Asymptotic Properties for Bayesian Neural Network in Besov Space - proof of Lemma 2

백승찬

서울대학교 통계학과 베이즈 연구실

마감: 2023년 2월 8일

목차

1 Lemma 2

Lemma 1.1 *(Theorem 4 of Ghosal and Van der Vaart[2007])*

Let $P^{(n)}_{(\theta)}$ *be product measures and* $d_n(\theta_0, \theta) = \frac{1}{n}\sum_{i=1}^{n}\int(\sqrt{p_{\theta_0,i}} - \sqrt{p_{\theta,i}})^2 d\mu_i$*. Suppose that for a sequence* $\epsilon_n \to 0$ *such that* $n{\epsilon_n}^2$ *is bounded away from zero, some* $k > 1$*, all sufficiently large* j *and sets* $\Theta_n \subset \Theta$ *which satisfies following conditions :*

$$\sup_{\epsilon > \epsilon_n} \log N(\epsilon/36, \{\theta \in \Theta_n : d_n(\theta, \theta_0) < \epsilon\}, d_n) \leq n{\epsilon_n}^2, \tag{1}$$

$$\frac{\Pi(\Theta - \Theta_n)}{\Pi(B_n^*(\theta_0, \epsilon_n; k))} = o(e^{-2n{\epsilon_n}^2}), \tag{2}$$

$$\frac{\Pi(\theta \in \Theta_n : j\epsilon_n < d_n(\theta, \theta_0) \leq 2j\epsilon_n)}{\Pi(B_n^*(\theta_0, \epsilon_n; k))} \leq e^{n{\epsilon_n}^2 j^2/4} \tag{3}$$

Then $P^{(n)}_{(\theta_0)}[\Pi(\theta : d_n(\theta, \theta_0) \geq M_n\epsilon_n|\mathbf{D}_n)] \to 0$ *for any sequence* $M_n \to \infty$*.*
Where

$$KL(f, g) = \mathbb{E}^f \left[\log \frac{f(X)}{g(X)}\right] = \int f \log \frac{f}{g} d\mu, \tag{4}$$

$$V_{k,0}(f, g) = \mathbb{E}^f \left[\left|\log \frac{f(X)}{g(X)} - KL(f, g)\right|^k\right], \tag{5}$$

$$B_n^*(\theta_0, \epsilon; k) = \left\{\theta \in \Theta : \frac{1}{n}\sum_{i=1}^{n} KL(P_{\theta_0,i}, P_{\theta,i}) \leq \epsilon^2, \frac{1}{n}\sum_{i=1}^{n} V_{k,0}(P_{\theta_0,i}, P_{\theta,i}) \leq C_k\epsilon^k\right\}. \tag{6}$$

Here, the C_k *is the constant satisfying*

$$\mathbb{E}\left[|\bar{X}_n - \mathbb{E}[\bar{X}_n]|^k\right] \leq C_k n^{-k/2} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[|X_i|^k\right] \tag{7}$$

for $k \geq 2$*.*

2 Brief explanation : 증명의 전반적인 개요 및 흐름

Lemma 2의 경우 S. Ghosal 과 A. Van der Vaart의 Convergence Rates of Posterior Distributions for Noniid Observations [2007]의 **Theorem 4**와 사실상 같기 때문에 이에 대한 증명도 위 논문에서 언급한 방식으로 따라가면 된다. 그런데 **Theorem 4**는 사실상 **Theorem 1**의 변형된 형태이며, **Theorem 1**도 사실상 S. Ghosal 과 A. Van der Vaart의 Convergence Rates of Posterior Distributions [2000]의 **Theorem 2.4**와 같다. 따라서 Ghosal[2007]의 **Theorem 1**을 증명한 뒤 이를 위한 몇개의 가정을 바꾼 **Theorem 4**를 증명하는 것으로 Asymptotic Properties for Bayesian Neural Network in Besov Space 의 **Lemma 2** 증명을 마무리 하도록 하겠다.

증명을 바로 시작하기에 앞서 증명에 대한 큰 틀을 설명해보려 한다. 증명은 크게 covering number의 정의와 사후확률의 분해로 나뉜다. 앞서 Ghosal[2007]의 **Theorem 1**의 증명을 먼저 한다고 하였으므로 이를 기준으로 설명하도록 하겠다. 먼저 사후확률의 분해를 설명하고자 한다. 또한 이후부터 편의상 $\mathbf{D}_n$을 $X^{(n)}$으로 표기하도록 하겠다. 우리가 보여야 하는 사실은 모든 $M_n \to \infty$에 대하여 $P_{\theta_0}^{(n)}[\Pi(\theta : d_n(\theta, \theta_0) \geq M_n\epsilon_n|X^{(n)})] \to 0$ 라는 것이다. 이때 편의상 $\Pi(\theta : d_n(\theta, \theta_0) \geq M_n\epsilon_n|X^{(n)})) \stackrel{let}{=} \Pi(\Theta_n|X^{(n)})$ 라고 하면 다음과 같은 분해를 생각해볼 수 있다 :

$$P_{\theta_0}^{(n)}\Pi(\Theta_n|X^{(n)}) = P_{\theta_0}^{(n)}\Pi(\Theta_n|X^{(n)})(1-\phi_n)I_{A_n} + P_{\theta_0}^{(n)}\Pi(\Theta_n|X^{(n)})\phi_n + P_{\theta_0}^{(n)}\Pi(\Theta_n|X^{(n)})(1-\phi_n)I_{A_n^c}$$

즉 $P_{\theta_0}^{(n)}\Pi(\Theta_n|X^{(n)})$ 을 세 개의 확률의 합으로 나타낼 수 있는데 각각의 확률이 적당히 큰 M_n 에 대하여 각각 0으로 수렴하게됨을 보이면 Ghosal[2007]의 **Theorem 1** 증명이 끝나게 된다.
Ghosal[2007]의 **Theorem 4**의 경우, 검정함수 ϕ_n의 존재성이 semimetric d_n을 Hellinger distance로 놓고 $P_\theta^{(n)}$를 product measure로 놓았을 때에도 만족됨을 Ghosal[2007]의 **Lemma 2**에서 확인 할 수 있는데, 이와 더불어 위에서 가정한 몇 개의 조건이 수정된 후에 Ghosal[2007]의 **Theorem 1**과 거의 같은 방식으로 **Theorem 4**의 증명을 확인 할 수 있다.

이제 증명의 대략적인 개요 및 흐름에 대한 설명이 끝났기에 증명을 위해 필요한 기본적인 전제조건들을 언급하고 다음 단계로 넘어 가고자 한다.

3 General Theorem - Section(2) of Ghosal[2007]

Assumption 1 *For each $n \in \mathbb{N}$ and $\theta \in \Theta$, let $P_\theta^{(n)}$ admit densities $p_\theta^{(n)}$ relative to a $\sigma-$finite measure $\mu^{(n)}$. Assume that $(x, \theta) \mapsto p_\theta^{(n)}(x)$ is jointly measurable relative to $\mathcal{A} \bigotimes \mathcal{B}$, where $\mathcal{B}$ is a σ - field on Θ. By Bayes's theorem, the posterior distribution is given by :*

$$\Pi_{(n)}(B|X^{(n)}) = \frac{\int_B p_\theta^{(n)}(X^{(n)}) d\Pi_{(n)}(\theta)}{\int_\Theta p_\theta^{(n)}(X^{(n)}) d\Pi_{(n)}(\theta)}, \qquad B \in \mathcal{B}. \tag{8}$$

Here, $X^{(n)}$ is an "observation", which, in our setup, will be understood to be generated according to $P_\theta^{(n)}$ for some given $\theta_0 \in \Theta$.

For each n, let d_n and e_n be semimetrics on Θ with the property that there exist universal constants $\xi > 0$ and $K > 0$ such that for every $\epsilon > 0$ and for each $\theta_1 \in \Theta$ with $d_n(\theta_1, \theta_0) > \epsilon$, there exists a test ϕ_n such that

$$P_{\theta_0}^{(n)} \phi_n \leq e^{-Kn\epsilon^2}, \qquad \sup_{\theta \in \Theta : e_n(\theta, \theta_1) < \epsilon\xi} P_{\theta_0}^{(n)}(1 - \phi_n) \leq e^{-Kn\epsilon^2}. \tag{9}$$

Typically, we have $d_n < e_n$ and in many cases we choose $d_n = e_n$, but using two semimetrics provides some added flexibility. Le Cam and Birge showed that the rate of convergence, in a minimax sense, of the best estimators of θ relative to the distance d_n can be understood in terms of the Le Cam dimension or local entropy function of the set Θ relative to d_n. For our purposes, this dimension is a function whose value at $\epsilon > 0$ is defined to be $\log N(\epsilon\xi, \{\theta : d_n(\theta, \theta_0) \leq \epsilon\}, e_n)$, that is, the logarithm of the minimum number of d_n - balls of radius $\epsilon\xi$ needed to cover an e_n-ball of radius ϵ around the true parameter θ_0. Birge and Le Cam showed that there exist estimators $\hat{\theta}_n = \hat{\theta}_n(X^{(n)})$ such that $d_n(\hat{\theta}_n, \theta_0) = O_P(\epsilon_n)$ under $P_{\theta_0}^{(n)}$, where

$$\sup_{\epsilon > \epsilon_n} \log N(\epsilon\xi, \{\theta : d_n(\theta, \theta_0) \leq \epsilon\}, e_n) \leq n\epsilon_n^2.$$

Further, under certain conditions ϵ_n is the best rate obtainalbe, given the model, and hence gives a minimax rate.

As in the i.i.d case, the behavior of posterior distributions depends on the size of the model measured by above inequality and the concentration rate of the prior Π_n at θ_0. For a given $k > 1$, let

$$B_n(\theta_0, \epsilon; k) = \{\theta \in \Theta : KL(p_{\theta_0}^{(n)}, p_\theta^{(n)}) \leq n\epsilon^2, V_{k,0}(p_{\theta_0}^{(n)}, p_\theta^{(n)}) \leq n^{k/2}\epsilon^k\}.$$

4 Convergence Rates of Posterior Distributions for Noniid Observations - Lemma 9

Lemma 4.1 *(Lemma 9 of Ghosal and Van der Vaart[2007])*
Let d_n and e_n be semimetrics on Θ for which tests satisfying the conditions of (9) exist. Suppose that for some nonincreasing function $\epsilon \mapsto N(\epsilon)$ and some $\epsilon_n \geq 0$,

$$N\left(\frac{\epsilon\xi}{2}, \{\theta \in \theta : d_n(\theta, \theta_0) < \epsilon\}, e_n\right) \leq N(\epsilon) \qquad for\ all \quad \epsilon > \epsilon_n. \tag{10}$$

Then for every $\epsilon > \epsilon_n$, there exist tests ϕ_n, $n \geq 1$, (depending on ϵ) such that $P_{\theta_0}^{(n)}\phi_n \leq N(\epsilon)\frac{e^{-Kn\epsilon^2}}{1-e^{-Kn\epsilon^2}}$ and $P_{\theta_0}^{(n)}(1-\phi_n) \leq e^{-Kn\epsilon^2 j^2}$ for all $\theta \in \Theta$ such that $d_n(\theta, \theta_0) > j\epsilon$ and for every $j \in \mathbb{N}$.

Proof. For a given $j \in \mathbb{N}$, choose a maximal set of points in $\Theta_j = \{\theta \in \Theta : j\epsilon < d_n(\theta, \theta_0) \leq (j+1)\epsilon\}$ with the property that $e_n(\theta, \theta^{'}) > j\epsilon\xi$ for every pair of points in the set. Because this set of points is a $j\epsilon\xi$-net over Θ_j for $e_n$and because $(j+1)\epsilon \leq 2j\epsilon$, this yields a set $\Theta_j^{'}$ of at most $N(2j\epsilon)$points, each at d_n-distance at least $j\epsilon$ from θ_0, and every $\theta \in \Theta_j$ is within e_n distance $j\epsilon\xi$ of at least one of these points. (If $\Theta_j$is empty, we take $\Theta_j^{'}$ to be empty also.) By assumption, for every point $\theta_1 \in \Theta_j^{'}$, there exists a test with the properties as in (9), but with ϵ replaced by $j\epsilon$.
Let ϕ_n be the maximum of all tests attached in this way to some point $\theta_1 \in \Theta_j^{'}$ for some $j \in \mathbb{N}$. Then

$$P_{\theta_0}^{(n)}\phi_n \leq e^{-Kn(j\epsilon)^2} \quad for\ \ all\ \ j \in \mathbb{N} \tag{11}$$

$$\leq \sum_{j=1}^{\infty} \sum_{\theta_1 \in \Theta_j^{'}} e^{-Knj^2\epsilon^2} \tag{12}$$

$$\leq \sum_{j=1}^{\infty} N(\epsilon)e^{-Knj^2\epsilon^2} \quad (\because N(2j\epsilon) \leq N(\epsilon).) \tag{13}$$

$$\leq \sum_{j=1}^{\infty} N(\epsilon)e^{-Knj\epsilon^2} \quad (\because f(x) = e^{-x} \ \ is\ \ decreasing\ \ function.) \tag{14}$$

$$= N(\epsilon)\tfrac{e^{-Kn\epsilon^2}}{1-e^{-Kn\epsilon^2}}. \tag{15}$$

위의 (13)은 (10)으로부터 다음과 같이 얻을 수 있다 :

$$N(j\epsilon\xi, \{\theta \in \Theta : j\epsilon < d_n(\theta, \theta_0) \leq (j+1)\epsilon)\}, e_n) \leq N(j\epsilon\xi, \{\theta \in \Theta : j\epsilon < d_n(\theta, \theta_0) \leq 2j\epsilon)\}, e_n) \leq N(2j\epsilon).$$

이때 covering number의 정의로부터 모든 $j \in \mathbb{N}$에 대해 $N(2j\epsilon)$이 $N(\epsilon)$ 보다 작거나 같게 되므로 위와 같은 결과를 보일 수 있다.

또한 모든 $j \in \mathbb{N}$ 에 대하여,

$$\sup_{\theta \in \Theta: e_n(\theta, \theta_1) < \epsilon\xi} P_{\theta_0}^{(n)}(1 - \phi_n) \leq \sup_{\theta \in \cup_{i>j} \Theta_i} e^{-Kn(i\epsilon)^2} \tag{16}$$

$$\leq \sup_{i>j} e^{-Kni^2\epsilon^2} \tag{17}$$

$$\leq e^{-Knj^2\epsilon^2} \tag{18}$$

where we have used the fact that for every $\theta \in \Theta_i$, there exists a test ϕ with $\phi_n \geq \phi$ and $P_{\theta_0}^{(n)}(1 - \phi_n) \leq e^{-Kni^2\epsilon^2}$.

5 Convergence Rates of Posterior Distributions for Noniid Observations - Lemma 10

Lemma 5.1 *(Lemma 10 of Ghosal and Van der Vaart[2007])*
For $k \geq 2$, every $\epsilon > 0$ and every probability measure $\bar{\Pi}_n$ supported on the set $B_n(\theta_0, \epsilon; k)$, we have, for every $C > 0$,

$$P_{\theta_0}^{(n)} \left(\int \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\bar{\Pi}_n(\theta) \leq e^{-(1+C)n\epsilon^2} \right) \leq \frac{1}{C^k (n\epsilon^2)^{k/2}}. \tag{19}$$

Proof.

1. 우선 젠센 부등식(Jensen Inequality)에 의해, 임의의 확률변수 X 에 대하여 $\log \mathbb{E}(X) \geq \mathbb{E}(\log(X))$가 성립됨을 알 수 있다. 이때 $\ell_{n,\theta} = \log \left(\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} \right)$ 라고 하면 젠센 부등식에 의해 다음이 성립한다 :

$$\log \int \left(\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} \right) d\bar{\Pi}_n(\theta) \geq \int \ell_{n,\theta} d\bar{\Pi}_n(\theta) = \int \log \left(\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} \right) d\bar{\Pi}_n(\theta). \tag{20}$$

2. 이제 (19) 에 있는 확률의 상한을 구하기 위해 확률안에 있는 부등식을 조작해보고자 한다. 이때 다음과 같은 형태를 고려할 수 있다 :

$$\int \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\bar{\Pi}_n(\theta) \leq e^{-(1+C)n\epsilon^2} \overset{\log}{\Rightarrow} \log \left(\int \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\bar{\Pi}_n(\theta) \right) \leq -(1+C)n\epsilon^2 \tag{21}$$

따라서 (99)에 의해 다음의 부등식 또한 성립함을 알 수 있다 :

$$\int \ell_{n,\theta} d\bar{\Pi}_n(\theta) \leq -(1+C)n\epsilon^2. \tag{22}$$

이 때 위 식의 양변에서 $KL(P_\theta^{(n)}, P_{\theta_0}^{(n)})$를 빼주게 되면 다음과 같다 :

$$\int \ell_{n,\theta} d\bar{\Pi}_n(\theta) - \int P_{\theta_0}^{(n)} \ell_{n,\theta} d\bar{\Pi}_n(\theta) \leq -(1+C)n\epsilon^2 - \int P_{\theta_0}^{(n)} \ell_{n,\theta} d\bar{\Pi}_n(\theta) \tag{23}$$

따라서 위의 관계식으로부터 다음과 같은 관계를 얻을 수 있게 된다 :

$$P_{\theta_0}^{(n)} \left(\int \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\bar{\Pi}_n(\theta) \leq e^{-(1+C)n\epsilon^2} \right) \leq P_{\theta_0}^{(n)} \left(\int \left(\ell_{n,\theta} - P_{\theta_0}^{(n)} \ell_{n,\theta} \right) d\bar{\Pi}_n(\theta) \leq -(1+C)n\epsilon^2 - \int P_{\theta_0}^{(n)} \ell_{n,\theta} d\bar{\Pi}_n(\theta) \right) \tag{24}$$

3. 확률측도 $\bar{\Pi}_n$의 토대가 $B_n(\theta_0, \epsilon; k)$ 에 의해 형성됨이 가정되어 있으므로 이를 이용하여 $-(1+C)n\epsilon^2 - \int P_{\theta_0}^{(n)} \ell_{n,\theta} d\bar{\Pi}_n(\theta)$ 의 상한을 구해보고자 한다. 이때 모든 $\theta \in B_n(\theta_0, \epsilon; k)$에 대하여 다음이 성립함을 $B_n(\theta_0, \epsilon; k)$의 정의로부터 알 수 있다 :

$$P_{\theta_0}^{(n)} \ell_{n,\theta} = P_{\theta_0}^{(n)} \log \left(\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} \right) = -KL(P_{\theta_0}^{(n)}, P_\theta^{(n)}) \geq -n\epsilon^2. \tag{25}$$

그러므로 다음과 같은 생각을 할 수 있게 된다 :

$$P_{\theta_0}^{(n)} \ell_{n,\theta} \geq -n\epsilon^2 \Rightarrow -\int P_{\theta_0}^{(n)} \ell_{n,\theta} d\bar{\Pi}_n(\theta) \leq \int n\epsilon^2 d\bar{\Pi}_n(\theta) = n\epsilon^2 \tag{26}$$

따라서 $-(1+C)n\epsilon^2 - \int P_{\theta_0}^{(n)} \ell_{n,\theta} d\bar{\Pi}_n(\theta)$ 의 상한을 구해보면 다음과 같음을 알 수 있다 :

$$-(1+C)n\epsilon^2 - \int P_{\theta_0}^{(n)} \ell_{n,\theta} d\bar{\Pi}_n(\theta) \leq -n(1+C)\epsilon^2 + n\epsilon^2 = -Cn\epsilon^2. \tag{27}$$

따라서 편의상 $\int \left(\ell_{n,\theta} - P_{\theta_0}^{(n)} \ell_{n,\theta} \right) d\bar{\Pi}_n(\theta) \overset{let}{=} S(\ell_{n,\theta}, P_{\theta_0}^{(n)})$ 라고 했을 때, 다음과 같은 결과를 얻을 수 있다 :

$$P_{\theta_0}^{(n)} \left(S(\ell_{n,\theta}, P_{\theta_0}^{(n)}) \leq -(1+C)n\epsilon^2 - \int P_{\theta_0}^{(n)} \ell_{n,\theta} d\bar{\Pi}_n(\theta) \right) \leq P_{\theta_0}^{(n)} \left(S(\ell_{n,\theta}, P_{\theta_0}^{(n)}) \leq -Cn\epsilon^2 \right). \tag{28}$$

4. 위의 결과를 이용하기에 앞서 마코프 부등식(Markov Inequality)에 의해, 임의의 확률변수 X 에 대하여, $\forall a > 0$에 대해 $P(|X| \geq a) \leq \frac{\mathbb{E}(|X|^n)}{a^n}$가 성립됨을 알 수 있다.
이때 위에서 얻은 $P_{\theta_0}^{(n)} \left(S(\ell_{n,\theta}, P_{\theta_0}^{(n)}) \leq -Cn\epsilon^2 \right)$ 의 상한을 마코프 부등식과 젠센 부등식을 이용하여 다음과 같이 구할 수 있다 :

$$P_{\theta_0}^{(n)} \left(\int \left(\ell_{n,\theta} - P_{\theta_0}^{(n)} \ell_{n,\theta} \right) d\bar{\Pi}_n(\theta) \leq -Cn\epsilon^2 \right) \leq P_{\theta_0}^{(n)} \left(\left| \int \left(\ell_{n,\theta} - P_{\theta_0}^{(n)} \ell_{n,\theta} \right) d\bar{\Pi}_n(\theta) \right| \leq Cn\epsilon^2 \right) \tag{29}$$

$$\leq \frac{\int \left| \int \left(\ell_{n,\theta} - P_{\theta_0}^{(n)} \ell_{n,\theta} \right) d\bar{\Pi}_n(\theta) \right|^k dP_{\theta_0}^{(n)}}{(Cn\epsilon^2)^k} \tag{30}$$

$$\leq \frac{\int \int \left| \left(\ell_{n,\theta} - P_{\theta_0}^{(n)} \ell_{n,\theta} \right) \right|^k d\bar{\Pi}_n(\theta) dP_{\theta_0}^{(n)}}{(Cn\epsilon^2)^k} \tag{31}$$

$$\leq \frac{n^{k/2}\epsilon^k}{C^k n^k \epsilon^{2k}} \tag{32}$$

$$= \frac{1}{C^k (n\epsilon^2)^{k/2}}. \tag{33}$$

이때 (32)는 $B_n(\theta_0, \epsilon; k)$의 정의에 의해 다음과 같은 이유로 성립함을 알 수 있다 :

$$
\begin{aligned}
\int\int \left|\left(\ell_{n,\theta} - P_{\theta_0}^{(n)}\ell_{n,\theta}\right)\right|^k d\bar{\Pi}_n(\theta) dP_{\theta_0}^{(n)} &= \int\int \left|\left(\ell_{n,\theta} - P_{\theta_0}^{(n)}\ell_{n,\theta}\right)\right|^k dP_{\theta_0}^{(n)} d\bar{\Pi}_n(\theta) && (34)\\
&= \int \mathbb{E}_{P_{\theta_0}^{(n)}} \left|\left(\ell_{n,\theta} - P_{\theta_0}^{(n)}\ell_{n,\theta}\right)\right|^k d\bar{\Pi}_n(\theta) && (35)\\
&\le \int \mathbb{E}_{P_{\theta_0}^{(n)}} \left|\left(\ell_{n,\theta} - P_{\theta_0}^{(n)}\ell_{n,\theta}\right)\right|^k |(-1)|^k \, d\bar{\Pi}_n(\theta) && (36)\\
&= \int V_{k,0}(P_{\theta_0}^{(n)}, P_{\theta}^{(n)}) d\bar{\Pi}_n(\theta) && (37)\\
&\le V_{k,0}(P_{\theta_0}^{(n)}, P_{\theta}^{(n)}) && (38)\\
&\le n^{k/2}\epsilon^k. && (39)
\end{aligned}
$$

6 Convergence Rates of Posterior Distributions for Noniid Observations - Theorem 1

Theorem 6.1 *(Theorem 1 of Ghosal and Van der Vaart[2007])*
Letd_n and e_n be semimetrics on Θ for which tests satisfying (9) exist. Let $\epsilon_n > 0, \epsilon_n \to 0, (n\epsilon_n^2)^{-1} = O(1), k > 1,$ and $\Theta_n \subset \Theta$ be such that for every sufficiently large $j \in \mathbb{N}$,

$$\sup_{\epsilon > \epsilon_n} \log N\left(\frac{\epsilon\xi}{2}, \{\theta \in \Theta_n : d_n(\theta, \theta_0) < \epsilon\}, e_n\right) \leq n\epsilon_n^2 \tag{40}$$

$$\frac{\Pi_n(\theta \in \Theta_n : j\epsilon_n < d_n(\theta, \theta_0) \leq 2j\epsilon_n)}{\Pi_n(B_n(\theta_0, \epsilon_n; k))} \leq e^{Kn\epsilon_n^2 j^2/2}. \tag{41}$$

Then for every $M_n \to \infty$, we have that

$$P_{\theta_0}^{(n)} \Pi_n(\theta \in \Theta_n : d_n(\theta, \theta_0) \geq M_n\epsilon_n | X^{(n)}) \to 0. \tag{42}$$

Proof.

1. (Lemma 9)에 의해, $N(\epsilon) = e^{n\epsilon_n^2}, \epsilon = M\epsilon_n, M \geq 2$ 라고 하면 다음의 조건을 만족하는 검정함수 ϕ_n의 존재성을 보일 수 있다 :

$$\text{There exist test} \quad \phi_n \quad \text{that satisfies} \quad \begin{cases} P_{\theta_0}^{(n)}\phi_n \leq e^{n\epsilon_n^2} \frac{e^{-KnM^2\epsilon_n^2}}{1-e^{-KnM^2\epsilon_n^2}} \\ P_{\theta_0}^{(n)}(1-\phi_n) \leq e^{-KnM^2\epsilon_n^2 j^2} \end{cases} \tag{43}$$

   for all $\theta \in \Theta$ such that $d_n(\theta, \theta_0) > M\epsilon_n j$ and for every $j \in \mathbb{N}$.

2. $M \geq 2$인 조건은 추후에 선택될 M이 다음과 같은 부등식을 만족하기에 충분히 크다는 것을 보장해주기 위해서 이다 :

$$KM^2 - 1 > KM^2/2 \Rightarrow KM^2 n\epsilon_n^2 - n\epsilon_n^2 > KM^2 n\epsilon_n^2/2 \tag{44}$$

$$\Rightarrow -KM^2 n\epsilon_n^2 + n\epsilon_n^2 < -KM^2 n\epsilon_n^2/2 \tag{45}$$

$$\Rightarrow e^{n\epsilon_n^2} e^{-KM^2 n\epsilon_n^2} < e^{-KM^2 n\epsilon_n^2/2}. \tag{46}$$

   이때 사후확률 $\Pi_n(d_n(\theta, \theta_0) \geq JM\epsilon_n | X^{(n)}) \leq 1$이라는 사실과 검정함수 ϕ_n이 0과 1사이임을 이용하면 다음과 같은 관계식을 생각할 수 있다 :

$$\Pi_n(d_n(\theta, \theta_0) \geq JM\epsilon_n | X^{(n)}) \leq 1 \Rightarrow \Pi_n(d_n(\theta, \theta_0) \geq JM\epsilon_n | X^{(n)})\phi_n \leq \phi_n \tag{47}$$

$$\Rightarrow P_{\theta_0}^{(n)}[\Pi_n(d_n(\theta, \theta_0) \geq JM\epsilon_n | X^{(n)})\phi_n] \leq P_{\theta_0}^{(n)}\phi_n \tag{48}$$

이때 (43)을 이용하면 다음과 같은 결과를 얻을 수 있게 된다 :

$$P_{\theta_0}^{(n)}[\Pi_n(d_n(\theta,\theta_0)\geq JM\epsilon_n|X^{(n)})\phi_n]\leq P_{\theta_0}^{(n)}\phi_n \tag{49}$$

$$\leq e^{n\epsilon_n^2}\tfrac{e^{-KnM^2\epsilon_n^2}}{1-e^{-KnM^2\epsilon_n^2}} \tag{50}$$

$$\leq \tfrac{e^{-KnM^2\epsilon_n^2/2}}{1-e^{-KnM^2\epsilon_n^2}} \tag{51}$$

$$\leq e^{-KnM^2\epsilon_n^2/2}+e^{-KnM^2\epsilon_n^2/2} \tag{52}$$

$$\leq 2e^{-KnM^2\epsilon_n^2/2}. \tag{53}$$

3. $\Theta_{n,j}=\{\theta\in\Theta_n : M\epsilon_n j<d_n(\theta,\theta_0)\leq M\epsilon_n(j+1)\}$ 라고 하게 되면 다음을 보일 수 있다 :

$$P_{\theta_0}^{(n)}\left[\int_{\Theta_{n,j}}(1-\phi_n)\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}}d\Pi_n(\theta)\right]=\int\int_{\Theta_{n,j}}(1-\phi_n)\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}}d\Pi_n(\theta)dP_{\theta_0}^{(n)} \tag{54}$$

$$=\int_{\Theta_{n,j}}\int(1-\phi_n)\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}}dP_{\theta_0}^{(n)}d\Pi_n(\theta)\quad (\because \text{Fubini's theorem}) \tag{55}$$

$$=\int_{\Theta_{n,j}}\int(1-\phi_n)P_\theta^{(n)}d\mu^{(n)}d\Pi_n(\theta)\quad (\because p_{\theta_0}^{(n)}=\frac{dP_{\theta_0}^{(n)}}{d\mu^{(n)}}) \tag{56}$$

$$=\int_{\Theta_{n,j}}P_\theta^{(n)}(1-\phi_n)d\Pi_n(\theta) \tag{57}$$

$$\leq\int_{\Theta_{n,j}}e^{-KnM^2\epsilon_n^2 j^2}d\Pi_n(\theta)\quad (\because (43)) \tag{58}$$

$$=e^{-KnM^2\epsilon_n^2 j^2}\Pi_n(\Theta_{n,j}) \tag{59}$$

4. 이제 어떠한 상수 $C>0$을 고정하도록 한다. 그러면 (Lemma 10)에 의해 다음과 같은 사건A_n은 적어도 $1-C^{-k}(n\epsilon_n^2)^{-k/2}$의 확률로 일어나게 됨을 알 수 있다 :

$$\int\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}}d\Pi_n(\theta)\ \geq\ \int_{B_n(\theta_0,\epsilon_n;k)}\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}}d\Pi_n(\theta)\ \geq\ e^{-(1+C)n\epsilon_n^2}\Pi_n(B_n(\theta_0,\epsilon_n;k)). \tag{60}$$

$$\because\int\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}}d\Pi_n(\theta)\leq e^{-(1+C)n\epsilon_n^2}\ \text{ satisfies w.p. at most }\ C^{-k}(n\epsilon_n^2)^{-k/2} \tag{61}$$

$$\Rightarrow\int\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}}d\Pi_n(\theta)\geq e^{-(1+C)n\epsilon_n^2}\ \text{ satisfies w.p. at least }\ 1-C^{-k}(n\epsilon_n^2)^{-k/2} \tag{62}$$

$$\Rightarrow\int_{B_n(\theta_0,\epsilon_n;k)}\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}}d\Pi_n(\theta)\geq\int_{B_n(\theta_0,\epsilon_n;k)}e^{-(1+C)n\epsilon_n^2}d\Pi_n(\theta)=e^{-(1+C)n\epsilon_n^2}\Pi_n(B_n(\theta_0,\epsilon_n;k)). \tag{63}$$

5. 이때 다음과 같은 집합 $\Theta_{M,J} = \{\theta \in \Theta : d_n(\theta, \theta_0) > JM\epsilon_n\}$ 을 다음과 같이 $\Theta_{n,j}$의 합으로 분해할 수 있다 :

$$\{\theta \in \Theta : d_n(\theta, \theta_0) > JM\epsilon_n\} = \bigcup_{j \geq J} \Theta_{n,j} = \bigcup_{j \geq J} \{\theta \in \Theta_n : M\epsilon_n j < d_n(\theta, \theta_0) \leq M\epsilon_n(j+1)\}. \quad (64)$$

또한 (41) 에서 j 대신에 Mj를 대입하고 난 뒤, 편의상 $\Theta_{2Mj} \overset{let}{=} \{\theta \in \Theta_n : Mj\epsilon_n < d_n(\theta, \theta_0) \leq 2Mj\epsilon_n\}$ 라고 하면 다음을 보일 수 있다 :

$$\frac{\Pi_n(\theta \in \Theta_n : Mj\epsilon_n < d_n(\theta, \theta_0) \leq 2Mj\epsilon_n)}{\Pi_n(B_n(\theta_0, \epsilon_n; k))} \leq e^{KnM^2\epsilon_n^2 j^2/2} \quad (65)$$

이제 $P_{\theta_0}^{(n)}[\Pi(\theta \in \Theta_n : d_n(\theta, \theta_0) > M\epsilon_n J | X^{(n)})(1 - \phi_n) I_{A_n}]$ 의 상한을 구해보도록 한다 :

$$P_{\theta_0}^{(n)}[\Pi(\Theta_{M,J} | X^{(n)})(1 - \phi_n) I_{A_n}] = \sum_{j \geq J} P_{\theta_0}^{(n)}[\Pi(\theta \in \Theta_n : M\epsilon_n j < d_n(\theta, \theta_0) \leq M\epsilon_n(j+1) | X^{(n)})(1 - \phi_n) I_{A_n}] \quad (66)$$

$$\leq \sum_{j \geq J} P_{\theta_0}^{(n)}[\Pi(\theta \in \Theta_n : M\epsilon_n j < d_n(\theta, \theta_0) \leq 2M\epsilon_n j | X^{(n)})(1 - \phi_n) I_{A_n}] \quad (67)$$

$$= \sum_{j \geq J} P_{\theta_0}^{(n)} \left[\frac{\int_{\Theta_{2Mj}} p_\theta^{(n)} d\Pi_n(\theta)}{\int_\Theta p_\theta^{(n)} d\Pi_n(\theta)} (1 - \phi_n) I_{A_n} \right] \quad (68)$$

$$= \sum_{j \geq J} P_{\theta_0}^{(n)} \left[\frac{\int_{\Theta_{2Mj}} \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\Pi_n(\theta)}{\int_\Theta \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\Pi_n(\theta)} (1 - \phi_n) I_{A_n} \right] \quad (69)$$

$$\leq \sum_{j \geq J} P_{\theta_0}^{(n)} \left[\frac{\int_{\Theta_{2Mj}} \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\Pi_n(\theta)}{\Pi_n(B_n(\theta_0, \epsilon_n; k))} e^{(1+C)n\epsilon_n^2} (1 - \phi_n) \right] \quad (\because (60)) \quad (70)$$

$$\leq \sum_{j \geq J} e^{-KnM^2\epsilon_n^2 j^2} \frac{\Pi_n(\Theta_{2Mj})}{\Pi_n(B_n(\theta_0, \epsilon_n; k))} e^{(1+C)n\epsilon_n^2} \quad (\because (59)) \quad (71)$$

$$\leq \sum_{j \geq J} e^{-KnM^2\epsilon_n^2 j^2} e^{KnM^2\epsilon_n^2 j^2/2} e^{(1+C)n\epsilon_n^2} \quad (72)$$

(73)

6. 지금까지는 우리는 충분히 큰 J에 대하여 다음이 성립함을 보였다 :

$$P_{\theta_0}^{(n)}[\Pi(\theta \in \Theta : d_n(\theta, \theta_0) > JM\epsilon_n \; |X^{(n)})(1-\phi_n)I_{A_n}] \leq \sum_{j \geq J} e^{-n\epsilon_n^2(KM^2j^2-1-C-\frac{1}{2}KM^2j^2)}.$$

이제 그동안 구한 결과들을 종합해보도록 한다 :

(i) $P_{\theta_0}^{(n)}[\Pi(\theta \in \Theta : d_n(\theta, \theta_0) > JM\epsilon_n \; |X^{(n)})(1-\phi_n)I_{A_n}] \leq \sum_{j \geq J} e^{-n\epsilon_n^2(\frac{1}{2}KM^2j^2-1-C)}$

(ii) $P_{\theta_0}^{(n)}[\Pi(\theta \in \Theta : d_n(\theta, \theta_0) > JM\epsilon_n \; |X^{(n)})\phi_n] \leq 2e^{-KM^2n\epsilon_n^2/2}$

(iii) $P_{\theta_0}^{(n)}[\Pi(\theta \in \Theta : d_n(\theta, \theta_0) > JM\epsilon_n \; |X^{(n)})(1-\phi_n)I_{A_n^c}] \leq P_{\theta_0}^{(n)}[I_{A_n^c}] \leq \frac{1}{C^k(n\epsilon_n^2)^{k/2}}$ ($\because$ Lemma 10)

따라서, 충분히 큰 M, J에 대하여 $n\epsilon_n^2 \to \infty$ 일때

$$P_{\theta_0}^{(n)}[\Pi(\theta \in \Theta : d_n(\theta, \theta_0) > JM\epsilon_n \; |X^{(n)})] \leq \sum_{j \geq J} e^{-n\epsilon_n^2(\frac{1}{2}KM^2j^2-1-C)} + 2e^{-KM^2n\epsilon_n^2/2} + \frac{1}{C^k(n\epsilon_n^2)^{k/2}} \to 0.$$

$$\therefore P_{\theta_0}^{(n)}[\Pi(\theta \in \Theta : d_n(\theta, \theta_0) \geq M_n\epsilon_n | X^{(n)})] \to 0 \quad as \quad M_n \to 0.$$

7 Convergence Rates of Posterior Distributions for Noniid Observations - Lemma 1

Lemma 7.1 *(Lemma 1 of Ghosal and Van der Vaart[2007])*
If $\frac{\Pi_n(\Theta\setminus\Theta_n)}{\Pi_n(B_n(\theta_0,\epsilon_n;k))} = o(e^{-2n\epsilon_n^2})$ *for some* $k > 1$*, then* $P_{\theta_0}^{(n)}\Pi_n(\Theta \setminus \Theta_n|X^{(n)}) \to 0$.

Proof.

1. $P_{\theta_0}^{(n)}\left(\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}}\right) = \int \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} dP_{\theta_0}^{(n)} = \int p_\theta^{(n)} d\mu^{(n)} = \int dP_\theta^{(n)} \leq 1$

2.

$$P_{\theta_0}^{(n)}\left[\int_{\Theta\setminus\Theta_n} \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\Pi_{(n)}(\theta)\right] = \int\int_{\Theta\setminus\Theta_n} \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\Pi_{(n)}(\theta) dP_{\theta_0}^{(n)} \tag{74}$$

$$= \int_{\Theta\setminus\Theta_n}\int \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} dP_{\theta_0}^{(n)} d\Pi_{(n)}(\theta) \quad (\because \text{Fubini's theorem}) \tag{75}$$

$$= \int_{\Theta\setminus\Theta_n}\int p_\theta^{(n)} d\mu^{(n)} d\Pi_{(n)}(\theta) \tag{76}$$

$$= \int_{\Theta\setminus\Theta_n}\int dP_\theta^{(n)} d\Pi_n(\theta) \tag{77}$$

$$= \int_{\Theta\setminus\Theta_n} P_\theta^{(n)} d\Pi_n(\theta) \tag{78}$$

$$\leq \int_{\Theta\setminus\Theta_n} d\Pi_n(\theta) \quad (\because [1.]) \tag{79}$$

$$= \Pi_n(\Theta \setminus \Theta_n). \tag{80}$$

3. $\Pi_n(\Theta \setminus \Theta_n|X^{(n)}) = \dfrac{\int_{\Theta\setminus\Theta_n} \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\Pi_{(n)}(\theta)}{\int_{\Theta} \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\Pi_{(n)}(\theta)}$ by Bayes Theorem.

4. $\int_\Theta \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\Pi_{(n)}(\theta) \geq e^{-(1+C)n\epsilon_n^2}\Pi_n(B_n(\theta_0,\epsilon_n;k))$ on event A_n by (60)

5. combining 2.,3.,4.

   From 2. $\Rightarrow \Pi_n(\Theta \setminus \Theta_n) \geq P_{\theta_0}^{(n)}\left[\int_{\Theta\setminus\Theta_n} \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\Pi_{(n)}(\theta)\right]$

From 4. 양변에 역수를 취하도록 한다 $\Rightarrow \dfrac{e^{(1+C)n\epsilon_n^2}}{\Pi_n(B_n(\theta_0,\epsilon_n;k))} \geq \dfrac{1}{\displaystyle\int_\Theta \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\Pi_{(n)}(\theta)}$

$$\Rightarrow P_{\theta_0}^{(n)}[\Pi_n(\Theta \setminus \Theta_n | X^{(n)}) I_{A_n}] = P_{\theta_0}^{(n)} \left[\frac{\displaystyle\int_{\Theta\setminus\Theta_n} \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\Pi_{(n)}(\theta)}{\displaystyle\int_\Theta \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\Pi_{(n)}(\theta)} I_{A_n} \right] \tag{81}$$

$$\leq \frac{e^{(1+C)n\epsilon_n^2}\Pi_n(\Theta\setminus\Theta_n)}{\Pi_n(B_n(\theta_0,\epsilon_n;k))} \tag{82}$$

$$= e^{(1+C)n\epsilon_n^2} o(e^{-2n\epsilon_n^2}) \tag{83}$$

$$\leq e^{(1+C)n\epsilon_n^2 - 2n\epsilon_n^2} o(1) \tag{84}$$

$$= o(1) e^{-n\epsilon_n^2(1-C)} \tag{85}$$

6. $P_{\theta_0}^{(n)}[\Pi_n(\Theta\setminus\Theta_n|X^{(n)}) I_{A_n^c}] \leq P_{\theta_0}^{(n)}[I_{A_n^c}] \leq \frac{1}{C^k(n\epsilon_n^2)^{k/2}}$ $(\because \textit{Lemma} \ \ 10)$

7. combining 5. , 6. $\Rightarrow P_{\theta_0}^{(n)}[\Pi_n(\Theta\setminus\Theta_n|X^{(n)})] \leq o(1)e^{-n\epsilon_n^2(1-C)} + \frac{1}{C^k(n\epsilon_n^2)^{k/2}}$

8. (i) $n\epsilon_n^2$ is bounded $\Rightarrow \frac{1}{C^k(n\epsilon_n^2)^{k/2}}$ 값이 0에 가깝게 해주는 충분히 큰 $C > 0$ 값을 찾아서 고정시킨다.

   (ii) $n\epsilon_n^2 \to \infty$ 인 경우 $C = 1$로 고정시킨다

$$\therefore P_{\theta_0}^{(n)}\Pi_n(\Theta\setminus\Theta_n|X^{(n)}) \to 0.$$

8 Convergence Rates of Posterior Distributions for Noniid Observations - Theorem 4

이제 앞서 Brief explanation 부분에서 설명한 것처럼, 검정함수 ϕ_n의 존재성이 semimetric d_n을 Hellinger distance로 놓고 $P_\theta^{(n)}$를 product measure로 놓았을 때에도 만족됨을 이용하여 (Ghosal[2007]의 **Lemma 2**), 이와 더불어 위에서 가정한 몇 개의 조건이 수정된 후에 Ghosal[2007]의 **Theorem 1**과 거의 같은 방식으로 **Theorem 4**의 증명을 해보도록 하겠다. 이를 위해 필요한 정의와 조건들을 언급하고 다음 단계로 넘어가고자 한다.

Definition 8.1 (Hellinger Distance) *Take measures $P_\theta^{(n)}$ as $P_\theta^{(n)} \equiv \bigotimes_{i=1}^{n} P_{\theta,i}$ on a product measurable space $\bigotimes_{i=1}^{n}(\mathfrak{X}_i, \mathcal{A}_i)$. Assume that distribution $P_{\theta,i}$ of the i-th component X_i possesses a density $p_{\theta,i}$ relative to a σ-finite measure μ_i on $(\mathfrak{X}_i, \mathcal{A}_i), i = 1, \dots, n$.*

$$d_n^2(\theta, \theta^{'}) = \frac{1}{n}\sum_{i=1}^{n}\int \left(\sqrt{p_{\theta,i}} - \sqrt{p_{\theta^{'},i}}\right)^2 d\mu_i \tag{86}$$

Lemma 8.2 *(Lemma 2 of Ghosal and Van der Vaart[2007])*
If $P_{\theta_0}^{(n)}$ are product measures and d_n is defined by (86), then there exist tests ϕ_n such that

$$\begin{cases} P_{\theta_0}^{(n)}\phi_n \le e^{-\frac{1}{2}nd_n^2(\theta_0,\theta_1)} \\ P_\theta^{(n)}(1-\phi_n) \le e^{-\frac{1}{2}nd_n^2(\theta_0,\theta_1)} \end{cases} \quad \text{for all } \theta \in \Theta \text{ such that } d_n(\theta,\theta_1) \le \tfrac{1}{18}d_n(\theta_0,\theta_1) \tag{87}$$

Theorem 8.3 *(Theorem 4 of Ghosal and Van der Vaart[2007])*
Let $P_{(\theta)}^{(n)}$ be product measures and $d_n(\theta_0, \theta) = \frac{1}{n}\sum_{i=1}^{n}\int(\sqrt{p_{\theta_0,i}} - \sqrt{p_{\theta,i}})^2 d\mu_i$. Suppose that for a sequence $\epsilon_n \to 0$ such that $n{\epsilon_n}^2$ is bounded away from zero, some $k > 1$, all sufficiently large j and sets $\Theta_n \subset \Theta$ which satisfies following conditions :

$$\sup_{\epsilon > \epsilon_n} \log N(\epsilon/36, \{\theta \in \Theta_n : d_n(\theta, \theta_0) < \epsilon\}, d_n) \le n{\epsilon_n}^2, \tag{88}$$

$$\frac{\Pi(\Theta - \Theta_n)}{\Pi(B_n^*(\theta_0, \epsilon_n; k))} = o(e^{-2n{\epsilon_n}^2}), \tag{89}$$

$$\frac{\Pi(\theta \in \Theta_n : j\epsilon_n < d_n(\theta, \theta_0) \le 2j\epsilon_n)}{\Pi(B_n^*(\theta_0, \epsilon_n; k))} \le e^{n{\epsilon_n}^2 j^2/4} \tag{90}$$

Then $P_{(\theta_0)}^{(n)}[\Pi(\theta : d_n(\theta, \theta_0) \ge M_n\epsilon_n | \mathbf{D}_n)] \to 0$ for any sequence $M_n \to \infty$.
Where

$$KL(f, g) = \mathbb{E}^f\left[\log\frac{f(X)}{g(X)}\right] = \int f \log\frac{f}{g} d\mu, \tag{91}$$

$$V_{k,0}(f,g) = \mathbb{E}^f\left[\left|\log\frac{f(X)}{g(X)} - KL(f,g)\right|^k\right], \tag{92}$$

$$B_n^*(\theta_0,\epsilon;k) = \left\{\theta\in\Theta : \frac{1}{n}\sum_{i=1}^n KL(P_{\theta_0,i},P_{\theta,i}) \le \epsilon^2, \frac{1}{n}\sum_{i=1}^n V_{k,0}(P_{\theta_0,i},P_{\theta,i}) \le C_k\epsilon^k\right\}. \tag{93}$$

Here, the C_k is the constant satisfying

$$\mathbb{E}\left[|\bar{X}_n - \mathbb{E}[\bar{X}_n]|^k\right] \le C_k n^{-k/2}\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[|X_i|^k\right] \tag{94}$$

for $k\ge 2$.

이제 **Theorem 4** 증명을 위한 2가지 레마를 설명하고자 한다.

Lemma 8.4 *(Transformed Lemma 9 of Ghosal and Van der Vaart[2007])*
Let $d_n(\theta_0,\theta) = \frac{1}{n}\sum_{i=1}^{n}\int(\sqrt{p_{\theta_0,i}} - \sqrt{p_{\theta,i}})^2 d\mu_i$*, and* $e_n(\theta_0,\theta) = \frac{1}{n}\sum_{i=1}^{n}\int(\sqrt{p_{\theta_0,i}} - \sqrt{p_{\theta,i}})^2 d\mu_i$ *be semimetrics on* Θ *for which tests satisfying the conditions of* ***Lemma 2*** *of Ghosal [2007] exist :*

$$\begin{cases} P_{\theta_0}^{(n)}\phi_n \le e^{-\frac{1}{2}nd_n^2(\theta_0,\theta_1)} \\ P_{\theta}^{(n)}(1-\phi_n) \le e^{-\frac{1}{2}nd_n^2(\theta_0,\theta_1)} \end{cases} \tag{95}$$

Suppose that for some nonincreasing functions $\epsilon \mapsto N'(\epsilon)$*, and some* $\epsilon_n \ge 0$*,*

$$N\left(\frac{\epsilon\xi}{36}, \{\theta \in \theta : d_n(\theta,\theta_0) < \epsilon\}, e_n\right) \le N'(\epsilon) \qquad for\ all \quad \epsilon > \epsilon_n. \tag{96}$$

Then for every $\epsilon > \epsilon_n$*, there exist tests* ϕ_n*,* $n \ge 1$*(depending on* ϵ*) such that*

$$\begin{cases} P_{\theta_0}^{(n)}\phi_n \le N'(\epsilon)\dfrac{e^{-\frac{1}{2}nd_n^2(\theta_0,\theta_1)}}{1-e^{-\frac{1}{2}nd_n^2(\theta_0,\theta_1)}} \\ P_{\theta}^{(n)}(1-\phi_n) \le e^{-\frac{1}{2}nd_n^2(\theta_0,\theta_1)j^2} \end{cases} \tag{97}$$

for all $\theta \in \Theta$ *such that* $d_n(\theta,\theta_1) \le \frac{1}{18}(\theta_0,\theta_1), d_n(\theta,\theta_0) > j$ *for every* $j \in \mathbb{N}$*.*

Proof. **Lemma 4.1** (Lemma 9 of Ghosal and Van der Vaart[2007]) 참조 : $K = \frac{1}{2}, d_n^2(\theta_0,\theta_1) = \epsilon^2$ 으로 대체

Lemma 8.5 *(Transformed Lemma 10 of Ghosal and Van der Vaart[2007])*
For $k \geq 2$, every $\epsilon > 0$ and every probability measure $\bar{\Pi}^*$ supported on the set $B_n^*(\theta_0, \epsilon; k)$, we have, for every $C > 0$,

$$P_{\theta_0}^{(n)}\left(\int \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\bar{\Pi}_n^*(\theta) \leq e^{-(1+C)n\epsilon^2}\right) \leq \frac{1}{C_k^*(n\epsilon^2)^{k/2}}. \tag{98}$$

Proof.

1. 우선 젠센 부등식(Jensen Inequality)에 의해, 임의의 확률변수 X 에 대하여 $\log \mathbb{E}(X) \geq \mathbb{E}(\log(X))$가 성립됨을 알 수 있다. 이때 $\ell_{n,\theta} = \log\left(\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}}\right)$ 라고 하면 젠센 부등식에 의해 다음이 성립한다 :

$$\log \int \left(\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}}\right) d\bar{\Pi}_n^*(\theta) \geq \int \ell_{n,\theta} d\bar{\Pi}_n^*(\theta) = \int \log\left(\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}}\right) d\bar{\Pi}_n^*(\theta). \tag{99}$$

2. 이제 (19) 에 있는 확률의 상한을 구하기 위해 확률안에 있는 부등식을 조작해보고자 한다. 이때 다음과 같은 형태를 고려할 수 있다 :

$$\int \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\bar{\Pi}_n^*(\theta) \leq e^{-(1+C)n\epsilon^2} \overset{\log}{\Rightarrow} \log\left(\int \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\bar{\Pi}_n^*(\theta)\right) \leq -(1+C)n\epsilon^2 \tag{100}$$

   따라서 (99)에 의해 다음의 부등식 또한 성립함을 알 수 있다 :

$$\int \ell_{n,\theta} d\bar{\Pi}_n^*(\theta) \leq -(1+C)n\epsilon^2. \tag{101}$$

   이 때 위 식의 양변에서 $KL(P_\theta^{(n)}, P_{\theta_0}^{(n)})$를 빼주게 되면 다음과 같다 :

$$\int \ell_{n,\theta} d\bar{\Pi}_n^*(\theta) - \int P_{\theta_0}^{(n)} \ell_{n,\theta} d\bar{\Pi}_n^*(\theta) \leq -(1+C)n\epsilon^2 - \int P_{\theta_0}^{(n)} \ell_{n,\theta} d\bar{\Pi}_n^*(\theta) \tag{102}$$

   따라서 위의 관계식으로부터 다음과 같은 관계를 얻을 수 있게 된다 :

$$P_{\theta_0}^{(n)}\left(\int \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\bar{\Pi}_n^*(\theta) \leq e^{-(1+C)n\epsilon^2}\right) \leq P_{\theta_0}^{(n)}\left(\int \left(\ell_{n,\theta} - P_{\theta_0}^{(n)}\ell_{n,\theta}\right) d\bar{\Pi}_n^*(\theta) \leq -(1+C)n\epsilon^2 - \int P_{\theta_0}^{(n)} \ell_{n,\theta} d\bar{\Pi}_n^*(\theta)\right) \tag{103}$$

3. **Lemma 5.1**(Lemma 10 of Ghosal and Van der Vaart[2007])과 비슷한 방식으로 전개하기 위해 다음과 같은 성질이 항상 만족됨을 증명하도록 한다 : (Chain Rule for KL-divergence)

Claim: Let P *and* Q be similar probability distributions. And define Kullback-Leibler divergence KL as :

$$KL(P,Q) = \sum_{x\in\Omega} P(x)\log\left(\frac{P(x)}{Q(x)}\right) \tag{104}$$

where Ω is sample space of P *and* Q. Then following statement is true :

$$KL(P(x,y),Q(x,y)) = KL(P(x),Q(x)) + \mathbb{E}_{x\sim P}[KL(P(y|x)),KL(Q(y|x))] \tag{105}$$

Proof.

$$KL(P(x,y),Q(x,y)) = \sum_x\sum_y P(x,y)\log\frac{P(x,y)}{Q(x,y)} = \sum_x\sum_y P(x,y)\log\frac{P(x)P(y|x)}{Q(x)Q(y|x)} \tag{106}$$

$$= \sum_x\sum_y P(x,y)\log\frac{P(x)}{Q(x)} + \sum_x\sum_y P(x,y)\log\frac{P(y|x)}{Q(y|x)} \tag{107}$$

$$= KL(P(x),Q(x)) + \mathbb{E}_{x\sim P}[KL(P(y|x)),KL(Q(y|x))] \tag{108}$$

Corollary 8.6 *If x and y are independent variables,*

$$KL(P(x,y),Q(x,y)) = KL(P(x),Q(x)) + KL(P(y),Q(y)) \tag{109}$$

위에서 얻은 사실을 바탕으로 다음과 같은 결과를 얻어낼 수 있다 :

$$KL\left(\bigotimes_{i=1}^{2} P_{\theta_0,i}, \bigotimes_{i=1}^{2} P_{\theta,i}\right) = KL(P_{\theta_0,1},P_{\theta,1}) + KL(P_{\theta_0,2},P_{\theta,2}) \tag{110}$$

현재상황은 관측치 $X^{(n)} = (X_1, X_2, \ldots, X_n)$에서 모든 X_i $(i = 1,\ldots,n)$가 서로 독립이므로 다음과 같은 확장이 가능하다 :

$$\sum_{i=1}^{n} KL(P_{\theta_0,i},P_{\theta,i}) = KL\left(\bigotimes_{i=1}^{n} P_{\theta_0,i}, \bigotimes_{i=1}^{n} P_{\theta,i}\right) = KL(P_{\theta_0}^{(n)}, P_{\theta}^{(n)}) \tag{111}$$

이제 위에서 얻은 결론을 토대로

$$P_{\theta_0}^{(n)}\left(\int\left(\ell_{n,\theta} - P_{\theta_0}^{(n)}\ell_{n,\theta}\right)d\bar{\Pi}_n^*(\theta) \le -(1+C)n\epsilon^2 - \int P_{\theta_0}^{(n)}\ell_{n,\theta}d\bar{\Pi}_n^*(\theta)\right) \tag{112}$$

의 상한을 구해보록 한다.

4. 모든 $\theta \in B_n^*(\theta_0, \epsilon; k)$ 에 대하여, (111)을 통해 다음이 성립됨을 확인할 수 있다 :

$$P_{\theta_0}^{(n)} \ell_{n,\theta} = P_{\theta_0}^{(n)} \log \left(\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} \right) = -KL(P_{\theta_0}^{(n)}, P_\theta^{(n)}) = -\sum_{i=1}^{n} KL(P_{\theta_0,i}, P_{\theta,i}) \geq -n\epsilon^2 \tag{113}$$

그러면 다음이 성립하게 된다 :

$$-(1+C)n\epsilon^2 - \int P_{\theta_0}^{(n)} \ell_{n,\theta} d\bar{\Pi}_n^*(\theta) \leq -n(1+C)\epsilon^2 + n\epsilon^2 = -Cn\epsilon^2. \tag{114}$$

5. 위의 결과를 이용하기에 앞서 마코프 부등식(Markov Inequality)에 의해, 임의의 확률변수 X 에 대하여, $\forall a > 0$에 대해 $P(|X| \geq a) \leq \frac{\mathbb{E}(|X|^n)}{a^n}$가 성립됨을 알 수 있다.

$$P_{\theta_0}^{(n)} \left(\int \left(\ell_{n,\theta} - P_{\theta_0}^{(n)} \ell_{n,\theta} \right) d\bar{\Pi}_n^*(\theta) \leq -Cn\epsilon^2 \right) \leq P_{\theta_0}^{(n)} \left(\left| \int \left(\ell_{n,\theta} - P_{\theta_0}^{(n)} \ell_{n,\theta} \right) d\bar{\Pi}_n^*(\theta) \right| \leq Cn\epsilon^2 \right) \tag{115}$$

$$\leq \frac{\int \left| \int \left(\ell_{n,\theta} - P_{\theta_0}^{(n)} \ell_{n,\theta} \right) d\bar{\Pi}_n^*(\theta) \right|^k dP_{\theta_0}^{(n)}}{(Cn\epsilon^2)^k} \tag{116}$$

$$\leq \frac{\int \int \left| \left(\ell_{n,\theta} - P_{\theta_0}^{(n)} \ell_{n,\theta} \right) \right|^k d\bar{\Pi}_n^*(\theta) dP_{\theta_0}^{(n)}}{(Cn\epsilon^2)^k} \tag{117}$$

$$= \frac{(C_k^{'} C_k) n^{k/2} \epsilon^k}{C^k n^k \epsilon^{2k}} \tag{118}$$

$$= \frac{1}{C_k^* (n\epsilon^2)^{k/2}}, \tag{119}$$

where $C_k^* = \dfrac{C^k}{C_k^{'} C_k}$

이때 (118)는 $B_n(\theta_0, \epsilon; k)$의 정의에 의해 다음과 같은 이유로 성립함을 알 수 있다 :

우선 가정에 의해 $V_{k,0}(P_{\theta_0,i}, P_{\theta,i})$와 $V_{k,0}(P^{(n)}_{\theta_0}, P^{(n)}_{\theta})$는 다음과 같이 정의됨을 알고 있다 :

(a) $V_{k,0}(P_{\theta_0,i}, P_{\theta,i}) = \mathbb{E}_{P_{\theta_0,i}} \left| \log\left(\frac{p_{\theta_0,i}}{p_{\theta,i}}\right) - P_{\theta_0,i} \log\left(\frac{p_{\theta_0,i}}{p_{\theta,i}}\right) \right|^k$

(b) $V_{k,0}(P^{(n)}_{\theta_0}, P^{(n)}_{\theta}) = \mathbb{E}_{P^{(n)}_{\theta_0}} \left| \log\left(\frac{p^{(n)}_{\theta_0}}{p^{(n)}_{\theta}}\right) - P^{(n)}_{\theta_0} \log\left(\frac{p^{(n)}_{\theta_0}}{p^{(n)}_{\theta,}}\right) \right|^k$

Definition 8.7 *(Application of Marcinkiewiz-Zygmund inequality)*

The mean $\bar{X}_n$ of n independent random variables satisfies

$$\mathbb{E}\left| \frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\right|^k \leq C^{'}_k n^{-k/2}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}|X_i|^k\right), \tag{120}$$

where $C^{'}_k$ is a constant depending only on k.

위에서 정의한 **Definition 8.7**에서 $X_i = \log\left(\frac{p_{\theta,i}}{p_{\theta_0,i}}\right) - P_{\theta_0,i}\log\left(\frac{p_{\theta,i}}{p_{\theta_0,i}}\right)$ 라고 하게 되면 다음이 성립됨을 알 수 있다:

(c)

$$\mathbb{E}_{P^{(n)}_{\theta_0}}\left| \frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}_{P^{(n)}_{\theta_0}}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\right|^k = \mathbb{E}_{P^{(n)}_{\theta_0}}\left| \frac{1}{n}\ell_{n,\theta} - \frac{1}{n}P^{(n)}_{\theta_0}\ell_{n,\theta} - \mathbb{E}_{P^{(n)}_{\theta_0}}\left(\frac{1}{n}\ell_{n,\theta} - \frac{1}{n}P^{(n)}_{\theta_0}\ell_{n,\theta}\right)\right|^k \tag{121}$$

$$= \mathbb{E}_{P^{(n)}_{\theta_0}}\left| \frac{1}{n}\ell_{n,\theta} - \frac{1}{n}P^{(n)}_{\theta_0}\ell_{n,\theta}\right|^k = \frac{1}{n^k}\mathbb{E}_{P^{(n)}_{\theta_0}}\left|\ell_{n,\theta} - P^{(n)}_{\theta_0}\ell_{n,\theta}\right|^k \tag{122}$$

$$= \frac{1}{n^k}\mathbb{E}_{P^{(n)}_{\theta_0}}\left|\log\left(\frac{p^{(n)}_{\theta_0}}{p^{(n)}_{\theta}}\right) - P^{(n)}_{\theta_0}\log\left(\frac{p^{(n)}_{\theta_0}}{p^{(n)}_{\theta,}}\right)\right|^k |(-1)|^k = \frac{1}{n^k}V_{k,0}(P^{(n)}_{\theta_0}, P^{(n)}_{\theta})|(-1)|^k \tag{123}$$

(d)

$$C^{'}_k n^{-k/2}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}|X_i|^k\right) = C^{'}_k n^{-k/2}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left|\log\left(\frac{p_{\theta,i}}{p_{\theta_0,i}}\right) - P_{\theta_0,i}\log\left(\frac{p_{\theta,i}}{p_{\theta_0,i}}\right)\right|^k\right) \tag{124}$$

$$= C^{'}_k n^{-k/2}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left|\log\left(\frac{p_{\theta_0,i}}{p_{\theta,i}}\right) - P_{\theta_0,i}\log\left(\frac{p_{\theta_0,i}}{p_{\theta,i}}\right)\right|^k |(-1)|^k\right) \tag{125}$$

$$= C^{'}_k n^{-k/2}\left(\frac{1}{n}\sum_{i=1}^{n}V_{k,0}(P_{\theta_0,i}, P_{\theta,i})|(-1)|^k\right) \tag{126}$$

그러면 (c),(d)를 (120)에 대입했을때 다음을 얻을 수 있다:

$$\frac{1}{n^k}V_{k,0}(P_{\theta_0}^{(n)}, P_\theta^{(n)})|(-1)|^k = C_k^{'} n^{-k/2}\left(\frac{1}{n}\sum_{i=1}^{n} V_{k,0}(P_{\theta_0,i}, P_{\theta,i})|(-1)|^k\right) \tag{127}$$

$$\leq C_k^{'} n^{-k/2} C^k \epsilon^k |(-1)|^k \tag{128}$$

이를 정리하면 다음과 같은 결과를 얻는다 :

$$V_{k,0}(P_{\theta_0}^{(n)}, P_\theta^{(n)}) \leq (C_k^{'} C_k) n^{k/2}\epsilon^k. \tag{129}$$

이제 다시 돌아가서 다음을 보일 수 있다 :

$$\int\int \left|\left(\ell_{n,\theta} - P_{\theta_0}^{(n)}\ell_{n,\theta}\right)\right|^k d\bar{\Pi}_n^*(\theta) dP_{\theta_0}^{(n)} = \int\int \left|\left(\ell_{n,\theta} - P_{\theta_0}^{(n)}\ell_{n,\theta}\right)\right|^k dP_{\theta_0}^{(n)} d\bar{\Pi}_n^*(\theta) \tag{130}$$

$$= \int \mathbb{E}_{P_{\theta_0}^{(n)}} \left|\left(\ell_{n,\theta} - P_{\theta_0}^{(n)}\ell_{n,\theta}\right)\right|^k d\bar{\Pi}_n^*(\theta) \tag{131}$$

$$\leq \int \mathbb{E}_{P_{\theta_0}^{(n)}} \left|\left(\ell_{n,\theta} - P_{\theta_0}^{(n)}\ell_{n,\theta}\right)\right|^k |(-1)|^k d\bar{\Pi}_n^*(\theta) \tag{132}$$

$$= \int V_{k,0}(P_{\theta_0}^{(n)}, P_\theta^{(n)}) d\bar{\Pi}_n^*(\theta) \tag{133}$$

$$\leq V_{k,0}(P_{\theta_0}^{(n)}, P_\theta^{(n)}) \tag{134}$$

$$\leq (C_k^{'} C_k) n^{k/2}\epsilon^k. \tag{135}$$

이제 (Theorem 4 of Ghosal and Van der Vaart[2007])에 대한 증명을 시작해보려 한다.

Theorem 8.8 *(Theorem 4 of Ghosal and Van der Vaart[2007])*

Let $P^{(n)}_{(\theta)}$ *be product measures and* $d_n(\theta_0, \theta) = \frac{1}{n}\sum_{i=1}^{n}\int(\sqrt{p_{\theta_0,i}} - \sqrt{p_{\theta,i}})^2 d\mu_i$. *Suppose that for a sequence* $\epsilon_n \to 0$ *such that* $n{\epsilon_n}^2$ *is bounded away from zero, some* $k > 1$, *all sufficiently large* j *and sets* $\Theta_n \subset \Theta$ *which satisfies following conditions :*

$$\sup_{\epsilon > \epsilon_n} \log N(\epsilon/36, \{\theta \in \Theta_n : d_n(\theta, \theta_0) < \epsilon\}, d_n) \le n{\epsilon_n}^2, \tag{136}$$

$$\frac{\Pi(\Theta - \Theta_n)}{\Pi(B_n^*(\theta_0, \epsilon_n; k))} = o(e^{-2n{\epsilon_n}^2}), \tag{137}$$

$$\frac{\Pi(\theta \in \Theta_n : j\epsilon_n < d_n(\theta, \theta_0) \le 2j\epsilon_n)}{\Pi(B_n^*(\theta_0, \epsilon_n; k))} \le e^{n{\epsilon_n}^2 j^2/4} \tag{138}$$

Then $P^{(n)}_{(\theta_0)}[\Pi(\theta : d_n(\theta, \theta_0) \ge M_n\epsilon_n|\mathbf{D}_n)] \to 0$ *for any sequence* $M_n \to \infty$.

Proof.

1. By **Lemma 8.4**(Transformed Lemma 9 of Ghosal and Van der Vaart[2007]),
   Put $N^{'}(\epsilon) = e^{n\epsilon_n^2}, d_n(\theta_0, \theta_1) = M\epsilon_n$, where $M \ge 2$. Then there exist test ϕ_n that satisfies

$$\begin{cases} P^{(n)}_{\theta_0}\phi_n \le e^{n\epsilon_n^2}\frac{e^{-\frac{1}{2}nM^2\epsilon_n^2}}{1-e^{\frac{1}{2}nM^2\epsilon_n^2}} \\ P^{(n)}_{\theta}(1-\phi_n) \le e^{-\frac{1}{2}nM^2\epsilon_n^2 j^2} \end{cases} \tag{139}$$

   for all $\theta \in \Theta$ such that $d_n(\theta, \theta_1) \le \frac{1}{18}d_n(\theta_0, \theta_1), \quad d_n(\theta, \theta_0) > M\epsilon_n j$

2. $M \ge 2$인 조건은 추후에 선택될 M이 다음과 같은 부등식을 만족하기에 충분히 크다는 것을 보장해주기 위해서 이다 :

$$\tfrac{1}{2}M^2 - 1 > \tfrac{1}{4}M^2 \Rightarrow \tfrac{1}{2}M^2 n\epsilon_n^2 - n\epsilon_n^2 > \tfrac{1}{4}M^2 n\epsilon_n^2 \tag{140}$$

$$\Rightarrow -\tfrac{1}{2}M^2 n\epsilon_n^2 + n\epsilon_n^2 < -\tfrac{1}{4}M^2 n\epsilon_n^2 \tag{141}$$

$$\Rightarrow e^{n\epsilon_n^2}e^{-\frac{1}{2}M^2 n\epsilon_n^2} < e^{-\frac{1}{4}M^2 n\epsilon_n^2}. \tag{142}$$

   이때 사후확률 $\Pi_n(d_n(\theta, \theta_0) \ge JM\epsilon_n|X^{(n)}) \le 1$이라는 사실과 검정함수 ϕ_n이 0과 1사이임을 이용하면 다음과 같은 관계식을 생각할 수 있다 :

$$\Pi_n(d_n(\theta, \theta_0) \ge JM\epsilon_n|X^{(n)}) \le 1 \Rightarrow \Pi_n(d_n(\theta, \theta_0) \ge JM\epsilon_n|X^{(n)})\phi_n \le \phi_n \tag{143}$$

$$\Rightarrow P^{(n)}_{\theta_0}[\Pi_n(d_n(\theta, \theta_0) \ge JM\epsilon_n|X^{(n)})\phi_n] \le P^{(n)}_{\theta_0}\phi_n \tag{144}$$

이때 (139)을 이용하면 다음과 같은 결과를 얻을 수 있게 된다 :

$$
\begin{aligned}
P_{\theta_0}^{(n)}[\Pi_n(d_n(\theta,\theta_0) \geq JM\epsilon_n | X^{(n)})\phi_n] &\leq P_{\theta_0}^{(n)}\phi_n & (145)\\
&\leq e^{n\epsilon_n^2}\frac{e^{-\frac{1}{2}nM^2\epsilon_n^2}}{1-e^{\frac{1}{2}nM^2\epsilon_n^2}} & (146)\\
&\leq \frac{e^{-\frac{1}{4}nM^2\epsilon_n^2}}{1-e^{-\frac{1}{2}nM^2\epsilon_n^2}} & (147)\\
&\leq e^{-\frac{1}{4}nM^2\epsilon_n^2} + e^{-\frac{1}{4}nM^2\epsilon_n^2} & (148)\\
&\leq 2e^{-\frac{1}{4}nM^2\epsilon_n^2}. & (149)
\end{aligned}
$$

3. $\Theta_{n,j} = \{\theta \in \Theta_n : M\epsilon_n j < d_n(\theta,\theta_0) \leq M\epsilon_n(j+1)\}$ 라고 하게 되면 다음을 보일 수 있다 :

$$
\begin{aligned}
P_{\theta_0}^{(n)}\left[\int_{\Theta_{n,j}}(1-\phi_n)\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}}d\Pi_n(\theta)\right] &= \int\int_{\Theta_{n,j}}(1-\phi_n)\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}}d\Pi_n(\theta)dP_{\theta_0}^{(n)} & (150)\\
&= \int_{\Theta_{n,j}}\int(1-\phi_n)\frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}}dP_{\theta_0}^{(n)}d\Pi_n(\theta) \quad (\because \text{Fubini's theorem}) & (151)\\
&= \int_{\Theta_{n,j}}\int(1-\phi_n)P_\theta^{(n)}d\mu^{(n)}d\Pi_n(\theta) \quad (\because p_{\theta_0}^{(n)} = \frac{dP_{\theta_0}^{(n)}}{d\mu^{(n)}}) & (152)\\
&= \int_{\Theta_{n,j}} P_\theta^{(n)}(1-\phi_n)d\Pi_n(\theta) & (153)\\
&\leq \int_{\Theta_{n,j}} e^{-\frac{1}{2}nM^2\epsilon_n^2 j^2}d\Pi_n(\theta) \quad (\because (139)) & (154)\\
&= e^{-\frac{1}{2}nM^2\epsilon_n^2 j^2}\Pi_n(\Theta_{n,j}) & (155)
\end{aligned}
$$

4. 이제 어떠한 상수 $C > 0$을 고정하도록 한다. 그러면 (Transformed Lemma 10 of Ghosal and Van der Vaart[2007])에 의해 다음과 같은 사건A_n^*은 적어도 $1 - \frac{1}{C_k^*(n\epsilon^2)^{k/2}}$의 확률로 일어나게 됨을 알 수 있다 :

$$\int \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\Pi_n(\theta) \ \geq \ \int_{B_n^*(\theta_0,\epsilon_n;k)} \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\Pi_n(\theta) \ \geq \ e^{-(1+C)n\epsilon_n^2} \Pi_n(B_n^*(\theta_0,\epsilon_n;k)). \tag{156}$$

$$\because \int \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\Pi_n(\theta) \leq e^{-(1+C)n\epsilon_n^2} \text{ satisfies w.p. at most } \frac{1}{C_k^*(n\epsilon^2)^{k/2}} \tag{157}$$

$$\Rightarrow \int \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\Pi_n(\theta) \geq e^{-(1+C)n\epsilon_n^2} \text{ satisfies w.p. at least } 1 - \frac{1}{C_k^*(n\epsilon^2)^{k/2}} \tag{158}$$

$$\Rightarrow \int_{B_n^*(\theta_0,\epsilon_n;k)} \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\Pi_n(\theta) \geq \int_{B_n^*(\theta_0,\epsilon_n;k)} e^{-(1+C)n\epsilon_n^2} d\Pi_n(\theta) = e^{-(1+C)n\epsilon_n^2} \Pi_n(B_n^*(\theta_0,\epsilon_n;k)). \tag{159}$$

5. 이때 다음과 같은 집합 $\Theta_{M,J} = \{\theta \in \Theta : d_n(\theta,\theta_0) > JM\epsilon_n\}$ 을 다음과 같이 $\Theta_{n,j}$의 합으로 분해할 수 있다 :

$$\{\theta \in \Theta : d_n(\theta,\theta_0) > JM\epsilon_n\} = \bigcup_{j \geq J} \Theta_{n,j} = \bigcup_{j \geq J} \{\theta \in \Theta_n : M\epsilon_n j < d_n(\theta,\theta_0) \leq M\epsilon_n(j+1)\}. \tag{160}$$

또한 (138) 에서 j 대신에 Mj를 대입하고 난 뒤, 편의상 $\Theta_{2Mj} \stackrel{let}{=} \{\theta \in \Theta_n : Mj\epsilon_n < d_n(\theta,\theta_0) \leq 2Mj\epsilon_n\}$ 라고 하면 다음을 보일 수 있다 :

$$\frac{\Pi_n(\theta \in \Theta_n : Mj\epsilon_n < d_n(\theta,\theta_0) \leq 2Mj\epsilon_n)}{\Pi_n(B_n(\theta_0,\epsilon_n;k))} \leq e^{\frac{1}{4}nM^2\epsilon_n^2 j^2} \tag{161}$$

이제 $P_{\theta_0}^{(n)}[\Pi(\theta \in \Theta_n : d_n(\theta,\theta_0) > M\epsilon_n J|X^{(n)})(1-\phi_n)I_{A_n^*}]$ 의 상한을 구해보도록 한다 :

$$P_{\theta_0}^{(n)}[\Pi(\Theta_{M,J}|X^{(n)})(1-\phi_n)I_{A_n^*}] = \sum_{j\geq J} P_{\theta_0}^{(n)}[\Pi(\theta \in \Theta_n : M\epsilon_n j < d_n(\theta,\theta_0) \leq M\epsilon_n(j+1)|X^{(n)})(1-\phi_n)I_{A_n^*}] \tag{162}$$

$$\leq \sum_{j\geq J} P_{\theta_0}^{(n)}[\Pi(\theta \in \Theta_n : M\epsilon_n j < d_n(\theta,\theta_0) \leq 2M\epsilon_n j|X^{(n)})(1-\phi_n)I_{A_n^*}] \tag{163}$$

$$= \sum_{j\geq J} P_{\theta_0}^{(n)}\left[\frac{\int_{\Theta_{2Mj}} p_\theta^{(n)} d\Pi_n(\theta)}{\int_{\Theta} p_\theta^{(n)} d\Pi_n(\theta)}(1-\phi_n)I_{A_n^*}\right] \tag{164}$$

$$= \sum_{j\geq J} P_{\theta_0}^{(n)}\left[\frac{\int_{\Theta_{2Mj}} \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\Pi_n(\theta)}{\int_{\Theta} \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\Pi_n(\theta)}(1-\phi_n)I_{A_n^*}\right] \tag{165}$$

$$\leq \sum_{j\geq J} P_{\theta_0}^{(n)}\left[\frac{\int_{\Theta_{2Mj}} \frac{p_\theta^{(n)}}{p_{\theta_0}^{(n)}} d\Pi_n(\theta)}{\Pi_n(B_n^*(\theta_0,\epsilon_n;k))} e^{(1+C)n\epsilon_n^2}(1-\phi_n)\right] \quad (\because (156)) \tag{166}$$

$$\leq \sum_{j\geq J} e^{-\frac{1}{2}nM^2\epsilon_n^2 j^2} \frac{\Pi_n(\Theta_{2Mj})}{\Pi_n(B_n^*(\theta_0,\epsilon_n;k))} e^{(1+C)n\epsilon_n^2} \quad (\because (155)) \tag{167}$$

$$\leq \sum_{j\geq J} e^{-\frac{1}{2}nM^2\epsilon_n^2 j^2} e^{\frac{1}{4}nM^2\epsilon_n^2 j^2} e^{(1+C)n\epsilon_n^2} \tag{168}$$

6. 지금까지는 우리는 충분히 큰 J에 대하여 다음이 성립함을 보였다 :

$$P_{\theta_0}^{(n)}[\Pi(\theta \in \Theta : d_n(\theta, \theta_0) > JM\epsilon_n \ |X^{(n)})(1-\phi_n)I_{A_n^*}] \leq \sum_{j \geq J} e^{-n\epsilon_n^2(\frac{1}{2}M^2j^2-1-C-\frac{1}{4}M^2j^2)}.$$

이제 그동안 구한 결과들을 종합해보도록 한다 :

(i) $P_{\theta_0}^{(n)}[\Pi(\theta \in \Theta : d_n(\theta, \theta_0) > JM\epsilon_n \ |X^{(n)})(1-\phi_n)I_{A_n^*}] \leq \sum_{j \geq J} e^{-n\epsilon_n^2(\frac{1}{4}M^2j^2-1-C)}$

(ii) $P_{\theta_0}^{(n)}[\Pi(\theta \in \Theta : d_n(\theta, \theta_0) > JM\epsilon_n \ |X^{(n)})\phi_n] \leq 2e^{-\frac{1}{4}M^2n\epsilon_n^2}$

(iii) $P_{\theta_0}^{(n)}[\Pi(\theta \in \Theta : d_n(\theta, \theta_0) > JM\epsilon_n \ |X^{(n)})(1-\phi_n)I_{(A_n^*)^c}] \leq P_{\theta_0}^{(n)}[I_{(A_n^*)^c}] \leq \frac{1}{C_k^*(n\epsilon_n^2)^{k/2}}$

($\because$ (Transformed Lemma 10 of Ghosal and Van der Vaart[2007]))

따라서, 충분히 큰 M, J에 대하여 $n\epsilon_n^2 \to \infty$ 일때

$$P_{\theta_0}^{(n)}[\Pi(\theta \in \Theta : d_n(\theta, \theta_0) > JM\epsilon_n \ |X^{(n)})] \leq \sum_{j \geq J} e^{-n\epsilon_n^2(\frac{1}{4}M^2j^2-1-C)} + 2e^{-\frac{1}{4}M^2n\epsilon_n^2} + \frac{1}{C_k^*(n\epsilon_n^2)^{k/2}} \to 0.$$