# Asymptotic Inference

## 1 Locally Asymptotically Normal Property

### 1.1 Sequence of Statistical Experiments

A sequence of statistical experiments is equivalent to a sequence of models $\left(\mathcal{X}^{(n)},\mathcal{A}^{(n)},\mathcal{P}^{(n)}\right)$. In the parametric case, $\mathcal{P}^{(n)}=\left\{P_\theta^{(n)}\mid\theta\in\Theta\right\}$, where $P_\theta^{(n)}$ is a probability measure on $\left(\mathcal{X}^{(n)},\mathcal{A}^{(n)}\right)$ and $\Theta\subseteq\mathbb{R}^k$ is an open set. We assume that $\theta$ is identifiable, that is, $\theta'\neq\theta\Rightarrow P_{\theta'}^{(n)}\neq P_\theta^{(n)}$. Usually, $\left(\mathcal{X}^{(n)},\mathcal{A}^{(n)}\right)=(\mathbb{R}^n,\mathcal{B}^n)$. The observed sample is noted by $\underline{X}^{(n)}=(X_1,...,X_n)$.

A local sequence of models stands for a local sequence of parameter values $\underline{\theta}^{(n)}=\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}^{(n)}$ centered in $\underline{\theta}$, where $\underline{\upsilon}(n)$ is a sequence of $k\times k$ non-singular matrices, such that $\|\underline{\upsilon}(n)\|\to 0$ when $n\to\infty$, and $\underline{\tau}^{(n)}\in\mathbb{R}^k$ $(\forall n\in\mathbb{N})$ is a bounded sequence, i.e. $\sup_n\quad\underline{\tau}^{(n)\prime}\underline{\tau}^{(n)}<\infty$. The selection of an adequate value for $\underline{\upsilon}(n)$ depends upon the considered model, but the difference among $\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}_1^{(n)}$, $\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}_2^{(n)}$ and $\underline{\theta}$ must always converge to zero for $n\to\infty$. Hence this choice establishes a relationship between the Euclidean neighborhoods of $\underline{\theta}\in\mathbb{R}^k$ and the neighborhoods of $P_\theta^{(n)}\in\mathcal{P}^{(n)}$. In the classic case, the typical choice is $\underline{\upsilon}(n)=\frac{1}{\sqrt{n}}I_{k\times k}$, or equivalently $\underline{\theta}^{(n)}=\underline{\theta}+\frac{1}{\sqrt{n}}\,\underline{\tau}^{(n)}$.

### 1.2 Log-Likelihoods

Consider the probability measures $P$ and $Q$ on $(\mathcal{X},\mathcal{A})$, both dominated by a $\sigma$-finite measure $\mu$ on $(\mathcal{X},\mathcal{A})$. If the densities $p=\mathrm{d}P/\mathrm{d}\mu$ and $q=\mathrm{d}Q/\mathrm{d}\mu$ are not null with probability one, the concept of the likelihood ratio of $Q$ with respect to $P$ is clearly defined by the ratio $q/p$. For a more general definition, we need to specify the Lebesgue decomposition of a measure with respect to another.

**Lemma 1 (Lebesgue decomposition of $Q$ with respect to $P$).** Let $P$ and $Q$ be two probability measures on $(\mathcal{X}, \mathcal{A})$. There are a non-negative function $f : \mathcal{X} \to \mathbb{R}$ and $N \in \mathcal{A}$ such that $P(N) = 0$ and $Q(A) = \int_A f \, \mathrm{d}P + Q(N \cap A)$, $A \in \mathcal{A}$.

**Proof.** Clearly, $P \ll P+Q$ and $Q \ll P+Q$, implying that the Radon-Nikodyn derivatives $\frac{\mathrm{d}P}{\mathrm{d}(P+Q)}$ and $\frac{\mathrm{d}Q}{\mathrm{d}(P+Q)}$ make sense. Let $N = \left\{x \mid \frac{\mathrm{d}P}{\mathrm{d}(P+Q)} = 0\right\}$ and

$$f(x) = \begin{cases} \frac{\mathrm{d}Q}{\mathrm{d}(P+Q)} \left(\frac{\mathrm{d}P}{\mathrm{d}(P+Q)}\right)^{-1} & x \in N^c \\ 0 & x \in N \end{cases}$$

The restrictions imposed by $N^c \cap \mathcal{A}$ in $P$ and $P+Q$ are therefore mutually absolutely continuous and there are versions of the Radon-Nikodyn derivatives such that $\frac{\mathrm{d}(P+Q)}{\mathrm{d}P} = \left(\frac{\mathrm{d}P}{\mathrm{d}(P+Q)}\right)^{-1}$. Thus, for all $A \in \mathcal{A}$, we have

$$\begin{aligned} Q(A) &= \int_{A \cap N^c} \mathrm{d}Q + \int_{A \cap N} \mathrm{d}Q \\ &= \int_{A \cap N^c} \frac{\mathrm{d}Q}{\mathrm{d}(P+Q)} \frac{\mathrm{d}(P+Q)}{\mathrm{d}P} \mathrm{d}P + Q(A \cap N) \\ &= \int_A f \, \mathrm{d}P + Q(A \cap N). \blacksquare \end{aligned}$$

**Remark 1.** The Lebesgue decomposition is essentially unique, namely if there are $(f_1, N_1)$ and $(f_2, N_2)$ satisfying the decomposition, then $f_1 = f_2$ $P$-almost surely and $Q(N_1 \triangle N_2) = 0$.

**Definition.** The function $f = \mathrm{d}Q/\mathrm{d}P$ is called the likelihood ratio of $Q$ with respect to $P$. Hence, the classic definition is adequate, since $\mathrm{d}Q/\mathrm{d}P = q/p$.

**Lemma 2.** There are versions $f_{P/Q}$ and $f_{Q/P}$ of $\mathrm{d}P/\mathrm{d}Q$ and $\mathrm{d}Q/\mathrm{d}P$, respectively, such that $f_{P/Q} = f_{Q/P}^{-1}$.

**Proof.** Consider $\left(f_{P/Q}, N_Q\right)$ and $\left(f_{Q/P}, N_P\right)$ defined by the Lebesgue decomposition of $P$ with respect to $Q$ and of $Q$ with respect to $P$, respectively. Let $\underline{x} \in \mathcal{X}$. For $\underline{x} \in N_P \cap N_Q$, both $f_{P/Q}(\underline{x})$ and $f_{Q/P}(\underline{x})$ can be modified in a arbitrary way so as to guarantee that $f_{P/Q}(\underline{x}) = f_{Q/P}^{-1}(\underline{x})$. For $\underline{x} \in N_P \cap N_Q^c$, $f_{Q/P}(\underline{x})$

can be modified in a arbitrary way so as to ensure that $f_{P/Q}(\underline{x}) = f^{-1}_{Q/P}(\underline{x})$. For $\underline{x} \in N_P^c \cap N_Q$, $f_{P/Q}(\underline{x})$ can be modified in a arbitrary way to ascertain that $f_{P/Q}(\underline{x}) = f^{-1}_{Q/P}(\underline{x})$. Finally, for $\underline{x} \in N_P^c \cap N_Q^c$, $P$ is equivalent to $Q$, and $f_{P/Q}$ and $f_{Q/P}$ are Radon-Nikodyn derivatives, so that $f_{P/Q}(\underline{x}) = f^{-1}_{Q/P}(\underline{x})$. ■

**Remark 1.** Note that $\mathrm{d}Q/\mathrm{d}P$ is the Radon-Nikodyn derivative with respect to $P$ of $Q(\cdot) - Q(\cdot \cap N)$, which is the part of $Q$ that is absolute continuous with respect to $P$, or equivalently $Q(\cdot \cap N)$ is the part of $Q$ that is singular with respect to $P$.

**Remark 2.** Let $N_P$ and $N_Q$ be the singularity point sets (of $Q$ with respect to $P$ and $P$ with respect to $Q$, respectively) defined by the Lebesgue decomposition. We adopt the following convention: (1) for $\underline{x} \in N_P \cap N_Q^c$, $\mathrm{d}Q/\mathrm{d}P = 0$ and $\mathrm{d}P/\mathrm{d}Q = \infty$; (2) for $\underline{x} \in N_Q \cap N_P^c$, $\mathrm{d}P/\mathrm{d}Q = 0$ and $\mathrm{d}Q/\mathrm{d}P = \infty$; and (3) for $\underline{x} \in N_P \cap N_Q$, $\mathrm{d}Q/\mathrm{d}P = \mathrm{d}P/\mathrm{d}Q = 1$.

### 1.3 The LAN Family

The definitions of the likelihood ratio $\mathrm{d}P^{(n)}_{\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}^{(n)}}/\mathrm{d}P^{(n)}_{\underline{\theta}}$ and the log-likelihood

$$\Lambda^{(n)}_{\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}^{(n)},\underline{\theta}}\left(\underline{X}^{(n)}\right) \equiv \log\left(\frac{\mathrm{d}P^{(n)}_{\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}^{(n)}}}{\mathrm{d}P^{(n)}_{\underline{\theta}}}\right)$$

naturally follow. As the likelihood ratio takes values at $\overline{\mathbb{R}}_+ = [0, \infty]$, the log-likelihood $\Lambda^{(n)}$ assumes values belonging to $\overline{\mathbb{R}} = [-\infty, \infty]$.

**Definition.** The sequence of experiments $\mathcal{E}^{(n)} = \left(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}^{(n)}\right)$, $n \in \mathbb{N}$, is LAN if, $\forall \underline{\theta} \in \Theta$, there are a sequence of random vectors $\underline{\Delta}^{(n)}(\underline{\theta})$ (central sequence) with dimension $k \times 1$ and a $k \times k$ positive-definite matrix $\underline{\Gamma}(\underline{\theta})$, continuous on $\underline{\theta}$, such that

$$(i) \qquad \Lambda^{(n)}_{\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}^{(n)},\underline{\theta}} = \underline{\tau}^{(n)\prime}\underline{\Delta}^{(n)}(\underline{\theta}) - \frac{1}{2}\underline{\tau}^{(n)\prime}\underline{\Gamma}(\underline{\theta})\underline{\tau}^{(n)} + o_P(1), \text{ under } P^{(n)}_{\underline{\theta}}.$$

$$(ii) \qquad \underline{\Delta}^{(n)}(\underline{\theta}) \xrightarrow{\mathcal{L}} N\left(\underline{0}, \underline{\Gamma}(\underline{\theta})\right), \text{ under } P^{(n)}_{\underline{\theta}},$$

for all bounded sequence $\underline{\tau}^{(n)} \in \mathbb{R}^k$

**Remark 1.** If $\underline{\Delta}_1^{(n)}(\underline{\theta}) - \underline{\Delta}_2^{(n)}(\underline{\theta}) = o_P(1)$, under $P_{\underline{\theta}}^{(n)}$, and $\underline{\Delta}_1^{(n)}(\underline{\theta})$ is a central sequence, then $\underline{\Delta}_2^{(n)}(\underline{\theta})$ is also a central sequence.

**Remark 2.** If $\Lambda_1^{(n)}$ and $\Lambda_2^{(n)}$ are two log-likelihoods computed using distinct versions of the likelihood ratio, then $P_{\underline{\theta}}^{(n)}\left(\Lambda_1^{(n)} \neq \Lambda_2^{(n)}\right) = 0$, $\forall n, \underline{\theta}$. The choice of the particular version has no influence on the LAN property of the sequence $\mathcal{E}^{(n)}$ nor on the determination of the central sequence.

**Remark 3.** From the conditions established in the definition,

$$\Lambda^{(n)}_{\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}^{(n)},\underline{\theta}} \xrightarrow{\mathcal{L}} N\left(-\frac{1}{2}\underline{\tau}^{(n)\prime}\underline{\Gamma}(\underline{\theta})\underline{\tau}^{(n)}, \underline{\tau}^{(n)\prime}\underline{\Gamma}(\underline{\theta})\underline{\tau}^{(n)}\right), \text{ under } P_{\underline{\theta}}^{(n)}.$$

Moreover, from Lemma 2,

$$\Lambda^{(n)}_{\underline{\theta},\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}^{(n)}} \xrightarrow{\mathcal{L}} N\left(\frac{1}{2}\underline{\tau}^{(n)\prime}\underline{\Gamma}(\underline{\theta})\underline{\tau}^{(n)}, \underline{\tau}^{(n)\prime}\underline{\Gamma}(\underline{\theta})\underline{\tau}^{(n)}\right), \text{ under } P_{\underline{\theta}}^{(n)}.$$

**Remark 4.** Consider the Gaussian location model $\left(\mathbb{R}^k, \mathcal{B}^k, \mathcal{P}\right)$, where $\mathcal{P} = \left\{N(\underline{\Gamma}\,\underline{\tau}\,, \underline{\Gamma}) \;\middle|\; \underline{\tau} \in \mathbb{R}^k\right\}$ nd $\underline{\Gamma}$ is a known $k \times k$ positive-definite matrix. Denote the measure $N(\underline{\Gamma}\,\underline{\tau}, \underline{\Gamma})$ on $\left(\mathbb{R}^k, \mathcal{B}^k\right)$ by $P_{\underline{\tau}}$. The density with respect to the Lebesgue measure then reads

$$p_{\underline{\theta}}(\underline{\delta}) = (2\pi \mid \underline{\Gamma} \mid)^{-k/2} \exp\left[-\frac{1}{2}(\underline{\delta} - \underline{\Gamma}\,\underline{\tau})'\underline{\Gamma}^{-1}(\underline{\delta} - \underline{\Gamma}\,\underline{\tau})\right].$$

Thus, for a observation $\underline{\Delta}$ described by the model, we have

$$\begin{aligned}
\frac{\mathrm{d}P_{\underline{\tau}}}{\mathrm{d}P_{\underline{0}}}(\underline{\Delta}) &= \frac{p_{\underline{\tau}}}{p_{\underline{0}}} \\
&= \left(\frac{2\pi \mid \underline{\Gamma} \mid}{2\pi \mid \underline{\Gamma} \mid}\right)^{k/2} \exp\left(-\frac{1}{2}\left[(\underline{\Delta} - \underline{\Gamma}\,\underline{\tau})'\underline{\Gamma}^{-1}(\underline{\Delta} - \underline{\Gamma}\,\underline{\tau}) - \underline{\Delta}'\underline{\Gamma}^{-1}\underline{\Delta}\right]\right) \\
&= \exp\left(\underline{\tau}'\underline{\Delta} - \frac{1}{2}\underline{\tau}'\underline{\Gamma}\,\underline{\tau}\right)
\end{aligned}$$

and

$$\log\frac{\mathrm{d}P_{\underline{\tau}}}{\mathrm{d}P_{\underline{0}}} = \underline{\tau}'\underline{\Delta} - \frac{1}{2}\underline{\tau}'\underline{\Gamma}\,\underline{\tau} \;\sim\; N\left(-\frac{1}{2}\underline{\tau}'\underline{\Gamma}\,\underline{\tau}, \underline{\tau}'\underline{\Gamma}\,\underline{\tau}\right)$$

under $P_{\underline{0}}$. In a more general framework,

$$\log \frac{\mathrm{d}P_{\underline{\tau}_1}}{\mathrm{d}P_{\underline{\tau}_2}} = (\underline{\tau}_1 - \underline{\tau}_2)'\underline{\Delta} - \frac{1}{2}(\underline{\tau}_1 - \underline{\tau}_2)'\underline{\Gamma}(\underline{\tau}_1 + \underline{\tau}_2).$$

The log-likelihood of LAN experiment sequences behaves asymptotically in the same way as the log-likelihood of a Gaussian location model.

# 2 Contiguity

Let $\left(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}\right)$ be a sequence of measurable spaces, and $P^{(n)}$ and $Q^{(n)}$ two sequences of probability measures on $\left(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}\right)$.

**Definition.** The sequence $Q^{(n)}$ is contiguous to the sequence $P^{(n)}$ if

$$\lim_{n\to\infty} P^{(n)}\left(A^{(n)}\right) = 0 \Longrightarrow \lim_{n\to\infty} Q^{(n)}\left(A^{(n)}\right) = 0,$$

for all $A^{(n)} \in \mathcal{A}^{(n)}$. If $P^{(n)}$ and $Q^{(n)}$ are mutually contiguous, they are denominated contiguous.

**Remark 1.** Contiguity is an asymptotic form of absolute continuity. Note that if $\mathrm{d}Q^{(n)}/\mathrm{d}P^{(n)}$ and $N^{(n)}$ provide a Lebesgue decomposition of $Q^{(n)}$ with respect to $P^{(n)}$, the contiguity of $Q^{(n)}$ to $P^{(n)}$ imposes that $Q^{(n)}\left(N^{(n)}\right) \to 0$.

**Remark 2.** Contiguity is also related to a weak form of asymptotic vicinity between the sequences $P^{(n)}$ and $Q^{(n)}$. Consider the total variation distance

$$d_1(P,Q) \equiv \frac{1}{2}\int \left|\frac{\mathrm{d}P}{\mathrm{d}(P+Q)} - \frac{\mathrm{d}Q}{\mathrm{d}(P+Q)}\right| \mathrm{d}(P+Q)$$

and the Hellinger distance

$$\mathcal{H}(P,Q) \equiv \sqrt{\frac{1}{2}\int\left(\sqrt{\mathrm{d}P} - \sqrt{\mathrm{d}Q}\right)^2}.$$

It is possible to show that $d_1\left(P^{(n)}, Q^{(n)}\right) \to 0 \Longleftrightarrow \mathcal{H}\left(P^{(n)}, Q^{(n)}\right) \to 0 \Rightarrow P^{(n)}$ and $Q^{(n)}$ are contiguous.

**Remark 3.** Contiguity is also a form of asymptotic undiscernability for tests. Let $\phi^{(n)}$ be a sequence of pure (nonrandomized) tests and $Q^{(n)}$ be contiguous to $P^{(n)}$. If $\limsup E_{P^{(n)}}\left[\phi^{(n)}\right] \leq 1-\delta,\ \delta > 0$ when $n \to \infty$, then there is $\delta' > 0$ such that $\limsup E_{Q^{(n)}}\left[\phi^{(n)}\right] \leq 1-\delta'$ when $n \to \infty$. Thus, if the size of the test for $P^{(n)}$ against $Q^{(n)}$ does not converge to one, the power does not converge to one, implying that there is no consistent test for $P^{(n)}$ against $Q^{(n)}$. Moreover,

$$\lim_{n\to\infty} E_{P^{(n)}}\left[\phi^{(n)}\right] = 0 \Longrightarrow \lim_{n\to\infty} E_{Q^{(n)}}\left[\phi^{(n)}\right] = 0.$$

Hence, as the size of the test for $P^{(n)}$ against $Q^{(n)}$ tends to zero, the asymptotic power also shrinks to zero.

**Remark 4.** If $Q^{(n)}$ is contiguous to $P^{(n)}$ and $\psi^{(n)} = o_P(h_n)$, then $\psi^{(n)} = o_Q(h_n)$, that is, a random vector has the same order of convergence in probability under two contiguous measures.

## 2.1 Le Cam's First Lemma

**Note.** For deriving the following result, it is important to keep in mind that $\mathrm{d}Q^{(n)}/\mathrm{d}P^{(n)}$ is the Radon-Nikodyn derivative of the absolute continuous part of $Q^{(n)}$ with respect to $P^{(n)}$.

**Le Cam's First Lemma (Hàjek, 1962).** Denote by $F^{(n)}$ the cumulative distribution function of $\mathrm{d}Q^{(n)}/\mathrm{d}P^{(n)}$ under $P^{(n)}$. If $F^{(n)}$ weakly converges to $F$ such that $\int x\,\mathrm{d}F(x) = 1$, then $Q^{(n)}$ is contiguous to $P^{(n)}$. Moreover, if $F(0) = 0$, then $Q^{(n)}$ and $P^{(n)}$ are (mutually) contiguous.

**Proof.** Let $A^{(n)} \in \mathcal{A}^{(n)}$ and $\epsilon > 0$ such that $1/\epsilon$ is a continuity points in $F$, then

$$\begin{aligned} Q^{(n)}\left(A^{(n)}\right) &= \int_{A^{(n)}} \frac{\mathrm{d}Q^{(n)}}{\mathrm{d}P^{(n)}}\mathrm{d}P^{(n)} + Q^{(n)}\left(A^{(n)} \cap N^{(n)}\right) \\ &\leq \int_{A^{(n)}\cap\left\{\frac{\mathrm{d}Q^{(n)}}{\mathrm{d}P^{(n)}}\leq 1/\epsilon\right\}} \frac{\mathrm{d}Q^{(n)}}{\mathrm{d}P^{(n)}}\mathrm{d}P^{(n)} \\ &\quad + \int_{A^{(n)}\cap\left\{\frac{\mathrm{d}Q^{(n)}}{\mathrm{d}P^{(n)}}> 1/\epsilon\right\}} \frac{\mathrm{d}Q^{(n)}}{\mathrm{d}P^{(n)}}\mathrm{d}P^{(n)} + Q^{(n)}\left(N^{(n)}\right) \end{aligned}$$

$$\leq \frac{1}{\epsilon} P^{(n)}\left(A^{(n)}\right) + 1 - \int_{\left\{\frac{\mathrm{d}Q^{(n)}}{\mathrm{d}P^{(n)}} \leq 1/\epsilon\right\}} \frac{\mathrm{d}Q^{(n)}}{\mathrm{d}P^{(n)}} \mathrm{d}P^{(n)},$$

since

$$\int_{\left\{\frac{\mathrm{d}Q^{(n)}}{\mathrm{d}P^{(n)}} \leq 1/\epsilon\right\}} \frac{\mathrm{d}Q^{(n)}}{\mathrm{d}P^{(n)}} \mathrm{d}P^{(n)} + \int_{\left\{\frac{\mathrm{d}Q^{(n)}}{\mathrm{d}P^{(n)}} > 1/\epsilon\right\}} \frac{\mathrm{d}Q^{(n)}}{\mathrm{d}P^{(n)}} \mathrm{d}P^{(n)} + Q^{(n)}\left(N^{(n)}\right) = 1.$$

Note that as $F$ has mean one and non-negative support, the Markov inequality dictates that $1 - \int_0^{1/\epsilon} \mathrm{d}F \leq \epsilon$. In this way, we can use Helly-Bray's theorem to demonstrate that

$$1 - \int_{\left\{\frac{\mathrm{d}Q^{(n)}}{\mathrm{d}P^{(n)}} \leq 1/\epsilon\right\}} \frac{\mathrm{d}Q^{(n)}}{\mathrm{d}P^{(n)}} \mathrm{d}P^{(n)} < \epsilon,$$

for $n$ large enough. Thus, taking $P^{(n)}\left(A^{(n)}\right) < \epsilon^2$, for instance, implies that $Q^{(n)}\left(A^{(n)}\right) < 2\epsilon$, which means that $Q^{(n)}$ is contiguous with respect to $P^{(n)}$. If $F(0) = 0$, then $P^{(n)}\left(\frac{\mathrm{d}P^{(n)}}{\mathrm{d}Q^{(n)}} \leq \epsilon\right) = P^{(n)}\left(\frac{\mathrm{d}Q^{(n)}}{\mathrm{d}P^{(n)}} \leq 1/\epsilon\right)$, and we can use the same argument to show mutual contiguity.

$$\begin{aligned}
P^{(n)}\left(A^{(n)}\right) &= \int_{A^{(n)} \cap \left\{\frac{\mathrm{d}P^{(n)}}{\mathrm{d}Q^{(n)}} \leq \epsilon\right\}} \frac{\mathrm{d}P^{(n)}}{\mathrm{d}Q^{(n)}} \mathrm{d}Q^{(n)} \\
&\quad + \int_{A^{(n)} \cap \left\{\frac{\mathrm{d}P^{(n)}}{\mathrm{d}Q^{(n)}} > \epsilon\right\}} \frac{\mathrm{d}P^{(n)}}{\mathrm{d}Q^{(n)}} \mathrm{d}Q^{(n)} + P^{(n)}\left(A^{(n)} \cap M^{(n)}\right) \\
&\leq \epsilon\, Q^{(n)}\left(A^{(n)}\right) + 1 - \int_{\left\{\frac{\mathrm{d}P^{(n)}}{\mathrm{d}Q^{(n)}} \leq \epsilon\right\}} \frac{\mathrm{d}P^{(n)}}{\mathrm{d}Q^{(n)}} \mathrm{d}Q^{(n)} \\
&= \epsilon\, Q^{(n)}\left(A^{(n)}\right) + 1 - \int_{\left\{\frac{\mathrm{d}Q^{(n)}}{\mathrm{d}P^{(n)}} \leq 1/\epsilon\right\}} \mathrm{d}P^{(n)} \\
&\leq \epsilon\, Q^{(n)}\left(A^{(n)}\right) + \epsilon.
\end{aligned}$$

Thus, if $Q^{(n)}\left(A^{(n)}\right) < \epsilon$, then $P^{(n)}\left(A^{(n)}\right) < \epsilon\,(1+\epsilon)$, so that $Q^{(n)}$ and $P^{(n)}$ are contiguous. ■

**Corollary.** If $\log \frac{\mathrm{d}Q^{(n)}}{\mathrm{d}P^{(n)}} \xrightarrow{\mathcal{L}} N\left(-\frac{1}{2}d^2, d^2\right)$ under $P^{(n)}$, then $P^{(n)}$ and $Q^{(n)}$ are (mutually) contiguous.

**Proof.** If $\log \frac{\mathrm{d}Q^{(n)}}{\mathrm{d}P^{(n)}} \xrightarrow{\mathcal{L}} N\left(-\frac{1}{2}d^2, d^2\right)$ under $P^{(n)}$, then $\frac{\mathrm{d}Q^{(n)}}{\mathrm{d}P^{(n)}}$ converges in distribution to a lognormal with parameters $(-\frac{1}{2}d^2, d^2)$ under $P^{(n)}$. But the mean of a lognormal with parameters $\left(\mu,\sigma^2\right)$ is $\exp\left(\mu+\frac{1}{2}\sigma^2\right)$, hence $\frac{\mathrm{d}Q^{(n)}}{\mathrm{d}P^{(n)}}$ has an asymptotic mean equal to one, i.e. $\int x\,\mathrm{d}F^{(n)}(x)=1$. ■

**Generalization (Le Cam, 1986).** If $\Lambda^{(n)} = \log\frac{\mathrm{d}Q^{(n)}}{\mathrm{d}P^{(n)}}$ is asymptotically normal under $P^{(n)}$, that is, there is $\mu^{(n)}$ and $\sigma^{(n)}$ such that

$$\frac{\Lambda^{(n)}-\mu^{(n)}}{\sigma^{(n)}} \xrightarrow{\mathcal{L}} N(0,1),$$

then $P^{(n)}$ and $Q^{(n)}$ are mutually contiguous if and only if

$(i)$ $\quad \limsup_{n\to\infty} \sigma^{(n)} < \infty$

$(ii)$ $\quad \lim_{n\to\infty} \mu^{(n)} + \frac{1}{2}\left(\sigma^{(n)}\right)^2 = 0$

**Corollary.** Among the LAN families, the sequences $P^{(n)}_{\underline{\theta}}$, $P^{(n)}_{\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}_1^{(n)}}$ and $P^{(n)}_{\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}_2^{(n)}}$ are mutually contiguous for all bounded sequences $\underline{\tau}_1^{(n)}$ and $\underline{\tau}_2^{(n)}$. Moreover, under $P^{(n)}_{\underline{\theta}}$,

$$\frac{\Lambda_{\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}^{(n)},\underline{\theta}}+\frac{1}{2}\underline{\tau}^{(n)\prime}\Gamma(\underline{\theta})\underline{\tau}^{(n)}}{\left(\underline{\tau}^{(n)\prime}\Gamma(\underline{\theta})\underline{\tau}^{(n)}\right)^{1/2}} \xrightarrow{\mathcal{L}} N(0,1).$$

## 2.2 Le Cam's Third Lemma

The intuition says that knowing the asymptotic distribution of a statistic $S^{(n)}$ under $P^{(n)}$ should be very informative to deriving its asymptotic behaviour under a sequence of a contiguous probability measure $Q^{(n)}$. This is the main concern of Le Cam's third lemma.

**Proposition.** If $\mathcal{L}\{(S^{(n)},\Lambda^{(n)})|P^{(n)}\} \to F$, where $F$ is a distribution on $(-\infty,\infty)\times(-\infty,\infty)$, and $Q^{(n)}$ is contiguous to $P^{(n)}$, then $\mathcal{L}\{S^{(n)}|Q^{(n)}\}\to G$, where $G(s)=\int_{-\infty}^{s}\int_{-\infty}^{\infty} e^{\lambda}\mathrm{d}F(s',\lambda)$.

**Proof.** Note that $\Lambda^{(n)}=\log\frac{\mathrm{d}Q^{(n)}}{\mathrm{d}P^{(n)}}$ and $Q^{(n)}(S^{(n)}\le s)=\int_{S^{(n)}\le s}\mathrm{d}Q^{(n)}$. Hence, using the Lebesgue decomposition, we have that

$$Q^{(n)}(S^{(n)}\le s)=\int_{S^{(n)}\le s} e^{\Lambda^{(n)}}\mathrm{d}P^{(n)}+Q^{(n)}\left(\{S^{(n)}\le s\}\cap N^{(n)}\right),$$

where the last term is negligible due to contiguity. Denote by $F^{(n)}(s,\lambda)$ the joint distribution function of $(S^{(n)},\Lambda^{(n)})$ under $P^{(n)}$. Then,

$$\begin{aligned}
Q^{(n)}(S^{(n)}\leq s) &= \int_{-\infty}^{s}\int_{-\infty}^{\infty} e^{\lambda}\mathrm{d}F^{(n)}(s',\lambda)+o(1)\\
&= \int_{-\infty}^{s}\int_{-c}^{c} e^{\lambda}\mathrm{d}F^{(n)}(s',\lambda)+\int_{-\infty}^{s}\int_{c}^{\infty} e^{\lambda}\mathrm{d}F^{(n)}(s',\lambda)\\
&\quad +\int_{-\infty}^{s}\int_{-\infty}^{-c} e^{\lambda}\mathrm{d}F^{(n)}(s',\lambda)+o(1)\\
&= C_1+C_2+C_3+o(1)
\end{aligned}$$

Because $e^{\lambda}$ is bounded for $(s',\lambda)\in(-\infty,s]\times[-c,c]$, the Helly-Bray's theorem can be applied to $C_1$ to show that it converges to $\int_{-\infty}^{s}\int_{-c}^{c}e^{\lambda}\mathrm{d}F(s',\lambda)$. Unfortunately, the same does not hold for $C_2$ and $C_3$ as $e^{\lambda}$ is not bounded in these two terms. Nevertheless, we show in the following that $\forall\epsilon>0$, there are $c_\epsilon$ and $N_\epsilon$ such that $C_1+C_2<\epsilon$, $\forall c\geq c_\epsilon$ and $n\geq N_\epsilon$. Suppose that $\exists\epsilon>0$ and sequences $c_j, n_j$ such that $\lim_{j\to\infty}c_j=0, \lim_{j\to\infty}n_j=\infty$, and

$$\begin{aligned}
Q^{(n)}\left(|\Lambda^{(n_j)}|>c_j\right) &= \int_{-\infty}^{\infty}\int_{-\infty}^{-c_j} e^{\lambda}\mathrm{d}F^{(n_j)}+\int_{-\infty}^{\infty}\int_{c_j}^{\infty} e^{\lambda}\mathrm{d}F^{(n_j)}\\
&\geq \int_{-\infty}^{s}\int_{-\infty}^{-c_j} e^{\lambda}\mathrm{d}F^{(n_j)}+\int_{-\infty}^{s}\int_{c_j}^{\infty} e^{\lambda}\mathrm{d}F^{(n_j)}\geq\epsilon.
\end{aligned}$$

By assumption, for $\xi\sim F$ and $j\to\infty$, $n_j\to\infty$ and $P(|\xi|>c_j)\to 0$. Because $F(n_j)$ weakly converges to $F$, $P^{(n_j)}\left(|\Lambda^{(n_j)}|>c_j\right)=P(|\xi|>c_j)+o(1)$, and hence $P^{(n_j)}\left(|\Lambda^{(n_j)}|>c_j\right)\to 0$. Then, by contiguity, $Q^{(n_j)}\left(|\Lambda^{(n_j)}|>c_j\right)\to 0$ when $j\to\infty$, contradicting the fact that $Q^{(n_j)}\left(|\Lambda^{(n_j)}|>c_j\right)\geq\epsilon$. From this perspective, $|Q^{(n)}(S^{(n)}\leq s)-C_1|\leq\epsilon$, $\forall\epsilon>0$, $\forall n\geq N_\epsilon$, and $\forall c\geq c_\epsilon$. In particular, for $c\to\infty$, $|Q^{(n)}(S^{(n)}\leq s)-G(s)|\leq\epsilon'$, where $G(s)=\int_{-\infty}^{s}\int_{-\infty}^{\infty}e^{\lambda}\mathrm{d}F(s',\lambda)$. ■

**Corollary (Le Cam's Third Lemma).** If $(S^{(n)},\Lambda^{(n)})'$ is asymptotic normal with mean $(\mu,-\frac{1}{2}d^2)'$ and variance

$$\Omega=\begin{pmatrix}\sigma^2 & \mu_{s\lambda}\\ \mu_{s\lambda} & d^2\end{pmatrix}$$

under $P^{(n)}$, then $P^{(n)}$ and $Q^{(n)}$ are contiguous, and $S^{(n)} \xrightarrow{\mathcal{L}} N(\mu + \mu_{s\lambda}, \sigma^2)$, under $Q^{(n)}$.

**Proof.** The contiguity is a direct consequence of Le Cam's first lemma. Note that $G(s) = \int_{-\infty}^{s} \int_{-\infty}^{\infty} e^{\lambda} \mathrm{d}\Phi(s', \lambda)$, where $\Phi$ is normal distributed with mean and variance given as above. Hence, the marginal distribution of $S^{(n)}$ is also normal and can be easily obtained by integrating this expression with respect to $\lambda$. ■

**Consequence 1.** If $\Lambda^{(n)} \xrightarrow{\mathcal{L}} N\left(-\frac{1}{2}d^2, d^2\right)$ under $P^{(n)}$, then

$$\begin{pmatrix} \Lambda^{(n)} \\ \Lambda^{(n)} \end{pmatrix} \xrightarrow{\mathcal{L}} N\left( \begin{pmatrix} -d^2/2 \\ -d^2/2 \end{pmatrix}, \begin{pmatrix} d^2 & d^2 \\ d^2 & d^2 \end{pmatrix} \right).$$

Hence, $\Lambda^{(n)}$ weakly converges to a normal distribution with mean $\frac{1}{2}d^2$ and variance $d^2$ under $Q^{(n)}$.

**Consequence 2.** For the LAN family, $\Lambda^{(n)} = \underline{\tau}'\underline{\Delta}^{(n)}(\underline{\theta}) - \frac{1}{2}\,\underline{\tau}'\underline{\Gamma}(\underline{\theta})\underline{\tau} + o_P(1)$ converges, under $P^{(n)}_{\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}}$, to a normal distribution with mean $\frac{1}{2}\underline{\tau}'\underline{\Gamma}(\underline{\theta})\underline{\tau}$ and variance $\underline{\tau}'\underline{\Gamma}(\underline{\theta})\underline{\tau}$, $\forall \underline{\tau} \in \mathbb{R}^k$. Accordingly, $\underline{\tau}'\underline{\Delta}^{(n)}(\underline{\theta}) \xrightarrow{\mathcal{L}} N\left(\underline{\tau}'\underline{\Gamma}(\underline{\theta})\underline{\tau}, \underline{\tau}'\underline{\Gamma}(\underline{\theta})\underline{\tau}\right)$ and $\underline{\Delta}^{(n)}(\underline{\theta}) \xrightarrow{\mathcal{L}} N\left(\underline{\Gamma}(\underline{\theta})\underline{\tau}, \underline{\Gamma}(\underline{\theta})\right)$, under $P^{(n)}_{\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}}$.

**Consequence 3.** Consider a test based on $S^{(n)}$ to distinguish between the null hypothesis $P^{(n)}$ and the alternative $Q^{(n)}$. The test that reject $P^{(n)}$ whenever $\left(S^{(n)} - \mu\right)/\sigma > z_\alpha$, where $S^{(n)} \xrightarrow{\mathcal{L}} N(\mu, \sigma^2)$, has asymptotic size $\alpha$. The power of the test is the probability of $\left(S^{(n)} - \mu\right)/\sigma > z_\alpha$ under $Q^{(n)}$, i.e.

$$\begin{aligned} \lim_{n\to\infty} Q^{(n)}\left[\frac{S^{(n)} - \mu}{\sigma} > z_\alpha\right] &= P\left[N(0,1) > \frac{\mu + \sigma z_\alpha - \mu - \mu_{s\lambda}}{\sigma}\right] \\ &= 1 - \Phi\left(z_\alpha - \mu_{s\lambda}/\sigma\right). \end{aligned}$$

The asymptotic power of the test increases with $\mu_{s\lambda}/\sigma$, so that the maximum is obtained when the correlation between $S^{(n)}$ and $\Lambda^{(n)}$ is one, i.e. the Neyman's test $S^{(n)} = \Lambda^{(n)} + o_p(1)$. As a consequence, the maximum asymptotic power is $1 - \Phi(z_\alpha - d)$ as $\rho = \mu_{s\lambda}/(d\sigma) = 1 \Rightarrow \mu_{s\lambda}/\sigma = d$.

**Consequence 4.** We can also look at the process of local log-likelihoods $\{\Lambda_{\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}^{(n)},\underline{\theta}} \,|\, \underline{\tau} \in \mathbb{R}^k\}$ as a process indexed by $\underline{\tau}$ ($\underline{\theta}$ fixed). Then, under $P_{\underline{\theta}}^{(n)}$,

$$\begin{pmatrix} \Lambda^{(n)}_{\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}_1,\underline{\theta}} \\ \vdots \\ \Lambda^{(n)}_{\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}_J,\underline{\theta}} \end{pmatrix} \xrightarrow{\mathcal{L}} N\left(-\frac{1}{2}\begin{pmatrix} \underline{\tau}_1'\underline{\Gamma}(\underline{\theta})\underline{\tau}_1 \\ \vdots \\ \underline{\tau}_J'\underline{\Gamma}(\underline{\theta})\underline{\tau}_J \end{pmatrix}, \left(\underline{\tau}_i'\underline{\Gamma}(\underline{\theta})\underline{\tau}_j\right)\right).$$

Note that this is equivalent to a Gaussian location model with covariance kernel $\underline{\Gamma}(\underline{\theta})$.

# 3 Convergence of Statistical Experiments

## 3.1 Statistical Decision Theory

Denote the statistical experiment by $\mathcal{E} = (\mathcal{X}, \mathcal{A}, \mathcal{P} = \{P_{\underline{\theta}} \,|\, \underline{\theta} \in \Theta\})$. The problem of statistical decision comprises a set of possible decisions $\mathcal{Z} = \{z\}$ with $\sigma$-algebra $\mathcal{B}^{\mathcal{Z}}$ and a loss function

$$\begin{aligned} W : \Theta \times \mathcal{Z} &\longrightarrow \overline{\mathbb{R}}_+ = [0, \infty] \\ (\underline{\theta}, z) &\longmapsto W_{\underline{\theta}}(z), \end{aligned}$$

which indicates the loss incurred by choosing the action $z$ given $P_{\underline{\theta}}$.

The decision rule is a Markovian kernel $\rho = \{\rho_{\underline{x}} \,|\, \underline{x} \in \mathcal{X}\}$, where $\rho_{\underline{x}}$ is a probability measure on $(\mathcal{Z}, \mathcal{B}^{\mathcal{Z}})$ such that $\rho_{\underline{x}}(B)$ is $\mathcal{A}$-measurable, $\forall B \in \mathcal{B}^{\mathcal{Z}}$. In this way, the loss given the decision rule $\rho$ is a random quantity $W_{\underline{\theta}}(Z)$. The conditional distribution of $Z$ given $\underline{X} = \underline{x}$ is $\rho_{\underline{x}}$, whereas $P_{\underline{\theta}}$ is the distribution of $\underline{X}$. Then, the conditional average loss given $\underline{X} = \underline{x}$, under $P_{\underline{\theta}}$, is $\int_{\mathcal{Z}} W_{\underline{\theta}}(z)\mathrm{d}\rho_{\underline{x}}(z)$. Finally, the unconditional average loss (henceforth risk function) is $R_W^{\rho}(\underline{\theta}) = \int_{\mathcal{X}}\int_{\mathcal{Z}} W_{\underline{\theta}}(z)\mathrm{d}\rho_{\underline{x}}(z)\mathrm{d}P_{\underline{\theta}}(x)$, under $P_{\underline{\theta}}$.

In a pure statistical point of view, a statistical experiment can be identified by the associated set of risk functions $R_W^{\rho}$. This means that comparing statistical experiments is equivalent to comparing the set of risk functions. Let

$$\mathcal{R}(\mathcal{P}, W) \equiv \{r : \Theta \to R \,|\, \exists\rho : R_W^{\rho}(\underline{\theta}) \le r(\underline{\theta}),\ \forall\underline{\theta}\}$$

denote the set of all functions dominated by $R^{\rho}_{W}(\underline{\theta})$ for all $\underline{\theta}$. Consider now its closure set

$$\overline{\mathcal{R}}(\mathcal{P},W) \equiv \left\{ r : \Theta \to R \,|\, r(\underline{\theta}) = \lim_{i\to\infty} r_i(\underline{\theta}),\ r_i \in \mathcal{R}(\mathcal{P},W),\ \forall \underline{\theta} \right\}.$$

Then, the comparison of two distribution families $\mathcal{P}_1$ and $\mathcal{P}_2$, on the same parametric space $\Theta$, can be based on the comparison of the sets $\{\overline{\mathcal{R}}(\mathcal{P}_1,W)|\mathcal{Z},W\}$ and $\{\overline{\mathcal{R}}(\mathcal{P}_2,W)|\mathcal{Z},W\}$, even when $(\mathcal{X}_1,\mathcal{A}_1)$ and $(\mathcal{X}_2,\mathcal{A}_2)$ are quite different. If these two sets are equal, then everything that is possible to do with $(\mathcal{X}_1,\mathcal{A}_1,\mathcal{P}_1)$, can also be done with $(\mathcal{X}_2,\mathcal{A}_2,\mathcal{P}_2)$.

## 3.2 Weak Convergence of Statistical Experiments

**Definition 1.** The deficiency of $\mathcal{P}_1$ with respect to $\mathcal{P}_2$, $\delta(\mathcal{P}_1,\mathcal{P}_2)$, is the smallest number $0 \leq \epsilon \leq 1$ such that for every arbitrary loss function $W$ such that $0 \leq W_{\underline{\theta}}(z) \leq 1$ and $\forall r_2 \in \mathcal{R}(\mathcal{P}_2,W)$, there is a $r_1 \in \overline{\mathcal{R}}(\mathcal{P}_1,W)$ satisfying $r_1(\underline{\theta}) \leq r_2(\underline{\theta}) + \epsilon$, $\forall \underline{\theta} \in \Theta$.

This means that $\delta(\mathcal{P}_1,\mathcal{P}_2)$ is the smallest quantity $\epsilon$, given $0 \leq W \leq 1$, that provides a risk function for the experiment $\mathcal{P}_1$ at least as good as whatever risk function for the experiment $\mathcal{P}_2$.

**Definition 2.** The Le Cam's distance $\Delta$ between $\mathcal{P}_1$ and $\mathcal{P}_2$ is $\Delta(\mathcal{P}_1,\mathcal{P}_2) \equiv \max\{\delta(\mathcal{P}_1,\mathcal{P}_2),\delta(\mathcal{P}_2,\mathcal{P}_1)\}$.

In summary, when using a loss function $W$ such that $0 \leq W \leq 1$, anything that it is done with an experiment $\mathcal{E}$ can be also done with an experiment $\mathcal{F}$ within $\Delta(\mathcal{E},\mathcal{F})$. Note that Le Cam's distance is actually a concept of pseudo-distance because two experiments $\mathcal{E}$ and $\mathcal{F}$ can be quite different but still such that $\Delta(\mathcal{E},\mathcal{F}) = 0$.

**Definition 3.** The sequence of experiments $\mathcal{E}^{(n)} = \left(\mathcal{X}^{(n)},\mathcal{A}^{(n)},\mathcal{P}^{(n)}\right)$ weakly converges to the experiment $\mathcal{E} = (\mathcal{X},\mathcal{A},\mathcal{P})$ if $\Delta\left(\mathcal{P}^{(n)}_{\Theta_0},\mathcal{P}_{\Theta_0}\right) \to 0$, $\forall \Theta_0 \subseteq \Theta$

with $\#\Theta_0 < \infty$, where $\mathcal{P}^{(n)}_{\Theta_0} = \left\{P^{(n)}_{\underline{\theta}} \mid \underline{\theta} \in \Theta_0\right\}$ and $\mathcal{P}_{\Theta_0} = \left\{P_{\underline{\theta}} \mid \underline{\theta} \in \Theta_0\right\}$ are the restrictions imposed by $\Theta_0$ in $\mathcal{P}^{(n)}$ and $\mathcal{P}$, respectively.

Note that this is a notion of convergence that is elementwise in $\underline{\theta} \in \Theta$, but uniform on $\mathcal{Z}$, $W$ and $\rho$.

## 3.3 The Fundamental Result

Let $\mathcal{E} = (\mathcal{X}, \mathcal{A}, \mathcal{P})$ be a statistical experiment and $\left\{\Lambda(\underline{t}, \underline{s}) \equiv \log \frac{\mathrm{d}P_{\underline{t}}}{\mathrm{d}P_{\underline{s}}} \,\middle|\, \underline{t}, \underline{s} \in \Theta\right\}$ the log-likelihood process associated to this experiment.

**Proposition (Le Cam, 1969).** The sequence $\mathcal{E}^{(n)} = \left(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}^{(n)}\right)$ weakly converges to $\mathcal{E} = (\mathcal{X}, \mathcal{A}, \mathcal{P})$ iff the log-likelihood process $\left\{\Lambda^{(n)}(\underline{\theta}, \underline{s}) \mid \underline{\theta}, \underline{s} \in \Theta\right\}$ weakly converges to $\{\Lambda(\underline{\theta}, \underline{s}) \mid \underline{\theta}, \underline{s} \in \Theta\}$ under $P^{(n)}_{\underline{\theta}}$. This means that, $\forall \underline{s}$,

$$\begin{pmatrix} \Lambda^{(n)}(\underline{t}_1, \underline{s}) \\ \vdots \\ \Lambda^{(n)}(\underline{t}_J, \underline{s}) \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} \Lambda(\underline{t}_1, \underline{s}) \\ \vdots \\ \Lambda(\underline{t}_J, \underline{s}) \end{pmatrix},$$

under $P^{(n)}_{\underline{s}}$, for all $J \in \mathbb{N}$ and $\underline{t}_1, \ldots, \underline{t}_J \in \Theta$.

## 3.4 The Case of the LAN Family

Consider a sequence of local experiments centered on $\underline{\theta}$,

$$\mathcal{E}^{(n)}_{\underline{\theta}} = \left(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \left\{P^{(n)}_{\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}} \mid \underline{\tau} \in \mathbb{R}^k\right\}\right),$$

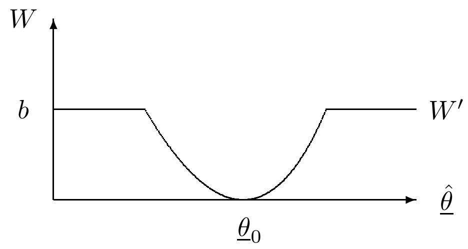
such that $\mathcal{E}^{(n)}_{\underline{\theta}} \xrightarrow{\mathcal{L}} \mathcal{E}_{\underline{\Gamma}(\underline{\theta})}$, where $\mathcal{E}_{\underline{\Gamma}(\underline{\theta})}$ is a Gaussian location model, i.e.

$$\begin{pmatrix} \Lambda^{(n)}_{\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}_1, \underline{\theta}} \\ \vdots \\ \Lambda^{(n)}_{\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}_J, \underline{\theta}} \end{pmatrix} \xrightarrow{\mathcal{L}} N\left(-\frac{1}{2}\begin{pmatrix} \underline{\tau}_1'\underline{\Gamma}(\underline{\theta})\underline{\tau}_1 \\ \vdots \\ \underline{\tau}_J'\underline{\Gamma}(\underline{\theta})\underline{\tau}_J \end{pmatrix}, \left(\underline{\tau}_i'\underline{\Gamma}(\underline{\theta})\underline{\tau}_j\right)\right),$$

under $P^{(n)}_{\underline{\theta}}$. For each hypothesis test, the risk function converges (elementwisely on $\underline{\tau}$ and uniformly on all possible tests) to a risk function associated to a Gaussian location model. In other words, if $R^* = E\left[\phi^*(\underline{\Delta})\right]$ is the risk function of a test $\phi^*(\underline{\Delta})$ for a limit Gaussian model, then there is a test sequence $\phi^*\left(\underline{\Delta}^{(n)}\right)$ with asymptotic risk function equal to $R^*$ if $\underline{\Delta}^{(n)}(\underline{\theta})$ converges to $\underline{\Delta}(\underline{\theta})$. Finally, note that the property of uniform convergence on all possible tests is crucial for constructing asymptotic most powerful tests.

# 4 Construction of LAN Estimators.

The problem of estimation is more complicated than of testing, because the loss function is not bounded. Thus, we need to truncate the loss function, $W' = \min\{b, W\}$, which clearly results in the loss of convexity.




## 4.1 Construction Methods

Let $\underline{\theta}_n^*$ be an auxiliary estimator satisfying two properties: $\underline{\upsilon}(n)$-convergence and locally asymptotic discreteness. The latter is important for deriving results on uniform convergence and it serves as an alternative to methods based on random perturbation. More formally,

**(A1) $\underline{\upsilon}(n)$-convergence.** For every $\underline{\theta}$, $\epsilon > 0$, there are $B(\underline{\theta}, \epsilon)$ and $N(\underline{\theta}, \epsilon)$ such that

$$P\left[\|\underline{\upsilon}^{-1}(n)\,(\underline{\theta}_n^* - \underline{\theta})\,\| \geq B(\underline{\theta}, \epsilon)\right] \leq \epsilon,\ \forall n \geq N(\underline{\theta}, \epsilon).$$

**(A2) local asymptotic discreteness.** For all $\underline{\theta}$, $b > 0$, there is $m_{\underline{\theta}}(b)$ such that the number of possible values of $\underline{\theta}_n^*$ inside the ball $\left\{\underline{t} : \|\underline{\upsilon}^{-1}(n)\,(\underline{t} - \underline{\theta})\,\| \leq b\right\}$ is inferior to $m_{\underline{\theta}}(b)$, $\forall n$.

**First Method.** At the vicinity of $\underline{\theta}$, and consequently of $\underline{\theta}_n^*$, the likelihood can be approximately written as

$$\Lambda^{(n)}_{\underline{\theta}^*_n+\underline{\upsilon}(n)\underline{t},\underline{\theta}^*_n} \sim \underline{t}'\underline{\Delta}^{(n)}\left(\underline{\theta}^*_n\right) - \frac{1}{2}\underline{t}'\underline{\Gamma}\left(\underline{\theta}^*_n\right)\underline{t}.$$

Maximizing this approximation (sort of local maximum likelihood estimation), we obtain $\hat{\underline{t}} = \underline{\Gamma}^{-1}\left(\underline{\theta}^*_n\right)\underline{\Delta}^{(n)}\left(\underline{\theta}^*_n\right)$ and thus $\hat{\underline{\theta}}^{(n)} = \underline{\theta}^*_n + \underline{\upsilon}(n)\underline{\Gamma}^{-1}\left(\underline{\theta}^*_n\right)\underline{\Delta}^{(n)}\left(\underline{\theta}^*_n\right)$. Note that $\hat{\underline{\theta}}^{(n)} = \underline{\theta}^*_n$ when $\underline{\theta}^*_n$ is such that $\underline{\Delta}^{(n)}\left(\underline{\theta}^*_n\right) = 0$ (maximum likelihood estimator). It is also interesting to note that this expression is quite similar to an iteration step of the Newton-Raphson algorithm.

**Second Method.** Assuming that $\underline{\Gamma}(\cdot)$ is continuous, we have that $\underline{\Gamma}\left(\underline{\theta}^*_n\right) \xrightarrow{P} \underline{\Gamma}(\underline{\theta})$. Hence we can try to fit a quadratic form to the local log-likelihood function in the vicinity of $\underline{\theta}^*_n$:

$$\Lambda^{(n)}_{\underline{\theta}^*_n+\underline{\upsilon}(n)\underline{t},\underline{\theta}^*_n} \sim \underline{t}'\underline{D} - \frac{1}{2}\underline{t}'\underline{\Gamma}\left(\underline{\theta}^*_n\right)\underline{t}.$$

For instance, a estimator can be derived for $\underline{t} = \underline{u}_k = (0,\ldots,0,1,0,\ldots,0)' \in \mathbb{R}^K$ with $k$ varying from 1 to $K$, which yields $K$ equations:

$$\Lambda^{(n)}_{\underline{\theta}^*_n+\underline{\upsilon}(n)\underline{u}_k,\underline{\theta}^*_n} \sim [\underline{D}]_k - \frac{1}{2}\left[\underline{\Gamma}(\underline{\theta}^*_n)\right]_{kk}.$$

This implies the following quadratic form

$$\underline{D}\left(\underline{\theta}^*_n\right) = \left(\Lambda^{(n)}_{\underline{\theta}^*_n+\underline{\upsilon}(n)\underline{u}_1,\underline{\theta}^*_n} + \frac{1}{2}\left[\underline{\Gamma}(\underline{\theta}^*_n)\right]_{11}, \ldots, \Lambda^{(n)}_{\underline{\theta}^*_n+\underline{\upsilon}(n)\underline{u}_K,\underline{\theta}^*_n} + \frac{1}{2}\left[\underline{\Gamma}(\underline{\theta}^*_n)\right]_{KK}\right)',$$

which is maximized by the LAN estimator $\hat{\underline{\theta}}^{(n)} = \underline{\theta}^*_n + \underline{\upsilon}(n)\underline{\Gamma}^{-1}\left(\underline{\theta}^*_n\right)\underline{D}\left(\underline{\theta}^*_n\right)$.

**Third Method.** This method is a generalization of the latter in the sense that it does not rely on the continuity of $\underline{\Gamma}(\cdot)$. We approximate $\underline{\Gamma}(\underline{\theta}^*_n)$ by $\underline{M}(\underline{\theta}^*_n)$, where

$$\begin{aligned}
[M(\underline{\theta}^*_n)]_{kl} &= \Lambda^{(n)}_{\underline{\theta}^*_n+\underline{\upsilon}(n)\underline{u}_k,\underline{\theta}^*_n} + \Lambda^{(n)}_{\underline{\theta}^*_n+\underline{\upsilon}(n)\underline{u}_l,\underline{\theta}^*_n} - \Lambda^{(n)}_{\underline{\theta}^*_n+\underline{\upsilon}(n)(\underline{u}_k+\underline{u}_l),\underline{\theta}^*_n} \\
&= \underline{u}'_k\underline{\Delta}(\underline{\theta}^*_n) - \frac{1}{2}\underline{u}'_k\underline{\Gamma}(\underline{\theta}^*_n)\underline{u}_k + \underline{u}'_l\underline{\Delta}(\underline{\theta}^*_n) - \frac{1}{2}\underline{u}'_l\underline{\Gamma}(\underline{\theta}^*_n)\underline{u}_l \\
&\quad -(\underline{u}'_k + \underline{u}'_l)\underline{\Delta}(\underline{\theta}^*_n) + \frac{1}{2}(\underline{u}'_k + \underline{u}'_l)\underline{\Gamma}(\underline{\theta}^*_n)(\underline{u}_k + \underline{u}_l)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}\underline{u}_k'\underline{\Gamma}(\underline{\theta}_n^*)\underline{u}_l + \frac{1}{2}\underline{u}_l'\underline{\Gamma}(\underline{\theta}_n^*)\underline{u}_k \\
&= [\underline{\Gamma}(\underline{\theta}_n^*)]_{kl}\,.
\end{aligned}
$$

Thus, the LAN estimator is $\hat{\underline{\theta}}^{(n)} = \underline{\theta}_n^* + \underline{\upsilon}(n)\underline{M}^{-1}(\underline{\theta}_n^*)\,\underline{D}(\underline{\theta}_n^*)$, where $\underline{D}(\underline{\theta}_n^*)$ is obtained according to the second method.

## 4.2 Asymptotic Equivalence

**Proposition.** Let $\mathcal{E}^{(n)} = \left(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}^{(n)}\right)$ satisfy the LAN property and consider an auxiliary estimator $\underline{\theta}_n^*$ of $\underline{\theta}$ satisfying conditions (A1) and (A2). Then, assuming that the vector $\left(\hat{\underline{\theta}}^{(n)}, \underline{M}^{(n)}\right)$ is a sequence of estimators and matrices constructed using the third method, we have that for all $\underline{\theta}$,

$$
(i) \qquad \underline{\upsilon}^{-1}(n)\left(\hat{\underline{\theta}}^{(n)} - \underline{\theta}\right) = \underline{\Gamma}^{-1}(\underline{\theta})\underline{\Delta}^{(n)}(\underline{\theta}) + o_P(1), \text{ under } P_{\underline{\theta}}^{(n)}.
$$

$$
(ii) \qquad \underline{\upsilon}^{-1}(n)\left(\hat{\underline{\theta}}^{(n)} - \underline{\theta}\right) \xrightarrow{\mathcal{L}} N\left(\underline{0}, \underline{\Gamma}^{-1}(\underline{\theta})\right), \text{ under } P_{\underline{\theta}}^{(n)}.
$$

**Proof.** Since $(ii)$ is a consequence of $(i)$, we need to show just (i). For a bounded sequence $\underline{t}_n$, we have that under $P_{\underline{\theta}}^{(n)}$

$$
\Lambda^{(n)}_{\underline{\theta}+\underline{\upsilon}(n)\underline{t}_n,\underline{\theta}} + \frac{1}{2}\left[\left(\underline{T}^{(n)} - \underline{t}_n\right)'\underline{\Gamma}(\underline{\theta})\left(\underline{T}^{(n)} - \underline{t}_n\right) - \underline{T}^{(n)\prime}\underline{\Gamma}(\underline{\theta})\underline{T}^{(n)}\right] = o_P(1),
$$

where $\underline{T}^{(n)} = \underline{\Gamma}^{-1}(\underline{\theta})\underline{\Delta}^{(n)}(\underline{\theta})$. Then, taking another bounded sequence $\underline{s}_n$,

$$
\begin{aligned}
\Lambda^{(n)}_{\underline{\theta}+\underline{\upsilon}(n)(\underline{t}_n+\underline{s}_n),\underline{\theta}+\underline{\upsilon}(n)\underline{s}_n} &+ \frac{1}{2}\left(\underline{T}^{(n)} - \underline{t}_n - \underline{s}_n\right)'\underline{\Gamma}(\underline{\theta})\left(\underline{T}^{(n)} - \underline{t}_n - \underline{s}_n\right) \\
&- \frac{1}{2}\left(\underline{T}^{(n)} - \underline{s}_n\right)'\underline{\Gamma}(\underline{\theta})\left(\underline{T}^{(n)} - \underline{s}_n\right) = o_P(1),
\end{aligned}
$$

under $P_{\underline{\theta}}^{(n)}$, cause the local sequences are mutually contiguous.

Consider now a random sequence $\underline{s}_n^*$ such that, for all $b > 0$ and $n > N(\epsilon)$, $P_{\underline{\theta}}^{(n)}\left[\|\underline{s}_n^*\| > b\right] \leq \epsilon$ and there is a finite number of possible values of $\underline{s}_n^*$ in the domain $\|\underline{s}_n^*\| \leq b$. Under (A1) and (A2), $\underline{s}_n^* = \underline{\upsilon}^{-1}(n)\left(\underline{\theta}_n^* - \underline{\theta}\right)$ satisfies these two conditions. Note that $\underline{s}_n^*$ is such that $\underline{\theta}_n^* = \underline{\theta} + \underline{\upsilon}(n)\underline{s}_n^*$ by construction. In this way, substituting $\underline{s}_n^*$ for $\underline{s}_n$ in the expression above yields

$$\begin{aligned}\Lambda^{(n)}_{\underline{\theta}_n^*+\underline{\upsilon}(n)\underline{t}_n,\underline{\theta}_n^*} &+ \frac{1}{2}\left(\underline{T}^{(n)}-\underline{t}_n-\underline{s}_n^*\right)'\underline{\Gamma}(\underline{\theta})\left(\underline{T}^{(n)}-\underline{t}_n-\underline{s}_n^*\right)\\ &- \frac{1}{2}\left(\underline{T}^{(n)}-\underline{s}_n^*\right)'\underline{\Gamma}(\underline{\theta})\left(\underline{T}^{(n)}-\underline{s}_n^*\right)=o_P(1),\end{aligned}$$

under $P^{(n)}_{\underline{\theta}}$, for all bounded sequences $\underline{t}_n$. In particular, taking $\underline{t}_n = \underline{u}_k$ and $\hat{\underline{T}}^{(n)} = \underline{\upsilon}^{-1}(n)\left(\hat{\underline{\theta}}^{(n)}-\underline{\theta}\right)$, we have

$$\begin{aligned}\Lambda^{(n)}_{\underline{\theta}_n^*+\underline{\upsilon}(n)\underline{t}_n,\underline{\theta}_n^*} &+ \frac{1}{2}\left(\hat{\underline{T}}^{(n)}-\underline{t}_n-\underline{s}_n^*\right)'\underline{M}(\underline{\theta}_n^*)\left(\hat{\underline{T}}^{(n)}-\underline{t}_n-\underline{s}_n^*\right)\\ &- \frac{1}{2}\left(\hat{\underline{T}}^{(n)}-\underline{s}_n^*\right)'\underline{M}(\underline{\theta}_n^*)\left(\hat{\underline{T}}^{(n)}-\underline{s}_n^*\right)=o_P(1).\end{aligned}$$

Combining the two last expressions with $\underline{t}_n = \underline{u}_k$, $\underline{u}_l$ and $\underline{u}_k+\underline{u}_l$ gives $\underline{M}(\underline{\theta}_n^*) = \underline{\Gamma}(\underline{\theta}) + o_P(1)$. Then, it is easy to show that $\hat{\underline{T}}^{(n)} = \underline{T}^{(n)} + o_P(1)$ using $\underline{t}_n = \underline{u}_k$, which completes the proof. ■

**Remark.** An estimator is superefficient if $\underline{\Gamma}^{-1}(\underline{\theta}) - \underline{\mathrm{I}}^{-1}(\underline{\theta})$ is positive-definite, where $\underline{\mathrm{I}}(\underline{\theta})$ is the Fisher information matrix.

## 4.3 Hàjek-Le Cam's Asymptotic Minimax Theorem

**Theorem (LAN Family).** Let the loss function $W$ be convex, non-negative and non-monotone. Let $a < E_{\underline{\theta}}\left[W\left(\underline{\Gamma}^{-1/2}(\underline{\theta})\,\underline{\xi}\right)\right]$, where $\underline{\xi} \sim N(\underline{0}, I_{k\times k})$. Then, there are a finite set $\mathcal{T} = \{\underline{\tau}_1,\ldots,\underline{\tau}_m\}$, $0 < b < \infty$ and an integer $N$ such that, $\forall n \geq N$,

$$\inf_{\underline{T}^{(n)}}\max_{\underline{\tau}\in\mathcal{T}} E\left[b\wedge W\left(\underline{T}^{(n)}-\underline{\tau}\right)\,\Big|\,\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}\right] > a,$$

where $\underline{T}^{(n)} = \underline{\Gamma}^{-1}(\underline{\theta})\underline{\Delta}^{(n)}(\underline{\theta})$ is an estimator of $\underline{\tau}$, which implies that $\underline{\theta}+\underline{\upsilon}(n)\underline{\tau}$ is estimated by $\underline{\theta}_n^* + \underline{\upsilon}(n)\underline{T}^{(n)}$.

**Remark.** $E\left[b\wedge W\left(\underline{T}^{(n)}-\underline{\tau}\right)\right]$ is the asymptotic risk of the LAN estimator under the bounded risk function $b\wedge W\left(\underline{T}^{(n)}-\underline{\tau}\right)$. Without taking into account the unboundness of the loss function $W$, the LAN estimator is the one that minimises the maximum risk.

### 4.4 Hàjek-Le Cam's Convolution Theorem.

**Theorem (LAN Family).** Let $\underline{T}^{(n)} \in \mathbb{R}^k$ be a sequence of random variables and $\mathrm{A}_{k\times k}$ be some constant matrix. Assume that $\mathcal{L}\left\{\underline{T}^{(n)} - \mathrm{A}\underline{\tau} \,\middle|\, \underline{\theta} + \underline{\upsilon}(n)\underline{\tau}\right\}$ weakly converges to $G$, which does not dependent on $\underline{\tau}$. Then, $G$ is also the distribution of $\mathrm{A}\underline{\Gamma}^{-1/2}(\underline{\theta})\underline{\xi} + \underline{U}$, where $\underline{\xi} \sim N(\underline{0}, \mathrm{I}_{k\times k})$ and independent of $\underline{U}$.

**Remark.** For the particular case of $\mathrm{A} = \mathrm{I}_{k\times k}$, the theorem establishes a more general result for the BAN property, since Cramér-Rao's result on efficient estimators is for a fixed sample size $n$ and there is no rigorous proof that it holds for $n \to \infty$. Moreover, for the LAN estimator described above, $\underline{U} = 0$ almost surely.

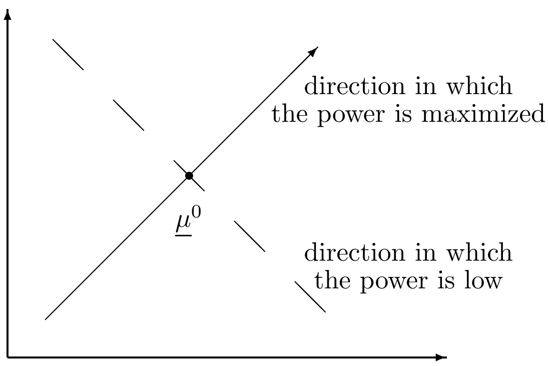
## 5 Asymptotic Locally Optimal Tests

For investigating the local optimality of hypothesis tests, it is necessary to consider a form of neighborhood centered on $\underline{\theta}_0$. Taking $\underline{\theta}_0 + \underline{\upsilon}(n)\underline{\tau}$,

$$\begin{cases} H_0^{(n)}: & \underline{\theta} = \underline{\theta}_0 \\ H_1^{(n)}: & \underline{\theta} \neq \underline{\theta}_0 \end{cases} \implies \begin{cases} H_0^{(n)}: & \underline{\tau} = 0 \\ H_1^{(n)}: & \underline{\tau} \neq 0. \end{cases}$$

### 5.1 Concepts of Optimality

**(1)**. A test $\phi^*$ is uniformly most powerful on the class $\mathcal{C}_\alpha$ of tests and against a specific alternative hypothesis $H_1$ if $\phi^* \in \mathcal{C}_\alpha = \{\phi \,|\, E_{\underline{\theta}}(\phi) \leq \alpha, \, \forall \underline{\theta} \in H_0\}$ is such that for every $\phi \in \mathcal{C}_\alpha$, $E_{\underline{\theta}}(\phi) \leq E_{\underline{\theta}}(\phi^*)$, $\forall \underline{\theta} \in H_1$. However, the requirements are too strong. Consider, for instance, $\underline{X} \sim N(\underline{\mu}, \Sigma)$ and the null hypothesis of $\underline{\mu} = \underline{\mu}^0$. Then, the most powerful test in one direction is, by construction, extremely weak in the orthogonal direction.




**(2).** A test $\phi^*$ is optimal in the maxmin sense on the class $\mathcal{C}_\alpha$ of tests and against the alternative hypothesis $H_1$ if $\phi^* \in \mathcal{C}_\alpha = \{\phi \,|\, E_{\underline{\theta}}(\phi) \leq \alpha, \, \forall \underline{\theta} \in H_0\}$ is such that $E_{\underline{\theta}}(\phi^*) \geq \beta(\alpha, H_0, H_1)$, $\forall \underline{\theta} \in H_1$, where

$$\beta(\alpha, H_0, H_1) \equiv \sup_{\phi \in \mathcal{C}_\alpha} \inf_{\underline{\theta} \in H_1} E_{\underline{\theta}}(\phi)$$

is the power function envelope. This means that the maxmin test has the best performance on the class $\mathcal{C}_\alpha$ in the less favorable alternative hypothesis.

**(3).** A test $\phi^*$ is the most stringent test on the class $\mathcal{C}_\alpha$ if its maximum deficit

$$r(\phi) \equiv \sup_{\underline{\theta} \in H_1} \sup_{\phi' \in \mathcal{C}_\alpha} E_{\underline{\theta}}(\phi') - E_{\underline{\theta}}(\phi)$$

achieves the minimum, viz. $r(\phi^*) \leq r(\phi)$, $\forall \phi \in \mathcal{C}_\alpha$. It is a sort of maxmin requirement but considering a more generous alternative hypothesis $H_1$.

## 5.2 Optimal Tests for the Gaussian Location Model

Consider $\underline{Z} \sim N(\underline{\mu}, \Sigma)$, where $\underline{\mu} \in \mathbb{R}^k$. The kind of test depends upon the concept of optimality that we are interested. For example, we already know that there is no uniformly most powerful test for the null hypothesis of $H_0 : \underline{\mu} = \underline{\mu}^0$ against the alternative of $H_1 : \underline{\mu} \neq \underline{\mu}^0$.

**(1).** For $H_0 : u'(\underline{\mu} - \underline{\mu}^0) \leq 0$, the uniformly most powerful test is

$$\phi^* = \begin{cases} 1 & \text{if} \quad (u'\Sigma u)^{-1/2}\, u'(\underline{z} - \underline{\mu}^0) > \mathcal{Z}_\alpha \\ 0 & \text{otherwise.} \end{cases}$$

For $k = 1$, the uniformly most powerful test consists in $\phi^* = 1$ if $z > \mu^0 + \sigma \mathcal{Z}_\alpha$ and $\phi^* = 0$ otherwise.

**(2).** For the null $H_0 : \underline{\mu} = \underline{\mu}^0$ against the alternative $H_1 : \underline{\mu} \neq \underline{\mu}^0$, the power envelope function is clearly $\beta(\alpha, H_0, H_1) = \alpha$, because $\underline{\mu} \neq \underline{\mu}^0$ allows $\underline{\mu}$ to be arbitrary close to $\underline{\mu}^0$. Under this perspective, the Hunt-Stein theorem shows that the test

$$\phi^* = \begin{cases} 1 & \text{if} \quad (\underline{z} - \underline{\mu}^0)'\Sigma^{-1}(\underline{z} - \underline{\mu}^0) > \chi^2_{k,1-\alpha} \\ 0 & \text{otherwise.} \end{cases}$$

is maxmin (on the class $\mathcal{C}_\alpha$) for every alternative hypothesis of the form $H_1^c$ : $(\underline{\mu} - \underline{\mu}^0)'\Sigma^{-1}(\underline{\mu} - \underline{\mu}^0) > c, \ \forall c > 0$. This test is also the most stringent.

**(3).** Let $\Omega_{k \times k-r}$ be a matrix of full rank $k - r$ and $\mathcal{M}(\Omega)$ denote the linear subspace of $\mathbb{R}^k$ spanned by the columns of $\Omega$. Then, for $\Omega_\perp$ such that $\mathcal{M}(\Omega) \perp \mathcal{M}(\Omega_\perp)$, consider the null

$$\begin{cases} H_0 : & \underline{\mu} - \underline{\mu}^0 \in \mathcal{M}(\Omega) \\ H_1 : & \underline{\mu} - \underline{\mu}^0 \notin \mathcal{M}(\Omega) \end{cases} \qquad \Longrightarrow \qquad \begin{cases} H_0 : & \Omega_\perp'(\underline{\mu} - \underline{\mu}^0) = 0 \\ H_1 : & \Omega_\perp'(\underline{\mu} - \underline{\mu}^0) \neq 0, \end{cases}$$

such that $r$ linear constraints are imposed on $\underline{\mu} - \underline{\mu}^0$. The most stringent test is then given by

$$\phi^* = \begin{cases} 1 & \text{if} \quad (\underline{z} - \underline{\mu}^0)'\Sigma^{-1/2}\, \mathbf{M}_{\Sigma^{-1/2}\Omega}\, \Sigma^{-1/2}(\underline{z} - \underline{\mu}^0) > \chi^2_{r,1-\alpha} \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathbf{M}_{\Sigma^{-1/2}\Omega}$ is the orthogonal projection matrix onto the subspace spanned by $(\Sigma^{-1/2}\Omega)_\perp$. The test statistic is simply

$$\left\| \mathbf{M}_{\Sigma^{-1/2}\Omega} \Sigma^{-1/2} \left(\underline{z} - \underline{\mu}^0\right) \right\|,$$

hence it gauges the length of the vector $(\underline{z} - \underline{\mu}^0)$ when obliquely projected onto the subspace spanned by $\Omega_\perp$ with slanting factor $\Sigma^{-1/2}$.

## 5.3 Locally Asymptotic Optimal Tests: The LAN Family

**(1).** Consider a sequence of local experiments $\theta^{(n)} = \theta + \upsilon(n)\tau$, where $\tau \in \mathbb{R}$. Since this sequence converges to a limit experiment $(\mathbb{R}, \mathcal{B}, \{N(\Gamma\tau, \Gamma) \,|\, \tau \in \mathbb{R}\})$, it is appropriate to write the local version of the test as

$$\begin{cases} H_0^{(n)}: & \theta \le \theta_0 \\ H_1^{(n)}: & \theta > \theta_0 \end{cases} \Longrightarrow \begin{cases} H_0^{(n)}: & \tau \le 0 \\ H_1^{(n)}: & \tau > 0 \end{cases} \Longrightarrow \begin{cases} H_0^{(n)}: & \Gamma\tau \le 0 \\ H_1^{(n)}: & \Gamma\tau > 0 \end{cases}$$

For a realization $\Delta$ of the limit experiment, the uniformly locally asymptotic most powerful test is $\phi^* = I\left[\Gamma^{-1/2}\Delta > \mathcal{Z}_\alpha\right]$. Note that $\Delta$ is centered under the less favorable case of the null ($\Gamma\tau = 0$). Thus, the test sequence

$$\phi^*_{(n)}(X^{(n)}) \equiv \phi^*\left(\Delta^{(n)}(\theta_0)\right) = I\left[\frac{(\Delta^{(n)}(\theta_0))}{\sqrt{\Gamma(\theta_0)}} > \mathcal{Z}_\alpha\right]$$

converges to a test satisfying the size condition

$$\limsup_{n\to\infty} \sup_{\theta \le \theta_0} E_{P_\theta^{(n)}}\left(\phi^*_{(n)}\right) \le \alpha.$$

**(2).** The locally asymptotic maxmin test for

$$\begin{cases} H_0^{(n)}: & \underline{\theta} = \underline{\theta}_0 \\ H_1^{(n)}: & \underline{\theta} \ne \underline{\theta}_0 \end{cases} \Longrightarrow \begin{cases} H_0^{(n)}: & \underline{\tau} = 0 \\ H_1^{(n)}: & \underline{\tau} \ne 0 \end{cases} \Longrightarrow \begin{cases} H_0^{(n)}: & \underline{\Gamma}\,\underline{\tau} = 0 \\ H_1^{(n)}: & \underline{\Gamma}\,\underline{\tau} \ne 0 \end{cases}$$

is $\phi^*(\underline{\Delta}) = I\left[\underline{\Delta}'\underline{\Gamma}^{-1}\underline{\Delta} > \chi^2_{k,1-\alpha}\right]$. Thus the sequence of locally maxmin tests is $\phi^*_{(n)}(X^{(n)}) = I\left[\underline{\Delta}^{(n)}(\underline{\theta}_0)'\underline{\Gamma}^{-1}(\underline{\theta}_0)\underline{\Delta}^{(n)}(\underline{\theta}_0) > \chi^2_{k,1-\alpha}\right]$. Examining the limit distribution of the test sequence under the alternative is possible to compute the asymptotic power. Under the alternative of $\underline{\Gamma}(\underline{\theta}_0)\underline{\tau} \ne 0$, $\underline{\Delta}^{(n)}(\underline{\theta}_0)$ weakly converges to a normal distribution with mean $\underline{\Gamma}(\underline{\theta}_0)\underline{\tau}$ and covariance matrix $\underline{\Gamma}(\underline{\theta}_0)$. Thus, the test statistic $\underline{\Delta}(\underline{\theta}_0)'\underline{\Gamma}^{-1}(\underline{\theta}_0)\underline{\Delta}(\underline{\theta}_0)$ weakly converges under the alternative to a noncentered $\chi^2$-distribution with $k$ degrees of freedom and non-centrality parameter $\underline{\tau}'\underline{\Gamma}(\underline{\theta}_0)\underline{\tau}$. Hence the asymptotic power is

$$\lim_{n\to\infty} E_{P^{(n)}_{\underline{\theta}_0 + \underline{\upsilon}(n)\underline{\tau}}}\left(\phi^*_{(n)}\right) = 1 - F_{\chi^2_k(\underline{\tau}'\underline{\Gamma}(\theta_0)\underline{\tau})}.$$

**(3).** Considering a local experiment with $\underline{\upsilon}(n)$ full ranked and the following local version of the hypothesis

$$\begin{cases} H_0^{(n)}: & \underline{\theta} - \underline{\theta}_0 \in \mathcal{M}(\Omega) \\ H_1^{(n)}: & \underline{\theta} - \underline{\theta}_0 \notin \mathcal{M}(\Omega) \end{cases} \Longrightarrow \begin{cases} H_{0,\underline{\upsilon}}^{(n)}: & \underline{\upsilon}(n)\underline{\tau} \in \mathcal{M}(\Omega) \\ H_{1,\underline{\upsilon}}^{(n)}: & \underline{\upsilon}(n)\underline{\tau} \notin \mathcal{M}(\Omega), \end{cases}$$

which is equivalent to test $H^{(n)}_{0;\underline{\upsilon}} : \underline{\Gamma}\,\underline{\tau} \in \mathcal{M}\left(\underline{\Gamma}\,\underline{\upsilon}^{-1}(n)\Omega\right)$. Then the test sequence

$$\begin{aligned}\phi^*_{\underline{\upsilon}}\left(\underline{\Delta}^{(n)}(\underline{\theta})\right) &= I\left[\underline{\Delta}^{(n)}(\underline{\theta})'\underline{\Gamma}^{-1/2}(\underline{\theta})\mathbf{M}_{\underline{\Gamma}^{1/2}\underline{\upsilon}^{-1}\Omega}\,\underline{\Gamma}^{-1/2}(\underline{\theta})\underline{\Delta}^{(n)}(\underline{\theta}) > \chi^2_{r,1-\alpha}\right] \\ &= I\left[\left\|\mathbf{M}_{\underline{\Gamma}^{1/2}\underline{\upsilon}^{-1}\Omega}\,\underline{\Gamma}^{-1/2}(\underline{\theta})\underline{\Delta}^{(n)}(\underline{\theta})\right\| > \chi^2_{r,1-\alpha}\right]\end{aligned}$$

converges uniformly in $\underline{\upsilon}(n)$ and elementwisely in $\underline{\tau}$ to the most stringent test of the limit experiment.

Note that the parameter vector $\underline{\theta}$ should be replaced by an estimator $\hat{\underline{\theta}}^{(n)}$ but without any effect on the convergence of $\phi^*$ under both the null and the alternative. In the following, we present the conditions for the nuisance parameter free property of the LAN tests.

**Proposition.** Consider a LAN estimator $\hat{\underline{\theta}}^{(n)}$ satisfying (A1), (A2) and the restriction imposed by the null, i.e. $\hat{\underline{\theta}}^{(n)} - \underline{\theta}_0 \in \mathcal{M}(\Omega)$. Then,

$$\phi^{(n)}_* = I\left[\underline{\Delta}\left(\hat{\underline{\theta}}^{(n)}\right)'\underline{\Gamma}^{-1/2}\left(\hat{\underline{\theta}}^{(n)}\right)\mathbf{M}_{\underline{\Gamma}^{1/2}\underline{\upsilon}^{-1}\Omega}\,\underline{\Gamma}^{-1/2}\left(\hat{\underline{\theta}}^{(n)}\right)\underline{\Delta}\left(\hat{\underline{\theta}}^{(n)}\right) > \chi^2_{r,1-\alpha}\right].$$

**Proof.** Assuming that $\underline{\Gamma}(\cdot)$ is continuous, $\underline{\Gamma}\left(\hat{\underline{\theta}}^{(n)}\right)$ converges in probability to $\underline{\Gamma}(\underline{\theta})$. Then, we just need to compute the pointwise limit of $\Delta^{(n)}\left(\hat{\underline{\theta}}^{(n)}\right)$ to fully characterize the asymptotic behaviour of the test. Consider the bounded (in probability) sequence $\underline{\tau}^{(n)} \equiv \underline{\upsilon}^{-1}(n)\left(\hat{\underline{\theta}}^{(n)} - \underline{\theta}\right)$, such that $\hat{\underline{\theta}}^{(n)} = \underline{\theta} + \underline{\upsilon}(n)\underline{\tau}^{(n)}$ by construction. Then,

$$\Lambda^{(n)}_{\hat{\underline{\theta}}^{(n)}/\underline{\theta}} = \underline{\tau}^{(n)'}\Delta^{(n)}(\underline{\theta}) - \frac{1}{2}\underline{\tau}^{(n)'}\underline{\Gamma}(\underline{\theta})\underline{\tau}^{(n)} + o_P(1)$$

and

$$\begin{aligned}\Lambda^{(n)}_{\underline{\theta}/\hat{\underline{\theta}}^{(n)}} &= -\underline{\tau}^{(n)'}\Delta^{(n)}\left(\underline{\theta} + \underline{\upsilon}(n)\underline{\tau}^{(n)}\right) - \frac{1}{2}\underline{\tau}^{(n)'}\underline{\Gamma}\left(\underline{\theta} + \underline{\upsilon}(n)\underline{\tau}^{(n)}\right)\underline{\tau}^{(n)} + o_P(1) \\ &= -\underline{\tau}^{(n)'}\Delta^{(n)}\left(\underline{\theta} + \underline{\upsilon}(n)\underline{\tau}^{(n)}\right) - \frac{1}{2}\underline{\tau}^{(n)'}\underline{\Gamma}(\underline{\theta})\underline{\tau}^{(n)} + o_P(1),\end{aligned}$$

as $\underline{\Gamma}\left(\hat{\underline{\theta}}^{(n)}\right) = \underline{\Gamma}(\underline{\theta}) + o_P(1)$. But $\Lambda^{(n)}_{\hat{\underline{\theta}}^{(n)}/\underline{\theta}} + \Lambda^{(n)}_{\underline{\theta},\hat{\underline{\theta}}^{(n)}} = 0$ by definition, which implies that

$$\underline{\tau}^{(n)'}\left[\Delta^{(n)}(\underline{\theta}) - \Delta^{(n)}\left(\underline{\theta} + \underline{\upsilon}(n)\underline{\tau}^{(n)}\right)\right] - \underline{\tau}^{(n)'}\underline{\Gamma}(\underline{\theta})\underline{\tau}^{(n)} = o_P(1).$$

Because $\underline{\tau}^{(n)}$ is bounded in probability, we can write that

$$\begin{aligned}\Delta^{(n)}(\underline{\theta}) - \Delta^{(n)}\left(\underline{\theta} + \underline{\upsilon}(n)\underline{\tau}^{(n)}\right) &= \underline{\Gamma}(\underline{\theta})\underline{\tau}^{(n)} + o_P(1) \\ &= \underline{\Gamma}(\underline{\theta})\underline{\upsilon}^{-1}(n)\left(\underline{\hat{\theta}}^{(n)} - \underline{\theta}\right) + o_P(1),\end{aligned}$$

under $P_{\underline{\theta}}^{(n)}$, $\forall \underline{\theta} \in \underline{\theta}_0 + \mathcal{M}(\Omega)$. Denoting the test statistic by $Q\left(\underline{\hat{\theta}}^{(n)}\right)$, the nuisance parameter effect on the test is

$$\begin{aligned}Q(\underline{\theta}) - Q\left(\underline{\hat{\theta}}^{(n)}\right) &= \left\|\mathbf{M}_{\underline{\Gamma}^{1/2}\underline{\upsilon}^{-1}\Omega}\,\underline{\Gamma}^{-1/2}(\underline{\theta})\underline{\Delta}(\underline{\theta})\right\| \\ &\quad - \left\|\mathbf{M}_{\underline{\Gamma}^{1/2}\underline{\upsilon}^{-1}\Omega}\,\underline{\Gamma}^{-1/2}\left(\underline{\hat{\theta}}^{(n)}\right)\underline{\Delta}\left(\underline{\hat{\theta}}^{(n)}\right)\right\| \\ &= \left\|\mathbf{M}_{\underline{\Gamma}^{1/2}\underline{\upsilon}^{-1}\Omega}\,\underline{\Gamma}^{-1/2}(\underline{\theta})\left[\underline{\Delta}(\underline{\theta}) - \underline{\Delta}\left(\underline{\hat{\theta}}^{(n)}\right)\right]\right\| + o_P(1) \\ &= \left\|\mathbf{M}_{\underline{\Gamma}^{1/2}\underline{\upsilon}^{-1}\Omega}\,\underline{\Gamma}^{1/2}(\underline{\theta})\underline{\upsilon}^{-1}(n)\left(\underline{\hat{\theta}}^{(n)} - \underline{\theta}\right)\right\| + o_P(1),\end{aligned}$$

using the fact that

$$\Delta^{(n)}(\underline{\theta}) - \Delta^{(n)}\left(\underline{\theta} + \underline{\upsilon}(n)\underline{\tau}^{(n)}\right) = \underline{\Gamma}(\underline{\theta})\underline{\upsilon}^{-1}(n)\left(\underline{\hat{\theta}}^{(n)} - \underline{\theta}\right) + o_P(1).$$

However, $\underline{\Gamma}^{1/2}(\underline{\theta})\underline{\upsilon}^{-1}(n)\left(\underline{\hat{\theta}}^{(n)} - \underline{\theta}\right)$ clearly belongs to $\mathcal{M}(\underline{\Gamma}^{1/2}\underline{\upsilon}^{-1}\Omega)$, so that $Q\underline{\theta}) - Q\left(\underline{\hat{\theta}}^{(n)}\right) = o_P(1)$. Thus, the asymptotic distribution of the test is the same whether the parameter $\underline{\theta}$ is estimated or not. ■

**Lemma.** Let $S^{(n)}(\underline{\theta})$ denote a sequence of random values such that, under $P_{\underline{\theta}}^{(n)}$, $S^{(n)}\left(\underline{\theta}^{(n)}\right)$ converges in probability to zero for every deterministic sequence $\underline{\theta}^{(n)}$ verifying $\sup_n \left\|\underline{\upsilon}^{-1}(n)\left(\underline{\theta}^{(n)} - \underline{\theta}\right)\right\| < \infty$. Then, for a sequence of estimators $\underline{\hat{\theta}}^{(n)}$ satisfying (A1) and (A2), $S^{(n)}\left(\underline{\hat{\theta}}^{(n)}\right) = o_P(1)$.

**Proof.** Let $P_\epsilon^{(n)}$ denote the probability of $\left\|S^{(n)}\left(\underline{\hat{\theta}}^{(n)}\right)\right\| \geq \epsilon$. Then for every $\epsilon > 0$ and $C > 0$, we have that

$$\begin{aligned}P_\epsilon^{(n)} &= P^{(n)}\left[\left\{\left\|S^{(n)}\left(\underline{\hat{\theta}}^{(n)}\right)\right\| \geq \epsilon\right\} \cap \left\{\left\|\underline{\upsilon}^{-1}(n)\left(\underline{\hat{\theta}}^{(n)} - \underline{\theta}\right)\right\| > C\right\}\right] \\ &\quad + P^{(n)}\left[\left\{\left\|S^{(n)}\left(\underline{\hat{\theta}}^{(n)}\right)\right\| \geq \epsilon\right\} \cap \left\{\left\|\underline{\upsilon}^{-1}(n)\left(\underline{\hat{\theta}}^{(n)} - \underline{\theta}\right)\right\| \leq C\right\}\right] \\ &\leq P^{(n)}\left[\left\{\left\|S^{(n)}\left(\underline{\hat{\theta}}^{(n)}\right)\right\| \geq \epsilon\right\} \cap \left\{\left\|\underline{\upsilon}^{-1}(n)\left(\underline{\hat{\theta}}^{(n)} - \underline{\theta}\right)\right\| > C\right\}\right]\end{aligned}$$

$$
\begin{aligned}
&+ P^{(n)}\left[\left\|\underline{\upsilon}^{-1}(n)\left(\underline{\hat{\theta}}^{(n)}-\underline{\theta}\right)\right\|\leq C\right]\\
\leq\;& P^{(n)}\left[\left\|S^{(n)}\left(\underline{\hat{\theta}}^{(n)}\right)\right\|\geq\epsilon\right]+P^{(n)}\left[\left\|\underline{\upsilon}^{-1}(n)\left(\underline{\hat{\theta}}^{(n)}-\underline{\theta}\right)\right\|\leq C\right]\\
\leq\;& P^{(n)}\left[\left\|S^{(n)}\left(\underline{\theta}_*^{(n)}\right)\right\|\geq\epsilon\right]+P^{(n)}\left[\left\|\underline{\upsilon}^{-1}(n)\left(\underline{\hat{\theta}}^{(n)}-\underline{\theta}\right)\right\|\leq C\right],
\end{aligned}
$$

where $\underline{\theta}_*^{(n)}$ is such that $\left\|S^{(n)}\left(\underline{\theta}_*^{(n)}\right)\right\| = \max\left\|S^{(n)}\left(\underline{\hat{\theta}}^{(n)}\right)\right\|$. But $\underline{\hat{\theta}}^{(n)}$ is $\underline{\upsilon}(n)$-convergent by (A1), which implies that $P^{(n)}\left[\left\|\underline{\upsilon}^{-1}(n)\left(\underline{\hat{\theta}}^{(n)}-\underline{\theta}\right)\right\|\leq C\right]\leq\epsilon^*/2$, for every $\epsilon^*>0$, $C>C(\epsilon^*)$ and $n>N(\epsilon^*)$. By the same token, (A2) imposes local asymptotic discreteness for the estimator $\underline{\hat{\theta}}^{(n)}$, which implies that $P^{(n)}\left[\left\|S^{(n)}\left(\underline{\theta}_*^{(n)}\right)\right\|\geq\epsilon\right]\leq\epsilon^*/2$ for sufficient large $n$. Hence we conclude that $P_\epsilon^{(n)} = P\left[\left\|S^{(n)}\left(\underline{\hat{\theta}}^{(n)}\right)\right\|\geq\epsilon\right]\leq\epsilon^*$, for every $\epsilon>0$, which means that $S^{(n)}\left(\underline{\hat{\theta}}^{(n)}\right)$ is $o_P(1)$. ■

# 6 The Case of Random Samples

## 6.1 The Hellinger Distance

**Definition.** The Hellinger distance $\mathcal{H}(P,Q)$ between two probability measures $P$ and $Q$ on $(\mathcal{X},\mathcal{A})$ stems from

$$
\mathcal{H}^2(P,Q)=\frac{1}{2}\int(\sqrt{\mathrm{d}P}-\sqrt{\mathrm{d}Q})^2=1-\gamma(P,Q),
$$

where $\gamma(P,Q)=\int\sqrt{\mathrm{d}P\mathrm{d}Q}$ stands for the Hellinger affinity.

The Hellinger concept of distance is very convenient for working with product measures. Consider, for instance, $P_0=\prod_{j\in J}p_{0,j}$ and $P_1=\prod_{j\in J}p_{1,j}$. The Hellinger affinity between the two product measures will then be equal to the product of the Hellinger affinities, viz. $\gamma(P_0,P_1)=\prod_{j\in J}\gamma(p_{0,j},p_{1,j})$, which implies that $\mathcal{H}^2(P_0,P_1)=1-\prod_{j\in J}\left(1-\mathcal{H}^2(p_{0,j},p_{1,j})\right)$.

**Properties.** Let $P_0$ and $P_1$ be two probability measures on $(\mathcal{X},\mathcal{A})$. Denote by $\gamma$ the Hellinger affinity between $P_0$ and $P_1$. If $P_1$ is not absolutely continuous with respect to $P_0$, denote by $\Upsilon_{P_1/P_0}$ the probability under $P_1$ of the singular set with respect to $P_0$, which implies that $1-\Upsilon_{P_1/P_0} = \int \frac{\mathrm{d}P_1}{\mathrm{d}P_0}(x)\mathrm{d}P_0$. Then, for $\varsigma(x) = \sqrt{\frac{\mathrm{d}P_1}{\mathrm{d}P_0}(x)} - 1$, we have that

$(i)$ $\quad E_{P_0}\left[\varsigma(x)\right] = \gamma - 1$

$(ii)$ $\quad E_{P_0}\left[\varsigma^2(x)\right] = 2(1-\gamma) - \Upsilon_{P_1/P_0}$

$(iii)$ $\quad Var_{P_0}\left[\varsigma(x)\right] = 1 - \gamma^2 - \Upsilon_{P_1/P_0}$

**Proof.** We demonstrate only the first property, because the others follow in the same way.

$$\begin{aligned}
E_{P_0}\left[\varsigma(x)\right] &= \int \varsigma(x)\mathrm{d}P_0(x) = \int \left[\sqrt{\frac{\mathrm{d}P_1}{\mathrm{d}P_0}(x)} - 1\right]\mathrm{d}P_0(x) \\
&= \int \sqrt{\frac{\mathrm{d}P_1}{\mathrm{d}P_0}(x)}\mathrm{d}P_0(x) - \int \mathrm{d}P_0(x) = \int \sqrt{\frac{\mathrm{d}P_1}{\mathrm{d}P_0}(x)}\mathrm{d}P_0(x) - 1 \\
&= \int \sqrt{\frac{\mathrm{d}P_1}{\mathrm{d}P_0}(x)}\frac{\mathrm{d}P_0}{\mathrm{d}\mu}(x)\mathrm{d}\mu(x) - 1 \\
&= \int \sqrt{\frac{\mathrm{d}P_1}{\mathrm{d}P_0}(x)\frac{\mathrm{d}P_0}{\mathrm{d}\mu}(x)\frac{\mathrm{d}P_0}{\mathrm{d}\mu}(x)}\mathrm{d}\mu(x) - 1 \\
&= \int \sqrt{\frac{\mathrm{d}P_1}{\mathrm{d}\mu}(x)\frac{\mathrm{d}P_0}{\mathrm{d}\mu}(x)}\mathrm{d}\mu(x) - 1 = \int \sqrt{\mathrm{d}P_1\mathrm{d}P_0} - 1 = \gamma - 1. \blacksquare
\end{aligned}$$

## 6.2 Differentiability in Quadratic Mean

Consider a random sample $X^{(n)} = (X_1, \ldots, X_n)$, where $X_i$'s are independent and identically $\mathcal{P} = \left\{p_{\underline{\theta}_0} \,|\, \underline{\theta}_0 \in \Theta \subseteq \mathbb{R}^k\right\}$ distributed. Under this perspective, the joint distribution of $X^{(n)}$ is trivially determined by the product of the marginal distributions, that is, $P^{(n)}_{\underline{\theta}_0} = \prod_{i=1}^n p^i_{\underline{\theta}_0}$.

Normally we characterize an experiment by the log-likelihood process. However, we do not know the variance of the log-likelihood process even though the log-likelihoods of each observation $x_i$ are independent and identically distributed. On the other hand, working with $z(\underline{\theta}) = \sqrt{\frac{\mathrm{d}P_{\underline{\theta}}}{\mathrm{d}P_{\underline{\theta}_0}}}$ makes things easier in this aspect for

$$\mathrm{Var}[z(\underline{\theta})] = 1 - \gamma^2(P_1, P_0) - \Upsilon_{P_1/P_0} < \infty.$$

In the following, we show the asymptotic relation of $z(\underline{\theta})$ and the log-likelihood process, such that weak convergence for $z(\underline{\theta})$ can be established.

**Definition.** The function $z(\cdot)$ is differentiable in quadratic mean (DQM) at the point $\underline{\theta}_0$ if there is a random vector $V(X)$ such that $E_{\underline{\theta}_0}[V(X)'V(X)] < \infty$ and

$$(i) \qquad \lim_{\underline{\theta}\to\underline{\theta}_0} E_{\underline{\theta}_0}\left[\frac{1}{\|\underline{\theta}-\underline{\theta}_0\|}\left(z(\underline{\theta}) - z(\underline{\theta}_0) - \frac{1}{2}(\underline{\theta}-\underline{\theta}_0)'V(X)\right)\right]^2 = 0$$

$$(ii) \qquad \lim_{\underline{\theta}\to\underline{\theta}_0} \frac{\Upsilon_{P_1/P_0}}{\|\underline{\theta}-\underline{\theta}_0\|^2} = 0.$$

**Remarks.** Three caveats are in order here. First, $z(\underline{\theta}_0) = 1$ by definition. Second, $V(X)$ plays the role of the gradient. Third, the classical requirements of Cramér imply the conditions for differentiability in quadratic mean, thus DQM constitutes a weaker assumption.

**Proposition (LAN Family).** Assume that $z(\cdot)$ is DQM at all points $\underline{\theta} \in \Theta$ and let $V_j = V(X_j)$. Then, for every bounded sequence $\underline{t}_n \in \mathbb{R}^k$, under $P^{(n)}_{\underline{\theta}}$, we have that

$$(i) \qquad \Lambda^{(n)}_{\underline{\theta}+n^{-1/2}\underline{t}_n,\underline{\theta}} = \underline{t}_n'\left(n^{-1/2}\sum_{j=1}^{n} V_j\right) - \frac{1}{2}\underline{t}_n' E_{\underline{\theta}}(V_1V_1')\underline{t}_n + o_P(1)$$

$$(ii) \qquad \underline{t}'\left(n^{-1/2}\sum_{j=1}^{n} V_j\right) \xrightarrow{\mathcal{L}} N\left(\underline{0}, \underline{t}'E_{\underline{\theta}}(V_1V_1')\underline{t}\right),\ \forall \underline{t}.$$

Note that $n^{-1/2}\sum_{j=1}^{n} V_j$ and $E_{\underline{\theta}}(V_1V_1')$ play the role of the central sequence and (noncentered) covariance matrix, respectively. The demonstration of this proposition relies on the two following lemmas.

**Lemma 1.** Let $\underline{\theta}_m - \underline{\theta}_0$ converges to zero when $m \to \infty$ in such a way that $\underline{\theta}_m - \underline{\theta}_0$ when normalized by $\|\underline{\theta}_m - \underline{\theta}_0\|$ goes to $u$. Then,

$$(i) \qquad E_{\underline{\theta}_0}\left[\frac{z(\underline{\theta}_m) - 1}{\|\underline{\theta}_m - \underline{\theta}_0\|}\right]^2 \longrightarrow \frac{1}{4}E_{\underline{\theta}_0}\left[u'VV'u\right]$$

$$(ii) \qquad \frac{1}{\|\underline{\theta}_m - \underline{\theta}_0\|^2}\,\mathcal{H}^2(p_{\underline{\theta}_m}, p_{\underline{\theta}_0}) \longrightarrow \frac{1}{8}E_{\underline{\theta}_0}\left[u'VV'u\right].$$

**Proof.** Using the second property of the Hellinger distance, yields

$$E_{\underline{\theta}_0}\left[\frac{z(\underline{\theta}_m)-1}{\|\underline{\theta}_m-\underline{\theta}_0\|}\right]^2 = \frac{2}{\|\underline{\theta}_m-\underline{\theta}_0\|^2}\,\mathcal{H}^2(p_{\underline{\theta}_m},p_{\underline{\theta}_0}) - \frac{1}{\|\underline{\theta}_m-\underline{\theta}_0\|^2}\,\Upsilon_{p_{\underline{\theta}_m}/p_{\underline{\theta}_0}}.$$

Because $z(\cdot)$ is DQM, the last term can be neglected. Using the first property of DQM, we have that

$$\frac{z(\underline{\theta}_m)-1}{\|\underline{\theta}_m-\underline{\theta}_0\|} - \frac{1}{2}\frac{(\underline{\theta}_m-\underline{\theta}_0)'}{\|\underline{\theta}_m-\underline{\theta}_0\|}V \overset{qm}{\longrightarrow} 0 \implies \frac{z(\underline{\theta}_m)-1}{\|\underline{\theta}_m-\underline{\theta}_0\|} \overset{qm}{\longrightarrow} \frac{1}{2}u'V,$$

which implies that

$$E_{\underline{\theta}_0}\left[\frac{z(\underline{\theta}_m)-1}{\|\underline{\theta}_m-\underline{\theta}_0\|}\right]^2 \longrightarrow \frac{1}{4}E_{\underline{\theta}_0}[u'VV'u].\blacksquare$$

**Lemma 2.** $E_{\underline{\theta}_0}[V]=0$, or equivalently $E_{\underline{\theta}_0}[u'V]=0$ for every normalized vector $u$.

**Proof.** The first property of the Hellinger distance posits that

$$E_{\underline{\theta}_0}\left[\frac{z(\underline{\theta}_m)-1}{\|\underline{\theta}_m-\underline{\theta}_0\|}\right] = -\frac{1}{\|\underline{\theta}_m-\underline{\theta}_0\|}\mathcal{H}^2(p_{\underline{\theta}_m},p_{\underline{\theta}_0}),$$

which converges in quadratic mean to $\frac{1}{2}E_{\underline{\theta}_0}[u'V]$ by the first condition of DQM. However,

$$\frac{1}{\|\underline{\theta}_m-\underline{\theta}_0\|}\mathcal{H}^2(p_{\underline{\theta}_m},p_{\underline{\theta}_0}) = \underbrace{\frac{1}{\|\underline{\theta}_m-\underline{\theta}_0\|^2}\mathcal{H}^2(p_{\underline{\theta}_m},p_{\underline{\theta}_0})}_{\to\frac{1}{8}E_{\underline{\theta}_0}[u'VV'u]}\underbrace{\|\underline{\theta}_m-\underline{\theta}_0\|}_{\to 0}\longrightarrow 0,$$

which implies that $E_{\underline{\theta}_0}[u'V]=0$. $\blacksquare$

**Proof of the Proposition.** Using the first condition of DQM, we can write that

$$z(\underline{\theta}_m)-1 = \frac{1}{2}(\underline{\theta}_m-\underline{\theta}_0)'V - \mathcal{H}^2(p_{\underline{\theta}_m},p_{\underline{\theta}_0}) + \|\underline{\theta}_m-\underline{\theta}_0\|R(\underline{\theta}_m),$$

where $R(\underline{\theta}_m)$ converges in quadratic mean to zero under $P^{(n)}_{\underline{\theta}_0}$ by construction. Then, for a random sample $X_j$ with $j=1,\dots,n$ and a local experiment $\underline{\theta}_m = \underline{\theta}_0 + n^{-1/2}\underline{t}_n$, we have that

$$\sum_{j=1}^{n}[z_j(\underline{\theta}_m)-1] = \frac{1}{2}\frac{\underline{t}_n'}{\sqrt{n}}\sum_{j=1}^{n}V_j - n\mathcal{H}^2(p_{\underline{\theta}_m},p_{\underline{\theta}_0}) + \frac{\|\underline{t}_n\|}{\sqrt{n}}\sum_{j=1}^{n}R_j(\underline{\theta}_m),$$

where the last term is negligible, since it goes to zero in quadratic mean. Using the second result of the first lemma, we have that

$$\sum_{j=1}^{n}[z_j(\underline{\theta}_m)-1]=\frac{1}{2}\frac{\underline{t}_n'}{\sqrt{n}}\sum_{j=1}^{n}V_j-\frac{1}{8}\underline{t}_n'E_{\underline{\theta}_0}[VV']\underline{t}_n+o_P(1).$$

Since $\log(1+x)\sim x-\frac{1}{2}x^2$, we have that

$$\begin{aligned}
\Lambda^{(n)}_{\underline{\theta}_m,\underline{\theta}_0} &= \sum_{j=1}^{n}\log\frac{\mathrm{d}P_{\underline{\theta}_m}}{\mathrm{d}P_{\underline{\theta}_0}}=2\sum_{j=1}^{n}\log z_j(\underline{\theta}_m)\\
&\sim 2\sum_{j=1}^{n}\left[(z_j(\underline{\theta}_m)-1)-\frac{1}{2}\left(z_j(\underline{\theta}_m)-1\right)^2\right]\\
&= 2\sum_{j=1}^{n}[z_j(\underline{\theta}_m)-1]-\sum_{j=1}^{n}[z_j(\underline{\theta}_m)-1]^2\\
&\stackrel{(a)}{=} 2\sum_{j=1}^{n}[z_j(\underline{\theta}_m)-1]-\frac{1}{4}E_{\underline{\theta}_0}\underline{t}_n'[VV']\underline{t}_n+o_P(1)\\
&\stackrel{(b)}{=} \frac{\underline{t}_n'}{\sqrt{n}}\sum_{j=1}^{n}V_j-\frac{1}{2}\underline{t}_n'E_{\underline{\theta}_0}[VV']\underline{t}_n+o_P(1).
\end{aligned}$$

Note that $(a)$ stems from the first lemma and (b) from the expression previously derived for $\sum_{j=1}^{n}[z_j(\underline{\theta}_m)-1]$. This proves the first part of the proposition. The second part is just a consequence of the first due to the LAN property. ■