Chapter 2

The Asymptotic Behavior of the Radon-Nikodym Derivative for Some Stochastic Processes

"... the epistemological value of the theory of probability is revealed only by limit theorems. Moreover, without limit theorems, it is impossible to understand the real content of the primary concept of all our sciences - the concept of probability ".

B.V. Gnedenko and A.N. Kolomogorov (1954)

2.1 Introduction

The likelihood ratio plays an important role in statistical inference. It forms the basis for estimation and testing of hypotheses. Asymptotic inference is based almost entirely on the likelihood ratio. Consistency of estimates and their asymptotic normality depend on the asymptotic behavior of the likelihood ratio. Consistency of estimates, for instance, is attainable only when the likelihood ratio converges almost surely to zero. Asymptotic efficiency of tests and estimators is related to the rate of convergence of the likelihood ratio to zero. In this chapter, we study the asymptotic behavior of the likelihood ratio for sequences of (a) independent identically distributed observations (b) independent but not identically distributed observations and (c) dependent observations, in the next section. For continuous processes, likelihood ratio or more precisely the Radon-Nikodym (R-N) derivative possesses properties similar to sequences of observations. These properties are studied for decompo-
sable processes in section 3 and for arbitrary stochastic processes in section 4. In section 5, we consider convergence in distribution of the log-likelihood ratio. Section 6 deals with Fisher information as a norm for the log-likelihood ratio, for deriving a.s. convergence results. Convergence of MLEs, so closely allied to that of likelihood ratios, is considered in section 7. Asymptotic Bayesian inference is discussed in section 8.

2.2.1 Sequences of observations

Consider a sequence \( \{X_n\} \) of real random variables (r.v.s) defined on a certain measurable space \((\Omega, \mathcal{A})\). Let \( \mathcal{B}_n \) be the \( \sigma \)-field induced by \((X_1, \ldots, X_n)\) and without loss of generality, we take \( \Omega = \mathbb{R}^\infty \), and for \( \mathcal{A} \), the minimal \( \sigma \)-field containing \( \bigcup \mathcal{B}_n \). Let \( P \) and \( Q \) be two probability measures on \((\Omega, \mathcal{A})\) and \( P_n \) and \( Q_n \) be the restrictions of \( P \) and \( Q \) respectively to sets of \( \mathcal{B}_n \). \( P \) and \( Q \) may be members of a certain family of distributions, \( \{ P_\theta, \theta \in \Theta \} \), \( \Theta \subseteq \mathbb{R} \). In that case \( P_\theta \) will be denoted by \( P \) and \( P_\theta \) by \( Q \).

For \( \omega \in \Omega \), let \( r_n(\omega) = \frac{dQ_n(\omega)}{dP_n(\omega)} \) be the likelihood ratio or the R-N derivative of \( Q_n \) w.r.t \( P_n \) defined zero if \( Q_n \) and \( P_n \) are mutually singular \((Q_n \perp \!\!\!\!\perp P_n)\). We assume that \( Q \) and \( P \) are mutually absolutely continuous \((P \|= Q)\) on \( \mathcal{B}_n \) with their common support not depending on the probability measure. For the sequence of observations then, \( \left\{ r_n(\omega), \mathcal{B}_n, P_\cdot \right\} \) is a martingale (see Doob (1953)). When \( Q \) is absolutely continuous w.r.t \( P \), \((Q \ll P)\) on \( \mathbb{R}^\infty \), \( \{r_n\} \) converges a.s.(\( P \)) to a bounded r.v. \( r_\infty > 0 \); when \( Q \perp P \), \( r_n \to 0 \), a.s. (\( P \)) or \( r_n \to \infty \), a.s.(\( Q \)) (see Neveu (1975)). Even though \( P \|= Q \) on \( \mathcal{B}_n \) for all \( n \), this does not imply that \( P \|= Q \) on \( \mathbb{R}^\infty \), unless \( \{r_n\} \) converges to
r_\infty in the first mean also (Neveu (1975)). Here onwards, 'E', 'var' and 'convergence' are w.r.t Q-measure, unless otherwise stated.

Now, \{\log r_n\} is a submartingale diverging a.s to \infty in the important case where Q \perp P on \mathbb{R}^\infty. Then it is possible that there exists a sequence \{b_n\} of constants with 0 < b_n \uparrow \infty, such that

$$b_n^{-1} \log r_n \rightarrow Z \quad \text{a.s.,} \quad (2.2.1)$$

where \(Z(\omega) \in [-\infty, \infty]\) and Z is possibly random. It may be seen that this result holds for partly singular measures, in the region of singularity.

a) I.I.D observations : Let \(\{X_n\}\) be a sequence of i.i.d observations with common probability density function f(x/Q) under Q and f(x/P) under P, w.r.t some \sigma-finite measure. Then

$$\log r_n = \sum_{i=1}^{n} \log \frac{f(x_i/Q)}{f(x_i/P)} = \sum_{i=1}^{n} Y_i, \text{ (say),} \quad (2.2.2)$$

where \(\{Y_i\}\) are i.i.d with \(E(Y_i) = K(Q,P)\), the Kullback-Leibler (K-L) information number. By strong law of large numbers,

$$\frac{1}{n} \log r_n \rightarrow K(Q,P), \quad \text{a.s.} \quad (2.2.3)$$

Thus in (2.2.1), \(b_n = n\) and Z is a constant.

b) Independent but not i.i.d observations : Let \(X_1\) possess the density \(f_1(x/Q)\) under Q and \(f_1(x/P)\) under P, \(i=1,...,n\). Then \(\log r_n = \sum_{i=1}^{n} Y_i\), where \(Y_i\) are independent but not i.i.d. Then \(E(Y_i) = E \left[ \log \frac{f_1(x/Q)}{f_1(x/P)} \right]\) is the K-L information contained in the \(i^{th}\) observation. Let

$$U_i = Y_i - E(Y_i), \text{ so that}$$

$$\log r_n - K_n(Q,P) = \sum_{i=1}^{n} U_i = S_n, \text{ (say),}$$

is a martingale sequence, where \(K_n(Q,P) = \sum_{i=1}^{n} E(Y_i)\). If \(\{b_n\}\) is such that 0 < \(b_n \uparrow \infty\) and \(\sum_{i=1}^{\infty} E(U_i^2) / b_i^2 < \infty\), then from Theorem
thereby giving

\[ b_n^{-1} \log r_n \to \text{hm } b_n^{-1} K_n(Q,P) = \bar{K}(Q,P), \text{ a.s.} \]

For example, choose \( b_n = \text{Var } S_n = \frac{1}{n} \sum E(U_i^2) \). It is clear that

\[ \sum b_n^{-1} E(U_i^2) < \infty. \]

Hence under the assumption: \( \text{Var } \log r_n = \text{Var } S_n \uparrow \infty \)

\[ b_n^{-1} \log r_n \to \text{hm } (\text{Var } \log r_n)^{-1} E(\log r_n), \text{ a.s.} \quad (2.2.4) \]

In general, under the assumptions

\[ (i) \quad 0 < b_n \uparrow \infty, \]

\[ (ii) \quad \sum b_n^{-1} E(U_i^2)/b_i < \infty, \]

and

\[ (iii) \quad \sum E(Y_i)/b_n \to \bar{K}(Q,P), \text{ a.s.} \]

Here \( K(Q,P) \) is a constant, finite or infinite. The last assertion follows from (i), since the set of a.s convergence of \( b_n^{-1} \log r_n \) is a tail event and by Kolmogorov 0-1 law, the limiting r.v. is degenerate a.s.

**Example 2.2.1**

Let \( \{X_i\} \) be a sequence of independent normal variables with \( E(X_i) = \theta, \) \( \text{Var } (X_i) = 1^{-1}. \) Under \( P, \) \( E(X_i/P) = \theta. \)

\[ E(Y_i) = \chi(\theta - \theta_0)^2/2, \]

\[ \text{Var } Y_i = \chi(\theta - \theta_0)^2, \]

so that by taking \( b_n = n^2 \), conditions (i) - (iii) are satisfied where \( K(Q,P) \) equals \( (\theta - \theta_0)^2/4. \)

**c) Dependent observations.** For dependent observations, the limiting
r.v. Z defined in section 2.2.1 may be random. Following Basawa and Scott (1983), we call such a stochastic process as Q-non-ergodic. (see also Chapter on Introduction).

At this juncture some explanation is necessary as to the choice of the words 'ergodic' and 'non-ergodic'. We quote Basawa and Scott (1983). "The term ergodic has a well-defined meaning in the theory of stationary processes, viz, that the $\sigma$-field of sets which are invariant under the shift transformation (the invariant $\sigma$-field) is trivial, that is, contains only sets of probability zero or one. Equivalently, such a process is ergodic if all random variables which are measurable w.r.t the invariant $\sigma$-field is a.s constant." However, in the more general context, i.e, including non-stationary processes, the basic ergodic theorems (cf. Loeve (1978) Vol.II, page 81) assert conditions under which there is convergence a.s for $\sum X_n/\sum Y_n$, where $\{X_n\}, \{Y_n\}$ denote a family of r.v.s. The convergence is possibly to a r.v. Degenerate and non-degenerate convergence depends on the invariant $\sigma$-field - if it is trivial, convergence is surely degenerate, as for example, for independent non-identically distributed sequences which are non-stationary. In general, however, for non-stationary sequences, the invariant $\sigma$-field may not be trivial and convergence, if it exists, is possibly to a r.v. Our 'ergodic' and 'non-ergodic' models actually correspond to trivial and non-trivial invariant $\sigma$-fields. The motivation for the ergodic hypothesis (Loeve, (1978), Chapter on ergodic theorems) suggests that we are seeking conditions under which there is degenerate convergence. Thus even though it is true that $\{X_n\}$ under (2.2.1) cannot be justifiably called non-ergodic
in the broad sense of Loeve, based on considerations of what the ergodic hypothesis originally intended to achieve, the terms appear appropriate.

**Lemma 2.2.1.** The support of $Z$ in (2.2.1) is contained in $[0, \infty)$.

**Proof:** Let $C_n(t) = \{ \omega : \log r_n/ b_n \leq t \}$. Then since $Q \parallel P$ on $\mathcal{B}_n$ for all $n$,

$$Q_n[ C_n(t) ] = \int_{b_n \log r_n - b_n t} d Q_n = \int r_n d P_n - b_n t \leq e^{b_n t} \int d P_n = e^{b_n t}.$$  \hspace{1cm} (2.2.7)

Since $b_n \uparrow \infty$, $Q_n[ C_n(t) ]$ converges to zero if $t < 0$. Therefore $Q(Z < 0) \to 0$ and the lemma is proved.

**2.2.2 Doob Decomposition and Convergence**

To evaluate the asymptotic behavior of the normed likelihood-ratio, we may utilise the fact that $\{ \log r_n, \mathcal{B}_n, Q \}$ is a submartingale possessing a unique Doob decomposition

$$\log r_n = M_n + A_n, \text{ say,}$$ \hspace{1cm} (2.2.8)

where $M_n$ is a martingale and $A_n$ is an increasing process given by

$$M_n = \sum_{k=0}^{n-1} [U_k - E(U_k | \mathcal{B}_{k-1})] \hspace{1cm} (2.2.9)$$

$$A_n = \sum_{k=0}^{n-1} E(U_k | \mathcal{B}_{k-1}), \hspace{0.5cm} A_0 = 0 \hspace{1cm} (2.2.10)$$

and

$$U_1 = \log r_1 - \log r_{1-1} \hspace{1cm} (2.2.11)$$

$E(A_n) = E(\log r_n) = K-L$ information in the observations $X_1, \ldots, X_n$. $A_n$ is therefore termed the "conditional K-L information".

Now, under the assumptions

$$(1) \hspace{0.5cm} M_n / A_n \to 0, \text{ a.s} \hspace{1cm} (2.2.12)$$

and
(2) there exists $0 < b_n \uparrow \infty$ such that

\[ b_n^{-1} A_n \rightarrow W > 0, \text{ a.s.,} \quad (2.2.13) \]

it follows that

\[ b_n^{-1} \log r_n \rightarrow W, \text{ a.s.} \]

The conditions (1) and (2) appear to hold for many stochastic processes.

**Example 2.2.2 (Galton-Watson branching process)**

Let \( \{X_n\} \) be successive generation sizes of a Galton-Watson branching process, with \( X_0 = 1 \) and the offspring distribution given by

\[
P_\Theta \left[ X_1 = j \right] = a_j \theta^j / A(\theta), \quad (\theta > 0), \quad A(\theta) = \sum_0^\infty a_j \theta^j.
\]

Then,

\[
E(X_1) = \mu(\theta) = \theta A'(\theta) / A(\theta),
\]

\[
\text{Var}(X_1) = \sigma^2(\theta) = \theta \mu'(\theta)
\]

where dot denotes the first derivative w.r.t \( \theta \). The log-likelihood ratio

of \( P_\Theta \) w.r.t. \( P_{\Theta_0} \) for the observations \( (X_0, X_1, \ldots, X_n) \) is given by

\[
\log r_n = \sum_1^n Y_i \log \theta / \theta_0 - N_n \log A(\theta) / A(\theta_0), \quad (2.2.15)
\]

where \( N_n = \sum_1^n X_{i-1} \) and \( Y_i \) denotes the number of offsprings for the \( i \)-th individual. If \( \mu'(\theta) > 1 \) and \( N_n \uparrow \infty \), a.s,

\[ X_n / \mu^n(\theta) \rightarrow W > 0, \text{ a.s.,} \]

where \( W \) is a r.v. with \( E(W) = 1 \). Taking

\[
b_n = (\mu^n(\theta) - 1) [ (N(\theta) - 1) \sigma^2(\theta) ]^{-1},
\]

\[ N_n / b_n \rightarrow W, \text{ a.s.} \]

It follows from the s.l.l.n for random sums of i.i.d r.v.s that

\[ \sum_1^n Y_i / N_n \rightarrow \mu(\theta), \text{ a.s} \]

and hence
\[
\lim_{n \to \infty} \frac{b_n \log r_n}{\mu(\theta) \log \theta_n - \log A(\theta)/A(\theta_n)} W_n \quad \text{a.s.} \quad (2.2.16)
\]

In this example, the Doob decomposition (2.2.8) gives

\[
M_n = \log \frac{\theta_n}{\theta_1} \sum_{i=1}^{n_i} [ Y_i - \mu(\theta) ],
\]

and

\[
A_n = [ \mu(\theta) \log \theta_n - \log A(\theta)/A(\theta_n) ] N_n.
\]

Obviously, conditions (2.2.12) and (2.2.13) are satisfied.

2.3. Decomposable Stochastic Processes

Consider the stochastic process \([ X_t, 0 \leq t \leq T \]). Let \( \mathcal{B}_t \) denote the \( \mathcal{F} \)-field induced by \([X_u, 0 \leq u \leq t] \). \( \{ X_t \} \) is said to be decomposable if increments \( X_t - X_s \) are independent for disjoint intervals \([s,t)\). If \( P_\theta = Q \) and \( P_{\theta_0} = P \) are the probability laws governing the stochastic processes, we assume that \( Q \ll P \) on \( \mathcal{B}_t \) for all \( t \). Denoting the R-N derivative \( dQ_t / dP_t \) on \( \mathcal{B}_t \) by \( r_t \), \( \{ r_t, \mathcal{B}_t, P \} \) is a martingale and \( \{ Y_t \} = \{ \log r_t \} \) is a submartingale w.r.t \( Q \).

For a decomposable process \( X_t, Y_t \) is also a decomposable process (see Gihman and Skorohod (1975) pp 524). If \( X_t \) does not have fixed discontinuities, \( Y_t \) is also free from fixed discontinuities. We shall assume this to be the case. Let \( A_t = E(Y_t) \) be the centering function. Then \( Y_t^c = Y_t - A_t \) is a centered decomposable process and hence is a martingale. This gives the Doob-Meyer (D-M) decomposition of \( Y_t = Y_t^c + A_t \), (since \( Y_t^c \) is a martingale \( A_t \) is increasing). If \( \langle Y_t^c \rangle \) is the quadratic variation process of the martingale \( Y_t^c \),

\[
\text{Var} \ Y_t^c = E \langle Y_t^c \rangle,
\]

which is equal to \( \text{var} \ Y_t = b_t \) (say), since \( A_t \) is non random. Since \( Y_t^c \) is decomposable, \( \langle Y_t^c \rangle \) is non-random (by definition of \( \langle Y_t^c \rangle \)) and hence...
\[ \langle Y_t^c \rangle = b_t. \]

By Basawa and Prakasa Rao (1980), if \( b_t \to \infty \)

\[ Y_t^c / b_t = Y_t^c / \langle Y_t^c \rangle \to 0, \text{ a.s.} \]

Therefore,

\[ Y_t / b_t \to \text{lim} A_t / b_t, \text{ a.s.} \quad (2.3.1) \]

if the latter exists, finite or infinite.

This result corresponds to the result (2.2.6) for sequences of independent r.v.s satisfying the assumptions (2.2.5). \( A_t \) may be viewed as the K-L information in the \( \{X_u, u \in [0,t]\} \). If the process \( X_t \) and hence \( Y_t \) has stationary independent increments,

\[ A_t = t \mathbb{E}(Y_t) \text{ and } b_t = t \text{Var}(Y_t), \]

and (2.3.1) becomes

\[ Y_t / t \to \mathbb{E}(Y_t), \text{ a.s.} \quad (2.3.2) \]

\( \mathbb{E}(Y_1) \) is the K-L information contained in \( \{X_u, 0 \leq u \leq 1\} \).

**Example 2.3.1 (Gaussian process)**

Consider the Gaussian process \( X_t \) given by the stochastic D.E

\[ dX_t = \mu(\theta,t)dt + B(t)dW_t, \quad 0 \leq t \leq T, \quad (2.3.3) \]

where \( \mu(\theta,t) \) and \( B(t) \) are known functions, \( \theta = \theta_0 \) under \( P \) and \( W_t \)

is the standard Wiener process. Then the R-N derivative \( \frac{dP_{\theta,T}}{dP_{\theta_0,T}} = r_T \)

for \( \{X_t, 0 \leq t \leq T\} \) is given by (see Liptser and Shiryayev (1978))

\[
\begin{align*}
\log r_T &= \int_0^T \frac{\mu(\theta,s) - \mu(\theta_0,s)}{B^2(s)} \, dX_s - 1/2 \int_0^T \frac{\mu^2(\theta,s) - \mu^2(\theta_0,s)}{B^2(s)} \, ds \\
&= \frac{1}{2} \int_0^T \left( \frac{\mu(\theta,s) - \mu(\theta_0,s)}{B(s)} \right)^2 \, ds + \int_0^T \frac{\mu(\theta,s) - \mu(\theta_0,s)}{B(s)} \, dW_s \quad (2.3.4)
\end{align*}
\]
Now the second term $U_T$ of (2.3.4) is a martingale with non-random quadratic variation process

$$\langle U_T \rangle = E \langle U_T \rangle - \int_0^T \frac{\mu(\theta, s) - \mu(\theta, s)}{B(s)} \, ds. \quad (2.3.5)$$

Hence, if $\langle U_T \rangle \uparrow \infty$,

$$\langle U_T \rangle \quad U_T \longrightarrow 0, \ a.s. \quad (2.3.6)$$

Since the first term of (2.3.4) equals $\langle U_T \rangle / 2$ and is non-random, corresponding to (2.3.1), we have

$$\log r_T / \langle U_T \rangle \longrightarrow 1/2, \ a.s. \quad (2.3.7)$$

If $\mu(\theta, t) = \mu(\theta)$ and $B(t) = B$, $\{X_t\}$ has stationary and independent increments. Then

$$E \langle U_T \rangle = T [\mu(\theta) - \mu(\theta)]^2 / B^2,$$

$$E(Y_1) = \{\mu(\theta) - \mu(\theta)\}^2 / 2B^2,$$

and (2.3.2) follows.

**Example 2.3.2 (Compound Poisson process)**

Let $X_t = Y_1 + \ldots + Y_N$, where $Y_i$'s are i.i.d with probability density function $f(y_i / \theta)$ under $P_{\theta}$ or more generally, with likelihood ratio

$$dP_{\theta} / d P_{\theta_0} = f(y / \theta) / f(y / \theta_0),$$

and $N_t$ is an independent Poisson process with parameter $\lambda(\theta)$ under $P_{\theta}$. For continuous observation in $[0, T]$ denoting $\log [f(y_i / \theta) / f(y_i / \theta_0)]$ by $U_{1T}$

$$\log r_T = N_T \log \frac{\lambda(\theta)}{\lambda(\theta_0)} - \frac{1}{T} \sum_{1}^{NT} U_{1T} \quad (2.3.8)$$

Since $N_T \longrightarrow \infty$ as $T \longrightarrow \infty$, by s.l.l.n

$$\frac{1}{N_T} \sum_{1}^{N_T} U_{1T} \longrightarrow K(\theta, \theta_0), \ a.s., \quad (2.3.9)$$

and

$$N_T/T \longrightarrow \lambda(\theta), \ a.s.$$
where \( K(\theta, \theta_0) = E_{\theta}(U) \).

Hence,
\[
T^{-1} \log r_T \rightarrow \bar{\lambda}(\theta) \log \bar{\lambda}(\theta) / \bar{\lambda}(\theta_0) - \left[ \bar{\lambda}(\theta) - \bar{\lambda}(\theta_0) \right] + \bar{\lambda}(\theta)K(\theta, \theta_0) \quad (2.3.10)
\]

We may note that (2.3.9) would hold if \( \Sigma_1 [U_1 - K(\theta, \theta_0)] \) is a martingale, a sufficient condition for which is that \( N_t \) is \( \mathcal{F}_t \)-adapted.

If \( N_t \) is a time dependent Poisson process with intensity \( \lambda(\theta, t) \) under \( P_\theta \), (2.3.8) is modified to
\[
\log r_T = \int_0^T \left\{ \log \frac{\bar{\lambda}(\theta, s)}{\bar{\lambda}(\theta_0, s)} \right\} dN_s - \int_0^T \left\{ \bar{\lambda}(\theta, s) - \bar{\lambda}(\theta_0, s) \right\} ds + \sum_{t=1}^{N_T} U_t, \quad (2.3.11)
\]

(see Akritas and Johnson (1981)).

Under \( P_\theta \), \( \int_0^T [dN_s - \bar{\lambda}(\theta, s) ds] \) and \( \sum_{t=1}^{N_T} [U_t - K(\theta, \theta_0)] \) are martingales. Hence
\[
V_T = \int_0^T \log \frac{\lambda(\theta, s)}{\lambda(\theta_0, s)} \left[ dN_s - \bar{\lambda}(\theta, s) ds \right] \quad (2.3.12)
\]
is a martingale with non-random quadratic variation process
\[
\langle V_T \rangle = \mathbb{E} [V_T] = \int_0^T \left\{ \log \frac{\lambda(\theta, s)}{\lambda(\theta_0, s)} \right\} \bar{\lambda}(\theta, s) ds. \quad (2.3.13)
\]

Assuming \( \int_0^T \bar{\lambda}(\theta, s) ds \uparrow \infty \) as \( T \uparrow \infty \) (implying \( N_T \uparrow \infty \), a.s) and \( \langle V_T \rangle \uparrow \infty \),
\[
\langle V_T \rangle^{-1} V_T \rightarrow 0 \text{, a.s.}
\]
and hence,
\[
\log r_T / \langle V_T \rangle \rightarrow g(\theta; \theta_0) = \lim (A_T/b_T), \text{ a.s.} \quad (2.3.14)
\]

where,
Example 2.3.3. Suppose $X_t$ has a continuous component given in Example 2.3.1 and a jump component given in Example 2.3.2 and these components are independent. Then the log R-N derivative of the process observed in $[0,T]$ is given by the sum of (2.3.4) and (2.3.11). Choosing either $\langle U_T \rangle$ given by (2.3.5) or $\langle V_T \rangle$ given by (2.3.13), as the norm for $\log r_T$, we get

\[
\lim_T (\log r_T / U_T) = 1/2 + g(\theta, \theta_0) \lim_T \langle V_T \rangle \overline{\langle U_T \rangle} \tag{2.3.16}
\]

Alternatively, we may proceed as follows. From (2.3.4) and (2.3.14), we can get

\[
E(\log r_T) = A_T + 1/2 \langle U_T \rangle.
\]

Since the components are independent, from (2.3.5) and (2.3.15), we get

\[
b_T = \text{Var} (\log r_T) = \langle U_T \rangle - A_T + \varphi^2 \int_0^T \lambda(\theta, s) ds.
\]

From (2.3.1),

\[
\lim_T (\log r_T / b_T) = 1/2 \lim_T \frac{\langle U_T \rangle}{b_T} + g(\theta, \theta_0) \lim_T \frac{\langle V_T \rangle}{b_T}. \tag{2.3.17}
\]

Remark 2.3.1: A decomposable process with finite Levy measure is the sum of two independent processes, a Gaussian process and a compound
Poisson process (Loeve (1978), Akritas and Johnson (1981)). Therefore, Example 2.3.3 also shows that for decomposable processes with finite Levy measure the normed log Radon-Nikodym derivative is always degenerate in the limit.

2.4. Arbitrary Stochastic Processes

Consider an arbitrary measurable stochastic process \(\{X_t, 0 \leq t \leq T\}\), trajectories of which are elements of a function space on which Borel field and probability measure \(P_0\) are defined. Let \(r_t\) denote the R-N derivative \(dP_t^u / dP_0^u\) for \(\{X_u, 0 \leq u \leq t\}\) where we have assumed that \(P_t^u \| P_0^u\). Then under \(P_0\), \(\log r_t\) is a sub-martingale. Then as per Doob Meyer decomposition (ref Liptser and Shiryayev (1978))

\[
\log r_t = Y_t^c + A_t, \quad (2.4.1)
\]

where \(Y_t^c\) is a martingale and \(A_t\) is the increasing process of \(\log r_t\), which is non-negative, \(\mathcal{G}_t\) adapted and non-decreasing. \(A_0 = 0, \ E(A_t) < \infty\) for all \(t \in [0,T]\). \(A_t\) was non-random for decomposable processes but may be random in general. \(E(A_t) = E(\log r_t) = K-L\) information contained in \(X_u\) \((0 \leq u \leq t)\). \(A_t\) may be viewed as the conditional K-L information process. Assume that

(1) \(\langle Y_t^c \rangle \uparrow \infty \) as \(t \uparrow \infty\), \(\quad (2.4.2)\)

(2) there exists a non-random function \(b_t, 0 \leq b_t < \infty\), such that

\[
\langle Y_t^c \rangle / b_t \rightarrow W, \ a.s, \quad (2.4.3)
\]

where \(0 < W < \infty\) is possibly random. Since from (1), by s.l.l.n for martingales,

\[
Y_t^c / \langle Y_t^c \rangle \rightarrow 0, \ a.s.
\]

Thus,
which may be a r.v.

For diffusion type processes and point processes, the R-N derivatives have been defined by many authors (cf. Liptser and Shiryayev (1978)). From these we can derive the D-M decomposition for the submartingale log $r_t$. We shall derive (2.4.4) for some of these processes.

a) Diffusion-type Processes: Let $\{X_t, 0 \leq t \leq T\}$ be a diffusion type process given by the stochastic D.E

$$dX_t = c(\theta, X_t, t)dt + B(X_t)dW_t, \ 0 \leq t \leq T, \ (2.4.5)$$

under $P_\Theta$, where $c(\theta, X_t, t)$ and $B(X_t)$ are $\mathfrak{F}_t$-adapted. We assume that there exists a solution of the above stochastic D.E which is a Markov process. Then the log R-N derivative is given by (Liptser and Shiryayev (1978) Vol.I, page 277)

$$\log r_T = \int_0^T \frac{c(\theta, X_s) - c(\theta^0, X_s)}{B^c(X_s)} dW_s + 1/2 \int_0^T \frac{c(\theta, X_s) - c(\theta^0, X_s)}{B(X_s)} ds. \ (2.4.6)$$

The first integral on the right side of (2.4.6) is a martingale (KaThanpur (1980)) and the second integral $A_t$ is non-decreasing, $\mathfrak{F}_t$-adapted with $A_0 = 0$, $E(A_t) < \infty$ for all $t \in [0, T]$ and hence represents the increasing process of the D-M decomposition. As such, we obtain that

$$V^-_T = \int_0^T \frac{c(\theta, X_s) - c(\theta^0, X_s)}{B(X_s)} dW_s, \ (2.4.7)$$
\[ A_T = \frac{1}{2} \left\{ \frac{c(\theta, X, s) - c(\theta_0, X, s)}{B(X, s)} \right\} ds, \quad (2.4.8) \]

\[ \langle Y_T \rangle = \left\{ \frac{c(\theta, X, s) - c(\theta_0, X, s)}{B(X, s)} \right\} ds = 2A_T. \quad (2.4.9) \]

If \( A_T \uparrow \infty \), \( A_T^{-1} \rightarrow c \), a.s.

Thus we obtain that if

1. \( A_T \uparrow \infty \) \( \Rightarrow \) \( E(A_T) \rightarrow W \), a.s,

\[ E(A_T)^{-1} \log r_T \rightarrow W, \text{ a.s} \] \quad (2.4.10)

**Example 2.4.1**

In (2.4.5) if \( c(\theta, X, s) = \theta X, \quad B(X, s) = 1 \) and \( X_0 = 0, \)

\[ A_T = \frac{1}{2}(\theta - \theta_0)^2 \int_0^T X_s^2 ds. \]

It is well known that (see Basawa and Scott (1983)) asymptotically

\[ E_\theta \left( \int_0^t X_s^2 ds \right) = \begin{cases} -\frac{t}{\theta} & \text{for } \theta < 0, \\ e^{\frac{\theta t}{2}} \theta^2/(\theta^2) & \text{for } \theta > 0. \end{cases} \]

Further \( E_\theta \left( \int_0^T X_s^2 ds \right)^{-1} \int_0^T X_s^2 ds \begin{cases} 1, & \text{for } \theta < 0, \\ \omega, & \text{for } \theta > 0, \end{cases} \text{ a.s.} \]

where \( W \) has a \( \chi^2 \) distribution with one degree of freedom. It follows that

\[ E(A_T) = \frac{1}{2}(\theta - \theta_0)^2 E(\int_0^T X_s^2 ds), \]

and as such

\[ E(A_T)^{-1} \log r_T \rightarrow \begin{cases} 1, & \text{for } \theta < 0, \\ \omega, & \text{for } \theta > 0, \end{cases} \text{ a.s} \] \quad (2.4.11)

**b) Point processes**: Let \( \{ X_t, 0 \leq t \leq T \} \) be a point process with
D-M decomposition $X_t = m_t(\theta) + c_t(\theta)$, where $m_t(\theta)$ is a martingale and $c_t(\theta)$ is the increasing process. $c_t(\theta)$ is called the compensator.

Let $c_t(\theta) = \int_0^t \lambda(\theta, X_s) \, ds$, where $\lambda(\theta, X_s)$ is $\mathcal{G}_s$-predictable. The log R-N derivative is then given by (Liptser and Shiryayev (1978), vol.II page 315)

$$Y_T = \log r_T = \int_0^T \log \frac{\lambda(\theta, X_s)}{\Lambda(\theta, X_s)} \, dX_s - [c_T(\theta) - c_T(\theta_0)]. \quad (2.4.12)$$

The D-M decomposition of $X_t$ yields the D-M decomposition for $Y_T = Y_T^c + A_T$, where

$$Y_T^c = \int_0^T \log \frac{\lambda(\theta, X_s)}{\Lambda(\theta, X_s)} \, dm_s(\theta), \quad (2.4.13)$$

and

$$A_T = \int_0^T \log \frac{\lambda(\theta, X_s)}{\Lambda(\theta, X_s)} \, ds = \left[ c_T(\theta) - c_T(\theta_0) \right]. \quad (2.4.14)$$

This assertion follows because $Y_T^c$ is a martingale and $A_T$ is non-decreasing and $\mathcal{G}_T$-adapted (since $\lambda(\theta, X, T)$ and $c_T(\theta)$ are so) with $A_0 = 0$ and $E(A_t) < \infty$ for $0 \leq t \leq T$. Hence

$$\langle Y_T^c \rangle = \int_0^T \left\{ \log \frac{\lambda(\theta, X_s)}{\Lambda(\theta, X_s)} \right\}^2 \, dm_s(\theta). \quad (2.4.15)$$

Since $\langle m_T(\theta) \rangle$ is shown to be equal to $\int_0^T (1 - \Delta c_s(\theta)) \, dc_s(\theta)$ by Liptser and Shiryayev (1978) (Vol.II, page 269), from (2.4.15),

$$\langle Y_T^c \rangle = \int_0^T \left\{ \log \frac{\lambda(\theta, X_s)}{\Lambda(\theta, X_s)} \right\}^2 \left\{ 1 - \Delta c_s(\theta) \right\} \, dc_s(\theta). \quad (2.4.16)$$
Here the relationship between $\langle Y_T^c \rangle$ and $A_T$ is not as explicit as in the case of diffusion-type processes and hence the limit of $[E(A_T)]^{-1} \log r_T$ could not be simplified based on conditions on $A_T$ such as in (2.4.10).

However, under some more restrictions on $c_t(\theta)$, we may obtain limits based on conditions on $c_t(\theta)$ alone. For example, suppose that $c_t(\theta)$ has a.s. continuous sample paths, with $\lambda(\theta,x,s) = \lambda_1(\theta) \lambda_2(x,s)$. Then $\Delta c_t(\theta) = 0$, and from (2.4.14) and (2.4.16),

$$\langle Y_T^c \rangle = \left[ \log \frac{\lambda_1(\theta)}{\lambda_1(\theta)} \right] T$$

and

$$A_T = \log \frac{\lambda_1(\theta)}{\lambda_1(\theta)} c_T(\theta) - [c_T(\theta) - c_T(\theta)]. \tag{2.4.17}$$

Thus if $c_T(\theta) \uparrow \infty$, $\frac{Y_T^c}{c_T(\theta)} \longrightarrow 0$, a.s. \tag{2.4.19}

$$A_T/c_T = \log \frac{\lambda_1(\theta)}{\lambda_1(\theta)} - \left[ 1 - \frac{\lambda_1(\theta)}{\lambda_1(\theta)} \right]. \tag{2.4.20}$$

Further, if there exists $b_T > 0$, non-random, such that

$$c_T(\theta) / b_T \longrightarrow W, \ a.s., \tag{2.4.21}$$

we obtain that

$$b_T^{-1} \log r_T \longrightarrow \left\{ \log \frac{\lambda_1(\theta)}{\lambda_1(\theta)} - \left[ 1 - \frac{\lambda_1(\theta)}{\lambda_1(\theta)} \right] \right\} W, \ a.s. \tag{2.4.22}$$

In particular, if $b_T = E[c_T(\theta)]$,

$$[E(A_T)]^{-1} \log r_T \longrightarrow W, \ a.s. \tag{2.4.23}$$

**Example 2.4.2.** Let $X_t$ be a pure birth process with $X_0 = 1$ and...
birth rate $\theta > 0$. The log R-N derivative is given by

$$\log r_T = \log \frac{\theta / \theta_0 - (\theta - \theta_0)}{\theta} \int_0^T x_s \, ds.$$ 

The compensator for this process is given by

$$c_t(\theta) = \theta \int_0^t x_s \, ds,$$

where,

$$\int_0^t x_s \, ds = \sum_{j=1}^\infty x_{\tau_j-1} (\tau_j - \tau_{j-1}) + \int_0^T (t - \tau_j) x_t \, ds,$$

$\tau_j$ denoting the jump points, is a stochastic process with a.s. continuous sample paths. The D-M decomposition of $Y_T = \log r_T$ yields

$$Y_T^c = \log \frac{\theta / \theta_0}{m_T(\theta)},$$

and

$$A_T = \log \frac{\theta / \theta_0}{c_T(\theta)} - [c_T(\theta) - c_T(\theta_0)].$$

Now, it is well-known that $(X_t - 1)^{-1} \int_0^t x_s \, ds \longrightarrow 1/\theta$, a.s. and $\theta^2(e^{\theta T} - 1)^{-1} (X_t - 1) \longrightarrow W$, a.s., where $W$ has an exponential density with unit mean. Hence

$$\theta^2 (e^{\theta t} - 1)^{-1} c_t(\theta) \longrightarrow W, \text{ a.s.}$$

It follows that

$$\theta^2 (e^{\theta T} - 1)^{-1} \log r_T \longrightarrow [\log \frac{\theta / \theta_0 - (1 - \theta_0^2)}{m_T(\theta)}]W, \text{ a.s.}$$

2.5 Convergence in Distribution

We have so far discussed the strong (a.s) convergence of the log-likelihood ratio (or the log R-N derivative) with norm $b_n$. It is evident in the discussion (see section 2.2) that this convergence depends in turn on that of the components of the Doob (or D-M) decomposition of the submartingale $\{\log r_n, \mathcal{F}_n, W\}$. We now study the convergence in distribution aspects, once again with the help of these components.
Consider the Doob decomposition (section 2.2)

\[ \log r_n = M_n + A_n. \]

Suppose we make the following assumption:

**Assumption A.** There exists \( 0 < b_n \uparrow \infty \) such that

\[ b_n^{-1} A_n \xrightarrow{\text{as}} W_p > 0, \]

and

\[ b_n^{-1} \langle M_n \rangle \xrightarrow{\text{as}} W_2, \quad 0 \leq W_2 \leq \infty, \]

By s.l.l.n for martingales,

\[ \langle M_n \rangle \xrightarrow{\text{s.l.l.n}} 0, \text{ a.s on } \langle \langle M_n \rangle \rangle \leq \infty. \]

It follows that under Assumption A, we obtain the convergence

\[ b_n^{-1} \log r_n \xrightarrow{\text{as}} W_1, \]  

(see also Hall and Heyde (1980)).

Now,

\[ b_n (\log r_n - A_n) = b_n M_n, \]

\[ = b_n \langle M_n \rangle \xrightarrow{\text{as}} W_2. \]

By martingale C.L.T, (see Basawa and Prakasa Rao (1980)) under appropriate conditional Lindeberg-Feller conditions, since

\[ b_n^{-1} \langle M_n \rangle \xrightarrow{\text{as}} W_2, \text{ bounded, and} \]

\[ \langle M_n \rangle \xrightarrow{\text{as}} N(0,1), \]

\[ b_n (\log r_n - A_n) \xrightarrow{\text{as}} N(0, W_2) \]  

(2.5.3)

This is a weighted normal law. Thus by the assumption that the increasing process \( A_n \) and \( \langle M_n \rangle \), both normed by \( b_n \) converge to appropriate non-negative random variables, we obtain a.s convergence (2.5.2) of \( \log r_n \) and convergence in law of \( M_n = \log r_n - A_n \) given by (2.5.3). Note that (2.5.3) may be converted into a convergence in law for \( \log r_n \) by making
additional assumptions on $A_n$. Thus assume that

Assumption B. $[A_n - E(A_n)] b_n \xrightarrow{L} W_3$. (2.5.4)

It follows that since $E(\log r_n) = E(A_n)$ (= K-L information in the sample)

$b_n [(\log r_n - E \log r_n) - (A_n - E(A_n))] \xrightarrow{L} N(0, W_2)$

Using a.s convergent constructions on an equivalent space (see for example, Pyke (1969), Billingsley (1971)), we may now obtain the following convergence in distribution

$b_n [(\log r_n - E(\log r_n)] \xrightarrow{L} W_3 + N(0, W_2)$. (2.5.5)

For i.i.d observations, $A_n = n K(Q, P) = E(A_n)$, where $K(Q, P)$

$= E_Q \left[ \log \frac{f(x/Q)}{f(x/P)} \right]$ represents the K-L information per observation,

$\langle M_n \rangle = n \sigma^2$, where $\sigma^2 = \text{Var} \left\{ \log \frac{f(x/Q)}{f(x/P)} \right\}$. Thus choosing $b_n = n$, we

find that $W_1 = K(Q, P)$, $W_2 = \sigma^2$ and $W_3 = 0$. The familiar a.s convergence

and convergence in law results follow.

For independent observations, $A_n = K_n(Q, P) = K-L$ information

in the sample. Choosing $b_n = \langle M_n \rangle$, we find $W_1 = K(Q, P)$ (section 2.2),

$W_2 = 1$ and $W_3 = 0$.

For dependent observations, $A_n$ and $\langle M_n \rangle$ are random. We

may choose $b_n = E_Q(A_n)$ or $b_n = E_Q \langle M_n \rangle$. However (2.5.5) remains

the most general result possible for convergence in distribution.

2.6 Fisher Information as Norm for Non-ergodic Models

We have already remarked in section 2.2 that the norm $b_n$ in

$Z_n$ may often be taken as the K-L information in the sample. However, the Fisher information serves equally well in many cases. This may be seen from the Taylor's series expansion of the log-likelihood function
Let $L_n(\theta_0)$ be the likelihood function. Expanding $L_n(\theta_0)$ as a Taylor's series around $\Theta$, we have

$$\log L_n(\theta) = \log L_n(\theta_0) + (\theta - \theta_0) \frac{\partial}{\partial \theta} \log L_n(\theta_0) + \frac{1}{2}(\theta - \theta_0)^2 \frac{\partial^2}{\partial \theta^2} \log L_n(\theta_0)$$

$$+ \frac{1}{3!} (\theta - \theta_0)^3 \frac{\partial^3}{\partial \theta^3} \log L_n(\theta_0)$$

(2.6.1)

where $\Theta_0$ lies between $\Theta$ and $\theta_0$.

It is well known that $U_n(\theta) = \frac{\partial}{\partial \theta} \log L_n(\theta)$ is a martingale with quadratic variation process

$$\langle U_n(\theta) \rangle = E \left\{ \left[ \frac{\partial}{\partial \theta} \log L_n(\theta) \right]^2 / \Theta_{n-1} \right\} = -E \left\{ \frac{\partial^2}{\partial \theta^2} \log L_n(\theta) / \Theta_{n-1} \right\},$$

(2.6.2)

where the expectations defined exist. Fisher information in the sample, say, $I_n(\theta)$, is given by

$$I_n(\theta) = E \langle U_n(\theta) \rangle.\quad (2.6.3)$$

We call $\langle U_n(\theta) \rangle$, the conditional Fisher information (as in Basawa and Prakasa Rao (1980)).

Further, by martingale strong convergence

$$\langle U_n(\theta) \rangle \rightarrow 0, \text{ a.s. (P_\theta)}, \text{ on } [\langle U_n(\theta) \rangle] \rightarrow \infty.\quad (2.6.4)$$

If $I_n^{-1}(\theta) U_n(\theta) \rightarrow W > 0, \text{ a.s. (P_\theta)},$ (2.6.5)

by martingale C.L.T,

$$\langle U_n(\theta) \rangle \rightarrow N(0,1).\quad (2.6.6)$$

(under appropriate conditional Lindberg-Feller conditions). We therefore have, from (2.6.1)
\[ \begin{align*}
\mathcal{I}_n^{-1}(\theta) \log r_n &= \mathcal{I}_n^{-1}(\theta) \left[ (\theta - \theta_0) \frac{\partial}{\partial \theta} \log L_n(\theta) < U_n(\theta) > -1 \right] \\
&= -\frac{1}{2} (\theta - \theta_0)^2 \frac{\partial^2}{\partial \theta^2} \log L_n(\theta) < U_n(\theta) > -1 \\
&- \frac{1}{3!} (\theta - \theta_0)^3 \frac{\partial^3}{\partial \theta^3} \log L_n(\theta) < U_n(\theta) > -1
\end{align*} \]

Hence if

1) \( \frac{\partial}{\partial \theta} \log L_n(\theta) < U_n(\theta) > -1 \xrightarrow{\text{a.s.}} -1, \) \( \text{a.s.} (P_\theta), \) (2.6.8)

and

2) \( \mathcal{I}_n^{-1}(\theta) \xrightarrow{\text{a.s.}} 0, \) \( \text{a.s.} (P_\theta) \) for all \( \theta_*, \) (2.6.9)

then, using (2.6.4),

\[ \mathcal{I}_n^{-1}(\theta) \log r_n \xrightarrow{\text{a.s.}} W, \text{a.s.} (P_\theta). \] (2.6.10)

Note: \( \mathcal{I}_n(\theta) \) may be replaced by an arbitrary sequence \( \{ b_n(\theta) \} \), \( \mathcal{I}_n \)

\( \xrightarrow{\text{a.s.}} \) \( \text{a.s.} (P_\theta) \), if (2.6.5) holds with \( \mathcal{I}_n(\theta) \) replaced by \( b_n(\theta) \).

2.7 Maximum Likelihood Estimates

The asymptotic behavior of the M.L.E depends on that of the likelihood ratio. Bhat and Prasad (1978) have shown the asymptotic weighted normal distribution of the M.L.E for non-ergodic models. Here we prove consistency and asymptotic weighted normality of MLEs for non-ergodic models essentially on the same lines.

It may be seen that the M.L.E \( \hat{\theta}_n \) is obtained as the solution of

\[ U_n(\hat{\theta}_n) = 0, \] (2.7.1)

where \( U_n(\theta) = \frac{\partial}{\partial \theta} \log L_n(\theta) \) is defined in section 2.6. It follows that

\[ \mathcal{I}_n^{-1}(\theta) \left[ U_n(\theta) - U_n(\hat{\theta}_n) \right] \xrightarrow{\text{a.s.}} 0, \) \( \text{a.s.} (P_\theta), \] (2.7.2)
on \([\cdot, \langle U_n(\theta) \rangle \rightarrow \infty]\), and if
\[
b_n(\theta) < U_n(\theta) \longrightarrow W > 0, \text{ a.s.} (P_{\theta}), \tag{2.7.3}
\]

\[
\langle U_n(\theta) \rangle^{-1/2} \left[ U_n(\theta) - U_n(\hat{\theta}_n) \right] \overset{d}{\longrightarrow} N(0,1). \tag{2.7.4}
\]

(under appropriate conditional Lindeberg-Feller conditions).

Now, using Taylor's series expansion for \(U_n(\hat{\theta}_n)\) around \(\theta\),
\[
U_n(\hat{\theta}_n) = U_n(\theta) + (\hat{\theta}_n - \theta) \frac{\partial}{\partial \theta} \log L_n(\theta) + \frac{1}{2}(\hat{\theta}_n - \theta)^2 \frac{\partial^2}{\partial \theta^2} \log L_n(\theta), \tag{2.7.5}
\]

where \(\theta_k\) lies between \(\hat{\theta}_n\) and \(\theta\).

\[
\langle U_n(\theta) \rangle^{-1} \left[ U_n(\theta) - U_n(\hat{\theta}_n) \right] = - (\hat{\theta}_n - \theta) \frac{\partial}{\partial \theta} \log L_n(\theta) \langle U_n(\theta) \rangle^{-1} - \frac{1}{2}(\hat{\theta}_n - \theta)^2 \frac{\partial^2}{\partial \theta^2} \log L_n(\theta) \langle U_n(\theta) \rangle^{-1}. \tag{2.7.6}
\]

Hence if
\[
(1) \frac{\partial}{\partial \theta} \log L_n(\theta) \langle U_n(\theta) \rangle^{-1} \longrightarrow -1, \text{ a.s.} (P_{\theta}), \tag{2.7.7}
\]

and
\[
(2) \frac{\partial^2}{\partial \theta^2} \log L_n(\theta) \langle U_n(\theta) \rangle^{1/2} \overset{d}{\longrightarrow} W \leq \infty, \text{ a.s.} (P_{\theta}), (2.7.8)
\]

by (2.7.2) and (2.7.4), it follows that
\[
\hat{\theta}_n \longrightarrow \theta, \text{ a.s.} (P_{\theta}) \tag{2.7.8}
\]

and
\[
\langle U_n(\theta) \rangle^{1/2} (\hat{\theta}_n - \theta) \overset{d}{\longrightarrow} N(0,1). \tag{2.7.9}
\]

By (2.7.3), we have
\[
b_n^2(\theta) (\hat{\theta}_n - \theta) \overset{d}{\longrightarrow} N(0, W^{-1}). \tag{2.7.10}
\]

2.8 Bayesian Inference
Here we shall be concerned with only the convergence aspects in Bayesian inference. The convergence is with respect to the measure $P_\theta$ on the sample space $(\mathcal{F}, \mathcal{A})$ and though this is really non-Bayesian in spirit, (see Basawa and Scott (1983)) we justify it on the grounds that to us, the parameter is not really random and the prior distribution represents merely our accumulated knowledge regarding the possible values of the parameter, prior to the experiment. (see also Zacks (1971) for justification) We discuss only those results which have a direct relevance here, in the sense that they are connected with the asymptotic behavior of the likelihood ratio. The results for the non-ergodic set-up follow the same lines as in the ergodic set-up with hardly any modification. We, therefore, omit the proofs and mention only the results. Thus our work is merely a review of the work done elsewhere.

a) Posterior distribution and the Berstein-von Mises Theorem

Prasad (1973) has shown that if $P_n^*(t|x_1,...,x_n)$ is the posterior density of $t = \frac{n^{1/2}}{} (\theta - \theta_n)$ where $\theta_n$ is the M.L.E, then, in the ergodic set-up under conditions which ensure that

$$\lim_{n \to \infty} n^{1/2} \left( \theta_n - \theta_0 \right) \overset{d}{\to} N(0, \Gamma^{-1}(\theta_0)),$$

and

$$n^{1/2} \left( \theta_n - \theta_0 \right) \overset{a.s.}{\to} N(0, \Gamma^2(\theta_0)),$$

for properly chosen priors, we may prove that

$$\lim_{n \to \infty} \int_{\mathbb{R}} H(t) P_n^*(t|x_1,...,x_n) - \Phi(\Gamma^{-1}(\theta_0), t) \ dt = 0, \ a.s.$$

where $\Phi(B,t)$ is the density of the normal distribution with mean 0 and variance B and $H(t)$ is a non-negative measurable function.

For non-ergodic models, where

$$b_n(\theta_0)(\theta_n - \theta_0) \overset{d}{\to} N(0, W^{-1}),$$

\(W\) is the asymptotic covariance matrix of $\theta_n$ as $n \to \infty$. The non-ergodic results then follow from the ergodic results by essentially the same arguments as in the ergodic case.
where \( W \) is the a.s limit of \( b_n^{-1}(\Theta_\ast) < U_n(\Theta_\ast) \), under conditions as in Prasad (1973), it can be proved that (see also Jegannathan (1980))
\[
\lim_{n \to \infty} \int_R |h(t)| p_n^*(t|x_1, \ldots, x_n) - \Phi(W^{-1}, t)|dt = 0, \quad \text{a.s.} \tag{2.8.1}
\]
where \( \Phi \) denotes the density of the weighted normal distribution \((N(0, W^{-1}))\)
and \( t = b_n^*(\Theta_\ast)(x - \Theta_\ast). \)

b) Bayes estimators: Bayes estimators are those which minimize the posterior risk corresponding to a given loss function. The loss function is generally assumed to be of the form
\[
\lambda(\Theta) \Pi(|T(x) - \Theta|) \quad \text{see Zacks (1971)}
\]
where \( T(x) \) is an estimator of \( \Theta \) and \( \Pi \) is a suitable convex function.

The posterior risk is then given by
\[
B_n(T) = \int \lambda(\Theta) \Pi(|T_n(x) - \Theta|) p_n(\Theta|x) d\Theta,
\]
where \( p_n(\Theta|x) \) represents the posterior distribution of \( \Theta \), given \( x \).

Following the methods used in Basawa and Prakasa Rao (1980), it can be shown that, if \( T_n \) is asymptotically equivalent to the MLE \( \hat{\Theta}_n \), which is consistent and asymptotically weighted normal, the Bayes estimator \( \Theta_n \) is also consistent and asymptotically weighted normal. Further, for every estimator \( T_n \)
\[
p \lim_{n \to \infty} B_n(T_n) \geq E_{\Theta}(1|W^{-1/2}Z|) \tag{2.8.2}
\]
and
\[
B_n(\Theta_n) \xrightarrow{P} E_{\Theta}(1|W^{-1/2}Z|),
\]
where \( Z \) is a standard normal variate independent of the mixing variable \( W \). Since the limiting risk \( B_n(\Theta_n) \) attains the lower bound in (2.8.2), \( \Theta_n \) is optimal in this sense. (see also Basawa and Scott (1983))

c) Bayes confidence intervals: Heyde and Johnstone (1979) have shown the asymptotic normality of stochastic processes in the sense of Bayesian
confidence intervals. They have shown that under conditions more or less as in Basawa and Scott (1983) or Bhat and Prasad (1979), the following result holds. Let $L_n(\theta)$ denote the likelihood function, and $\sigma_n$ the positive square root of $[- L''_n(\hat{\theta}_n)]^{-1}$, where dot denotes differential coefficient and $\hat{\theta}_n$ is the M.L.E. Suppose the conditions A1 - A5 of Heyde and Johnstone (1979) hold. Then, if $-\infty < b < a < \infty$, the posterior probability that $\hat{\theta}_n + b \sigma_n < \theta < \hat{\theta}_n + a \sigma_n$, viz.,

$$\int_{\hat{\theta}_n + b \sigma_n}^{\hat{\theta}_n + a \sigma_n} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi} \sigma_i} \exp \left(-\frac{1}{2} \left(\frac{x_i - \theta}{\sigma_i}\right)^2\right) \, d\theta$$

tends in $P_{\Theta_0}$-probability to

$$(2\pi)^{-1/2} \int_{b}^{a} e^{-u^2/2} \, du$$
as $n \to \infty$. 

---