).

4 Convergence of Posterior Distributions and Bayes Estimates

In this section, we study the asymptotic behaviour of posterior distributions and Bayes estimates. The properties of the LRP established in the previous section and the results obtained in IH and Chapter 1 are used to prove our results. As mentioned in Remark 2.1, the results of IH and Chapter 1 follow also from Assertions (I), (II) and (III) which are modified versions of the conditions of IH.

We consider prior densities which are positive and continuous at (θ_0, φ_0) and have polynomial majorants in θ and exponential majorants in φ. Let Π be the class of all such priors and $\mathfrak{L}$ be the class of continuous functions $l : \mathbb{R} \times \mathbb{R}^d \to [0, \infty)$ satisfying the following:

(i) $l(0,0) = 0$, $l(x,y) = l(-x,-y)$ $\forall x \in \mathbb{R}, y \in \mathbb{R}^d$.

(ii) The sets $\{(x,y) : l(x,y) < c\}$ are convex $\forall c > 0$ and are bounded if $c > 0$ is sufficiently small.

(iii) $l(x,y) \leq B(1 + |x|^{b_1}) \exp[b_2\|y\|]$ for some $B, b_1, b_2 > 0$.

(iv) There exist $H_0, \gamma > 0$ such that for $H \geq H_0$,

$$\sup\{l(x,y) : |x| \leq H^\gamma, \|y\| \leq H^\gamma\} - \inf\{l(x,y) : |x| \geq H, \|y\| \geq H\} \leq 0.$$

We immediately have the strong consistency of posterior distributions and Bayes esimates. The following result, a consequence of Theorem I.10.2 of IH, gives the asymptotic distributions of the Bayes estimates.

THEOREM 4.1. *Let $(\widetilde{\theta}_n, \widetilde{\varphi}_n)$ be a Bayes estimate of (θ, φ) with respect to a prior $\pi \in \Pi$ and loss function $l(n(x - \theta_0), n^{1/2}(y - \varphi_0))$ where $l \in \mathfrak{L}$. Assume Conditions* (A1)-(A8) *stated in Section* 2 *and suppose that the random function*

$$\psi(s,t) = \int_{\mathbb{R}^{1+d}} l(s-u, t-v)\xi(u,v)\,du\,dv$$

attains its absolute minimum at a unique point $\tau = \tau(\theta_0, \varphi_0)$, where

$$\xi(u,v) = Z(u,v) / \int_{\mathbb{R}^{1+d}} Z(u',v')\,du'\,dv'.$$

Then

$$(n(\widetilde{\theta}_n - \theta_0), n^{1/2}(\widetilde{\varphi}_n - \varphi_0)) \xrightarrow{d} \tau, \tag{4.1}$$

and for any continuous function $w(u,v)$ satisfying

$$|w(u,v)| \leq B(1 + |u|^{b_1}) \exp[b_2\|v\|]$$

for some $B, b_1, b_2 > 0$, we have

$$\lim_{n\to\infty} Ew(n(\widetilde{\theta}_n - \theta_0), n^{1/2}(\widetilde{\varphi}_n - \varphi_0)) = Ew(\tau). \tag{4.2}$$

Also the diameter of the set of all normalized Bayes estimates converges to zero in probability.

If further, the loss function $l \in \mathfrak{L}$ is of the form

$$l(x, y) = l_1(x) + l_2(y),$$

then $n(\widetilde{\theta}_n - \theta_0)$ and $n^{1/2}(\widetilde{\varphi}_n - \varphi_0)$ are asymptotically independent.

To investigate whether the posterior distribution, suitably normalized and centered, converges to a limit, we use Theorem 3.4 of Chapter 1. In this case, $Z(u, v)$ is of product form and this criterion for existence of a posterior limit is satisfied if and only if $\xi^{(1)}(u) := Z^{(1)}(u) / \int Z^{(1)}(u')du'$ is of the form $g(u + W)$ for some random variable W and a fixed probability density g on $\mathbb{R}$. In view of the investigations carried out in Section 4 of Chapter 1, we can answer whether the posterior converges or not in the following cases:

CASE 1. The set Γ is empty.

(i) If both Γ^+ and Γ^- are nonempty, then the criterion is not satisfied and hence a limit of posterior does not exist.

(ii) If one of Γ^+ and Γ^- is empty, then it is immediate that the criterion is satisfied and hence a posterior limit exists. An important example of this kind is

$$f(x; \theta, \varphi) = \begin{cases} e^{\varphi} \exp[-e^{\varphi}(x - \theta)], & \text{if } x \geq \theta, \\ 0, & \text{otherwise.} \end{cases}$$

where $\theta, \varphi \in \mathbb{R}$. Indeed, any density in Example 1 with $f(x) < 0$ for $x < 0$ falls in this category.

CASE 2. The set Γ is nonempty but both Γ^+ and Γ^- are empty.

In the case $r = 1$, as shown in Chapter 1, a limit of posterior does not exist. An important example of this kind is the change point problem described in Example 3 of Section 2.

5 A Convolution Theorem

Let $\{(\mathfrak{X}^n, \mathcal{A}^n), P^n_{\theta,\varphi}; \theta \in \Theta, \varphi \in \Phi\}$, $n \geq 1$, be a sequence of statistical experiments where Θ and Φ are open subsets of $\mathbb{R}$ and $\mathbb{R}^d$ $(d \geq 1)$ respectively. We fix $\theta_0 \in \Theta$ and $\varphi_0 \in \Phi$ and as in Section 3, set

$$Z_n(u, v) = \frac{dP^n_{\theta_0 + u/n, \varphi_0 + n^{-1/2}v}}{dP^n_{\theta_0, \varphi_0}}, \quad u \in U_n, v \in V_n,$$

U_n, V_n being as in Section 3. Here dP/dQ denotes the derivative of the absolutely continuous component of P with respect to Q. We assume that either of the following representations holds:

$$\text{(C1)} \qquad Z_n(u,v) = \begin{cases} \exp\{uc + v'\Delta_n - (1/2)v'Iv\} + \varepsilon_n, & \text{if } u < \sigma_n, \\ 0, & \text{if } u > \sigma_n, \end{cases} \tag{5.1}$$

where $c > 0$ is a constant, I is a positive definite matrix and ε_n, Δ_n and σ_n are random variables (all depening on (θ_0, φ_0)) such that

$$\varepsilon_n \xrightarrow{P^n_{\theta_0,\varphi_0}} 0, \quad \text{and } \mathcal{L}((\Delta_n, \sigma_n)|P^n_{\theta_0,\varphi_0}) \Rightarrow \mathcal{L}(\Delta, \sigma)$$

where $\Delta \sim N_d(0, I)$, σ has an exponential distribution with mean $1/c$ and Δ and σ are independent.

$$\text{(C2)} \qquad Z_n(u,v) = \begin{cases} \exp\{uc + v'\Delta_n - (1/2)v'Iv\} + \varepsilon_n, & \text{if } u > -\sigma_n, \\ 0, & \text{if } u < -\sigma_n, \end{cases} \tag{5.2}$$

where c, I, ε_n, Δ_n and σ_n are as in (C1).

It is to be noted that the above representation is a sort of combination of local asymptotic normality in v and local asymptotic exponentiality (as discussed in IH, Sec. V.5) in u.

If we consider i.i.d. observations with a common density, Theorem 3.1 established in Section 3 gives an asymptotic expansion of the likelihood ratio $Z_n(u,v)$ under Assumptions (A1)–(A5) of Section 2. This implies that the asymptotic expansion (5.1) holds if the sets Γ and Γ^- are empty and the asymptotic expansion (5.2) holds if both Γ and Γ^+ are empty.

Below we consider only the case when (5.1) holds. The treatment of the other case is similar.

We define experiments

$$\mathcal{E}^n = \{(\mathfrak{X}^n, \mathcal{A}^n), P^n_{\theta_0+u/n, \varphi_0+n^{-1/2}v}; u \geq 0, v \in \mathbb{R}^d\},\ n \geq 1. \tag{5.3}$$

One can then easily show that the sequence of experiments $\mathcal{E}^n$ converges to an experiment $\mathcal{E} = \{Q_{u,v}; u \geq 0, v \in \mathbb{R}^d\}$ where $Q_{u,v}$ is the product of a distribution with density $c\exp[-c(x-u)]\chi\{x \geq u\}$ and $N_d(v, I)$. This may be proved by using Proposition II.2.3 of Millar (1983). Using the Hájek-Le Cam-Millar asymptotic minimax theorem (see Millar (1983), Sec. III.1), a lower bound to the local asymptotic minimax risk (for a natural class of loss functions including the class $\mathfrak{L}$ of Section 4) can be obtained as in Samanta (1988, Ch. 4). This bound is attained by the Bayes estimates. In the remaining part of this section, we shall obtain a convolution theorem characterizing the possible limiting distribution of a sequence of regular estimators $\{(S_n, T_n)\}$ of (θ, φ) which satisfies (5.4) below. We consider the class of estimators $\{(S_n, T_n)\}$ of (θ, φ) for which

$$\mathcal{L}((n(S_n - \theta_0 - u/n), n^{1/2}(T_n - \varphi_0 - n^{-1/2}v))|P^n_{\theta_0+u/n,\varphi_0+n^{-1/2}v}) \Rightarrow G \tag{5.4}$$

for all $u \geq 0$ and $v \in \mathbb{R}^d$ where G is some probabilty distribution on $\mathbb{R}^{d+1}$ not depending on (u, v).

THEOREM 5.1. *Suppose that the asymptotic expansion in* (5.1) *holds. Then for any estimator* $\{(S_n, T_n)\}$ *satisfying* (5.4), *the limiting distribution* G *of* $(n(S_n - \theta_0), n^{1/2}(T_n - \varphi_0))$ *under* $P^n_{\theta_0,\varphi_0}$ *is a convolution of* $Q_{0,0}$ *and some probability distribution* μ *depending on* $\{(S_n, T_n)\}$*:*

$$G = Q_{0,0} * \mu.$$

This theorem may be proved using arguments similar to that used in the proof of Theorem V.5.2 of IH. We, however, present a proof that uses the ideas of Millar (1983, Sec. III.2).

To prove this theorem, we shall use the following result which is a modified version of the convolution theorem given in Millar (1983, Sec. III.2).

THEOREM 5.2. *Let* $\mathcal{E}^n = \{(S^n, \mathcal{S}^n), Q^n_{u,v}; u \geq 0, v \in \mathbb{R}^d\}$, $n \geq 1$, *be a sequence of statistical experiments converging to the experiment* $\mathcal{E} = \{(\mathbb{R}^{d+1}, \mathcal{B}^{d+1}), Q_{u,v}; u \geq 0, v \in \mathbb{R}^d\}$. *Suppose that* R_n *is a sequence of statistics on* $(S^n, \mathcal{S}^n)$ *taking values in* $\mathbb{R}^{d+1}$. *Assume further that*

(i) *there is a family of probabilities* $\{G_{u,v}; u \geq 0, v \in \mathbb{R}^d\}$ *on* $(\mathbb{R}^{d+1}, \mathcal{B}^{d+1})$ *such that for all* $u \geq 0$, $v \in \mathbb{R}^d$,

$$\mathcal{L}(R_n|Q^n_{u,v}) \Rightarrow G_{u,v}.$$

(ii) $Q_{u,v}(A \times B) = Q_{0,0}(A \times B - (u, v))$ *and* $G_{u,v}(A \times B) = G_{0,0}(A \times B - (u, v))$ *for all* $u \geq 0$, $v \in \mathbb{R}^d$, $A \in \mathcal{B}$ *and* $B \in \mathcal{B}^d$.

(iii) $Q_{0,0}$ *is concentrated on* $[0, \infty) \times \mathbb{R}^d$ *and is absolutely continuous with respect to the Lebesgue measure. Also the number* $(0, 0)$ *belongs to the support of* $Q_{0,0}$.

Then there is a probability μ on $\mathbb{R}^{d+1}$ such that

$$G_{0,0} = Q_{0,0} * \mu.$$

PROOF. let $\mathcal{E}_0 = \{(\mathbb{R}^{d+1}, \mathcal{B}^{d+1}), G_{u,v}; u \geq 0, v \in \mathbb{R}^d\}$. Then by an argument given in Millar (1983, p. 98), there exists a Markov kernel K_0 of $(\mathbb{R}^{d+1}, \mathcal{B}^{d+1})/(\mathbb{R}^{d+1}, \mathcal{B}^{d+1})$ such that

$$G_{u,v} = K_0 Q_{u,v} \quad \text{for all } (u,v) \in [0,\infty) \times \mathbb{R}^d. \tag{5.5}$$

Let $\mathcal{K}_0$ be the collection of all Markov kernels K of $(\mathbb{R}^{d+1}, \mathcal{B}^{d+1})/(\mathbb{R}^{d+1}, \mathcal{B}^{d+1})$ such that (5.5) holds. Then $\mathcal{K}_0$ is a compact, convex subset of a topological vector space. For all $g \in [0,\infty) \times \mathbb{R}^d$, define a map $K \mapsto gK$ as

$$gK(x, A) = K(x + g, A + g), \quad x \in \mathbb{R}^{d+1}, A \in \mathcal{B}^{d+1}.$$

Then the family $\{g : g \in [0,\infty) \times \mathbb{R}^d\}$ is a commuting family of continuous linear mappings which leaves $\mathcal{K}_0$ invariant. Therefore, by the Markov-Kakutani fixed point theorem (see, e.g., Dunford and Schwartz, p. 456), there exists $K \in \mathcal{K}_0$ such that

$$gK = K \quad \text{for all } g \in [0,\infty) \times \mathbb{R}^d,$$

i.e., for every $\mu \in V(\mathcal{E})$ (see Millar (1983) for defintion), for every Borel set $A \in \mathcal{B}^{d+1}$ and $g \in [0,\infty) \times \mathbb{R}^d$,

$$\int gK(x, A)\mu(dx) = \int K(x, A)\mu(dx).$$

Since $V(\mathcal{E}) = L^1(\nu)$ for some probability ν with support $[0,\infty) \times \mathbb{R}^d$ which is equivalent to the Lebesgue measure on $[0,\infty) \times \mathbb{R}^d$, this implies that for all $g \in [0,\infty) \times \mathbb{R}^d$ and $A \in \mathcal{B}^d$,

$$K(x, A) = K(x + g, A + g) \quad \text{a.e. } x \in [0,\infty) \times \mathbb{R}^d. \tag{5.6}$$

We shall now use (5.6) to show that there exists a probability μ on $\mathbb{R}^{d+1}$ such that

$$K(x, A + x) = \mu(A), \quad A \in \mathcal{B}^{d+1} \tag{5.7}$$

for all x outside a null set N_0. To show this, one can use straightforward arguments involving lifting theorems (see, e.g., Ionescu Tulcea and Ionescu Tulcea (1969), p. 122). We, however, give an elementary proof. In view of (5.6), it follows from Fubini's theorem that there exists a null set N such that for $x \notin N$, $x \in [0,\infty) \times \mathbb{R}^d$,

$$K(x, A) = K(x + g, A + g) \quad \text{for all } A \in \mathcal{B}^{d+1} \quad \text{and for a.e. } g \in [0,\infty) \times \mathbb{R}^d.$$

We now choose a sequence $\alpha_n \to 0$, $\alpha_n \in [0,\infty) \times \mathbb{R}^d$, $\alpha_n \notin N$ for all n. Then for all $n \geq 1$, there is a null set N_n such that for all $g \notin N_n$, $g \in [0,\infty) \times \mathbb{R}^d$,

$$K(\alpha_n + g, A + g) = K(\alpha_n, A) \quad \text{for all } A \in \mathcal{B}^{d+1}.$$

Hence for all $x \notin N_n + \alpha_n$, $x \in [0,\infty) \times \mathbb{R}^d + \alpha_n$,

$$K(x, A + x) = K(\alpha_n, A + \alpha_n) \quad \text{for all } A \in \mathcal{B}^{d+1}.$$

Let $N_0 = \bigcup_{n\geq 1}(N_n + \alpha_n)$. Then N_0 is a null set and for any $x, y \in [0,\infty) \times \mathbb{R}^d$, $x, y \notin N_0$, we have

$$K(x, A + x) = K(y, A + y) \quad \text{for all } A \in \mathcal{B}^{d+1}.$$

This proves (5.7). We thus have

$$K(x, A) = K(x, (A - x) + x) = \mu(A - x) \quad \text{for all } A \in \mathcal{B}^{d+1}, x \notin N_0,$$

and therefore

$$G_{u,v}(A) = \int K(x, A) Q_{u,v}(dx) = \int \mu(A - x) Q_{u,v}(dx).$$

Since μ is a probability, this proves the theorem. □

Proof of Theorem 5.1. Consider $\mathcal{E}^n = \{P^n_{\theta_0+u/n, \varphi_0+n^{-1/2}v}; u \geq 0, v \in \mathbb{R}^d\}$ and $\mathcal{E} = \{Q_{u,v}; u \geq 0, v \in \mathbb{R}^d\}$ where $Q_{u,v}$ is as defined earlier in this section. It is easy to see that all the conditions of Theorem 5.2 are satisfied and hence

$$G = Q_{0,0} * \mu. \qquad \square$$

6 Asymptotic Independence

Theorem 3.2 of Section 3 states that under Assumptions (A2)–(A5) and (A7) of Section 2, the limiting LRP $Z(u, v)$ is of the form

$$Z(u, v) = Z^{(1)}(u) Z^{(2)}(v),$$

where $Z^{(1)}(\cdot)$ and $Z^{(2)}(\cdot)$ are two independent processes as described in the theorem. This result has been used in the previous sections to obtain the following interesting results:

Chapter 3

Multiparameter Densities with Singularities

1 Introduction

In the previous chapter, we have considered the estimation problem with density $f(x;\theta,\varphi)$ where the family is smooth with respect to φ while the paramter θ exhibits nonregularity in the sense that the densities are discontinuous at certain points depending on θ. In the present chapter, we shall consider a very similar problem. Here, the family of densities $f(x;\theta,\varphi)$ is again smooth with respect to φ but the densities have singularities (see Example 5 of Section 2.1) at certain points depending on θ.

Let us recall the definition of singularity.

Let f be a density on $\mathbb{R}$ which admits the following representation in a neighbourhoood of a point z :

$$f(x) = \begin{cases} p(x)|x-z|^{\alpha}, & \text{if } x > z, \\ q(x)|x-z|^{\alpha}, & \text{if } x < z, \end{cases} \tag{1.1}$$

where $-1 < \alpha < 1$, $\alpha \neq 0$ and the functions $p(x)$ and $q(x)$, which possibly involve certain paramters, are continuous with $p(z)+q(z) > 0$. In this case, f is said to have a singularity of the first type if $0 < \alpha < 1$ and a singularity of the third type if $-1 < \alpha < 0$. The case $\alpha = 0$ corresponds to the case with discontinuous densities. Important examples of this type are the gamma density

$$f(x) = (1/\Gamma(\alpha))e^{-x}x^{\alpha-1}, \qquad x > 0,$$

and the Weibull density

$$f(x) = \alpha x^{\alpha-1}\exp[-x^{\alpha}], \qquad x > 0.$$

A singularity of the second type occur if the density admits the following representaion in a neighbourhood of a point z :

$$f(x) = \begin{cases} \psi(x)\exp[a(x)|x-z|^{\alpha/2}], & \text{if } x < z, \\ \psi(x)\exp[b(x)|x-z|^{\alpha/2}], & \text{if } x > z, \end{cases} \tag{1.2}$$

where $0 < \alpha < 1$, $\psi(z) > 0$, the functions $\psi(x), a(x)$ and $b(x)$ are continuous and at least one of the quantities $a(z)$ and $b(z)$ is nonzero. An example is given by the density

$$f(x) = C\exp[-|x|^{\alpha}].$$

IH (Ch. VI) considered the problem of estimating a location parameter θ in a location shift model $f(x;\theta) = f(x-\theta)$ where f is a density with singularities at certain points. This case was also studied by Woodroofe (1974) and Polfeldt (1970). There are, however, many important examples of multiparameter family of densities where in addition to the "nonregular" parameter θ, there is also a vector of parameters φ with respect to which the problem is regular (with θ being fixed). For important examples of this kind see, for example, Smith (1985) and Cheng and Iles (1987). We here consider the case with densities having singularities in presence of a "regular" parameter φ. As in IH, for simplicity, we assume θ to be a location parameter and consider the case with only one point of singularity; extension to a more general case does not involve additional difficulties and does not lead to substantially new phenomena. Also, most of the examples of practical importance involves a location parameter and a single point of singularity.

Let $X_1, X_2, \ldots$ be i.i.d. observations with values in $\mathbb{R}$. We assume that each observation X_i possesses a distribution $P_{\theta,\varphi}$ with a density $f(x;\theta,\varphi)$ which is of the form

$$f(x;\theta,\varphi) = g(x-\theta;\varphi), \tag{1.3}$$

where $g(x;\varphi)$ is a density with singularity at the point $x=0$. We study the estimation problem using a method similar to that in IH. We first obtain the properties of the likelihood ratios. It is shown that the likelihood ratios satisfy certain conditions similar to those in Theorem I.10.2 of IH. The general results of IH are then used to obtain asymptotic behaviour of the Bayes estimates. In Section 2, we consider the case with a singularity of the first and third type. Section 3 deals with the singularity of the second type. In Section 2, we also examine whether a limit of a suitably centered (and normalized) posterior exists. We use Theorem 3.4 of Chapter 1 and conclude that a limit of posterior does not exist in this case.

2 Singularities of the First and Third Type

In this section, we consider the case with a density $g(x-\theta;\varphi)$ where g admits the representation (1.1) in a neighbourhood $(-\varepsilon_0, \varepsilon_0)$ (say) of $z=0$ with $p(x) = p(x;\varphi)$ and $q(x) = q(x;\varphi)$. We make the following assumptions:

(A0) For all $\varphi \in \Phi$,

$$\inf_{\|(u,v)\|>\varepsilon} \int |g(x-u, \varphi+v) - g(x;\varphi)| > 0, \quad \varepsilon > 0.$$

(A1) The functions $p(x;\varphi)$ and $q(x;\varphi)$ are continuously differentiable with respect to both x and φ in their respective domains of definition.

(A2) The functions $h(x;\varphi) := \log g(x;\varphi)$ is thrice differentiable with respect to φ for all $x \neq 0$.

We set

$$\begin{aligned} h_i(x;\varphi) &= (\partial/\partial\varphi_i)h(x;\varphi), \\ h_{ij}(x;\varphi) &= (\partial^2/\partial\varphi_i\partial\varphi_j)h(x;\varphi), \\ h_{ijk}(x;\varphi) &= (\partial^3/\partial\varphi_i\partial\varphi_j\partial\varphi_k)h(x;\varphi), \end{aligned}$$

$i,j,k = 1,\ldots,d$.

(A3) There exist $\varepsilon_i > 0$ and functions $H_i(x;\varphi)$, $i = 1,2,3$, such that H_1 is square integrable and H_2, H_3 are integrable with respect to $g(x;\varphi)$ and for all x and all $i,j,k = 1,2,\ldots,d$,

(a) $\sup\{|h_i(y;\varphi)| : |y-x| < \varepsilon_1\} \leq H_1(x;\varphi)$,

(b) $\sup\{|h_{ij}(y;\varphi)| : |y-x| < \varepsilon_2\} \leq H_2(x;\varphi)$

and (c) $\sup\{|h_{ijk}(y;\varphi')| : |y-x| < \varepsilon_3, \|\varphi' - \varphi\| < \varepsilon_3\} \leq H_3(x;\varphi)$

(A4) The function $h_i(x;\varphi)$, $i = 1,\ldots,d$, are differentiable with respect to x on $\{x \neq 0\}$. Also for any $\varepsilon > 0$, there exists a $g(x;\varphi)$-integrable function $H_4(x;\varphi)$ and $\varepsilon_4 > 0$ such that for $|x| > \varepsilon$ and $i = 1,\ldots,d$,

$$\sup_{|y-x|<\varepsilon_4} |(\partial/\partial y)h_i(y;\varphi)| \leq H_4(x;\varphi).$$

(A5) The functions $(\partial^2/\partial x\partial\varphi_i)\log p(x;\varphi)$ is bounded in $x \in (0,\varepsilon_0)$ and the function $(\partial^2/\partial x\partial\varphi_i)\log q(x;\varphi)$ is bounded in $x \in (-\varepsilon_0, 0)$.

(A6) For any $\varepsilon > 0$, the functions

$$\begin{aligned} J(\varphi;\varepsilon) &= \int_{|x|>\varepsilon} ((\partial/\partial x)g^{1/2}(x;\varphi))^2 dx, \\ J_i(\varphi;\varepsilon) &= \int_{|x|>\varepsilon} ((\partial/\partial x)g^{1/2}(x;\varphi))((\partial/\partial\varphi_i)g^{1/2}(x;\varphi))dx, \\ \text{and} \quad J_{ij}(\varphi;\varepsilon) &= \int_{|x|>\varepsilon} ((\partial/\partial\varphi_i)g^{1/2}(x;\varphi))((\partial/\partial\varphi_j)g^{1/2}(x;\varphi))dx \end{aligned}$$

are (finite and) continuous in φ. Moreover, the matrix

$$\begin{pmatrix} J & J_*' \\ J_* & J_{**} \end{pmatrix}$$

is positive definite, where $J_* = (J_1,\ldots,J_d)'$ and $J_{**} = ((J_{ij}))$. Further, for $i,j = 1,\ldots,d$, $I_{ij}(\varphi) := \int h_{ij}(x;\varphi)g(x;\varphi)dx$ is finite.

(A7) The quantities $\sup\{p(x;\varphi) : 0 < x < \varepsilon_0\}$, $\sup\{q(x;\varphi) : -\varepsilon_0 < x < 0\}$, $\int_0^{\varepsilon_0}((p'(x;\varphi))^2/p(x;\varphi))dx$ and $\int_{-\varepsilon_0}^0((q'(x;\varphi))^2/q(x;\varphi))dx$, $J(\varphi;\varepsilon)$, $\varepsilon > 0$, and $I_{ij}(\varphi)$, $i,j = 1,\ldots,d$, possess exponential majorants in $\|\varphi\|$; here the prime stands for differentiation with respect to x.

(A8) For $i = 1,\ldots,d$, $h_i(x;\varphi)$ is bounded in a neighbourhood of zero.

If $\Theta \times \Phi$ is unbounded, we impose the following additional condition:

(A9) For some $\gamma > 0$ and $C > 0$,

$$\int g^{1/2}(x;\varphi)g^{1/2}(x-u;\varphi+v)dx \leq C|u|^{-\gamma}\exp[-\gamma\|v\|]$$

as $\max\{|u|,\|v\|\} \to \infty$.

REMARK 2.1. In IH, certain smoothness condition on p and q are included in the definition of singularity. The condition is stated in (1.2) of Section VI.1 of IH. It is to be noted that this condition is satisfied with $\lambda = 2$ under the assumptions stated above.

Some important exmples which fall within our framework are as follows:

(i) Gamma,

$$f(x;\theta,\varphi) = (1/\Gamma(\alpha))e^{\alpha\varphi}(x-\theta)^{\alpha-1}\exp[-e^{\varphi}(x-\theta)], \quad x > \theta,\ \theta \in \mathbb{R},\ \varphi \in \mathbb{R}.$$

(ii) Weibull,

$$f(x;\theta,\varphi) = \alpha e^{\varphi}(x-\theta)^{\alpha-1}\exp[-e^{\varphi}(x-\theta)^{\alpha}], \quad x > \theta,\ \theta \in \mathbb{R},\ \varphi \in \mathbb{R}.$$

For some more exmples, see Smith (1985). It is to be noted that in these examples, one may also consider α to be unknown. Also, we here use a logarithmic reparametrization for the usual scale parameter; this is necessary as argued in Example 1 of Section 1.2 and Remark 2.1 of Chapter 2.

Fix a parameter point $(\theta_0,\varphi_0) \in \Theta \times \Phi$ which may be regarded as the "true parameter point". We first study the behaviour of the likelihood ratio process (LRP)

$$Z_n(u,v) = \prod_{i=1}^{n} \frac{f(X_i;\theta_0 + n^{-1/(1+\alpha)}u, \varphi_0 + n^{-1/2}v)}{f(X_i;\theta_0,\varphi_0)}, \tag{2.1}$$

where $u \in U_n := n^{1/(1+\alpha)}(\Theta - \theta_0)$ and $v \in V_n := n^{1/2}(\Phi - \varphi_0)$. Since the P_{θ_0,φ_0}-distribution of $Z_n(u,v)$ does not depend on θ_0, we shall assume below that $\theta_0 = 0$. Below all the probability statements refer to the distribution $P_{(0,\varphi_0)}$.

The following properties of the LRP will be used to prove our asymptotic results:

(I) There exist $a_1, a_2, B > 0$ such that

$$\begin{aligned} & E|Z_n^{1/2}(u_1, v_1) - Z_n^{1/2}(u_2, v_2)|^2 \\ \leq\ & B(1 + R_1^{a_1}) \exp[a_2 R_2](|u_1 - u_2|^{1+\alpha} + \|v_1 - v_2\|^2) \end{aligned}$$

for all $u_1, u_2 \in [-R_1, R_1]$ and $v_1, v_2 \in \{v : \|v\| \leq R_2\}$.

(II) For all $u \in U_n$ and $v \in V_n$, we have

$$EZ_n^{1/2}(u, v) \leq \exp[-g_n(|u|, \|v\|)],$$

where $\{g_n(\cdot, \cdot)\}$ is a sequence of functions from $[0, \infty) \times [0, \infty)$ into $[0, \infty)$ satisfying

(a) For each $n \geq 1$, $g_n(x, y)$ is increasing to infinity in each component,

(b) For any $N_1, N_2 \geq 0$,

$$\lim_{\substack{n \to \infty \\ \max(x,y) \to \infty}} x^{N_1} e^{N_2 y} \exp[-g_n(x, y)] = 0.$$

(III) The finite dimensional distributions of the process $Z_n(u, v)$ converge to those of a process $Z(u, v)$. (The expression for $Z(u, v)$ is given in Theorem 2.2.)

REMARK 2.2. Assertions (I), (II) and (III) above can also be shown to hold uniformly in (θ_0, φ_0) belonging to compact subsets of $\Theta \times \Phi$ under appropriate uniform versions of Assumptions (A0)–(A9). Consequently, one can obtain a uniform version of Theorem 2.3.

We now prove assertions (I), (II) and (III). Let the Hellinger distance between $P_{\theta,\varphi}$ and $P_{\theta+u,\varphi+v}$ be denoted by $r_2((\theta, \varphi), (\theta + u, \varphi + v))$ and set $k_n = n^{-1/(1+\alpha)}$. As shown in IH (p. 54), we have

$$E|Z_n^{1/2}(u_1, v_1) - Z_n^{1/2}(u_2, v_2)|^2 \leq n r_2^2((k_n u_1, \varphi_0 + n^{-1/2} v_1), (k_n u_2, \varphi_0 + n^{-1/2} v_2)). \tag{2.2}$$

To prove Assertion (I), it is now enough to show that

$$r_2^2((\theta, \varphi), (\theta + u, \varphi + v)) \leq C(\varphi)(|u|^{1+\alpha} + \|v\|^2) \tag{2.3}$$

where $C(\varphi)$ grows at most like an exponential function in $\|\varphi\|$. The expression in LHS in (2.3) is actually free from θ, and so we suppose $\theta = 0$. For definiteness, let

$u > 0$. Using the inequality $(a+b)^2 \le 2(a^2+b^2)$, we have

$$\begin{aligned}
& r_2^2((0,\varphi),(u,\varphi+v)) \\
= & \int |f^{1/2}(x;0,\varphi) - f^{1/2}(x;u,\varphi+v)|^2 dx \\
\le & \ 2\int |g^{1/2}(x-u;\varphi) - g^{1/2}(x,\varphi)|^2 dx \\
& +2\int |g^{1/2}(x;\varphi+v) - g^{1/2}(x;\varphi)|^2 dx. \qquad (2.4)
\end{aligned}$$

In view of a well known result on regular families (see, e.g., (II.2.6) of IH) and (A7), the second term in RHS of (2.4) is bounded by $C_1(\varphi)\|v\|^2$ where $C_1(\varphi)$ is exponential in $\|\varphi\|$. Also, for $\varepsilon < \varepsilon_0$ and u sufficiently small,

$$\begin{aligned}
& \int |g^{1/2}(x-u;\varphi) - g^{1/2}(x;\varphi)|^2 dx \\
= & \int_0^u |g^{1/2}(x-u;\varphi) - g^{1/2}(x;\varphi)|^2 dx \\
& + \int_u^\varepsilon |g^{1/2}(x-u;\varphi) - g^{1/2}(x;\varphi)|^2 dx \\
& + \int_{-\varepsilon}^0 |g^{1/2}(x-u;\varphi) - g^{1/2}(x;\varphi)|^2 dx \\
& + \int_{|x|>\varepsilon} |g^{1/2}(x-u;\varphi) - g^{1/2}(x;\varphi)|^2 dx \\
= & \int_0^u ||x-u|^{\alpha/2} q^{1/2}(x-u;\varphi) - |x|^{\alpha/2} p^{1/2}(x;\varphi)|^2 dx \\
& + \int_u^\varepsilon ||x-u|^{\alpha/2} p^{1/2}(x-u;\varphi) - |x|^{\alpha/2} p^{1/2}(x;\varphi)|^2 dx \\
& + \int_{-\varepsilon}^0 ||x-u|^{\alpha/2} q^{1/2}(x-u;\varphi) - |x|^{\alpha/2} q^{1/2}(x;\varphi)|^2 dx \\
& + \int_{|x|>\varepsilon} |g^{1/2}(x-u;\varphi) - g^{1/2}(x;\varphi)|^2 dx. \qquad (2.5)
\end{aligned}$$

The first term in RHS of (2.5) is dominated by

$$(2/(1+\alpha))|u|^{1+\alpha}[\sup_{-\varepsilon_0<x<0} q(x;\varphi) + \sup_{0<x<\varepsilon_0} p(x;\varphi)]$$

and the coefficient of $|u|^{1+\alpha}$ in this expression has an exponential majorant in φ by (A7).

The second term in RHS of (2.5) is dominated by

$$\begin{aligned}
& 2\int_u^\varepsilon |p^{1/2}(x-u;\varphi) - p^{1/2}(x;\varphi)|^2 |x-u|^\alpha dx \\
& +2(\sup_{0<x<\varepsilon_0} p(x;\varphi)) \int_0^\varepsilon (|x-u|^{\alpha/2} - |x|^{\alpha/2})^2 dx. \qquad (2.6)
\end{aligned}$$

By (A7) and the arguments given in IH (p. 282), the first term has the required bound and so it suffices to show that the intergal in the second term of (2.6) is $O(|u|^{1+\alpha})$. This can be checked by elementary computations. In a similar manner, the third term in RHS of (2.5) has a bound $C_2(\varphi)|u|^{1+\alpha}$ where $C_2(\varphi)$ has at most an exponential growth.

By the Cauchy-Schwartz inequality and changing the order of integration, for sufficiently small u, we have

$$\begin{aligned} &\int_{|x|>\varepsilon} |g^{1/2}(x-u;\varphi) - g^{1/2}(x;\varphi)|^2 dx \\ = \;& u^2 \int_{|x|>\varepsilon} |(1/u) \int_{x-u}^{x} (\partial/\partial t) g^{1/2}(t;\varphi) dt|^2 dx \\ \leq \;& u^2 \int_{|x|>\varepsilon} (1/u) \int_{x-u}^{x} |(\partial/\partial t) g^{1/2}(t;\varphi)|^2 dt dx \\ \leq \;& u^2 \int_{|t|>\varepsilon/2} |(\partial/\partial t) g^{1/2}(t;\varphi)|^2 dt = u^2 J(\varphi;\varepsilon/2), \end{aligned} \tag{2.7}$$

where $J(\varphi,\varepsilon/2)$ is as in (A7). This completes the proof of Assertion (I).

Let us now prove Assertion (II). We note that as in IH (p. 53), we have

$$EZ_n^{1/2}(u,v) \leq \exp[-(n/2) r_2^2((0,\varphi_0),(k_n u, \varphi_0 + n^{-1/2}v))]. \tag{2.8}$$

We shall show that for all (u,v) satisfying $(u,\varphi_0+v) \in \Theta \times \Phi$,

$$r_2^2((0,\varphi_0),(u,\varphi_0+v)) \geq a(|u|^{1+\alpha} + \|v\|^2)/(1+|u|^{1+\alpha}+\|v\|^2), \tag{2.9}$$

where $a = a(\varphi_0) > 0$. The following result will be used in our proof:

LEMMA 2.1. *Under Assumptions* (A2) *and* (A6), *we have for any* $\varepsilon > 0$,

$$\begin{aligned} \int_{|x|>\varepsilon} &|g^{1/2}(x-u;\varphi_0+v) - g^{1/2}(x;\varphi_0) + u(\partial/\partial x) g^{1/2}(x;\varphi_0) \\ &- \sum_{i=1}^{d} v_i (\partial/\partial\varphi_i) g^{1/2}(x;\varphi_0)|^2 dx = o(|u|^2 + \|v\|^2). \end{aligned} \tag{2.10}$$

The proof of Lemma 2.1 is similar to that of Lemma 3.1 of Chapter 1. Now for sufficiently small and positive u,

$$\begin{aligned} &r_2^2((0,\varphi_0),(u,\varphi_0+v)) \\ \geq \;& \int_0^u |g^{1/2}(x-u;\varphi_0+v) - g^{1/2}(x;\varphi_0)|^2 dx \\ &+ \int_{|x|>\varepsilon} |g^{1/2}(x-u;\varphi_0+v) - g^{1/2}(x;\varphi_0)|^2 dx. \end{aligned} \tag{2.11}$$

By Lemma 2.1, the second term in RHS of (2.11) is equal to

$$u^2J(\varphi_0;\varepsilon) + 2u\sum_{i=1}^{d} v_iJ_i(\varphi_0;\varepsilon) + \sum_{i=1}^{d}\sum_{j=1}^{d} v_iv_jJ_{ij}(\varphi_0;\varepsilon) + o(u^2 + \|v\|^2),$$

which is bounded below by $\eta(u^2 + \|v\|^2)$ for some $\eta > 0$ in view of the assumption of positive definiteness in (A6).

It is now enough to show that for u and $\|v\|$ sufficiently small,

$$\int_0^u |g^{1/2}(x-u;\varphi_0+v) - g^{1/2}(x;\varphi_0)|^2 dx \geq a_1 u^{1+\alpha} \tag{2.12}$$

for some $a_1 > 0$. We write LHS of (2.12) as

$$\begin{aligned}
&\int_0^u ||x-u|^{\alpha/2}q^{1/2}(x-u;\varphi_0+v) - |x|^{\alpha/2}p^{1/2}(x;\varphi_0)|^2 dx \\
= &\int_0^u ||x-u|^{\alpha/2}q^{1/2}(0;\varphi_0+v) - |x|^{\alpha/2}p^{1/2}(0;\varphi_0)|^2 dx \\
&+\int_0^u ||x-u|^{\alpha/2}(q^{1/2}(x-u;\varphi_0) - q^{1/2}(0;\varphi_0)) \\
&\qquad - |x|^{\alpha/2}(p^{1/2}(x;\varphi_0) - p^{1/2}(0;\varphi_0))|^2 dx \\
&+2\int_0^u (|x-u|^{\alpha/2}(q^{1/2}(x-u;\varphi_0+v) - q^{1/2}(0;\varphi_0)) \\
&\qquad - |x|^{\alpha/2}(p^{1/2}(x;\varphi_0) - p^{1/2}(0;\varphi_0))) \\
&\qquad \times(|x-u|^{\alpha/2}q^{1/2}(0;\varphi_0) - |x|^{\alpha/2}p^{1/2}(0;\varphi_0))dx.
\end{aligned} \tag{2.13}$$

The first term in RHS of (2.13) is dominated by $a_2u^{1+\alpha}$ for some $a_2 > 0$. By (A1), the functions $p(x;\varphi)$ and $q(x;\varphi)$ are jointly continuous in (x,φ) and therefore, for any $\delta > 0$, the second term is bounded by $\delta u^{1+\alpha}$. Hence by an application of the Cauchy-Schwartz inequality, the third term is $o(u^{1+\alpha})$. Therefore, to prove (2.12), it suffices to show that the first term is bounded below by $a_3u^{1+\alpha}$ for some $a_3 > 0$. To avoid triviality, assume that both $p(0;\varphi_0)$ and $q(0;\varphi_0)$ are strictly positive. We note that there exists a constant $0 < c < 1$ such that whenever $0 < x < cu$, we have

$$|x-u|^{\alpha/2} - r^{1/2}|x|^{\alpha/2} \geq |x|^{\alpha/2} \tag{2.14}$$

where $r = p(0;\varphi_0)/q(0;\varphi_0)$. The inequality (2.12) now immediately follows from (2.13) and (2.14) and the observations made above. This proves (2.9) for all sufficiently small $|u|$ and $\|v\|$. Now by (A0), there exist $a > 0$ such that (2.9) holds for all u, v. From (2.8), we thus have $EZ_n^{1/2}(u,v) \leq \exp[-g_n(|u|,\|v\|)]$ where $g_n(x,y) = (a/2)(x^{1+\alpha}+y^2)/(1+(x^{1+\alpha}+y^2)/n)$. The sequence of functions $\{g_n(x,y)\}$ satisfies Conditions (a) and (b) of Assertion (II) if $\Theta \times \Phi$ is bounded (see Lemma

I.5.3 of IH). When $\Theta \times \Phi$ is unbounded, we further assume (A9) and the result follows as in Lemma I.5.4 of IH.

To prove Assertion (III), we first obtain an approximate expression for $Z_n(u,v)$. We consider a sequence $A_n \to \infty$ sucn that $k_n A_n \to 0$ and

$$\int_0^{k_n A_n|u|} \sup_{|w|\le k_n A_n|u|} |p(x-w;\varphi_0)-p(x;\varphi_0)||x|^\alpha dx$$
$$+\int_{-k_n A_n|u|}^0 \sup_{|w|\le k_n A_n|u|} |q(x-w;\varphi_0)-q(x;\varphi_0)||x|^\alpha dx = o(1/n).$$

Since the functions p and q are continuous, one can always choose such a sequence A_n. Define a random measure $\nu_n(B)$ on $[-A_n|u|, A_n|u|]$ as the number of variables X_i, $i=1,\ldots,n$, whose values belong to the set $k_n B$.

THEOREM 2.1. *Under Assumptions* (A1)–(A6) *and* (A8), *we have the following asymptotic representation:*

$$\begin{aligned}\log Z_n(u,v) \;&=\; \alpha \int_{-A_n|u|}^{A_n|u|} \log|1-(u/x)|(\nu_n(dx)-E\nu_n(dx))\\ &\quad+\log(q(0;\varphi_0)/p(0;\varphi_0))\int_0^u \nu_n(dx)\\ &\quad-[q(0;\varphi_0)\int_{-\infty}^0 (|1-(u/x)|^\alpha - 1 - \alpha\log|1-(u/x)|)|x|^\alpha dx\\ &\quad+p(0;\varphi_0)\int_0^\infty (|1-(u/x)|^\alpha - 1 - \alpha\log|1-(u/x)|)|x|^\alpha dx\\ &\quad+(1/(1+\alpha))(q(0;\varphi_0)-p(0;\varphi_0))]\\ &\quad+v'\Delta_n - (1/2)v'I(\varphi_0)v + r_n(u,v),\end{aligned} \tag{2.15}$$

where $\Delta_n = n^{-1/2}\sum_{i=1}^n h^{(1)}(X_i;\varphi_0)$, $h^{(1)} = (h_1,\ldots,h_d)'$ *and* $r_n(u,v)$ *converges in probability to zero.*

To prove Theorem 2.1, we note that

$$\begin{aligned}\log Z_n(u,v) \;&=\; \sum_{i=1}^n \log\frac{g(X_i-k_n u;\varphi_0)}{g(X_i;\varphi_0)} + \sum_{i=1}^n \log\frac{g(X_i-k_n u;\varphi_0+n^{-1/2}v)}{g(X_i-k_n u;\varphi_0)}\\ &=\; I_{1n}+I_{2n}\ \text{(say)}.\end{aligned} \tag{2.16}$$

As shown in IH, Ch. VI, we have

$$I_{1n} \;=\; \alpha\int_{-A_n|u|}^{A_n|u|} \log|1-(u/x)|(\nu_n(dx)-E\nu_n(dx))$$

$$
\begin{aligned}
&+\log(q(0;\varphi_0)/p(0;\varphi_0))\int_0^u \nu_n(dx) \\
&-[q(0;\varphi_0)\int_{-\infty}^0 (|1-(u/x)|^\alpha - 1 - \alpha\log|1-(u/x)|)|x|^\alpha dx \\
&+p(0;\varphi_0)\int_0^\infty (|1-(u/x)|^\alpha - 1 - \alpha\log|1-(u/x)|)|x|^\alpha dx \\
&+(1/(1+\alpha))(q(0;\varphi_0)-p(0;\varphi_0))] + o_p(1). \qquad (2.17)
\end{aligned}
$$

Expanding in Taylor's series, we have

$$
\begin{aligned}
I_{2n} &= n^{-1/2}\sum_{i=1}^n v'h^{(1)}(X_i - k_n u;\varphi_0) + (1/(2n))\sum_{i=1}^n v'h^{(2)}(X_i - k_n u;\varphi_0)v \\
&\quad +(1/(6n^{3/2}))\sum_{i=1}^n \sum_{1\le j,k,l\le d} v_j v_k v_l h_{jkl}(X_i - k_n u;\varphi_n), \qquad (2.18)
\end{aligned}
$$

where $h^{(2)} = ((h_{ij}))$ and φ_n lies on the line segment joining φ_0 and $\varphi_0 + n^{-1/2}v$. We rewrite the expression in RHS of (2.18) as

$$
I_{2n} = n^{-1/2}\sum_{i=1}^n v'h^{(1)}(X_i;\varphi_0) - (1/2)v'Iv + \varepsilon_{1n} + \varepsilon_{2n} + \varepsilon_{3n} \qquad (2.19)
$$

where

$$
\begin{aligned}
\varepsilon_{1n} &= n^{-1/2}\sum_{i=1}^n v'h^{(1)}(X_i - k_n u;\varphi_0) - n^{-1/2}\sum_{i=1}^n v'h^{(1)}(X_i;\varphi_0) \\
\varepsilon_{2n} &= (1/(2n))\sum_{i=1}^n v'h^{(2)}(X_i - k_n u;\varphi_0)v + (1/2)v'I(\varphi_0)v \\
\text{and } \varepsilon_{3n} &= (1/(6n^{3/2}))\sum_{i=1}^n \sum_{1\le j,k,l\le d} v_j v_k v_l h_{jkl}(X_i - k_n u;\varphi_n).
\end{aligned}
$$

It now remains to prove that $\varepsilon_{in} = o_p(1)$ for $i = 1, 2, 3$. We prove these in three steps.

STEP 1. We show that $E\varepsilon_{1n}^2 = o(1)$.

Note that

$$
\begin{aligned}
&E\varepsilon_{1n}^2 = E[v'h^{(1)}(X_1 - k_n u;\varphi_0) - h^{(1)}(X_1;\varphi_0)]^2 \\
&+(n-1)[E\{v'h^{(1)}(X_1 - k_n u;\varphi_0) - v'h^{(1)}(X_1;\varphi_0)\}]^2. \qquad (2.20)
\end{aligned}
$$

Under Assumptions (A2) and (A3), the first term in RHS of (2.19) converges to zero by the Dominated Convergence Theorem (DCT). Thus it is enough to show that for $i = 1, \ldots, d$

$$
E[h_i(X_1 - k_n u;\varphi_0) - h_i(X_1;\varphi_0)] = o(n^{-1/2}). \qquad (2.21)
$$

Fix $C > |u|$ and $\varepsilon < \varepsilon_0$. Then we have

$$\begin{aligned}
&|E[h_i(X_1 - k_n u; \varphi_0) - h_i(X_1; \varphi_0)]| \\
\leq\ & \int_{-Ck_n}^{Ck_n} (|h_i(x - k_n u; \varphi_0)| + |h_i(x; \varphi_0)|) g(x; \varphi_0) dx \\
&+ \int_{|x|>\varepsilon} |h_i(x - k_n u; \varphi_0) - h_i(x; \varphi_0)| g(x; \varphi_0) dx \\
&+ \int_{Ck_n}^{\varepsilon} |h_i(x - k_n u; \varphi_0) - h_i(x; \varphi_0)| g(x; \varphi_0) dx \\
&+ \int_{-\varepsilon}^{-Ck_n} |h_i(x - k_n u; \varphi_0) - h_i(x; \varphi_0)| g(x; \varphi_0) dx. \qquad (2.22)
\end{aligned}$$

By Assumption (A8), the first term in RHS of (2.22) is $O(1/n)$. By Assumption (A4) and an application of the Mean Value Theorem, the second term in RHS of (2.22) is dominated by $k_n u \int_{|x|>\varepsilon} H_4(x; \varphi_0) g(x; \varphi_0) dx$ which is $o(n^{-1/2})$. Now for $x \in (-\varepsilon, -Ck_n)$, both x and $x - k_n u$ are in $(-\varepsilon_0, 0)$ and therefore $g(x; \varphi_0) = |x|^\alpha q(x; \varphi_0)$ and $(\partial/\partial x) h_i(x; \varphi_0) = (\partial^2/\partial x \partial \varphi_i) \log q(x; \varphi_0)$. Thus the third term is $o(n^{-1/2})$ in view of Assumoption (A5). In a similar manner, one can show that the fourth term is also $o(n^{-1/2})$.

STEP 2. We show that $\varepsilon_{2n} = o_p(1)$.

For $i = 1, \ldots, n$, set

$$Y_{in} = v' h^{(2)}(X_i - k_n u; \varphi_0) v - E[v' h^{(2)}(X_i - k_n u; \varphi_0) v].$$

By Assumptions (A2) and (A3)(b) and DCT,

$$E[v' h^{(2)}(X_i - k_n u; \varphi_0) v] \to -v' I(\varphi_0) v,$$

and therefore

$$\begin{aligned}
&E[|Y_{1n}| \chi\{|Y_{1n}| > a\}] \\
\leq\ & E[(d\|v\|^2 H_2(X_1; \varphi_0) + |Ev' h^{(2)}(X_1 - k_n u; \varphi_0) v|) \\
&\times \chi\{d\|v\|^2 H_2(X_1; \varphi_0) + |Ev' h^{(2)}(X_1 - k_n u; \varphi_0) v| > a\}
\end{aligned}$$

which converges to zero as $a \to \infty$. Thus by a version of the Weak Law of Large Numbers (WLLN) for triangular arrays, the result follows.

STEP 3. By Assumption (A3)(c) and the WLLN, $\varepsilon_{3n} = o_p(1)$.

The theorem is now proved. □

We now obtain the limit of the LRP $Z_n(u,v)$. Let ν^+ and ν^- be two independent nonhomogeneous Poisson processes with rate functions $p(0;\varphi_0)x^\alpha$ and $q(0;\varphi_0)x^\alpha$ repectively. We define a stochastic process ν as follows:

$$\nu(x) = \begin{cases} \nu^+(x), & \text{if } x \geq 0, \\ \nu^-(-x), & \text{if } x < 0. \end{cases}$$

Let Δ be an $N_d(0, I(\varphi_0))$ random variable independent of ν.

THEOREM 2.2. *Under the assumptions of Theorem 2.1, the finite dimensional distributions of the LRP $Z_n(u,v)$ converge to those of the process*

$$Z(u,v) = Z^{(1)}(u)Z^{(2)}(v), \tag{2.23}$$

where

$$\begin{aligned} \log Z^{(1)}(u) \;=\; & \alpha \int_{-\infty}^{\infty} \log|1-(u/x)|(\nu(dx) - E\nu(dx)) \\ & + \log(q(0;\varphi_0)/p(0;\varphi_0)) \int_0^u \nu(dx) \\ & - [q(0;\varphi_0) \int_{-\infty}^{0} (|1-(u/x)|^\alpha - 1 - \alpha \log|1-(u/x)|)|x|^\alpha dx \\ & + p(0;\varphi_0) \int_0^{\infty} (|1-(u/x)|^\alpha - 1 - \alpha \log|1-(u/x)|)|x|^\alpha dx \\ & + (1/(1+\alpha))(q(0;\varphi_0) - p(0;\varphi_0))], \quad u \in \mathbb{R}, \end{aligned} \tag{2.24}$$

and

$$Z^{(2)}(v) = \exp[v'\Delta_n - (1/2)v'I(\varphi_0)v], \quad v \in \mathbb{R}^d. \tag{2.25}$$

In view of Theorem 2.1, the result follows from the following lemma:

LEMMA 2.2. *For any bounded intervals $I_0, I_1, \ldots, I_k$, the distribution of the vector $(\nu_n(I_1), \ldots, \nu_n(I_k), \Delta_n)$ converges to that of $(\nu(I_1), \ldots, \nu(I_k), \Delta)$ and further $\lim_{n\to\infty} E\nu_n(I_0) = E\nu(I_0)$.*

PROOF. As argued in the proof of Lemma VI.3.1 of IH, it is enough to prove for intervals $I_j = [\alpha_j, \beta_j)$, $j = 1, \ldots, k$, $0 \leq \alpha_1 < \beta_1 \leq \alpha_2 < \cdots \leq \beta_k$. Let $\psi(s)$ $s \in \mathbb{R}^d$ be the characteristic function (c.f.) of $h^{(1)}(X_1;\varphi_0)$. By (A8),

$$\exp[in^{-1/2}s'h^{(1)}(x;\varphi_0)] = 1 + o(1)$$

uniformly in a neighbourhood of 0. Also, by continuity of c.f., $\psi(n^{-1/2}s) = 1 + o(1)$. By the Cental Limit Theorem,

$$\psi^n(n^{-1/2}s) \to \exp[-(1/2)s'I(\varphi_0)s].$$

Now

$$
\begin{aligned}
& E[\exp\{\sum_{j=1}^{k} it_j\nu_n(I_j) + is'\Delta_n\}] \\
= \;& \{\sum_{j=1}^{k}(\exp(it_j)-1)\int_{k_nI_j}\exp[in^{-1/2}s'h^{(1)}(x;\varphi_0)]g(x;\varphi_0)\,dx \\
& \quad + \int \exp[in^{-1/2}s'h^{(1)}(x;\varphi_0)]g(x;\varphi_0)\,dx\}^n \\
= \;& \psi^n(n^{-1/2}s)(1+(1+o(1))\sum_{j=1}^{k}(\exp(it_j)-1)P\{X_j \in k_nI_j\})^n \\
= \;& \psi^n(n^{-1/2}s)(1+\frac{p(0;\varphi_0)}{n(1+\alpha)}\sum_{j=1}^{k}(\exp(it_j)-1)(\beta_j^{1+\alpha}-\alpha_j^{1+\alpha})+o(1/n))^n \\
\to \;& E(\exp[i\sum_{j=1}^{k} t_j\nu(I_j)])E(\exp[-(1/2)s'I(\varphi_0)s]), \qquad (2.26)
\end{aligned}
$$

which completes the proof of the first part by virtue of the Levy Continuity Theorem. *The second part is contained in Lemma VI.3.1 of IH.* □

In the remaining part of this section, we study the asymptotic behaviour of posterior distributions and Bayes estimates. The properties of the LRP (Assertions (I), (II) and (III)) established above and the results obtained in IH and Chapter 1 are used to prove our results. It is to be noted that the results of IH and Chapter 1 follow also from Assertions (I), (II) and (III) which are modified versions of the conditions of Theorem I.10.2 of IH. See, in this connection, Remark 2.1 of Chapter 2.

We consider prior densities which are positive and continuous at (θ_0, φ_0) and have polynomial majorants in θ and exponential majorants in $\|\varphi\|$. Let Π be the class of all such priors and $\mathfrak{L}$ be the class of continuous functions $l : \mathbb{R}\times\mathbb{R}^d \to [0,\infty)$ satisfying the following:

(i) $l(0,0) = 0$, $l(x,y) = l(-x,-y)$ $\forall x \in \mathbb{R}, y \in \mathbb{R}^d$.

(ii) The sets $\{(x,y) : l(x,y) < c\}$ are convex $\forall c > 0$ and are bounded if $c > 0$ is sufficiently small.

(iii) $l(x,y) \le B(1+|x|^{b_1})\exp[b_2\|y\|]$ for some $B, b_1, b_2 > 0$.

(iv) There exist $H_0, \gamma > 0$ such that for $H \ge H_0$,

$$\sup\{l(x,y) : |x| \le H^{\gamma}, \|y\| \le H^{\gamma}\} - \inf\{l(x,y) : |x| \ge H, \|y\| \ge H\} \le 0.$$

We immediately have the strong consistency of posterior distribution and Bayes esimates. The following result, a consequence of Theorem I.10.2 of IH, gives the asymptotic distribution of Bayes estimates.

THEOREM 2.3. *Let $(\widetilde{\theta}_n, \widetilde{\varphi}_n)$ be a Bayes estimate of (θ, φ) with respect to a prior $\pi \in \Pi$ and loss function $l(n^{1/(1+\alpha)}(x-\theta_0), n^{1/2}(y-\varphi_0))$ where $l \in \mathfrak{L}$. Assume Conditions* (A0)-(A9) *and let the random function*

$$\psi(s,t) = \int_{\mathbb{R}^{1+d}} l(s-u, t-v)\xi(u,v)dudv$$

attain its absolute minimum at a unique point $\tau = \tau(\theta_0, \varphi_0)$, where

$$\xi(u,v) = Z(u,v) / \int_{\mathbb{R}^{1+d}} Z(u', v')du'dv'.$$

Then

$$(n^{1/(1+\alpha)}(\widetilde{\theta}_n - \theta_0), n^{1/2}(\widetilde{\varphi}_n - \varphi_0)) \xrightarrow{d} \tau, \tag{2.27}$$

and for any continuous function $w(u,v)$ satisfying

$$|w(u,v)| \leq B(1+|u|^{b_1})\exp[b_2\|v\|]$$

for some $B, b_1, b_2 > 0$, we have

$$\lim_{n\to\infty} Ew(n^{1/(1+\alpha)}(\widetilde{\theta}_n - \theta_0), n^{1/2}(\widetilde{\varphi}_n - \varphi_0)) = Ew(\tau). \tag{2.28}$$

Also the diameter of the set of all normalized Bayes estimates with respect to the prior π and loss l converges to zero in probability.

If further the loss function $l \in \mathfrak{L}$ is of the form

$$l(x,y) = l_1(x) + l_2(y),$$

then $n^{1/(1+\alpha)}(\widetilde{\theta}_n - \theta_0)$ and $n^{1/2}(\widetilde{\varphi}_n - \varphi_0)$ are asymptotically independent.

To investigate whether the posterior distribution, suitably normalized and centered, converges to a limit, we use Theorem 3.4 of Chapter 1. In this case, $Z(u,v)$ is of product form and the criterion for existence of posterior limit is satisfied if and only if $\xi^{(1)}(u) := Z^{(1)}(u)/\int Z^{(1)}(u')du'$ is of the form $\psi(u+W)$ for some random variable W and a fixed probability density $\psi(\cdot)$ on $\mathbb{R}$. We consider only a simple but important case where $q(0;\varphi_0) = 0$ and $p(0;\varphi_0) > 0$; most of the examples of practical importance are of this type. As argued in Example 4.3 of Chapter 1, $\xi^{(1)}(u)$ is not of the form stated above. Consequently there does not exist a limit of the posterior distribution.

We immediately have the strong consistency of posterior distribution and Bayes esimates. The following result, a consequence of Theorem I.10.2 of IH, gives the asymptotic distribution of Bayes estimates.

THEOREM 2.3. *Let $(\widetilde{\theta}_n, \widetilde{\varphi}_n)$ be a Bayes estimate of (θ, φ) with respect to a prior $\pi \in \Pi$ and loss function $l(n^{1+\alpha}(x-\theta_0), n^{1/2}(y-\varphi_0))$ where $l \in \mathfrak{L}$. Assume Conditions* (A0)-(A9) *and let the random function*

$$\psi(s,t) = \int_{\mathbb{R}^{1+d}} l(s-u, t-v)\xi(u,v)dudv$$

attain its absolute minimum at a unique point $\tau = \tau(\theta_0, \varphi_0)$, where

$$\xi(u,v) = Z(u,v) / \int_{\mathbb{R}^{1+d}} Z(u', v')du'dv'.$$

Then

$$(n^{1+\alpha}(\widetilde{\theta}_n - \theta_0), n^{1/2}(\widetilde{\varphi}_n - \varphi_0)) \xrightarrow{d} \tau, \tag{2.27}$$

and for any continuous function $w(u,v)$ satisfying

$$|w(u,v)| \leq B(1+|u|^{b_1})\exp[b_2\|v\|]$$

for some $B, b_1, b_2 > 0$, we have

$$\lim_{n\to\infty} Ew(n^{1+\alpha}(\widetilde{\theta}_n - \theta_0), n^{1/2}(\widetilde{\varphi}_n - \varphi_0)) = Ew(\tau). \tag{2.28}$$

Also the diameter of the set of all normalized Bayes estimates with respect to the prior π and loss l converges to zero in probability.

If further the loss function $l \in \mathfrak{L}$ is of the form

$$l(x,y) = l_1(x) + l_2(y),$$

then $n^{1+\alpha}(\widetilde{\theta}_n - \theta_0)$ and $n^{1/2}(\widetilde{\varphi}_n - \varphi_0)$ are asymptotically independent.

To investigate whether the posterior distribution, suitably normalized and centered, converges to a limit, we use Theorem 3.4 of Chapter 1. In this case, $Z(u,v)$ is of product form and the criterion for existence of posterior limit is satisfied if and only if $\xi^{(1)}(u) := Z^{(1)}(u)/\int Z^{(1)}(u')du'$ is of the form $\psi(u+W)$ for some random variable W and a fixed probability density $\psi(\cdot)$ on $\mathbb{R}$. We consider only a simple but important case where $q(0;\varphi_0) = 0$ and $p(0;\varphi_0) > 0$; most of the examples of practical importance are of this type. As argued in Example 4.3 of Chapter 1, $\xi^{(1)}(u)$ is not of the form stated above. Consequently there does not exist a limit of the posterior distribution.

3 Singularities of the Second Type

We now consider the case with a density having a singulaity of the second type. The treatment of this case is similar to that in Section 2.

In addition to Assumptions (A0), (A2), (A3), (A4), (A6), (A8) and (A9) of Section 2, we make the following assumptions:

(A1)′ The functions $a(x;\varphi)$, $b(x;\varphi)$ and $\psi(x;\varphi)$ are continuously differentiable in (x,φ) in their respective domains of definition.

(A2)′ In their respective domains of definition, for all $i = 1,\ldots,d$, the functions $(\partial^2/\partial x\partial\varphi_i)\log\psi(x;\varphi)$, $(\partial^2/\partial x\partial\varphi_i)\log a(x;\varphi)$ and $(\partial^2/\partial x\partial\varphi_i)\log b(x;\varphi)$ are bounded.

(A3)′ The quantities $J(x;\varphi)$, $\varepsilon > 0$, $I_{ij}(\varphi)$, $i,j = 1,\ldots,d$, $\sup\{a(x;\varphi) : -\varepsilon_0 < x < 0\}$, $\sup\{b(x;\varphi) : 0 < x < \varepsilon_0\}$, $\sup\{g(x;\varphi) : -\varepsilon_0 < x < \varepsilon_0\}$, $\sup\{(\partial/\partial x)a(x;\varphi) : -\varepsilon_0 < x < 0\}$, $\sup\{(\partial/\partial x)b(x;\varphi) : 0 < x < \varepsilon_0\}$, $\int_{-\varepsilon_0}^{\varepsilon_0}(|(\partial/\partial x)\psi(x;\varphi)|^2/\psi(x;\varphi))dx$, $\int_{-\varepsilon_0}^{0}|(\partial/\partial x)a(x;\varphi)|^2dx$ and $\int_0^{\varepsilon_0}|(\partial/\partial x)b(x;\varphi)|^2dx$ have exponential majorants in φ.

REMARK 3.1. In IH, certain smoothness conditions on a, b and ψ are included in the definition of singularity. The conditions are stated in (1.5) and (1.6) of Section VI.1 of IH. It is to be noted that these conditions are satisfied with $\lambda = 2$ under the assumptions stated above (see pp. 282-283 of IH).

An example which falls in this framework is

$$f(x;\theta,\varphi) = C(\varphi)\exp[-e^{\varphi}|x-\theta|^{\alpha}], \quad x \in \mathbb{R},\ \theta \in \mathbb{R},\ \varphi \in \mathbb{R}.$$

We fix a parameter point $(\theta_0,\varphi_0) \in \Theta \times \Phi$. Let the likelihood ratio $Z_n(u,v)$ be defined as in (2.1). We first prove Assetions (I), (II) and (III) stated in Section 2. As mentioned in Remark 2.2, uniform versions of these assertions can be shown to be valid under suitable uniform versions of the assumptions. As before, we assume $\theta_0 = 0$ and proceed in a manner similar to that in Section 2.

As argued in Section 2 (see inequalities (2.2), (2.3) and (2.4)), it is enough to show that there exists a function $B(\varphi)$, exponential in $\|\varphi\|$ such that

$$\int |g^{1/2}(x-u;\varphi) - g^{1/2}(x;\varphi)|^2 dx \le B(\varphi)|u|^{1+\alpha}. \tag{3.1}$$

For definiteness, let $u > 0$. As in (2.6), for $\varepsilon < \varepsilon_0$ and u sufficiently small, LHS of (3.1) can be expressed as

$$\int_0^u |\psi^{1/2}(x-u;\varphi)\exp[(1/2)a(x;\varphi)|x-u|^{\alpha/2}]$$

$$
\begin{aligned}
&- \psi^{1/2}(x;\varphi)\exp[(1/2)b(x;\varphi)|x|^{\alpha/2}]|^2dx \\
&+ \int_u^{\varepsilon_0} |\psi^{1/2}(x-u;\varphi)\exp[(1/2)b(x;\varphi)|x-u|^{\alpha/2}] \\
&\quad - \psi^{1/2}(x;\varphi)\exp[(1/2)b(x;\varphi)|x|^{\alpha/2}]|^2dx \\
&+ \int_{-\varepsilon_0}^{0} |\psi^{1/2}(x-u;\varphi)\exp[(1/2)a(x;\varphi)|x-u|^{\alpha/2}] \\
&\quad - \psi^{1/2}(x;\varphi)\exp[(1/2)a(x;\varphi)|x|^{\alpha/2}]|^2dx \\
&+ \int_{|x|>\varepsilon} |g^{1/2}(x-u;\varphi) - g^{1/2}(x;\varphi)|^2dx. \qquad (3.2)
\end{aligned}
$$

The first term in (3.2) is dominated by

$$
\begin{aligned}
&3\int_0^u (\psi^{1/2}(x-u;\varphi) - \psi^{1/2}(x;\varphi))^2dx \\
&+3\int_0^u \psi(x-u;\varphi)(\exp[(1/2)b(x-u;\varphi)|x-u|^{\alpha/2}]-1)^2dx \\
&+3\int_0^u \psi(x;\varphi)(\exp[(1/2)a(x;\varphi)|x-u|^{\alpha/2}]-1)^2dx. \qquad (3.3)
\end{aligned}
$$

In view of Remark 3.1 and Assumption (A3)$'$, the first term in (3.3) has an appropriate bound while the other two term can be tackled using the inequality $|e^x - 1| \le |x|(e^x+1)$, $x \in \mathbb{R}$.

Using the Mean Value Theorem and Assumptions (A1)$'$ and (A3)$'$, we get an appropriate bound for the second and the third term in (3.2). The treatment of the fourth term is same as that in (2.5). This completes the proof of (3.1).

In order to prove Assertion (II), we proceed as in Section 2 (see (2.8)– (2.12)) and note that LHS of (2.12) can be written as

$$
\begin{aligned}
&\int_0^u |\psi^{1/2}(x-u;\varphi_0)\exp[(1/2)a(x;\varphi_0)|x-u|^{\alpha/2}] \\
&\quad - \psi^{1/2}(x;\varphi_0)\exp[(1/2)b(x;\varphi_0)|x|^{\alpha/2}]|^2dx. \qquad (3.4)
\end{aligned}
$$

Using the expansion $e^x = 1 + x + O(x^2)$ (as $x \to 0$) and the fact that

$$
\int_0^u |\psi^{1/2}(x-u;\varphi_0+v) - \psi^{1/2}(x;\varphi_0)|^2dx = o(u^2 + \|v\|^2),
$$

the expression in (3.4) can be written as

$$
(1/4)\int_0^u ||x-u|^{\alpha/2}b(x-u;\varphi_0+v) - |x|^{\alpha/2}a(x;\varphi_0)|^2dx + o(u^{1+\alpha} + \|v\|^2).
$$

The result now follows as in Section 2.

Before proving Assertion (III), we obtain an approximate expression for $Z_n(u, v)$. We define a random measure $\mu_n(B)$ on $[-A_n|u|, A_n|u|]$ as the number of variables $X_i \in k_n B$.

THEOREM 3.1. *Under Assumptions* (A2)–(A4), (A6), (A8), (A1)′, (A2)′, $Z_n(u,v)$ *admits the following asymptotic representation*:

$$\begin{aligned}
Z_n(u,v) \quad = \quad & n^{-\alpha/(2(1+\alpha))} \int_{-A_n|u|}^{A_n|u|} (d(x-u;\varphi_0)|x-u|^{\alpha/2} \\
& \qquad - d(x;\varphi_0)|x|^{\alpha/2})(\mu_n(dx) - E\mu_n(dx)) \\
& -(1/2)\psi(0;\varphi_0)|u|^{1+\alpha}\frac{\Gamma(1+\alpha/2)\Gamma((1-\alpha)/2)}{2^{1+\alpha}\pi^{1/2}(1+\alpha)} \\
& \quad \times(a^2(0;\varphi_0) + b^2(0;\varphi_0) - 2a(0;\varphi_0)b(0;\varphi_0)\cos(\pi\alpha/2)) \\
& +v'\Delta_n - (1/2)v'I(\varphi_0)v + r_n(u,v), \qquad (3.5)
\end{aligned}$$

where

$$d(x;\varphi_0) = \begin{cases} a(x;\varphi_0), & \text{if } x < 0, \\ b(x;\varphi_0), & \text{if } x > 0, \end{cases}$$

$\Delta_n = n^{-1/2}\sum_{i=1}^n h^{(1)}(X_i;\varphi_0)$ *and* $r_n(u,v)$ *converges in probability to zero.*

To prove Theorem 3.1, we express $\log Z_n(u,v)$ as in (2.15). An approximation for the first term I_{1n} is obtained in Lemma VI.4.2 of IH. Proceeding as in Section 2, one can show that the second term I_{2n} can be approximated by $v'\Delta_n - (1/2)v'I(\varphi_0)v$.

We now obtain the limit of the LRP $Z_n(u,v)$. Let a random process

$$b(u) = \begin{cases} b_1(u), & \text{if } x \geq 0, \\ b_2(-u), & \text{if } x < 0, \end{cases}$$

where b_1 and b_2 are independent standard Wiener processes. Define a random measure μ on $\mathbb{R}$ by

$$\mu(U) = \int_U db, \quad U \in \mathcal{B}.$$

Let Δ be an $N_d(0, I(\varphi_0))$ radom variable independent of μ.

THEOREM 3.2. *Under the assumptions of Theorem* 3.1, *the finite dimensional distributions of the LRP* $Z_n(u,v)$ *converges to those of the process*

$$Z(u,v) = Z^{(1)}(u)Z^{(2)}(v), \qquad (3.6)$$

where

$$\begin{aligned} Z^{(1)}(u) &= \int_{-\infty}^{\infty}(d(x-u;\varphi_0)|x-u|^{\alpha/2} - d(x;\varphi_0)|x|^{\alpha/2})\mu(dx) \\ &\quad -(1/2)\psi(0;\varphi_0)|u|^{1+\alpha}\frac{\Gamma(1+\alpha/2)\Gamma((1-\alpha)/2)}{2^{1+\alpha}\pi^{1/2}(1+\alpha)} \\ &\quad (a^2(0;\varphi_0)+b^2(0;\varphi_0)-2a(0;\varphi_0)b(0;\varphi_0)\cos(\pi\alpha/2)), \quad u \in \mathbb{R} \end{aligned} \tag{3.7}$$

and

$$Z^{(2)}(v) = \exp[v'\Delta_n - (1/2)v'I(\varphi_0)v], \quad v \in \mathbb{R}^d. \tag{3.8}$$

In view of Theorem 3.1, the result follows from the following lemma.

LEMMA 3.1. *For any bounded intervals $I_1, I_2, \ldots, I_k$, the distribution of the vector $(n^{-\alpha/(2(1+\alpha))}(\mu_n(I_1) - E\mu_n(I_1)), \ldots, n^{-\alpha/(2(1+\alpha))}(\mu_n(I_k) - E\mu_n(I_k)), \Delta_n)$ converges to that of $(\psi^{1/2}(0;\varphi_0)\mu(I_1), \ldots, \psi^{1/2}(0;\varphi_0)\mu(I_k), \Delta)$.*

PROOF. It is enough to prove the result for disjoint intervals. Let $\beta_n(t)$, $\beta(t)$, $t \in \mathbb{R}^k$ and $\gamma(s)$, $s \in \mathbb{R}^d$ be the c.f.'s of $((\chi\{X_1 \in k_nI_j\} - P\{X_1 \in I_j\}), j = 1, \ldots, k)$, $\psi^{1/2}(0;\varphi_0)(\mu(I_k), \ldots, \mu(I_k))$ and $h^{(1)}(X_1;\varphi_0)$ respectively. Let δ stand for the number $\alpha/(2(1+\alpha))$. By Lemma VI.4.3 of IH (or by dirtect computations), we have

$$(\beta_n(n^{-\delta}t))^n \to \beta(t).$$

It can be easily shown that

$$\sum_{j=1}^{k}(e^{it_j} - 1)P\{X_1 \in k_nI_j\} = \beta_n(t)\exp[i\sum_{j=1}^{k} t_jP\{X_1 \in k_nI_j\}] - 1.$$

Also, by Assumption (A8), one can show that

$$\int_{k_nI_j} \exp[in^{-1/2}s'h^{(1)}(x;\varphi_0)]g(x;\varphi_0)dx = P\{X_1 \in k_nI_j\} + o(1/n), \quad j = 1, \ldots, k.$$

Therefore the joint c.f. can be written as

$$\begin{aligned} &\{\exp[-in^{-\delta}\sum_{j=1}^{k} t_jP\{X_1 \in k_nI_j\}](\sum_{j=1}^{k}(\exp[in^{-\delta}t_j] - 1)P\{X_1 \in k_nI_j\} \\ &\quad + o(1/n) + \gamma(n^{-1/2}s))\}^n \\ = \ &\{\exp[-in^{-\delta}\sum_{j=1}^{k} t_jP\{X_1 \in k_nI_j\}](\beta_n(n^{-\delta}t)\exp[in^{-\delta}\sum_{j=1}^{k} t_jP\{X_1 \in k_nI_j\}] \\ &\quad - 1 + \gamma(n^{-1/2}s) + o(1/n))\}^n. \end{aligned}$$

Since $\gamma(n^{-1/2}s) = 1 - (1/(2n))s'I(\varphi_0)s + o(1/n)$, $\exp[in^{-\delta}\sum_{j=1}^{k} t_j P\{X_1 \in k_n I_j\}] = 1 + o(1)$ and $\beta_n(n^{-\delta}t) = 1 + o(1)$, the c.f. is equal to

$$\begin{aligned} & \{\beta_n(n^{-\delta}t) + (1+o(1))(-(1/(2n))s'I(\varphi_0)s + o(1/n))\}^n \\ = \; & (\beta_n(n^{-\delta}t))^n (1 - (1/(2n))s'I(\varphi_0)s + o(1/n))^n, \end{aligned}$$

which converges to $\beta(t)\exp[-(1/2)s'I(\varphi_0)s]$. This completes the proof. □

Asymptotic properties of Bayes estimates can now be obtained using the general results of IH. The statement of the result is exactly same as Theorem 2.3.

4 Discussions

In the definition of singularities, it is assumed that α lies strictly between 0 and 1 (first and second type) or -1 and 0 (third type). It is well known that the case $\alpha > 1$ corresponds to the regular situations. The boundary case $\alpha = 1$ leads to the "almost smooth" family (described in Example 3 of Section 1.2); the proper normalizing factor in this case is $(n \log n)^{-1/2}$. In presence of an additional "regular" paramter φ, this case can be treated in a manner similar to that in this chapter under appropriate assumptions.

In the present chapter, we have considered a multiparamter family of densities involving two different kinds of parameters θ and φ with different normalizers. The LRP turned out to be a product of two independent processes – one involving the normalized parameter (u) corresponding to θ only and the other involving that (v) corresponding to φ. As a consequence, we have the follwing results:

(1) The joint posterior density of $n^{1/(1+\alpha)}(\theta - \theta_0)$ and $n^{1/2}(\varphi - \varphi_0)$ given the observations $X_1, \ldots, X_n$ converges weakly (as a random density on $\mathbb{R} \times \mathbb{R}^d$) to a random density which is of the product type with probability one. Moreover, as random functions, the two factors are independent. (See Theorem 3.3 of Chapter 1.)

(2) The joint asymptotic distribution of the normalized version of the Bayes estimates $n^{1/(1+\alpha)}(\tilde{\theta}_n - \theta_0)$ and $n^{1/2}(\tilde{\varphi}_n - \varphi_0)$ is of the product type if the loss function is of the form

$$l(x, y) = l_1(x) + l_2(y).$$

In other words, the (normalized) Bayes estimates are asymptotically independent.

Smith (1985) obtained the joint asymptotic distribution of the maximum likelihood estimates (or corrected MLE's) of θ and φ which also turned out to be of the product type.

In view of these observations, we again note that the estimation problem of θ and φ, when considered together, are asymptotically independent in the sense of Section 1.6.

In Chapter 2, the same phenomenon was observed in the context of multiparamter discontinuous densities. Thus we may expect in general that in presence of two kinds of parameters with normalizing factors of different orders, the limiting LRP can be expressed as a product of two independent processes as above. As a consequence, the estimation problems of θ and φ would then be asymptotically independent.

Chapter 4

Expansion of Entropy Risk: Reference Prior

1 Introduction

Let $X_1, X_2, \ldots, X_n$ be independent observations each having a distribution P_θ with a density $f(x;\theta)$ with respect to a fixed dominating measure where $\theta \in \Theta$, an open subset of $\mathbb{R}$. Consider a prior on Θ having a density $\pi(\cdot)$ with respect to the Lebesgue measure. Lindley's (1965) measure of information $I(\pi; X^n)$ in $X^n = (X_1, \ldots, X_n)$ about θ is given by the average relative entropy or Kullback-Leibler distance between the posterior ditribution of θ given X^n and the prior π (see Sec. 2). This measure is also equal to the average (with respect to π) relative entropy distance between the joint distribution of X^n given θ and the marginal distribution of X^n and indeed is the Bayes risk when one estimates the density of X^n given θ using the entropy loss (see Aitchison (1975)).

In Section 2 of this chapter, we obtain an asymptotic expansion of this Bayes risk (or information) $I(\pi; X^n)$ for a family of nonregular cases. Since such an expansion leads to a posterior convergence result in the entropy distance which in turn implies L^1-convergence, we restrict our attention to the case which, by results of Chapter 1, is essentially the only case where a posterior convergence holds. Our treatment is similar to that of Clarke and Barron (1990a, b) who obtained an expansion of the entropy risk for the regular cases. Results similar to those of Clarke and Barron (1990b) were obtained earlier by Ibragimov and Has'minskii (1973). For extensions to non i.i.d. cases, see, for example, Polson (1988). Rissanen (1986, 1987) obtained certain related results.

The reference prior method for development of noninformative priors was initiated by Bernardo (1979) and developed further in a number of papers including Berger and Bernardo (1989, 1992a, b, c), Berger, Bernardo and Mendoza (1989), Ghosh and Mukherjee (1992) and Chang and Eaves (1990). Bernardo (1979) considered the measure $I(\pi; X^n)$ as a measure of information in X^n about θ and argued that the larger the measure, the less informative is the prior. The reference prior is thus defined as a π that maximizes this measure (in an asymptotic sense). Since this measure is also the Bayes risk with respect to the entropy loss, maximizing $I(\pi; X^n)$ would lead to an (asymptotically) least favourable π and therefore under reasonable conditions, the corresponding Bayes estimate is (asymptotically) minimax. Also, the reference priors usually have the property that the corresponding procedures match with some standard frequentist procedures upto a certain order.

In Section 3, we use the asymptotic expansion obtained in Section 2 to find a prior that maximizes $I(\pi; X^n)$ in an asymptotic sense. Explicit forms are derived in some important examples.

2 Expansion of Bayes Risk for Entropy Loss

Let $X_1, X_2, \ldots$ be i.i.d. observations with a distribution P_θ having a density $f(x;\theta)$, $\theta \in \Theta$ where Θ is an open interval in $\mathbb{R}$. We shall occasionally abbreviate P_θ as θ. We assume that the densities $f(x;\theta)$ satisfy the following conditions:

(A1) Uniformly on compact subsets of Θ,

$$\inf_{|h|>\varepsilon} r_2^2(\theta, \theta+h) > 0, \quad \varepsilon > 0,$$

where $r_2^2(\theta, \theta+h) = \int (f^{1/2}(x;\theta) - f^{1/2}(x;\theta+h))^2 dx$ is the squared Hellinger distance between $f(\cdot;\theta)$ and $f(\cdot;\theta+h)$.

(A2) The density $f(x;\theta)$ is supported on an interval $S(\theta) := [a_1(\theta), a_2(\theta)]$, where it is possible that $a_1(\theta) = -\infty$ or $a_2(\theta) = \infty$, but not both. If an endpoint is finite, $f(x;\theta)$ has a discontinuity at that point. In the region $a_1(\theta) < x < a_2(\theta)$, $f(x;\theta)$ is continuously differentiable in θ with derivative $f'(x;\theta)$. Moreover, $S(\theta)$ is either increasing or decreasing in θ.

(A3) If not infinite, $a_1(\theta)$ and $a_2(\theta)$ are continuously differentiable in θ and $a_k'(\theta) \neq 0$, $k = 1, 2$.

(A4) The limits

$$p(\theta) = \lim_{x \downarrow a_1(\theta)} f(x;\theta), \qquad q(\theta) = \lim_{x \uparrow a_2(\theta)} f(x;\theta)$$

exist $\forall \theta \in \Theta$, and the above convergences are uniform on compact subsets of Θ. Moreover, $p(\theta)$ and $q(\theta)$ are continuous in θ.

(A5) The function $\int |f'(x;\theta)| dx$ is finite and continuous in θ.

(A6) The functions $p(\theta)$, $q(\theta)$ and $\int |f'(x;\theta)| dx$ have polynomial majorants.

(A7) As $|u| \to \infty$, for some $\gamma > 0$,

$$\int f^{1/2}(x;\theta) f^{1/2}(x;\theta+u) dx \le B(\theta)|u|^{-\gamma}$$

where $B(\theta)$ is bounded on compact subsets of Θ.

(A8) On the set $a_1(\theta) < x < a_2(\theta)$, $\log f(x;\theta)$ is twice continuously differentiable in θ, for any θ and there is a neighbourhood N_θ of θ such that

$$\sup_{t\in N_\theta} |(\partial^2/\partial t^2)\log f(x;t)| \le H_\theta(x),$$

where $E_\theta H_\theta(X_1)$ is finite and continuous in θ.

We restrict our attention to the case where $S(\theta)$ is decreasing; treatment for the other case is similar. Let $X_{1:n} \le \cdots \le X_{n:n}$ be the order statistics based on $X_1, \ldots, X_n$. We define a random variable W_n by

$$W_n = \min\{a_1^{-1}(X_{1:n}), a_2^{-1}(X_{n:n})\}$$

(The first term is ignored if $a_1(\theta) \equiv -\infty$, and the second term is ignored if $a_2(\theta) \equiv \infty$.) Note that $X_{1:n} \downarrow a_1(\theta)$ and $X_{n:n} \uparrow a_2(\theta)$ a.s. $[P_\theta]$, and hence for almost all samples, W_n is defined for all n, sufficiently large. Now the likelihood function $p^n(X^n;\theta) = \prod_{i=1}^n f(X_i;\theta)$ is positive if and only if $\theta \le W_n$. Also define $c(\theta) = E_\theta((\partial/\partial\theta)\log f(X_1;\theta))$ and observe that

$$c(\theta) = a_1'(\theta)p(\theta) - a_2'(\theta)q(\theta) > 0. \tag{2.1}$$

Let $\sigma_n = \sigma_n(\theta) = n(W_n - \theta)$. We make one more assumption.

(A9) For any compact subset K of Θ,

$$\sup_{\theta\in K}\sup_{n\ge 1} E_\theta\sigma_n < \infty.$$

We shall give useful sufficient conditions for this assumption at the end of this section.

Let $\pi(\theta)$ be a (proper) prior density on Θ. The Lindley's measure of information $I(\pi;X^n)$ is defined to be the expected Kullback-Leibler distance between the posterior and the prior, i.e.,

$$I(\pi;X^n) = H(\pi) - H_{X^n}(\pi),$$

where

$$\begin{aligned} H(\pi) &= -\int \pi(\theta)\log\pi(\theta)d\theta, \\ H_{X^n}(\pi) &= EH(\theta|X^n), \\ H(\theta|X^n) &= -\int \pi(\theta|X^n)\log\pi(\theta|X^n)d\theta \end{aligned}$$

and $\pi(\theta|X^n)$ stands for the posterior obtained from $\pi(\theta)$.

We shall derive the following theorem which is the main result of this section:

THEOREM 2.1. *Assume Conditions* (A1)–(A9). *Let K be a compact subset of Θ and $\pi(\theta)$ be a proper prior which is positive, continuous and concentrated on K. Then as $n \to \infty$,*

$$I(\pi; X^n) = \log(n/e) + \int_K \pi(\theta) \log(c(\theta)/\pi(\theta))d\theta + o(1). \tag{2.2}$$

The proof of this theorem will be a long one and we shall break it into several parts. We shall show that

$$\liminf_{n\to\infty}[I(\pi; X^n) - \log n] \geq -1 + \int_K \pi(\theta) \log(c(\theta)/\pi(\theta))d\theta \tag{2.3}$$

and

$$\limsup_{n\to\infty}[I(\pi; X^n) - \log n] \leq -1 + \int_K \pi(\theta) \log(c(\theta)/\pi(\theta))d\theta. \tag{2.4}$$

We first establish (2.3), the proof of which is much easier than that of (2.4). We observe the following entropy maximization property of the (negative of) exponential distribution (see, e.g., p. 217 of Rao (1973)).

LEMMA 2.1. *Let $\mathfrak{F}$ be the class of all densities on $(-\infty, 0]$ having expectation $-\mu$, $\mu > 0$. Let $g(x) = \mu^{-1}\exp[x/\mu]\chi\{x \leq 0\}$. Then for all $f \in \mathfrak{F}$,*

$$-\int f(x) \log f(x)dx \leq -\int g(x) \log g(x)dx. \tag{2.5}$$

PROOF. Clearly $g \in \mathfrak{F}$. Since $\int_{-\infty}^0 f(x)\log(f(x)/g(x))dx \geq 0$, for any $f \in \mathfrak{F}$, we have

$$\begin{aligned} -\int f(x) \log f(x)dx &\leq -\int f(x) \log g(x)dx \\ &= \log\mu + 1. \end{aligned} \tag{2.6}$$

But $-\int g(x)\log g(x)dx = \log\mu + 1$, which proves (2.5). □

REMARK 2.1. An analogous version of Lemma 2.1 for densities on the positive half line is also true.

Let $m_n(x^n) = \int p^n(x^n; \theta)\pi(\theta)d\theta$ be the marginal density of X^n. The following result is an important step in the proof of Theorem 2.1, and is also of interest in its own right.

PROPOSITION 2.1. *Under Assumptions* (A1)–(A7),

(i) $\log(m_n(X^n)/p^n(X^n;\theta)) + \log n + \log(c(\theta)/\pi(\theta)) - c(\theta)\sigma_n \xrightarrow{P_\theta^n} 0$,

(ii) $n(E(t|X^n) - W_n) \xrightarrow{P_\theta^n} -1/c(\theta)$,

where t stands for a dummy variable for the parameter.

PROOF. By the general results derived in IH (Ch. V),

$$p^n(X^n;\theta+u/n)/p^n(X^n;\theta) = \exp[c(\theta)u]\chi\{u<\sigma_n\} + o_p(1), \tag{2.7}$$

where the convergences are uniform in $|u| \le H$ for any $H > 0$.

Now

$$\begin{aligned}
&\int \left|\pi(\theta+u/n)\frac{p^n(X^n;\theta+u/n)}{p^n(X^n;\theta)} - \pi(\theta)\exp[c(\theta)u]\chi\{u<\sigma_n\}\right|du \\
\le\ & \int_{|u|\le H} |\pi(\theta+u/n)-\pi(\theta)|\frac{p^n(X^n;\theta+u/n)}{p^n(X^n;\theta)}du \\
&+ \int_{|u|\le H} \pi(\theta)\left|\frac{p^n(X^n;\theta+u/n)}{p^n(X^n;\theta)} - \exp[c(\theta)u]\chi\{u<\sigma_n\}\right| du \\
&+ \int_{|u|>H} \pi(\theta+u/n)\frac{p^n(X^n;\theta+u/n)}{p^n(X^n;\theta)}du \\
&+ \int_{|u|>H} \pi(\theta)\exp[c(\theta)u]\chi\{u<\sigma_n\}du.
\end{aligned} \tag{2.8}$$

Let $\varepsilon > 0$ and $\delta > 0$ be given. We can get $H > 0$ such that $P\{\sigma_n > H\} < \delta/4$ and $(\pi(\theta)/c(\theta))\exp[-c(\theta)H] < \varepsilon/4$. Thus the last term is less than $\varepsilon/4$ with probability greater than $1-\delta/4$. The third term can also be made less than $\varepsilon/4$ with probability greater than $1-\delta/4$ by Lemma I.5.2 of IH, provided H is chosen to be large enough. Now, for such an H, choose n large enough so that the first two terms are less than $\varepsilon/4$ with probability greater than $1-\delta/4$, by virtue of (2.7). Thus

$$\int \left|\pi(\theta+u/n)\frac{p^n(X^n;\theta+u/n)}{p^n(X^n;\theta)} - \pi(\theta)\exp[c(\theta)u]\chi\{u<\sigma_n\}\right|du \xrightarrow{P_\theta^n} 0 \tag{2.9}$$

which also leads to the posterior approximation:

$$\int \left|\frac{\pi(\theta+u/n)p^n(X^n;\theta+u/n)}{nm_n(X^n;\theta)} - c(\theta)\exp[c(\theta)(u-\sigma_n)]\chi\{u<\sigma_n\}\right|du \xrightarrow{P_\theta^n} 0. \tag{2.10}$$

From (2.9), we immediately have

$$n\, m_n(X^n)/p^n(X^n;\theta) - (\pi(\theta)/c(\theta))\exp[c(\theta)\sigma_n] \xrightarrow{P_\theta^n} 0 \tag{2.11}$$

which is equivalent to (i). Part (ii) is proved using a decomposition similar to (2.8).

□

REMARK 2.2. An almost sure version of Proposition 2.1 can also be obtained under some further assumptions, see Samanta (1988, Ch. 3).

We now prove (2.3). Let $v = n(t - W_n)$. Then the posterior density of v is

$$\pi_n^*(v) = n^{-1}\pi_n(W_n + v/n),$$

where π_n is the posterior density of t. Note that $\pi_n^*(v)$ is concentrated on $(-\infty, 0]$. Thus $I(\pi; X^n)$ is equal to

$$\begin{aligned}
& -\int_K \pi(\theta)\log\pi(\theta)d\theta + \int m_n(x^n)\int_K \pi_n(\theta)\log\pi_n(\theta)d\theta dx^n \\
= & -\int_K \pi(\theta)\log\pi(\theta)d\theta + \int m_n(x^n)\int_{-\infty}^0 \pi_n^*(v)[\log n + \log\pi_n^*(v)]dv dx^n \\
= & -\int_K \pi(\theta)\log\pi(\theta)d\theta + \log n + \int m_n(x^n)\int_{-\infty}^0 \pi_n^*(v)\log\pi_n^*(v)dv dx^n \\
\geq & -\int_K \pi(\theta)\log\pi(\theta)d\theta + \log n - \int m_n(x^n)(1 + \log E(-v|X^n))dx^n \qquad (2.12)
\end{aligned}$$

by Lemma 2.1. By Proposition 2.1 (ii),

$$E(-v|X^n) \xrightarrow{P_\theta^n} (c(\theta))^{-1}.$$

Further, $E(-v|X^n) = n(W_n - E(t|X^n)) = \sigma_n - n(E(t|X^n) - \theta)$ (t is a dummy variable for the parameter) is uniformly integrable by (A9) and Theorem I.5.2 of IH. (Here the uniformity is also with respect to θ belonging to compacts.) Hence $\log E(-v|X^n)$ is uniformly integrable from above and so

$$\begin{aligned}
& \liminf_{n\to\infty}(I(\pi; X^n) - \log(n/e) + \int_K \pi(\theta)\log\pi(\theta)d\theta) \\
\geq & -\limsup_{n\to\infty}\int_K \pi(\theta)\int \log E(v|x^n)p^n(x^n;\theta)dx^n d\theta \\
\geq & -\int_K \pi(\theta)\log(c(\theta))^{-1}d\theta = \int_K \pi(\theta)\log c(\theta)d\theta, \qquad (2.13)
\end{aligned}$$

which is equivalent to (2.3). □

We shall now prove (2.4). Fix a $\theta \in \Theta$ and let

$$\begin{aligned}
R_n &= -\log(m_n(x^n)/p^n(x^n;\theta)) - \log n, \\
\psi_n(\theta) &= K(P_\theta^n; m_n) - \log n = E_\theta R_n, \\
\psi(\theta) &= \log(c(\theta)/\pi(\theta)) - 1
\end{aligned}$$

where $K(P_\theta^n; m_n)$ stands for the Kullback-Leibler information number between the two probability measures P_θ^n and m_n. Note that

$$I(\pi; X^n) = \int_K \psi_n(\theta)\pi(\theta)d\theta. \qquad (2.14)$$

The proof of (2.4) follows from the following result:

LEMMA 2.2. *Under Assumptions* (A1)–(A7), *we have*

(a) $\limsup_{n\to\infty} \psi_n(\theta) \le \psi(\theta)$,

(b) $\psi_n(\theta)$ *is uniformly dominated by an integrable function on* K *from above.*

The following result will be used to prove Lemma 2.2.

LEMMA 2.3. *Let* $\theta \in \Theta$. *Then under the true parameter* θ, R_n *is uniformly integrable from above, i.e., there exists a uniformly integrable sequence of random variables* R_n' *such that* $R_n \le R_n'$, $n \ge 1$.

Below, we need the following notion purely for technical reasons.

DEFINITION 2.1. Let P and Q be two probability measures on a measurable space $(\Omega, \mathcal{A})$ and let h stand for the Radon-Nikodym derivative of the absolutely continuous part of P with respect to Q. The *modified Kullback-Leibler information number* $K^*(Q; P)$ of Q with respect to P is defined by

$$K^*(Q;P) = -\int_{\{h>0\}} \log h(\omega) Q(d\omega).$$

Unlike the usual Kullback-Leibler information number $K(Q;P)$, $K^*(Q;P)$ can be negative and hence cannot be viewed as a distance measure. We do not claim any statistical significance of it. Nevertheless, it has the following obvious but important property, which we need later:

Let P^n and Q^n be two symmetric product probabilities. Then

$$K^*(Q^n;P^n) = nK^*(Q;P)(Q\{h>0\})^{n-1}. \tag{2.15}$$

PROOF OF LEMMA 2.3. For any $A > 0$,

$$\begin{aligned} R_n &= -\log[(1/p^n(x^n;\theta)) \int_K p^n(x^n;t)\pi(t)dt] - \log n \\ &\le -\log[\int_{\{-A\le u\le \sigma_n(A)\}} \frac{p^n(x^n;\theta+u/n)}{p^n(x^n;\theta)}\pi(\theta+u/n)du] \end{aligned} \tag{2.16}$$

where $\sigma_n(A) = \min\{\sigma_n, A\}$.

Now applying Jensen's inequality, the RHS of (2.16) can be bounded by

$$\begin{aligned} &-(A+\sigma_n(A))^{-1}\int_{\{-A\le u\le\sigma_n(A)\}} \log(p^n(x^n;\theta+u/n)/p^n(x^n;\theta))du \\ &-\log(A+\sigma_n(A)) - \log\inf\{\pi(\theta+u/n) : |u|\le A\}. \end{aligned} \tag{2.17}$$

The second term in (2.17) is uniformly bounded and the last term converges to $\log \pi(\theta) > -\infty$. Hence it is enough to show that the first term is bounded by an uniformly integrable function from above. Now

$$[\log(p^n(X^n;\theta+u/n)/p^n(X^n;\theta)) - c(\theta)u]\chi\{u<\sigma_n\} \xrightarrow{P_\theta^n} 0 \tag{2.18}$$

for any fixed u. Therefore, by Fubini's theorem, the above convergence also holds in the joint probability $(\nu \times P_\theta^n)$ where ν is the uniform distribution on $[-A, A]$. So it is now enough to show that

$$[\log(p^n(X^n;\theta+u/n)/p^n(X^n;\theta)) - c(\theta)u]\chi\{u<\sigma_n\}$$

is uniformly integrable with respect to $\nu \times P_\theta^n$. Indeed, we then have

$$\int\int_{[-A,A]} |\log \frac{p^n(x^n;\theta+u/n)}{p^n(x^n;\theta)} - c(\theta)u|\chi\{u<\sigma_n\}du\, p^n(x^n;\theta)dx^n \to 0,$$

which implies that

$$\int_{[-A,A]} \log \frac{p^n(x^n;\theta+u/n)}{p^n(x^n;\theta)}\chi\{u<\sigma_n\}du - \int_{[-A,A]} c(\theta)u\,\chi\{u<\sigma_n\}du$$

is uniformly integrable with respect to P_θ^n. Since the last term is always so, the result will then follow. Now

$$\begin{aligned}
& [-\log(p^n(x^n;\theta+u/n)/p^n(x^n;\theta)) + c(\theta)u]\chi\{u<\sigma_n\} \\
= \;& 2[-\log(\frac{p^n(x^n;\theta+u/n)}{p^n(x^n;\theta)})^{1/2} + (\frac{p^n(x^n;\theta+u/n)}{p^n(x^n;\theta)})^{1/2} - 1]\chi\{u<\sigma_n\} \\
& -[2(\frac{p^n(x^n;\theta+u/n)}{p^n(x^n;\theta)})^{1/2} + c(\theta)u - 2]\chi\{u<\sigma_n\}. \qquad (2.19)
\end{aligned}$$

The second term has a bounded second moment and hence is uniformly integrable. the first term is nonnegative and being a continuous function of the normalized likelihood ratio, is itself weakly convergent with repect to P_θ^n for each fixed u, and hence is also with respect to $(\nu \times P_\theta^n)$-probability. The weak limit here is

$$(-c(\theta)u + 2\exp[c(\theta)u/2] - 2)\chi\{u<\sigma\},$$

where σ has an exponential distribution with parameter $c(\theta)$. The expectation of the limit is

$$(2A)^{-1}\{-\int_{[-A,A]}\int_0^\infty c(\theta)u\,\chi\{u<s\}c(\theta)\exp[-c(\theta)s]dsdu$$

$$+2\int_{[-A,A]}\int_0^\infty \exp[c(\theta)u/2]\chi\{u<s\}c(\theta)\exp[-c(\theta)s]dsdu$$
$$+2\int_{[-A,A]}\int_0^\infty \chi\{u<s\}c(\theta)\exp[-c(\theta)s]dsdu\}$$
$$= (2A)^{-1}\{\int_{[-A,0]}(-c(\theta)u+2\exp[c(\theta)u/2]-2)du$$
$$+\int_{[0,A]}(-c(\theta)u+2\exp[c(\theta)u/2]-2)\exp[-c(\theta)u]du\}. \qquad (2.20)$$

By a well known fact, the uniform integrability verification now boils down to showing that the limit of expectations coincides with the expression in (2.20). Fix $u \geq 0$ and note that expectation of the first term in (2.19) given θ is equal to

$$\int_{p^n(x^n;\theta+u/n)>0} p^n(x^n;\theta)\log(p^n(x^n;\theta+u/n)/p^n(x^n;\theta))dx^n$$
$$+2\int(p^n(x^n;\theta+u/n)p^n(x^n;\theta))^{1/2}dx^n - 2P_\theta\{\sigma_n > u\}$$
$$= nK^*(\theta;\theta+u/n)[1-\int_{a_1(\theta)}^{a_1(\theta+u/n)} f(x;\theta)dx - \int_{a_2(\theta+u/n)}^{a_2(\theta)} f(x;\theta)dx]^{n-1}$$
$$+2[1-r_2^2(\theta,\theta+u/n)/2]^n - 2P_\theta\{\sigma_n > u\}$$
$$= n(-c(\theta)u/n+o(1/n))(1-c(\theta)u/n)^{n-1}$$
$$+2(1-c(\theta)u/(2n)+o(1/n))^n - 2P_\theta\{\sigma_n > u\}.$$

This expression converges to

$$-c(\theta)u\exp[-c(\theta)u] + 2\exp[-c(\theta)u/2] - 2\exp[-c(\theta)u]$$

and the convergence is uniform in u belonging to compact subsets. Similarly for $u < 0$, the limit is

$$-c(\theta)u + 2\exp[-c(\theta)u/2] - 2$$

and the convergence is uniform on compacts again. Clearly, the last two statements are sufficient to imply the desired conclusion. □

Proof of Lemma 2.2. We first prove Statement (a). As a consequence of Proposition 2.1,

$$R_n \xrightarrow{d} \log(c(\theta)/\pi(\theta)) - c(\theta)\sigma \qquad (2.21)$$

where σ has an exponential distribution with parameter $c(\theta)$. By Lemma 2.3, we have

$$\limsup_{n\to\infty} E_\theta R_n \leq \log(c(\theta)/\pi(\theta)) - c(\theta)E\sigma$$
$$= \log(c(\theta)/\pi(\theta)) - 1 \qquad (2.22)$$

as desired.

It remains to prove Statement (b). From (2.17),

$$\begin{aligned}\psi_n(\theta) &= \leq A^{-1}\iint_{[-A,\sigma_n(A)]} p^n(x^n;\theta)|\log\frac{p^n(x^n;\theta+u/n)}{p^n(x^n;\theta)}|du\,dx^n \\ &\quad -\log A - \log\inf\{\pi(t): t\in K\}. \end{aligned} \tag{2.23}$$

If $p^n(x^n;\theta+u/n)>0$,

$$\begin{aligned}\log\frac{p^n(x^n;\theta+u/n)}{p^n(x^n;\theta)} &= (u/n)\sum_{i=1}^{n}(\partial/\partial\theta)\log f(x_i;\theta) \\ &\quad +(u^2/2n^2)\sum_{i=1}^{n}(\partial^2/\partial\theta^2)\log f(x_i;\theta^*)\end{aligned}$$

where θ^* lies between θ and $\theta+u/n$. Consequently by (A8), for all sufficiently large n,

$$|\log\frac{p^n(x^n;\theta+u/n)}{p^n(x^n;\theta)}|\leq(|u|/n)\sum_{i=1}^{n}\frac{|f'(x_i;\theta)|}{f(x_i;\theta)}+(u^2/2n^2)\sum_{i=1}^{n}H_\theta(x_i),$$

which yields

$$\begin{aligned}\psi_n(\theta) &\leq \int_{[-A,A]}\int_{p^n(x^n;\theta+u/n)>0} p^n(x^n;\theta)|\log\frac{p^n(x^n;\theta+u/n)}{p^n(x^n;\theta)}|dx^n du \\ &\quad -\log A-\log\inf\{\pi(t): t\in K\} \\ &\leq \int_{[-A,A]}|u|\int|f'(x;\theta))|dx\,du \\ &\quad +(1/2)\int_{[-A,A]}u^2\int H_\theta(x)f(x;\theta)dxdu \\ &\quad -\log A-\log\inf\{\pi(t): t\in K\} \\ &\leq 2A\int|f'(x;\theta)|dx+2A^3\int H_\theta(x)f(x;\theta)dx \\ &\quad -\log A-\log\inf\{\pi(t): t\in K\}.\end{aligned} \tag{2.24}$$

By assumption, the terms are continuous in θ and hence bounded on compacts. Thus (b) is proved. □

From Proposition 2.1, we get the L^1-convergence of the posterior, i.e.,

$$\int_{-\infty}^{0}|\pi_n^*(v|X^n)-c(\theta)\exp[c(\theta)v]|dv\xrightarrow{P_\theta^n}0$$

where $v=n(t-W_n)$, and π_n^* is the posterior density of v. This can be equivalently written as

$$\int_{-\infty}^{0}|\pi_n(t|X^n)-nc(\theta)\exp[nc(\theta)(t-W_n)]|dt\xrightarrow{P_\theta^n}0. \tag{2.25}$$

We shall now show an information theoretic version of (2.25).

THEOREM 2.2. *Assume* (A1)–(A8) *and further suppose that*

(A9)′ $\sigma_n(\theta)$ *is uniformly (in n and θ in compacts) integrable.*

Then

$$\iiint_{-\infty}^{W_n}(\pi_n(t|x^n)\log(\pi_n(t|x^n)/\rho_n(t;\theta,x^n))dt)\,p^n(x^n;\theta)\pi(\theta)d\theta \to 0, \qquad (2.26)$$

where $\rho_n(t;\theta,x^n) = nc(\theta)\exp[nc(\theta)(t-W_n)]$.

PROOF. The expression in the right hand side of (2.26) is equal to

$$I(\pi;x^n) + \int_K \pi(\theta)\log\pi(\theta)d\theta - \log n - \int_K \log c(\theta)\pi(\theta)d\theta$$
$$- \int_K\int n(E(t|x^n) - W_n)c(\theta)p^n(x^n;\theta)dx^n\pi(\theta)d\theta. \qquad (2.27)$$

By Proposition 2.1 (ii) and (A9)′, the last term converges to one. Using Theorem 2.1, we now get the result. □

At this stage, it should be clear why we considered a particular case of the general set up of IH (Ch. V) for discontinuous densities. The reason is that such an expansion leads to posterior convergence in entropy and hence also in L^1-distance. By the results obtained in Chapter 1, the case we have considered is essentially the only case where a posterior limit can be obtained.

As we have promised, we now give sufficient conditions for (A9) to hold. For the other conditions, note that (A1)–(A7) are required by the general theory of IH (Ch. V) and (A8) is usually satisfied.

PROPOSITION 2.2. *Let* $g_\theta(v) = P_\theta(S(\theta+v))$. *If*

$$\sup_{\theta\in K}\int_0^\infty g_\theta(v)dv < \infty, \qquad (2.28)$$

then (A9) *holds. If for some* $\delta > 0$,

$$\sup_{\theta\in K}\int_0^\infty v^\delta g_\theta(v)dv < \infty, \qquad (2.29)$$

then (A9)′ *holds.*

PROOF. We shall prove only the first assertion, the second one is similar. We have

$$\begin{aligned} E_\theta\sigma_n(\theta) &= \int_0^\infty P_\theta^n\{\sigma_n(\theta) > v\}dv \\ &= \int_0^\infty (P_\theta(S(\theta+v/n))^n dv \\ &= \int_0^\infty n(g_\theta(v))^n dv. \end{aligned} \qquad (2.30)$$

Now $n(g_\theta(v))^n \le g_\theta(v)$ if and only if $g_\theta(v) \le n^{-1/(n-1)}$, which is true if and only if $v > c_n$, where $g_\theta(c_n) = n^{-1/(n-1)}$. Clearly, $c_n \to 0$ since $n^{-1/(n-1)} \to 1$. Thus

$$E_\theta \sigma_n(\theta) \le \int_0^{c_n} n(g_\theta(v))^n dv + \int_0^\infty g_\theta(v) dv. \tag{2.31}$$

By assumption, the last term in (2.31) is finite whereas the first term is

$$\int_0^{nc_n} (1 - \int_{a_1(\theta)}^{a_1(\theta+u/n)} f(x;\theta)dx)^n du. \tag{2.32}$$

Find $\varepsilon > 0$ and $\delta > 0$ such that

$$\begin{aligned}
&\inf_{\theta\in K} \inf_{a_1(\theta)<x<a_1(\theta)+\delta} f(x;\theta) \ge \varepsilon > 0,\\
&\inf_{\theta\in K} \inf_{a_2(\theta)-\delta<x<a_2(\theta)} f(x;\theta) \ge \varepsilon > 0,\\
&\inf_{\theta\in K} \inf_{0<u<\delta} a_1'(\theta+u) \ge \varepsilon > 0,\\
&\inf_{\theta\in K} \inf_{0<u<\delta} (-a_2'(\theta)) \ge \varepsilon > 0;
\end{aligned}$$

this is possible by the assumed conditions in the set up. Then the term in (2.32) is less than

$$\int_0^\infty (1 - 2\varepsilon^2 u/n)^n du \le \int_0^\infty \exp[-2\varepsilon^2 u] du < \infty,$$

which proves the result. □

REMARK 2.3. The condition in (2.29) is satisfied clearly if Θ is bounded above. In the very important case when $a_1(\theta) = \theta$ and $a_2(\theta) = \infty$, if we have $\sup_{\theta\in K} E_\theta(X_1 - \theta)^{1+\delta} < \infty$ for some $\delta > 0$, then (2.29) is satisfied. This is simply by

$$g_\theta(v) = P_\theta\{X_1 - \theta > v\} \le v^{-(1+\delta)} E_\theta(X_1 - \theta)^{1+\delta}. \tag{2.33}$$

3 Reference Prior

Reference priors are proposed by Bernardo (1979) as noninformative priors which can be thought as a reference point against which any partcular subjective prior belief can be judged. These are obtained by maximizing the expected Kullback-Leibler divergence between the posterior and the prior in an asymptotic sense. However, Bernardo's original suggestion needs to be slightly modified for technical reasons and we shall follow essentially Ghosh and Mukherjee (1992).

For any (possibly improper) prior $\pi(\theta)$ on Θ and $K \subset \Theta$ compact, we shall write $\pi|_K$ for the proper prior defined by

$$\pi|_K(A) = \pi(A)/\pi(K), \qquad A \subset K.$$

DEFINITION 3.1. A prior density π^* on Θ is called a *reference prior* if

(i) π^* is positive and continuous in Θ.

(ii) for all compact $K \subset \Theta$ and for all prior density π positive and continuous in Θ, we have

$$J(\pi^*|_K) \geq J(\pi|_K)$$

where $J(q) = \int_K q(\theta) \log(c(\theta)/\pi(\theta))d\theta$.

Condition (ii) in Definition 3.1 comes from Bernardo's (1979) motivation (the present form is suggested by Ghosh and Mukherjee (1992)). We agree that the first one is included for technical reasons; nevertheless, it is an impartiality requirement. One would hardly like to call a prior as a reference prior if (i) is not satisfied.

As is immediate, the prior $\pi^*(\theta) \propto c(\theta)$ is the reference prior. (In case $S(\theta)$ is increasing, $\pi^*(\theta) \propto |c(\theta)|$.) In the particular case of location family $f(x;\theta) = f(x-\theta)$, $\theta \in \mathbb{R}$ with $f(0+) > 0$, $f(0-) = 0$, and so $c(\theta) = f(0+)$. Consequently, $\pi^*(\theta)$ is the improper uniform prior as it is expected. In case of truncation model

$$f(x;\theta) = g(x)/\bar{G}(\theta), \qquad x \geq \theta$$

where g is a smooth positive density and $\bar{G}(x) = \int_x^\infty g(y)dy$, we have

$$c(\theta) = g(\theta)/\bar{G}(\theta).$$

Consequently, the reference prior $\pi^*(\theta)$ is proportional to the hazard rate for the density g.

Appendix A

LEMMA 1. *For an $f \in L^1(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$, let f_x stand for the L^1-function defined by $f_x(y) = f(y-x)$. Then the mapping $(x, f) \mapsto f_x$ from $\mathbb{R}^d \times L^1(\mathbb{R}^d)$ is continuous in x and isometry in f, and so is jointly continuous.*

For a proof, see Rudin (1966, Theorem 9.5, p. 183).

LEMMA 2. *Let $\mathcal{P}$ be the space of all absolutely continuous probabilities on $\mathbb{R}^d$ equiped with the total variation distance. Let $(\Omega, \mathcal{A})$ be a measurable space and $\xi : \Omega \to \mathcal{P}$ be a map. Then ξ is measuable if and only if, for all $A \in \mathcal{B}^d$, the map ξ_A defined by*

$$\xi_A(\omega) = \xi(\omega)(A)$$

is Borel measurable.

PROOF. Since $Q \mapsto Q(A)$ from $\mathcal{P} \to \mathbb{R}$ is continuous, the only if part is trivial.

For the if part, let $\mathcal{F}$ be countable field generating $\mathcal{B}^d$. Then for any $Q, Q_0 \in \mathcal{P}$, the total variation distance $\|Q - Q_0\|$ satisfies

$$\|Q - Q_0\| = \sup\{|Q(A) - Q_0(A)| : A \in \mathcal{F}\}$$

by a well known fact in measure theory (see, e.g., Theorem 13.D of Halmos (1974)).

Now

$$\{\omega : \|\xi(\omega) - Q_0\| \leq r\} = \bigcap_{A \in \mathcal{F}} \{\omega : |\xi_A(\omega) - Q_0(A)| \leq r\} \in \mathcal{A}$$

by hypothesis, which completes the proof. □

LEMMA 3. *Let $\mathcal{P}$ be the space of all absolutely continuous probabilities on $\mathbb{R}^d$ and $Q \in \mathcal{P}$. Let $\mathfrak{M}(Q) = \{Q_x : x \in \mathbb{R}^d\}$ where $Q_x \in \mathcal{P}$ is defined by $Q_x(A) = Q(A - x)$ for all $A \in \mathcal{B}^d$. Define a map $\psi : \mathfrak{M}(Q) \to \mathbb{R}^d$ by $\psi(Q_x) = x$. Then ψ is a homeomorphism and $\mathfrak{M}(Q)$ is closed in $\mathcal{P}$.*

PROOF. First let us check that ψ is well defined. If not, there are $x_1, x_2 \in \mathbb{R}^d$, $x_1 \neq x_2$ such that $Q_{x_1} = Q_{x_2}$. Put $\delta = x_1 - x_2$. Then $Q(A) = Q(A + \delta)$ for all $A \in \mathcal{B}^d$. If A is bounded, by repeated application, we have $Q(A) = Q(A + n\delta) \to 0$, so $Q(A) = 0$ for all bounded sets. This contradicts the countable additivity of Q and so ψ is well defined, and at the same time it is a bijection.

To show ψ is continuous, let $Q_{x_n} \to Q_x$. We have to show that $x_n \to x$. First note that x_n is bounded. If not, let $\{m\}$ be a subsequence of $\{n\}$ along which it converges to plus or minus infinity. Then for any bounded set $A \in \mathcal{B}^d$,

$$Q_x(A) = \lim_{m\to\infty} Q(A - x_m) = 0$$

which is again a contradiction. By Lemma A.1, $\psi^{-1} : \mathbb{R}^d \to \mathcal{P}$ is continuous. So all the subsequential limits of $\{x_n\}$ must be equal to x. Thus $x_n \to x$, and so ψ is continuous, and so a homeomorphism.

The fact that $\mathfrak{M}(Q)$ is closed in $\mathcal{P}$ follows from essentially the same arguments.
□

THEOREM 1. *Let $\xi(t)$ be a real valued random function defined on a closed subset F in $\mathbb{R}^d$. Assume that $\xi(t)$ is a measurable and separable process and there exist numbers $m \geq \alpha > d$ and a function $H : \mathbb{R}^d \to \mathbb{R}$ bounded on compact sets such that for all $x, h \in \mathbb{R}^d$ with $x, x+h \in F$, we have*

$$\begin{aligned} E|\xi(x)|^m &\leq H(x) \\ E|\xi(x+h) - \xi(x)|^m &\leq H(x)\|h\|^\alpha. \end{aligned}$$

Then with probability one, the sample paths of $\xi(t)$ are continuous on F. Moreover, set

$$\omega(\delta; \xi, R) = \sup\{|\xi(x) - \xi(y)| : x, y \in F, \|x\|, \|y\| \leq R, \|x - y\| \leq \delta\}.$$

Then

$$E\omega(\delta; \xi, R) \leq B_0(\sup_{\|x\|\leq R} H(x))^{1/m} R^{d/m} \delta^{(\alpha-d)/m}$$

where the constants B_0 depends on m, α and d only.

For a proof, see Theorem A.19 of IH.

THEOREM 2. *Let $\xi_n(t)$, $n \geq 1$ and $\xi(t)$ be real valued measurable functions defined on a compact set $F \in \mathbb{R}^d$ and $w(t)$ be a measurable function on F. Assume that the following conditions are satisfied:*

(1) $\sup_{n\geq 1} E(\int_F |w(t)||\xi_n(t)|dt) < \infty$.

(2) *There exist $H, \alpha > 0$ such that* $\sup_{n\geq 1} E|\xi_n(t) - \xi_n(s)| \leq H\|t - s\|^\alpha$.

(3) *Finite dimensional distributions of $\xi_n(t)$ converge to those of $\xi(t)$.*

Then for any $t_1, \ldots, t_k \in \mathbb{R}^d$,

$$(\xi_n(t_1), \ldots, \xi_n(t_k), \int_F w(t)\xi_n(t)dt) \xrightarrow{d} (\xi(t_1), \ldots, \xi(t_k), \int_F w(t)\xi(t)dt).$$

In particular,

$$\int_F w(t)\xi_n(t)dt \xrightarrow{d} \int_F w(t)\xi(t)dt.$$

The proof is a minor modification of that of Theorem A.22 of IH.

THEOREM 3. *A subset* $\Gamma \subset L^1(\mathbb{R}^d)$ *has a norm-compact closure if and only if*

(a) $\sup\{\|f\| : f \in \Gamma\} < \infty$,

(b) $\lim_{\|x\| \to 0} \sup\{\int |f(x+y) - f(y)|dy : f \in \Gamma\} = 0$,

(c) $\lim_{\lambda \to \infty} \sup\{\int_{\|y\| > \lambda} |f(y)|dy : f \in \Gamma\} = 0$.

For a proof, see Dunford and Schwartz (1957, pp. 298-301).

COROLLARY 1. *Let* $\xi_n, n \geq 1$ *be a sequence of random densities on* $\mathbb{R}^d$. *Then* $\{\xi_n\}$ *is tight if and only if given* $\varepsilon > 0$, $\eta > 0$, *there exist* $M, \delta > 0$ *and* $n_0 \geq 1$ *such that*

(a) $P\{\int |\xi_n(x+y) - \xi_n(y)|dy > \varepsilon\} < \eta$ *for all* $\|x\| < \delta$, $n \geq n_0$

(b) $P\{\int_{\|y\| > M} \xi_n(y)dy > \varepsilon\} < \eta$ *for all* $n \geq n_0$.

COROLLARY 2. *Let* $\xi_n, n \geq 1$ *be a sequence of random densities on* $\mathbb{R}^d$ *satisfying the conditions of Corollary* A.1. *Let* ξ *be a random density such that for all* $A_1, \ldots, A_k \in \mathcal{B}^d$,

$$(\int_{A_1} \xi_n(x)dx, \ldots, \int_{A_k} \xi_n(x)dx) \xrightarrow{d} (\int_{A_1} \xi(x)dx, \ldots, \int_{A_k} \xi(x)dx).$$

Then as random elements in $L^1(\mathbb{R}^d)$, ξ_n *converges to* ξ *weakly.*

Appendix B

Conditions (IH) have several implications regarding the asymptotic properties of the Bayes estimates and the MLE. These results are derived in Sections I.5 and I.10 of IH. Here we state the main results alongwith the numberings of IH; details and proofs are available in IH (Sec. I.5 and I.10). We continue the notaions of Chapter 1.

THEOREM I.5.2. *Assume Conditions* (IH 1) *and* (IH 2), *the prior* $\pi \in \Pi$ *and loss* $l \in \mathfrak{L}$ *(however, continuity of* l *is not needed) and let* $\{\widetilde{\theta}_n\}$ *denote the Bayes estimates. Then for any* $N \geq 0$,

$$\lim_{\substack{H\to\infty \\ n\to\infty}} H^N P_\theta^n\{\|\varphi_n^{-1}(\widetilde{\theta}_n - \theta_0)\| > H\} = 0. \tag{1}$$

REMARK 1. It follows from $\{\widetilde{\theta}_n\}$ is weakly consistent for θ_0, $\{\varphi_n^{-1}(\widetilde{\theta}_n - \theta_0)\}$ (i.e., the sequence of normalized Bayes estimates) is stochastically bounded and all its powers are uniformly integrable. Further, if $\sum_{n=1}^\infty \|\varphi_n\|^s < \infty$ for some $s > 0$, then $\{\widetilde{\theta}_n\}$ is strongly consistent also (provided almost sure convergence is meaningful).

A main step in proving Theorem I.5.2 is the following lemma (Lemma I.5.2 of IH). Since we have slightly changed the conditions of IH and this lemma gives a very important estimate which is needed in many places in this work, we briefly sketch the proof, referring IH for the details.

For this, define $Q_n(H) = \int_{\|u\|>H} \xi_n(u)du$ and set $g_{in}(y) = g_n(0,\dots,0,y,0,\dots,0)$, $y \geq 0$, where y occurs at the ith component, $i = 1,\dots,d$.

LEMMA I.5.2. *For any* $m \geq 0$,

$$\lim_{\substack{H\to\infty \\ n\to\infty}} H^m EQ_n(H) = 0. \tag{2}$$

PROOF. We prove the result only for $\pi \equiv 1$, the general case can be easily derived from this. It is enough to prove the result where the Euclidean norm $\|\cdot\|$ is replaced by the l^∞-norm $|\cdot|$ (and consequently the definition of $Q_n(H)$). We see that there are constants $B, b > 0$ such that for each $i = 1,\dots,d$,

$$P\{I_{in}(H) > \exp[-bg_{in}(H)]\} \leq B(1 + H^B)\exp[-bg_{in}(H)]\}, \tag{3}$$

where $I_{in}(H) = \int_{\{H \le |u_i| < H+1\}} Z_n(u)du$. Indeed, this follows essentially from the arguments given by IH. Rest of the proof now follows from the obvious modifications of the steps of IH. □

To find out the asymptotic distribution of Bayes estimates, define

$$\begin{aligned} \xi(u) &= Z(u)/(\int Z(v)dv) \\ \psi(s) &= \int l(s-u)\xi(u)du \end{aligned}$$

and assume that the random function $\psi(\cdot)$ attains its minimum at a unique point $\tau = \tau(\theta_0)$. This assumption is rather mild; any convex loss with a unique minimum satisfies it. One such example is $l(u) = \|u\|^p$, $p \ge 1$. The limiting distribution is given by the next result.

THEOREM I.10.2. *Under Conditions* (IH), *we have*

$$\varphi_n^{-1}(\widetilde{\theta}_n - \theta_0) \xrightarrow{d} \tau \tag{4}$$

and for any continuous function w with a polynomial majorant

$$\lim_{n\to\infty} Ew(\varphi_n^{-1}(\widetilde{\theta}_n - \theta_0)) = Ew(\tau). \tag{5}$$

Moreover, $\operatorname{diam}(\Lambda_n) \xrightarrow{p} 0$ *where Λ_n denotes the set of all normalized Bayes estimates with respect to prior π and loss l.*

THEOREM I.5.1. *Suppose the LRP $Z_n(\cdot)$ has continuous sample paths and Conditions* (IH 1)$'$ *and* (IH 2) *are satisfied. Then for any $N \ge 0$ we have*

$$\lim_{\substack{H\to\infty \\ n\to\infty}} H^N P\{\|\varphi_n^{-1}(\widehat{\theta}_n - \theta_0)\| > H\} = 0. \tag{6}$$

REMARK 2. It follows from the above estimate that $\{\widehat{\theta}_n\}$ is weakly consistent for θ_0, $\{\varphi_n^{-1}(\widetilde{\theta}_n - \theta_0)\}$ (i.e., the sequence of normalized MLEs) is tight and all its powers are uniformly integrable. Further, if $\sum_{n=1}^{\infty} \|\varphi_n\|^s < \infty$ for some $s > 0$, then $\{\widetilde{\theta}_n\}$ is strongly consistent also (provided almost sure convergence makes sense).

THEOREM I.10.1. *Assume Conditions* (IH)$'$, *suppose that $Z_n(\cdot)$ and $Z(\cdot)$ has continuous sample paths and $Z(\cdot)$ attains its maximum at a unique point $\widehat{u}$. Then as $n \to \infty$*

$$\varphi_n^{-1}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} \widehat{u} \tag{7}$$

and for any continuous function w with a polynomial majorant,

$$\lim_{n\to\infty} Ew(\varphi_n^{-1}(\widehat{\theta}_n - \theta_0)) = Ew(\widehat{u}). \tag{8}$$

Further the diameter of the set of all normalized MLEs converges to zero in probability.

REMARK 3. Let $C(\mathbb{R}^d)$ be the space of all continuous functions on $\mathbb{R}^d$ with the topology of uniform convergence on compacts. Fix any sequence $\{K_m\}$ of compact sets increasing to $\mathbb{R}^d$. For $f \in C(\mathbb{R}^d)$ let f_m denote the restriction of f to K_m. By the canonical identification of $C(\mathbb{R}^d)$ with a closed subset of the prroduct space $\prod_{m=1}^{\infty} C(K_m)$ through the mapping $f \mapsto (f_1, f_2, \ldots)$, it follows from the proof of Theorem I.10.1 of IH that the process $Z_n(\cdot)$ converge to $Z(\cdot)$ in $C(\mathbb{R}^d)$ and $Z(\cdot) \in C_0(\mathbb{R}^d)$ (the class of functions in $C(\mathbb{R}^d)$ vanishing at infinity) with probability one. This fact may have some independent interest.

References

AITCHISON, J. (1975). Goodness of prediction fit. *Biometrika* **62**, 547-554.

BASU, A. P., GHOSH, J. K. and JOSHI, S. N. (1988). On estimating change-point in a failure rate. In *Statistical Decision Theory and Related Topics IV* **2**, 239-252. (S. S. Gupta and J. O. Berger, eds.). Springer-Verlag, New York.

BERGER, J. O. (1985). *Statistical Decion Theory and Bayesian Analysis.* Second ed. Springer-Verlag, New York.

BERGER, J. O. and BERNARDO, J. M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *J. Amer. Statist. Assoc.* **84**, 200-207.

BERGER, J. O. and BERNARDO, J. M. (1992a). On the development of reference priors (with discuusion). In *Bayesian Statistics* **4**, (J. M. Bernardo et al., eds.) Oxford University Press, 35-60.

BERGER, J. O. and BERNARDO, J. M. (1992b). Ordered group reference priors with application to a multinomial problem. *Biometrika* **79**, 25-37.

BERGER, J. O. and BERNARDO, J. M. (1992c). Reference priors in a variance component problem. In *Proceedings of the Indo-USA Workshop on Bayesian Analysis in Statistics and Econometrics* (P. Goel, ed.) Springer, New York, 323-340.

BERGER, J. O., BERNARDO, J. M. and MENDOZA, M. (1989). On priors that maximize expected information. In *Recent Developments in Statistics and Their Applications* (J. P. Klein and J. C. Lee, eds.) Freedom Academy Publishing, Seoul, 1-20.

BERNARDO, J. M. (1979). Reference posterior distributions for Bayesian inference (with discussions). *J. Roy. Statist. Soc. Ser. B* **41**, 113-147.

BERNSTEIN, S. (1917). *Theory of Probability* (in Russian).

BICKEL, P. and YAHAV, J. (1969). Some contributions to the asymptotic theory of Bayes solutions. *Z. Wahrsch. Verw. Gebiete* **11**, 257-275.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

CHANG T. and EAVES, D. (1990). Reference priors for the orbit in a group model. *Ann. Statist.* **18**, 1070-1090.

CHEN, C. F. (1985). On asymptotic normality of limiting density functions with Bayesian implications. *J. Roy. Statist. Soc. Ser. B* **47**, 540-546.

CHENG, R. C. H. and ILES, T. C. (1987). Corrected maximum likelihood estimates in nonregular problems. *J. Roy. Statist. Soc. Ser. B* **49**, 95-101.

CHERNOFF, H. and RUBIN, H. (1956). The estimation of the location of a discontinuity in a density. In *Proc. Third. Berkeley Symp.* **1**, University of California Publications. 19-38.

CLARKE, B. and BARRON, A. R. (1990a). Information theoretic asymptotics of Bayes Methods. *IEEE Trans. Inform. Theory* **36**, 453-471.

CLARKE, B. and BARRON, A. R. (1990b). Entropy risk and Bayesian central limit theorem. Tech. Report.

CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press.

DAWID, A. P. (1970). On the limiting normality of posterior distribution. *Proc. Canad. Phil. Soc. Ser. B* **67**, 625-633.

DIACONIS, P. and FREEDMAN, D. (1986). On the consistency of Bayes estimates. (with discussions) *Ann. Statist.* **14**, 1-67.

DUNFORD, N. and SCHWARTZ, J. T. (1957). *Linear Operators. Part I: General Theory.* Wiley Interscience, New York.

ERMAKOV, M. S. (1977). Asymptotic behavior of statistical estimates of parameters of multidimensional discontinuous density. *Zap. LOMI* **74**, 83-107.

GHOSH, J. K. (1985). Efficiency of estimates-Part I. *Sankhyā Ser. A* **47**, 310-325.

GHOSH, J. K. (1993). *Higher Order Asymptotics.* NSF-CBMS Lecture Notes 4, IMS, California.

GHOSH, J. K., JOSHI, S. N. and MUKHOPADHYAY, C. (1992a). A Bayesian study of reliability change point problem. (Under preparation).

GHOSH, J. K., JOSHI, S. N. and MUKHOPADHYAY, C. (1992b). Bayesian asymptotics for the reliability change point problem. (Under preparation).

GHOSH, J. K. and MUKHERJEE, R. (1992). Non-informative priors (with discussion). In *Bayesian Statistics* 4 (J. M. Bernardo et al., eds.) 195-210.

HÁJEK, J. (1970). A characterization of limiting distribution of regular estimates. *Z. Wahrsch. Verw. Gebiete* **14**, 323-330.

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests.* Academic, New York.

HALMOS, P. (1974). *Measure Theory.* Springer-Verlag.

HEYDE, C. C. and JOHNSTONE, I. M. (1979). On asymptotic posterior normality for stochastic processes. *J. Roy. Statist. Soc., Ser B*, **41**, 184-189.

HILL, B. (1963). The three parameter lognormal distribution and Bayesian analysis of a point source epidemic. *J. Amer. Statist. Assoc.* 58, 72-84.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1970). On the asymptotic behavior of generalized Bayes estimators, *Dokl. Acad. Nauk SSSR* **194**, 257-260.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1971). On limiting behavior of maximum likelihood and Bayesian estimators, *Dokl. Akad. Nauk SSSR* **198**, 520-523.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1972a). Asymptotic behavior of statistical estimators in the smooth case 1. *Theory Probab. Appl.* **17**, 445-462.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1972b). The asymptotic behavior of statistical estimators for samples with a discontinuous density. *Mat. Sb.* **87 (129)**, 554-558.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1973a). Asymptotic analysis of statistical estimators for the almost smooth case. *Theory Probab. Appl.* **18**, 241-252.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1973b). On approximation of statistical estimators in terms of sums of independent random variables. *Dokl. Akad. Nauk SSSR* **210**, 1273-1276.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1973c). On the information contained in a sample about a parameter. In *Second International Symposium on Information Theory.* Akademiai, Kiado, Budapest. 295-309.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1974). The asymptotic behavior of statistical estimates of shift parameter for samples with a continuous density with singularities. *Zap. Nauchn. Seminar LOMI Akad. Nauk SSSR* **44**, 67-93.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1975a). Local asymptotic normality for non-identically distributed observations. *Theory Probab. Appl.* **20**, 246-260.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1975b). Properties of maximum likelihood and Bayes estimators for non-identically distributed observations. *Theory Probab. Appl.* **20**, 689-697.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1976). Asymptotic beehavior of estimators of a location parameter for samples with unbounded density. *Zap. Nauchn. Seminar LOMI* **55**, 175-184.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1977). On the estimation of multidimensional parameter. *Theory Probab. Appl.* **22**, 640-641.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer-Verlag, New York.

IONESCU TULCEA, A. and IONESCU TULCEA, C. (1969). *Topics in the Theory of Lifting.* Springer-Verlag, Berlin.

JOHNSON, N. L. AND KOTZ, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions*-1. John Wiley, New York.

JOHNSON, R. A. (1967). On asymptotic expansion for posterior distribution. *Ann. Math. Statist.* **38**, 1899-1906.

JOHNSON, R. A. (1970). Asymptotic expansions associated with posterior distribution. *Ann. Math. Statist.* **42**, 1241-1253.

LAPLACE, P. S. (1774). Mémoire sur la probabilité de causes par les évenements. *Mem. Acad. R. Sci. Presentés par Divers Savanas.* **6**, 621-656. (Translated in *Stat. Sci.* **1**, 359-378.)

LE CAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *University of California Publications in Statistics* **1**, 277-330.

LE CAM, L. (1958). Les properties asymptotiques des solutions de Bayes. *Publ. Inst. Statist. Univ. Paris* **7**, 17-35.

LE CAM, L. (1970). Remarks on the Bernstein-von Mises theorem. Unpublished manuscript.

LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New York.

LE CAM, L. and YANG, G. L. (1990). *Asymptotics in Statistics: Some Basic Concepts.* Springer-Verlag, New York.

LINDLEY, D. V. (1956). On a measure of information provided by an experiment. *Ann. Math. Statist.* **27**, 986-1005.

MASE, S. (1992). Uniform LAN condition of planar Gibbsian point processes and optimality of maximum likelihood estimators of soft-core potential functions. *Probab. Theory Relat. Fields* **92**, 51-67.

MILLAR, P. W. (1983). The minimax priciple in asymptotic theory. In *Ecole d'Été de Probabilités de Saint-Flour XI*-1981 (P. L. Hennequin, ed.). *Lecture Notes in Mathematics.* **976**, Springer-Verlag, Berlin.

NGUYEN, H. T., ROGERS, G. S. and WALKER, E. A. (1984). Estimation in change-point hazard rate models. *Biometrika* **71**, 299-305.

PFLUG, G. C. (1982a). The limiting log-likelihood process for discontinuous density families. *Z. Wahrsch. Verw. Gebiete* **64**, 15-35.

PFLUG, G. C. (1982b). The limiting log-likelihood process for discontinuous multiparameter density families. In *Probabilty and Statistical Inference.* (W. Grossman et al., eds.) 287-295, D. Reidel Publishing Company.

POLFELDT, T. (1970). Asymptotic results in non-regular estimation. *Skand. Aktaur. Suppl.* 1-2.

POLFELDT, T. (1970a). The order of minimum variance in a non-regular case. *Ann. Math. Statist.* **41**, 667-672.

POLFELDT, T. (1970b). Minimum variance order when estimating the location of an irregularity in the density. *Ann. Math. Statist.* **41**, 673-679.

POLLARD, D. (1990). *Empirical Processes: Theory and Applications.* NSF-CBMS Regional Conference Series in Probability and Statistics **2**, IMS, Calif.

POLSON, N. G. (1988). On the expected amount of information from a nonlinear model. *J. Roy. Statist. Soc. Ser. B.* **54**, 889-895.

PRAKASA RAO, B. L. S. (1968). Estimation of the location of the cusp of a continuous density. *Ann. Math. Statist.* **39**, 76-87.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications. Second edition.* John Wiley, New York.

RISSANEN, J. (1986). Stochastic complexity and modelling. *Ann. Statist.* **14**, 1080-1100.

RISSANEN, J. (1987). Stochastic complexity. *J. Roy. Statist. Soc. Ser. B* **49**, 223-239.

RUBIN, H. (1961). The estimation of discontinuities in multivariate densities and related problems in stochastic processes. In *Proc. Fourth Berkeley Symp.* **1**, University of California Publications. 563-574.

RUDIN, W. (1966). *Real and Complex Analysis.* Second ed. McGraw-Hill, New York.

SAMANTA, T. (1988). *Some Contributions to the Asymptotic Theory of Estimation in Non-regular Cases.* Unpublished Ph. D. Thesis, Indian Statistical Institute.

SAMANTA, T. (1988a). Local asymptotic minimax estimation in non-regular cases. *Sankhyā Ser. A*, **51**, 343-369.

SCHWARTZ, L. (1965). On Bayes procedures. *Z. Wahrsch. Verw. Gebiete* **4**, 10-26.

SMITH, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72**, 69-90.

SWEETING, T. J. and ADEKOLA, A. O. (1987). Asymptotic posterior normality for stochastic processes. *J. Roy. Statist. Soc. Ser. B*, **49**, 215-222.

VON MISES, R. (1931). *Wahrscheinlichkeitsrechnung.* Springer, Berlin.

WALKER, A. M. (1969). On the asymptotic behaviour of posterior distributions. *J. Roy. Statist. Soc. Ser. B* **31**, 80-88.

WEISS, L. and WOLFOWITZ, J. (1973). Maximum likelihood estimation of a translation parameter of a truncated distribution. *Ann. Statist.* **1**, 944-947.

WEISS, L. and WOLFOWITZ, J. (1974). *Maximum Probability Estimators and Related Topics.* Springer-Verlag, Berlin.

WOODROOFE, M. (1972). Maximum likelihood estimation of a translation parameter of a truncated distribution. *Ann. Math. Statist.* **43**, 113-122.

WOODROOFE, M. (1974). Maximum likelihood estimation of a translation parameter of a truncated distribution II. *Ann. Statist.* **2**, 474-488.

WOODROOFE, M. (1992). Integrable expansion for posterior distributions for one-parameter exponential families. *Statist. Sinica* **2**, 91-111.

Index