Definition. Maximum Likelihood Estimator (MLE)
For $X_i \overset{iid}{\sim} f(x|\theta)$, the MLE is $\hat{\theta} = \arg\max_{\theta} f(x^n|\theta)$ for $\theta \in \Omega \subset \mathbb{R}^k$

1 Consistency of the MLE

This is a summary/direct restatement of Wald [1949] which shows that $\hat{\theta}$ is consistent. It consists of a set of assumptions, some introductory lemmas, and the proof of the main theorem.

Consider the following 8 assumptions

1. The distribution function $F_\theta(x)$ is either discrete for all θ or absolutely continuous for all θ.
2. Let
$$\begin{aligned} f(x|\theta,p) &= \sup_{|\theta-\theta'|\le p} f(x|\theta') \\ \phi(x|r) &= \sup_{|\theta|>r|} f(x|\theta) \\ f^*(x|\theta,p) &= \max(1, f(x|\theta,p)) \\ \phi^*(x|\theta,p) &= \max(1, \phi(x|r)) \end{aligned}$$
Assume that for sufficiently small p and sufficiently large r the expected values
$$\int_{-\infty}^{\infty} \ln f^*(x|\theta,p)\, dF_{\theta_0}(x) \quad \text{and} \quad \int_{-\infty}^{\infty} \ln \phi^*(x|r)\, dF_{\theta_0}(x)$$
are finite, $\theta_0 = \theta_{\text{true}}$.
3. If $\lim_{i\to\infty} \theta_i = \theta$ then $\lim_{i\to\infty} f(x|\theta_i) = f(x|\theta)$ for all x, except a set which may depend on θ (but not $\langle\theta_i\rangle$) and has P_{θ_0} probability 0.
4. If $\theta_0 \neq \theta_1$, there exists x such that $F_{\theta_0}(x) \neq F_{\theta_1}(x)$
5. If $\lim_{i\to\infty} |\theta_i| = \infty$ then $\lim_{i\to\infty} f(x|\theta_i) = 0$ except for a set of P_{θ_0} with probability 0 (and independent of $\langle\theta_i\rangle$)
6. $\int_{-\infty}^{\infty} |\log f(x|\theta_0)| dF_{\theta_0}(x) < \infty$
7. Ω is a closed subset of $\mathbb{R}^k$
8. $f(x|\theta,p)$ is measurable in x for any θ, p (not necessary for discrete θ)

Summary Wald's hypotheses can be succinctly stated as:

1. For all θ there exists $p = p(\theta)$ $E_\theta \sup_{|\phi-\theta|>p} \ln \dfrac{f(X|\phi)}{f(X|\theta)} < 0$ **(Wald's integrability condition)**
2. $\forall \theta, \delta > 0$ small enough, $E_\theta \ln \dfrac{f(X|\theta)}{\sup_{(\theta'-\theta)<\delta} f(X|\theta)} < \infty$
3. $p(x|\theta) \to 0$ as $||\theta|| \to \infty$

A more thorough explanation of these conditions and the theorem itself is available in Geyer.

Lemma 1. *For any $\theta \neq \theta_0$, we have $E \log f(X|\theta) < E \log f(X|\theta_0)$, where X is a chance variable with the distribution $F(X, \theta_0)$*

Proof:
Assumption 2 means that the expected values exist.
Assumption 6 means $E|\log f(X, \theta_0) < \infty$.
If $E \log f(X, \theta) = -\infty$, the lemma obviously holds.
If $E \log f(X, \theta) > -\infty$, then $E|\log f(X, \theta)| < \infty$.
Let $u = \log f(X, \theta) - \log f(X, \theta_0)$. $E|u| < \infty$. We know that for any RV u that is non-constant with probability 1 with finite expectation, $Eu < \log Ee^u$.
Since $Ee^u \leq 1$, $Eu < \log Ee^u \leq 0$ and thus $Eu < 0$.

Definition. Kullback-Leibler Information
Note that

$$E \log f(X|\theta) - E \log f(X|\theta_0) = E \log \frac{f(X|\theta)}{f(X|\theta_0)}$$

$$\int \log \frac{f_\theta(x)}{f_{\theta_0}(x)} P_{\theta_0}(dx) = \lambda(\theta)$$

By Jensen's Inequality, $\lambda(\theta) \leq 0$ for all θ and $\lambda(\theta) = \lambda(\theta_0)$ iff $f_\theta(x) = f_{\theta_0}(x)$ for almost all x (P_{θ_0}). So λ is non-positive on Θ and has a maximum at θ_0, where $\lambda(\theta_0) = 0$.

Lemma 2. $\lim_{p=0} E \log f(X, \theta, p) = E \log f(X, \theta)$

Proof:

Let $f^*(x, \theta, p) = \begin{cases} f(x, \theta, p) & f(x, \theta, p) \geq 1 \\ 1 & \textit{otherwise} \end{cases}$ *and* $f^*(x, \theta) = \begin{cases} f(x, \theta) & f(x, \theta) \geq 1 \\ 1 & \textit{otherwise} \end{cases}$

$f^{**}(x, \theta, p) = \begin{cases} f(x, \theta, p) & f(x, \theta, p) \leq 1 \\ 1 & \textit{otherwise} \end{cases}$ *and* $f^{**}(x, \theta) = \begin{cases} f(x, \theta) & f(x, \theta) \leq 1 \\ 1 & \textit{otherwise} \end{cases}$

$$\lim_{p \to 0} \log f^*(x, \theta, p) = \log f^*(x, \theta) \ \textit{a.e.} \quad \textit{(by Assumption 3)}$$

$$\lim_{p \to 0} E \log f^*(X, \theta, p) = E \log f^*(X, \theta) \ \textit{(since } \log f^*(x, \theta, p) \textit{ is incr in } p + \textit{Assumption 2)}$$

$$|\log f^{**}(x, \theta, p)| \leq |\log f^{**}(x, \theta)|$$

$$\lim_{p \to 0} \log f^{**}(x, \theta, p) = \log f^{**}(x, \theta) \ \textit{for } x \ \textit{a.e.}$$

$$\lim_{p \to 0} E \log f^{**}(X, \theta, p) = E \log f^{**}(X, \theta) \ \textit{(follows from the two previous lines).}$$

$$\textit{Since } \lim_{p \to 0} E \log f^*(X, \theta, p) = E \log f^*(X, \theta) \ \textit{and}$$
$$\lim_{p \to 0} E \log f^{**}(X, \theta, p) = E \log f^{**}(X, \theta), \ \textit{we are done.}$$

Lemma 3.

$$\lim_{r\to\infty} E \log \phi(X|r) = -\infty$$

Proof:

Assumption 5 implies $\lim_{r\to\infty} \log \phi(X|r) = -\infty$ *a.e.*

Assumption 2 says $E \log \phi^*(X, r) < \infty$.

$\log \phi(x, r) - \log \phi^*(x, r)$ *and* $\log \phi^*(x, r)$ *are decreasing functions of* r*, so the lemma follows by the monotone convergence theorem.*

The three lemmas posed above are necessary in order to prove the following theorems.

Theorem 4. *Let* ω *be any closed subset of the parameter space* Ω *which does not contain the true parameter point* θ_0*. Then*

$$\lim_{n\to\infty} \frac{\sup_{\theta\in\omega} f(X_1,\theta) f(X_2,\theta)\cdots f(X_n,\theta)}{f(X_1,\theta_0) f(X_2,\theta_0)\cdots f(X_n,\theta_0)} = 0 \textit{ with probability } 1$$

$$\textit{(more compact notation)}\ \lim_{n\to\infty} \frac{\sup_{\theta\in\omega} f(X^n|\theta)}{f(X^n|\theta_0)} = 0 \textit{ with probability } 1$$

Proof:

Let r_0 *be a positive number chosen such that*

$$E \log \phi(X, r_0) < E \log f(X, \theta_0) \tag{1}$$

The existence of such a positive number follows from Lemma 3. Let ω_1 *be the subset of* ω *consisting of all points* θ *of* ω *for which* $|\theta| \le r_0$*. With each point* θ *in* ω_1 *we associate a positive value* p_0 *such that*

$$E \log f(X, \theta, p_\theta) < E \log f(X, \theta_0) \tag{2}$$

The existence of such a p_θ *follows from Lemma 1 and Lemma 2. Since the set* ω_1 *is closed and bounded (and thus compact), there exist a finite number of points* $\theta_1, \cdots, \theta_h \in \omega_1$ *such that* $S(\theta_1, p_{\theta_1}) + \cdots + S(\theta_h, p_{\theta_h})$ *contains* ω_1 *as a subset. Here* $S(\theta, p)$ *denotes the sphere with center* θ *and radius* p*. Clearly,*

$$0 \le \sup_{\theta\in\omega} f(x_1,\theta)\cdots f(x_n,\theta) \le \sum_{i=1}^{h} f(x_1,\theta_i,p_{\theta_i})\cdots f(x_n,\theta_i,p_{\theta_i}) + \phi(x_1,r_0)\cdots\phi(x_n,r_0). \tag{3}$$

$$0 \le \sup_{\theta\in\omega} f(X^n|\theta) \le \sum_{i=1}^{h} f(X^n|\theta_i,p_{\theta_i}) + \phi(X^n|r_0)\ \textit{(more compact notation)}$$

The theorem is proved if we can show that

$$\lim_{n\to\infty} \frac{f(X_1,\theta_i,p_{\theta_i})\cdots f(X_n,\theta_i,p_{\theta_i})}{f(X_1,\theta_0)\cdots f(X_n,\theta_0)} = 0 \quad \textit{with probability 1}, (i = 1, ..., h) \tag{4}$$

and

$$\lim_{n\to\infty} \frac{\phi(X_1,r_0)\cdots\phi(X_n,r_0)}{f(X_1,\theta_0)\cdots f(X_n,\theta_0)} = 0 \quad \textit{with probability 1.} \tag{5}$$

(cont'd)
That is, we divide the whole expression in Equation (3) by $f(X^n|\theta_0)$ and consider each term of the sum separately.
Using Equation (1), Equation (2), and the strong LLN, we can rewrite Equation (4) and Equation (5) as

$$\lim_{n\to\infty}\sum_{\alpha=1}^{n}[\log f(X_\alpha,\theta_i,p_{\theta_i})-\log f(X_\alpha,\theta_0)]=-\infty \quad \textit{with probability 1}, (i=1,...,h)$$

and

$$\lim_{n\to\infty}[\log\phi(X_\alpha,r_\theta)-\log f(X_\alpha,\theta_0)]=-\infty \quad \textit{with probability 1.}$$

This essentially shows that the conditions imply that the log likelihood

$$l_n(\theta)=\frac{1}{n}\sum_{i=1}^{n}\log\frac{f_\theta(X_i)}{f_{\theta_0}(X_i)}$$

converges to λ (the KL Information) uniformly from above:

$$\lim_{n\to\infty}\sup_{\theta\in\omega}l_n(\theta)\le\sup_{\theta\in\omega}\lambda(\theta) \text{ almost surely}$$

However, the MLE doesn't always exist. We may be interested in what happens when we have an approximate MLE (AMLE) derived from a sequence of estimators $\{\overline{\theta}_n\}$. If the MLE exists, then this sequence should converge to the MLE, but if it doesn't exist, what does the AMLE converge to?

Theorem 5 (Consistency of the approximate MLE)**.** *Let $\overline{\theta}_n(x_1,...,x_n)$ be a function of the observations $x_1,...,x_n$ such that*

$$\frac{f(x_1,\overline{\theta}_n)\cdots f(x_n,\overline{\theta}_n)}{f(x_1,\theta_0)\cdots f(x_n,\theta_0)}\ge c>0 \tag{6}$$

for all n and $x_1,...,x_n$. Then

$$\lim_{n\to\infty}\overline{\theta}_n=\theta_0 \textit{ with probability } 1.$$

Since a maximum likelihood estimate $\hat{\theta}_n(x_1,\cdots,x_n)$, if it exists, obviously satisfies Equation (6) with $c=1$, this establishes the consistency of $\hat{\theta}_n(x_1,\cdots,x_n)$ as an estimate of θ.

Proof:
It is sufficient to prove that for any $\epsilon > 0$ the probability is one that all limit points $\overline{\theta}$ of the sequence $\{\overline{\theta}_n\}$ satisfy the inequality $|\overline{\theta} - \theta_0| \leq \epsilon$. The event that there exists a limit point $\overline{\theta}$ of the sequence $\{\overline{\theta}_n\}$ such that $|\overline{\theta} - \theta_0| > \epsilon$ implies that

$$\sup_{|\theta-\theta_0|\geq\epsilon} f(x_1,\theta)\cdots f(x_n,\theta) \geq f(x_1,\overline{\theta}_n)\cdots f(x_n,\overline{\theta}_n) \text{ for infinitely many } n.$$

But then

$$\frac{\sup_{|\theta-\theta_0|\geq\epsilon} f(x_1,\theta)\cdots f(x_n,\theta)}{f(x_1,\overline{\theta}_n)\cdots f(x_n,\overline{\theta}_n)} \geq c > 0 \text{ for infinitely many } n.$$

Since, according to Theorem 1, this is an event with probability zero, we have shown that the probability is one that all limit points $\overline{\theta}$ of $\{\overline{\theta}_n\}$ satisfy the inequality $|\overline{\theta} - \theta_0| \leq \epsilon$. This completes the proof.

This proves strong convergence, which is stronger than weak convergence (in the sense that consistency is a weak property - a property of distribution functions - and not a strong property - a property of infinite sequences of observed sums).
This proof was then extended by Wolfowitz [1949], to use only the weak law of large numbers. Wald's proof uses the strong LLN, but the same result (consistency) holds with the weak LLN, which is applicable to a larger class of variables.

Theorem 6 (Wolfowitz's Extension). *Let η and ϵ be given, arbitrarily small, positive numbers. Let $S(\theta_0, \eta)$ be the open sphere with center θ_0 and radius η, and let $\Omega(\eta) = \Omega - S(\theta_0, \eta)$. Let Wald's assumptions 1-8 hold.*
There exists a number $h(\eta)$, $0 < h < 1$, and another positive number $N(\eta, \epsilon)$ such that, for any $n > N(\eta, \epsilon)$,

$$P_0 \left\{ \frac{\sup_{\theta\in\Omega(\eta)} \prod_{i=1}^n f(X_i,\theta)}{\prod_1^N f(X_i,\theta_0)} > h^n \right\} < \epsilon$$

where P_0 is the probability of the relation in braces according to $f(x, \theta_0)$.

See Wolfowitz [1949] for the proof.

2 Cramèr: Asymptotic properties of MLEs

This section is based heavily on Cramér [1946].

Definition. Likelihood function A sample of n values from a variable has likelihood

$$L(x_1, ..., x_n, \theta) = \begin{cases} f(x_1,\theta)\cdots f(x_n,\theta) & \text{if } x \text{ is continuous} \\ p(x_1,\theta)\cdots p(x_n,\theta) & \text{if } x \text{ is discrete} \end{cases}$$

Definition. Maximum likelihood estimation When $X^n = (x_1, ..., x_n)$ are observed, we obtain the maximum likelihood estimate $\hat{\theta} = \arg\max_\theta f(X^n|\theta)$ as the solution to the likelihood equation

$$\frac{\partial \log L}{\partial \theta} = 0$$

where θ is conditioned on the sample values.

When an efficient estimate $\hat{\theta}$ of θ exists, the likelihood equation will have a unique solution equal to $\hat{\theta}$:

$$\frac{\partial \log L}{\partial \theta} = \sum_{1}^{n} \frac{\partial \log f(x_i|\theta)}{\partial \theta} = \frac{\partial \log g}{\partial \theta} = k(\hat{\theta} - \theta).$$

When a sufficient estimate $\hat{\theta}$ of θ exists, the likelihood equation reduces to

$$\frac{\partial \log L}{\partial \theta} = \frac{\partial \log g(\hat{\theta}, \theta)}{\partial \theta} = 0$$

Definition. Fisher Information
The Fisher information is the variance of the score:

$$0 \le \mathcal{I}(\theta) = E_{\theta_0}\left[\left(\frac{\partial}{\partial \theta}\log f(X,\theta)\right)^2\bigg|_{\theta=\theta_0}\right] = \int_{\mathbb{R}} \left(\frac{\partial}{\partial \theta}\log f(x,\theta)\right)^2 f(x,\theta)dx\bigg|_{\theta=\theta_0}$$

If $\log f(x,\theta)$ is twice differentiable with respect to θ, and under certain regularity conditions, the FI may also be written as

$$\mathcal{I}(\theta) = -E_{\theta_0}\left[\frac{\partial^2}{\partial \theta^2}\log f(X,\theta)\bigg|_{\theta=\theta_0}\right]$$

Theorem 7. *Asymptotic Normality of the MLE*
Let X_1^n be iid f_θ where $\Omega \subset \mathbb{R}$ is an open interval, all f_θ have common support, $0 < \mathcal{I}(\theta) < \infty$, $\int f(x|\theta)dx$ is twice differentiable under the integral, and

$$E \sup_{\theta \in B(\theta_0,\epsilon)} \Bigg| \underbrace{\frac{\partial^3}{\partial \theta^3}\log f(x|\theta)}_{exists} \Bigg| < \infty.$$

Suppose $\hat{\theta}_n$ is a consistent sequence of roots of the likelihood equation. Then

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} N\left(0, \mathcal{I}(\theta)^{-1}\right)$$

Proof:
Write $L(\theta|X^n) = \log f(X^n|\theta)$ (we regard X^n as fixed)
Taylor expansion gives

$$\underbrace{0 =}_{\hat{\theta} \text{ is MLE}} \ L'(\hat{\theta}_n) = L'(\theta_0) + (\hat{\theta} - \theta_0)L''(\theta_0) + \frac{1}{2}(\hat{\theta} - \theta_0)^2 L'''(\theta^*) \qquad \theta_n^* \in \left[\theta_0, \hat{\theta}_n\right]$$

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{\overbrace{\frac{1}{\sqrt{n}}L'(\theta_0)}^{*1}}{\underbrace{-\frac{1}{n}L''(\theta_0)}_{*2} - \underbrace{\frac{1}{2n}(\hat{\theta} - \theta_0)L'''(\theta^*)}_{*3}} := H_n$$

(cont'd)
We examine each labeled component separately:

$$*1: \qquad \frac{1}{\sqrt{n}}L'(\theta_0) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \frac{f'(X_i|\theta_0)}{f(X_i|\theta_0)} - \underbrace{E\left[\frac{f'(X_1|\theta_0)}{f(X_1|\theta_0)}\right]}_{=0}\right) \underbrace{\xrightarrow{d}}_{\textit{(by CLT)}} \quad N(0,\mathcal{I}(\theta_0))$$

$$*2: \qquad -\frac{1}{n}L''(\theta_0) \underbrace{\xrightarrow{p}}_{(byLLN)} \quad \mathcal{I}(\theta_0)$$

$$*3: \qquad (\hat{\theta}-\theta_0)\cdot\frac{1}{2}\cdot\frac{1}{n}\left|L'''(\theta^*)\right|$$

$$\frac{1}{n}\left|L'''(\theta^*)\right| \leq \frac{1}{n}\sum_{i=1}^n \sup_{\theta\in B(\theta_0,\epsilon)}\left|\frac{\partial^3}{\partial\theta^3}\log f(X_i|\theta)\right| \leq E\sup\left|\frac{\partial^3}{\partial\theta^3}\right| < \infty$$

$$\textit{But } (\hat{\theta}-\theta_0) \xrightarrow{p} 0$$

$$\textit{So } \frac{1}{2}\cdot\frac{1}{n}\left|L'''(\theta^*)\right|\cdot(\hat{\theta}-\theta_0) \to \quad 0.$$

Combining the above results and using Slutsky's theorem, we have

$$H_n \xrightarrow{d} -\frac{1}{\mathcal{I}(\theta_0)}N(0,\mathcal{I}(\theta_0)) \stackrel{d}{=} N\left(0,\frac{1}{\mathcal{I}(\theta_0)}\right) \ \textit{under } \theta_0$$

Definition. Expected and Observed Fisher Information

$$\mathcal{I}_1(\hat{\theta}_n) \text{ is the "expected Fisher information"}$$
$$-\frac{1}{n}L''(\hat{\theta}_n) \text{ is the "observed Fisher Information"}$$

Theorem 8. *Asymtotic Efficiency of MLE Under the assumptions of Theorem 7, under θ_0,*

$$(\hat{\theta}_n-\theta_0)\sqrt{-L''(\hat{\theta}_n)} \xrightarrow{d} N(0,1)$$

<u>*Note:*</u>
For a sequence of estimators $\{\hat{\theta}_n^*\}$ with $\sqrt{n}(\hat{\theta}_n^*-\theta_0) \xrightarrow{d} N\left(0,V(\theta_0)\right)$, under θ_0,

$$\frac{\frac{1}{\mathcal{I}_1(\theta_0)}}{V(\theta_0)} \textit{ is called the } \boldsymbol{\textit{asymptotic efficiency}} \textit{ of } \{\hat{\theta}_n^*\}$$

It is possible (even in regular problems) to find some θ_0 for which the asymptotic efficiency is larger than one (super-efficiency). There are theorems that say that we cannot have too many super-efficiency points.

3 Asymptotic Behavior of Posterior Distributions

This section is from Walker [1969]. Consider a prior $\pi(\theta)$ and a parametric family $f(\cdot|\theta)$. Impose the following conditions:

A1. $\Omega \subset \mathbb{R}$ is a closed set

A2. $\sup f(\cdot|\theta)$ is independent of θ

A3. if $\theta_1 \neq \theta_2$ then $\mu\{x : f(x|\theta_1) \neq f(x|\theta_2)\} > 0$ (μ is a σ-finite measure)

A4. if $\delta > 0$ is sufficiently small, $E_{\theta_0} \sup\limits_{\theta:|\theta-\theta'|<\delta} |\log f(x|\theta)| \to 0$ as $\delta \to 0$

A5. if Ω not bounded, then for any $\theta_0 \in \Theta$ and sufficiently large Δ,

$$\log f(x|\theta) - \log f(x|\theta_0) < K_\Delta(x,\theta) \quad \text{whenever } |\theta| > \Delta,$$

where $\lim\limits_{\Delta\to\infty} \int K_\Delta(x|\theta_0) f(x|\theta_0) d\mu \;\; < 0$ (It is allowable for the limit to be $-\infty$ here)

These conditions are introduced to ensure that when θ_0 is the true value of θ, $L_n(\theta) - L_n(\theta_0)$ is sufficiently small, with probability converging to 1 as $n \to \infty$. for all values of θ not near to θ_0. In the next conditions, let θ_0 be an interior point of Θ

B1. $\log f(x|\theta)$ is twice differentiable with respect to θ in some neighborhood of θ_0

B2. $\mathcal{I}(\theta_0) = \int f(x|\theta_0) \left(\frac{\partial \log f(x|\theta_0)}{\partial \theta_0} \right)^2 d\mu$. Then $0 < \mathcal{I}(\theta_0) < \infty$.

B3. $\int \frac{\partial f(x|\theta_0)}{\partial \theta_0} d\mu = \int \frac{\partial^2 f(x|\theta_0)}{\partial \theta_0^2} d\mu = 0$

B4. If $|\theta - \theta_0| < \delta$, where δ is sufficiently small, then

$$\left| \frac{\partial^2 \log f(x|\theta)}{\partial \theta^2} - \frac{\partial^2 f(x|\theta_0)}{\partial \theta_0^2} \right| < M_\delta(x,\theta_0) \text{ where } \lim_{\delta\to 0} \int M_\delta(x,\theta_0) f(x|\theta_0) d\mu = 0$$

This is equivalent to $E_{\theta_0} \sup\limits_{\theta:|\theta-\theta_0|<\delta} \left| \frac{\partial^2}{\partial \theta^2} \log f(x|\theta) \right| < \infty$.

These conditions make $L_n''(\cdot)$ behave smoothly for values of θ near θ_0; asymptotic normality of $\hat{\theta}_n$ is used only to show that $\hat{\theta}_n - \theta_0 = O_p(\sqrt{n})$. These conditions also ensure that when $\theta = \theta_0$, $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, 1/\mathcal{I}(\theta_0))$.

C1. π is continuous at $\theta = \theta_0$ and $\pi(\theta_0) > 0$

Denote the MLE by $\hat{\theta}_n, \hat{\theta}$

Theorem 9. *Asymptotic Normality of the Posterior Distribution*
Assume the regularity conditions above. Suppose the X_is are iid $f(\theta_i)$. Let $\sigma_n = \left(-L_n''(\hat{\theta}_n)\right)^{-1/2}$. Then for all a, b,

$$\int_{\hat{\theta}_n + b\sigma_n}^{\hat{\theta}_n + a\sigma_n} \pi_n(\theta|X^n) d\theta \xrightarrow{p} \frac{1}{\sqrt{2\pi}} \int_b^a \exp\left(-\frac{1}{2} y^2\right) dy$$

This theorem arises due to several preliminary results:

Lemma 10. *Let $N_0(\delta) = \{\theta : |\theta - \theta_0| < \delta\} = B(\theta_0, \delta) \subset \Omega$. Then there exists a positive number $k(\delta)$ that depends on δ such that*

$$\lim_{n\to\infty} P\left[\sup_{\theta\in\Theta - N_0(\delta)} n^{-1}\{L_n(\theta) - L_n(\theta_0) < -k(\delta)\right] = 1 \tag{7}$$

This obviously implies that $\hat{\theta}_n$ is weakly consistent, that is, $p\lim_{n\to\infty} \hat{\theta}_n = \theta_0$

Lemma 11.

$$\frac{1}{n} L''(\hat{\theta}_n) \xrightarrow{p} -\mathcal{I}(\theta_0) \tag{8}$$

Lemma 12. $L_n(\theta_0) - L_n(\hat{\theta}) = O_p(1)$

Corollary 13. $E(\Theta|X^n) \xrightarrow{p} \theta_0$ *provided* $\left|\int \theta\pi(\theta)d\theta\right| < \infty$

Theorem 14. *Assume that $\tilde{\theta} = E(\theta|X^n)$ is the Bayes estimator under squared error loss with prior density ω and that the conditions of Walker's theorem hold. Then $\sqrt{n}(\tilde{\theta} - \theta_0) \xrightarrow{d} N(0, \mathcal{I}(\theta_0)^{-1})$, that is, $\tilde{\theta}$ is consistent and asymptotically efficient.*

References

H. Cramér. *Mathematical Methods of Statistics*. Princeton University Press, 1946. ISBN 978-0-691-08004-8. Google-Books-ID: _db1jwEACAAJ.

C. J. Geyer. The Wald Consistency Theorem. URL https://www.stat.umn.edu/geyer/8112/notes/wald.pdf.

A. Wald. Note on the Consistency of the Maximum Likelihood Estimate. *The Annals of Mathematical Statistics*, 20(4):595–601, 1949. ISSN 0003-4851. URL https://www.jstor.org/stable/2236315. Publisher: Institute of Mathematical Statistics.

A. M. Walker. On the Asymptotic Behaviour of Posterior Distributions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 31(1):80–88, 1969. ISSN 2517-6161. doi: 10.1111/j.2517-6161.1969.tb00767.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1969.tb00767.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1969.tb00767.x.

J. Wolfowitz. On Wald's Proof of the Consistency of the Maximum Likelihood Estimate. *The Annals of Mathematical Statistics*, 20(4):601–602, 1949. ISSN 0003-4851. doi: 10.1214/aoms/1177729953. URL https://www.jstor.org/stable/2236316. Publisher: Institute of Mathematical Statistics.