# ECON 481-3
# LECTURE 8: LOCAL ASYMPTOTIC NORMALITY

Ivan A. Canay
Northwestern University

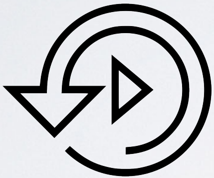
# Past & Future

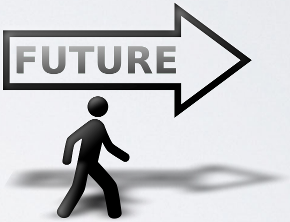
## So Far

- Absolute Continuity and LR
- Contiguity and Le Cam's 1st Lemma
- Le Cam's 3rd Lemma
- Wilcoxon Signed Ranked Test

## TODAY

- Local Asymptotic Normality
- Differentiability in Quadratic Mean
- Limit Distribution under Contig. Alt.
- Symmetric Location Model





# Statistical Experiment

- **Observed data**: a sample $X_1, \dots, X_n$ from a distribution $P_\theta$ on some measurable space $(\mathcal{X}, \mathcal{A})$ indexed by a parameter $\theta$ in $\Theta$ open.

- The sample is a single observation from the product $P_\theta^n$ of $n$ copies of $P_\theta$ and the statistical model (also called **statistical experiment**) is completely described as the collection of probability measures

$$\mathbf{P} = \{P_\theta^n : \theta \in \Theta\} .$$

- **Today**: study conditions under which a statistical experiment can be approximated by a **Gaussian experiment** after a suitable reparametrization.

- **Starting Point**: understand the properties of a Gaussian experiment.

# Normal Location Model

## Example (Normal Location Model)

Suppose $P_\theta = N(\theta, \sigma^2)$, where $\sigma^2$ is **known**. In this case,

$$\begin{aligned}\log\left[dP^n_{\theta_0+h/\sqrt{n}}/dP^n_{\theta_0}\right] &= -\frac{1}{2\sigma^2}\sum_{i=1}^n\left(X_i-\theta_0-\frac{h}{\sqrt{n}}\right)^2+\frac{1}{2\sigma^2}\sum_{i=1}^n(X_i-\theta_0)^2\\ &= \frac{1}{\sigma^2}\sum_{i=1}^n(X_i-\theta_0)\frac{h}{\sqrt{n}}-\frac{h^2}{2\sigma^2}\\ &= \frac{h}{\sigma^2}n^{1/2}(\bar{X}_n-\theta_0)-\frac{h^2}{2\sigma^2}\\ &= h\Delta_n-\frac{1}{2}h^2 I_{\theta_0}\ , \end{aligned} \tag{1}$$

where

$$\Delta_n = n^{1/2}(\bar{X}_n-\theta_0)/\sigma^2 \sim N(0, I_{\theta_0}) \text{ and } I_{\theta_0} = 1/\sigma^2\ ,$$

under $P_{\theta_0}$. It follows that

$$\log\left[dP^n_{\theta_0+h/\sqrt{n}}/dP^n_{\theta_0}\right] \sim N\left(-\frac{1}{2}\frac{h^2}{\sigma^2}, \frac{h^2}{\sigma^2}\right) \text{ under } P_{\theta_0}\ . \tag{2}$$

# Local Asymptotic Normality

- **Traditional regularity conditions** for maximum likelihood theory involve existence of two or three derivatives of the density function, together with domination assumptions to justify differentiation under integrals.
- Le Cam (1970) noted that such conditions are unnecessarily stringent. He showed that the traditional conditions can be replaced by a simple assumption of differentiability in quadratic mean (QMD).
- Le Cam showed that QMD implies a quadratic approximation property for the log-likelihoods known as **Local Asymptotic Normality (LAN)**.

## Definition (LAN)

The statistical experiment is called LAN at $\theta_0 \in \Theta$ if there exist a sequence of stochastic vectors $\Delta_{n,\theta_0}$ and a nonsingular $(k \times k)$ matrix $I_{\theta_0}$ such that $\Delta_{n,\theta_0} \xrightarrow{d} N(0, I_{\theta_0})$ under $P^n_{\theta_0}$ and such that,

$$\log\left[\frac{dP^n_{\theta_0+h/\sqrt{n}}}{dP^n_{\theta_0}}\right] = h\Delta_{n,\theta_0} - \frac{1}{2}h' I_{\theta_0} h + o_{P_{\theta_0}}(1)\ .$$

# Traditional Arguments

- Let $X_1, \ldots, X_n$ be i.i.d. from a density $p_\theta = dP_\theta/d\mu$ st $\theta \mapsto p_\theta$ is **twice differentiable**.

- Let $\ell_\theta(x) = \log(p_\theta(x))$, with derivatives $\dot{\ell}_\theta(x)$ and $\ddot{\ell}_\theta(x)$ with respect to $\theta$. Taylor expansion at $x$:

$$\log\left(p_{\theta+h/\sqrt{n}}(x)\right) = \log(p_\theta(x)) + \frac{h}{\sqrt{n}}\dot{\ell}_\theta(x) + \frac{h^2}{2n}\ddot{\ell}_\theta(x) + o_x(h^2/n) \, ,$$

where the subscript $x$ in the reminder term denotes its dependence on $x$. It then follows,

$$\sum_{i=1}^{n} \log\left(\frac{p_{\theta+h/\sqrt{n}}}{p_\theta}(X_i)\right) = h\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\dot{\ell}_\theta(X_i) - \frac{h^2}{2}\left(-\frac{1}{n}\sum_{i=1}^{n}\ddot{\ell}_\theta(X_i)\right) + o_{p_\theta}(1) \, ,$$

(assuming $n$ reminders $o_x(h^2/n)$ are negligible (i.e., $o_{p_\theta}(1)$)).

- **Expected score**: $E_\theta \dot{\ell}_\theta = 0$. **Fisher information**: $-E_\theta \ddot{\ell}_\theta = E_\theta \dot{\ell}_\theta^2 = I_\theta$.

- **First term**: can be rewritten as $h\Delta_{n,\theta}$, where $\Delta_{n,\theta} \xrightarrow{d} N(0, I_\theta)$ under $P_\theta$ by the CLT.
  **Second Term**: is asymptotically equivalent to $-\frac{h^2}{2} I_\theta$ by the LLN.

# QUESTIONS?

# Differentiability in Quadratic Mean

**Le Cam**: all you need to get LAN is a single condition that only involves a **first derivative**.

## Definition (QMD)

A model $\mathbf{P} = \{P_\theta : \theta \in \Theta\}$ is called **differentiable in quadratic mean** (or Hellinger differentiable or QMD) at $\theta$ if there exists a vector of measurable functions $\eta_\theta = (\eta_{\theta,1}, \dots, \eta_{\theta,k})'$ such that, as $h \to 0$,

$$\int \left[ \sqrt{p_{\theta+h}} - \sqrt{p_\theta} - \frac{1}{2} h' \eta_\theta \sqrt{p_\theta} \right]^2 d\mu = o(\|h\|^2) \, ,$$

where $p_\theta$ is the density of $P_\theta$ w.r.t. some measure $\mu$.

Usually, $\frac{1}{2} h \, \eta_\theta(x) \sqrt{p_\theta(x)}$ is the derivative of the map $h \mapsto \sqrt{p_{\theta+h}(x)}$ at $h = 0$ for (almost) every $x$:

$$\frac{\partial}{\partial\theta} \sqrt{p_\theta} = \frac{1}{2\sqrt{p_\theta}} \frac{\partial}{\partial\theta} p_\theta = \frac{1}{2} \left( \frac{\partial}{\partial\theta} \log p_\theta \right) \sqrt{p_\theta} \, ,$$

so the function $\eta_\theta(x)$ is $\dot{\ell}_\theta(x) = (\frac{\partial}{\partial\theta} \log p_\theta)$, the score function of the model.

# A Distribution that is not QMD

$$\int \left[ \sqrt{p_{\theta+h}} - \sqrt{p_\theta} - \frac{1}{2} h \, \eta_\theta \sqrt{p_\theta} \right]^2 d\mu = o(h^2) \; .$$

## Example (Uniform Distribution)

The family of uniform distributions on $[0, \theta]$ is nowhere differentiable in quadratic mean. The reason is that the support depends too much on the parameter.

# Local Asymptotic Normality

- **Le Cam**: QMD is exactly what we need to get LAN.

**Theorem**

*Suppose that $\Theta$ is an open subset of $\mathbf{R}^k$ and that the model $(P_\theta : \theta \in \Theta)$ is differentiable in quadratic mean at $\theta$. Then (1) $E_\theta \dot{\ell}_\theta = 0$, (2) the Fisher information matrix $I_\theta = E_\theta \dot{\ell}_\theta \dot{\ell}_\theta{}'$ exists, and (3)*

$$\log \prod_{i=1}^{n} \frac{p_{\theta+h/\sqrt{n}}}{p_\theta}(X_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h' \dot{\ell}_\theta(X_i) - \frac{1}{2} h' I_\theta h + o_{p_\theta}(1)\,.$$

- Okay, but QMD looks like hard to check. Are there simple **sufficient conditions** for QMD to hold?
- Usually one proceeds by showing differentiability of the map $\theta \mapsto \sqrt{p_\theta(x)}$ for almost every $x$ plus $\mu$-equi-integrability (which in turn will imply a convergence theorem for integrals).
- These conditions are stated in the following lemma.

# Useful Lemma

### Lemma

*For every* $\theta$ *in an open subset of* $\mathbf{R}^k$ *let* $p_\theta$ *be a* $\mu$*-probability density. Assume that the map* $\theta \mapsto s_\theta \equiv \sqrt{p_\theta(x)}$ *is continuously differentiable for every* $x$. *Assume the elements of the matrix*

$$I_\theta = \int (\dot{p}_\theta / p_\theta)(\dot{p}_\theta / p_\theta)' p_\theta d\mu$$

*are well defined and continuous in* $\theta$, *then the map* $\theta \mapsto \sqrt{p_\theta}$ *is differentiable in quadratic mean with* $\eta_\theta = \dot{p}_\theta / p_\theta$.

- We will prove this result in 2 steps
- **Step 1**. Argue that $\dot{p}_\theta$ exists and find expression for $\dot{s}_\theta$
- **Step 2**. Invoke Vitali's theorem to show QMD.

# Step 1: $\dot{p}_\theta$ exists and find $\dot{s}_\theta$

$$s_\theta \equiv \sqrt{p_\theta(x)} \quad \text{and} \quad \eta_\theta = \dot{p}_\theta / p_\theta$$

# Step 2: use Vitali's theorem

**Theorem (Vitali's)**

*If* ① $f_n(x) \to f(x)$ *for* $\mu$*-almost every* $x$ *(both real-valued measurable functions) and*

$$② \limsup_{n\to\infty} \int f_n^2(x)d\mu(x) \leqslant \int f^2(x)d\mu(x) < \infty ,$$

*it follows that*

$$\lim_{n\to\infty} \int |f_n(x) - f(x)|^2 d\mu(x) = 0 .$$

- Need to check the **two conditions**.
- **First**: $\theta \mapsto s_\theta = \sqrt{p_\theta}$ is continuously differentiable, so

$$f_h(x) = \frac{1}{h}\Big(s_{\theta+h}(x) - s_\theta(x)\Big) \to f(x) = \dot{s}_\theta(x) \;\text{ as }\; h \to 0 . \tag{3}$$

- **Second**: to prove ② we need to do some work.

# Step 2: use Vitali's theorem

(2) $\limsup_{h\to 0} \int f_h^2(x) d\mu(x) \leqslant \int f^2(x) d\mu(x) < \infty$ with $f_h(x) = \frac{1}{h}\Big(s_{\theta+h}(x) - s_\theta(x)\Big)$ and $f(x) = \dot{s}_\theta(x)$ .

# Step 2: use Vitali's theorem

(2) $\limsup_{h\to 0} \int f_h^2(x)d\mu(x) \leqslant \int f^2(x)d\mu(x) < \infty$ with $f_h(x) = \frac{1}{h}\Big(s_{\theta+h}(x) - s_\theta(x)\Big)$ and $f(x) = \dot{s}_\theta(x)$ .

# Complete the proof

**Final Step**: Apply Vitali's Theorem with $f_h(x) = \frac{1}{h}(s_{\theta+h}(x) - s_\theta(x))$ and $f(x) = \dot{s}_\theta(x)$,

$$\lim_{h \to 0} \int \left[ \frac{1}{h}(s_{\theta+h}(x) - s_\theta(x)) - \dot{s}_\theta(x) \right]^2 d\mu = 0 \,.$$

Replacing $s_\theta = \sqrt{p_\theta}$ and $\dot{s}_\theta(x) = \frac{1}{2}\eta_\theta(x)\sqrt{p_\theta}$ completes the proof.

# Example: Location Model

## Example (Location Model)

Let $\{p_\theta(x) = f(x-\theta) : \theta \in \Theta\}$ be a location model, where $f(\cdot)$ is continuously differentiable. Let

$$\dot{\ell}_\theta(x) = \frac{\dot{p}_\theta}{p_\theta} = \frac{-f'(x-\theta)}{f(x-\theta)}$$

if $f(x-\theta) > 0$ and $f'(x-\theta)$ exists and zero otherwise. Assume

$$I_0 = \int \dot{\ell}_0^2(x) f(x) dx < \infty \, .$$

Since in this model the Fisher information is equal to $I_0$ for all $\theta$ (just set $y = x - \theta$ in the integral for $I_\theta$), and thus continuous in $\theta$, it follows that the family is QMD.

# QUESTIONS?

# Limit Dist. under Contiguous Alternatives

- LAN: convenient tool in the study of the behavior of statistics under **contiguous alternatives**.
- **LAN**. $dP^n_{\theta+h/\sqrt{n}}$ and $dP^n_\theta$ are mutually contiguous:

$$\log L_n = \log\left(\frac{dP^n_{\theta+h/\sqrt{n}}}{dP^n_\theta}\right) \xrightarrow{d} N\left(-\frac{1}{2}h' I_\theta h\ ,\ h' I_\theta h\right) \text{ under } P_\theta\ .$$

- **LAN + Le Cam's third lemma**: limit distributions of statistics under the parameters $\theta + h/\sqrt{n}$ from limit behavior under $\theta$.
- **General scheme**: (1) many sequences of statistics $T_n$ allow an approximation of the type,

$$\sqrt{n}(T_n - \mu_\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_\theta(X_i) + o_{P_\theta}(1)\ .$$

(2) By the LAN theorem, the sequence of log likelihood ratios can be approximated by

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n h'\dot{\ell}_\theta(X_i) - \frac{1}{2}h' I_\theta h + o_{P_\theta}(1)\ .$$

# Limit Dist. under Contiguous Alternatives

$$\sqrt{n}(T_n - \mu_\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\theta(X_i) + o_{p_\theta}(1) \text{ and } \log L_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h' \dot{\ell}_\theta(X_i) - \frac{1}{2} h' I_\theta h + o_{p_\theta}(1)$$

The sequence of **joint** averages

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} \psi_\theta(X_i) \\ h\dot{\ell}_\theta(X_i) \end{pmatrix}$$

is asymptotically multivariate normal under $P_\theta$ by the CLT

$$\begin{pmatrix} \sqrt{n}(T_n - \mu_\theta) \\ \log L_n \end{pmatrix} \xrightarrow{d} N\left( \begin{pmatrix} 0 \\ -1/2 h' I_\theta h \end{pmatrix}, \begin{pmatrix} E_\theta \psi_\theta \psi_\theta' & E_\theta \psi_\theta h' \dot{\ell}_\theta \\ E_\theta \psi_\theta' h \dot{\ell}_\theta & h' I_\theta h \end{pmatrix} \right).$$

By Le Cam's third lemma we get the limit distribution of $\sqrt{n}(T_n - \mu_\theta)$ under $\theta + h/\sqrt{n}$ which depends on

$$\tau = E_\theta \left[ \psi_\theta h' \dot{\ell}_\theta \right].$$

# Symmetric Location Model

- **SML**: $P_\theta$ is the distribution with density $f(x-\theta)$ on the real line, symmetric about $\theta$.
- **Data**: $X_1, \ldots, X_n$ from $P_\theta$ and wish to test the null $H_0 : \theta = 0$.
- **New assumption**: $\mathbf{P} = \{P_\theta : \theta \in \Theta\}$ is differentiable in quadratic mean. We know $p_\theta(x) = f(x-\theta)$ is QMD at $\theta = 0$ if $f(\cdot)$ is absolutely continuous with finite Fisher information,

$$I_0 = \int \dot{\ell}_0^2(x) f(x) dx \, ,$$

where

$$\dot{\ell}_\theta(x) = \frac{-f'(x-\theta)}{f(x-\theta)} \, .$$

- By the LAN Theorem

$$\log L_n = \log(dP_{\theta_n}/dP_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} -h \frac{f'(X_i)}{f(X_i)} - \frac{1}{2} h^2 I_0 + o_p(1) \, .$$

- We can now easily derive the local asymptotic power of many tests by using Le Cam's third lemma.

## T-TEST

$$\psi_\theta(X_i) = X_i/\sigma \quad \text{and} \quad h\dot{\ell}_\theta(x) = h\frac{-f'(x-\theta)}{f(x-\theta)} \, .$$

# SIGN-TEST

$$\psi_\theta(X_i) = I\{X_i > 0\} - \frac{1}{2} \quad \text{and} \quad h\dot{\ell}_\theta(x) = h\frac{-f'(x-\theta)}{f(x-\theta)} \ .$$

# Wilcoxon Signed-rank Test

$$\psi_\theta(X_i) = G(|X_i|)\,\text{sign}(X_i) \quad \text{and} \quad h\dot{\ell}_\theta(x) = h\frac{-f'(x-\theta)}{f(x-\theta)}\ .$$

# THE END!