Reminder:

Let $(\Omega, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$ be a statistical model, dominated by μ, $\Theta \subseteq \mathbb{R}^k$, $k \in \mathbb{N}$.

$\ell(\vartheta, x) := \dfrac{d\mathbb{P}_\vartheta}{d\mu}(x)$	Likelihood function
$\ln(\ell(\vartheta, x))$	Log-Likelihood function
$x \mapsto \dfrac{d}{d\vartheta} \ln(\ell(\vartheta, x)) \Big\vert_{\vartheta=\vartheta_0} =: \dot{\ell}(\cdot, \vartheta_0)$	Score function
$I(\vartheta_0) = \mathbb{E}_{\vartheta_0}[\dot{\ell}(\cdot, \vartheta_0)(\dot{\ell}(\cdot, \vartheta_0))^t]$	Fisher-Information (matrix) in ϑ_0.

$\ell(\vartheta, x)$ $L_2(\mathbb{P}_{\vartheta_0})$-differentiable in $\vartheta_0 \in \Theta$

$\Rightarrow$ (a) $\mathbb{E}_{\vartheta_0}[\dot{\ell}(\cdot, \vartheta_0)] = 0$

(b) $I(\vartheta_0) = \mathbb{E}_{\vartheta_0}[(\dot{\ell}(\cdot, \vartheta_0))^2]$

$= \mathrm{Var}_{\vartheta_0}(\dot{\ell}(\cdot, \vartheta_0))$

$= -\mathbb{E}_{\vartheta_0}\left[\dfrac{\partial^2}{\partial \vartheta^2} \ln(\ell(\vartheta, \cdot))\Big\vert_{\vartheta=\vartheta_0}\right].$

$\boxed{\text{In the special case of } k=1,}$

<u>Goal:</u> Asymptotic distribution of the MLE

$\hat{\vartheta}$ MLE for ϑ, if

$$\ell(\hat{\vartheta}(x), x) = \sup_{\tilde{\vartheta} \in \Theta} \ell(\tilde{\vartheta}, x)$$

for $\mathbb{P}_\vartheta$ - almost all $x \in \Omega$ and all $\vartheta \in \Theta$.

<u>Here:</u> Product experiments $(\mathbb{P}_\vartheta^n)_{\vartheta \in \Theta \subseteq \mathbb{R}^k}$

$\ell(\vartheta, x_i)$ one-dimensional likelihood function

$$\ell(\vartheta, x) = \prod_{i=1}^{n} \ell(\vartheta, x_i) \quad \text{joint likelihood function.}$$

$$\ln(\ell(\vartheta, x)) = \sum_{i=1}^{n} \log(\ell(\vartheta, x_i)) \quad \text{joint log-likelihood-function.}$$

<u>First, we consider $k = 1$.</u>

<u>General assumptions:</u>

(A1) Product experiment $(\mathbb{P}_\vartheta^n)_{\vartheta \in \Theta \subseteq \mathbb{R}^1}$

(A2) Θ contains an open set $\mathcal{U}$ of which the true parameter value ϑ_0 is an interior point.

(A3) $\forall \vartheta \neq \vartheta_0 : \mathbb{P}_\vartheta \neq \mathbb{P}_{\vartheta_0}$ (Identifiability)

Lemma (Theorem 3.2 in Lehmann & Casella, 1998)

Under (A1) and (A3),

$$\mathbb{P}^n_{\vartheta_0}\left(\ell(\vartheta_0, X) > \ell(\vartheta, X) \right) \xrightarrow[(n\to\infty)]{} 1$$

for any fixed $\vartheta \neq \vartheta_0$.

Proof: The inequality inside $\mathbb{P}^n_{\vartheta_0}(\dots)$ is equivalent to

$$n^{-1} \sum_{i=1}^{n} \log\left\{ \frac{\ell(\vartheta, X_i)}{\ell(\vartheta_0, X_i)} \right\} < 0. \qquad (*)$$

Law of large numbers $\Rightarrow$ LHS of $(*)$ tends in probability toward

$$n^{-1}\, \mathbb{E}_{\vartheta_0}\left[\log\left\{ \frac{\ell(\vartheta, X)}{\ell(\vartheta_0, X)} \right\} \right].$$

Since $u \mapsto -\log(u)$ is strictly convex, Jensen's inequality shows that

$$\mathbb{E}_{\vartheta_0}\left[\log\left\{ \frac{\ell(\vartheta, X)}{\ell(\vartheta_0, X)} \right\} \right]$$

$$< \log \mathbb{E}_{\vartheta_0}\left[\frac{\ell(\vartheta, X)}{\ell(\vartheta_0, X)} \right] = 0,$$

and the result follows.

Theorem (thm. 3.7 in Lehmann & Casella, 1998):

Let (A1) - (A3) be valid.

Assume that $\ell(\vartheta, x)$ is differentiable w.r.t. ϑ on $\mathcal{M}$, for almost all $x \in \Omega$.

Then, with probability 1 as $n \to \infty$, the likelihood equation

$$\dot{\ell}(x, \vartheta) = 0$$

$$\Leftrightarrow \sum_{i=1}^{n} \frac{\partial / \partial\vartheta \; \ell(\vartheta, x_i)}{\ell(\vartheta, x_i)} = 0 \qquad (**)$$

has a root $\hat{\vartheta}_n \equiv \hat{\vartheta}_n(x)$ such that $\hat{\vartheta}_n(X)$ tends to ϑ_0 in probability

(Consistency of $\hat{\vartheta}_n$).

Proof: Let $a > 0$ be small enough so that $(\vartheta_0 - a, \vartheta_0 + a) \subset \mathcal{M}$, and let

$$S_n(a) = \{x \in \Omega : \ln(\ell(\vartheta_0, x)) > \ln(\ell(\vartheta_0 \pm a, x))\}$$

For any $x \in S_n(a)$, there exists a value $\hat{\vartheta}_n(a) \in (\vartheta_0 - a, \vartheta_0 + a)$ at which $\ln(\ell(\cdot, x))$ has a local maximum, such that $\dot{\ell}(x, \hat{\vartheta}_n(a)) = 0$.

Moreover, by the preceding lemma,

$$\mathbb{P}^n_{\vartheta_0}(S_n(a)) \to 1 \quad \text{for any such } a$$

$\Rightarrow \quad \exists \{a_n\}_n \searrow 0$ such that $\mathbb{P}^n_{\vartheta_0}(S_n(a_n)) \xrightarrow[(n\to\infty)]{} 1$.

Let $\hat{\vartheta}^*_n = \hat{\vartheta}_n(a_n)$ for $x \in S_n(a_n)$ and equal to an arbitrary constant otherwise.

Then
$$\mathbb{P}^n_{\vartheta_0}(\dot{\ell}(X, \hat{\vartheta}^*_n) = 0) \geq \mathbb{P}^n_{\vartheta_0}(S_n(a_n)) \to 1$$

$\Rightarrow$ For any fixed $a > 0$ and n sufficiently large, we get

$$\mathbb{P}^n_{\vartheta_0}(|\hat{\vartheta}^*_n - \vartheta_0| < a)$$
$$\geq \mathbb{P}^n_{\vartheta_0}(|\hat{\vartheta}^*_n - \vartheta_0| < a_n)$$
$$\geq \mathbb{P}^n_{\vartheta_0}(S_n(a_n)) \to 1. \quad \blacksquare$$

<u>Remark:</u> Under the assumptions of the Theorem, if the probability for the likelihood equation (**) having multiple roots tends to zero $(n \to \infty)$, then the sequence $\{\hat{\vartheta}_n\}$ is a

consistent sequence of estimators of ϑ.

If, in addition, the parameter space is an open (not necessarily finite) interval, then with probability tending to 1, $\hat{\vartheta}_n$ is the MLE.

<u>We therefore introduce two additional assumptions:</u>

(CA4) For $n \to \infty$, the probability of the likelihood equation (**) having multiple roots, tends to zero.

(CA5) Θ is an open (not necessarily finite) interval.

<u>Theorem (Thm. 3.10 in Lehmann & Casella, 1998):</u>

Assume (CA1) − (CA5), and let the one-dimensional likelihood function $\ell(\vartheta, x_1)$ be three times continuously differentiable wrt. $\vartheta \in \Theta$ and $L_2(\mathbb{P}_{\vartheta_0})$-differentiable in any ϑ_0. Assume that $0 < I(\vartheta) < \infty$ ($I(\vartheta)$ one-dim. Fisher information!) and that for any given $\vartheta_0 \in \Theta$ there exists a positive number c and a function $M(x)$ (both of which may depend on ϑ_0) such that

$$\left| \frac{\partial^3}{\partial \vartheta^3} \ln(\ell(\vartheta, x_1)) \right| \leq M(x) \qquad \text{(Condition M)}$$

for all x in the support of $\mathbb{P}_{\vartheta_0}$,

$\vartheta_0 - c < \vartheta < \vartheta_0 + c$ with $\mathbb{E}_{\vartheta_0}[M(\cdot)] < \infty$.

Then the MLE is asymptotically normally distributed and asymptotically Cramér-Rao efficient, i.e.,

$$\sqrt{n}\,(\hat{\vartheta}_n(X) - \vartheta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I^{-1}(\vartheta_0)).$$

under ϑ_0 for $n \to \infty$.

<u>Proof:</u> For any fixed $x \in \Omega$, we expand $\dot{\ell}(x, \hat{\vartheta}_n)$ about ϑ_0:

$$\dot{\ell}(x, \hat{\vartheta}_n) = \dot{\ell}(x, \vartheta_0) + (\hat{\vartheta}_n - \vartheta_0)\,\ddot{\ell}(x, \vartheta_0) + 1/2\,(\hat{\vartheta}_n - \vartheta_0)^2\,\dddot{\ell}(x, \vartheta_n^*), \quad [***]$$

where ϑ_n^* lies between ϑ_0 and $\hat{\vartheta}_n$.

Since the LHS of [***] equals 0, we have

$$\sqrt{n}\,(\hat{\vartheta}_n - \vartheta_0) = \frac{n^{-1/2}\,\dot{\ell}(x, \vartheta_0)}{-n^{-1}\,\ddot{\ell}(x, \vartheta_0) - 1/2\, n^{-1}(\hat{\vartheta}_n - \vartheta_0)\,\dddot{\ell}(x, \vartheta_n^*)}.$$

We show that under ϑ_0

[i] $n^{-1/2}\,\dot{\ell}(X, \vartheta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I(\vartheta_0))$, $n \to \infty$

[ii] $-n^{-1}\,\ddot{\ell}(X, \vartheta_0) \xrightarrow[(n\to\infty)]{} I(\vartheta_0)$ in probability

[iii] $n^{-1}\,\dddot{\ell}(X, \vartheta_n^*)$ is bounded in probability.

The assertion then follows from the Cramér–Slutzky theorem.

To show [i], we notice that

$$n^{-1/2}\, \dot{\ell}(X, \vartheta_0) = \sqrt{n}\, \frac{1}{n} \sum_{i=1}^{n} \dot{\ell}(X_i, \vartheta_0)$$

$$= \sqrt{n}\, \overline{Y}_n \quad , \text{where}$$

$$\overline{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i \quad \text{with} \quad Y_i = \dot{\ell}(X_i, \vartheta_0).$$

Since $\forall\, 1 \le i \le n: \ \mathbb{E}_{\vartheta_0}[Y_i] = 0$ and

$$\mathrm{Var}_{\vartheta_0}(Y_i) = I(\vartheta_0),$$

the CLT yields the assertion under [i].

For the proof of [ii], we write

$$-n^{-1}\, \ddot{\ell}(X, \vartheta_0) =$$

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\left[\frac{\partial}{\partial \vartheta}\, \ell(\vartheta, X_i)\big|_{\vartheta_0}\right]^2 - \ell(\vartheta_0, X_i) \cdot \frac{\partial^2}{\partial \vartheta^2}\, \ell(\vartheta, X_i)\big|_{\vartheta_0}}{\ell^2(\vartheta_0, X_i)}$$

By the law of large numbers, this tends in probability to

$$\mathbb{E}_{\vartheta_0}\left[\frac{\{\partial/\partial\vartheta\, \ell(\vartheta, x_i)|_{\vartheta_0}\}^2}{\ell^2(\vartheta_0, x_i)}\right] - \mathbb{E}_{\vartheta_0}\left[\frac{\frac{\partial^2}{\partial\vartheta^2}\, \ell(\vartheta, x_i)|_{\vartheta_0}}{\ell(\vartheta_0, x_i)}\right]$$

$$= I(\vartheta_0) - 0 = I(\vartheta_0).$$

Finally, [iii] is established by noting

$$n^{-1}\, \dddot{\ell}(x, \vartheta) = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial^3}{\partial\vartheta^3} \ln(\ell(\vartheta, x_i)).$$

Therefore, due to (Condition M)

$$|n^{-1}\, \dddot{\ell}(\vartheta_n^*)| \le n^{-1}\sum_{i=1}^{n} M(x_i)$$ with probability tending to 1. The RHS tends in probability to $E_{\vartheta_0}[M(\cdot)]$, completing the proof. ☑

<u>Remark:</u> The latter theorem can be proved under weaker differentiability conditions, too.

For $k > 1$, calculations become more cumbersome, but an analogous result nevertheless holds.

We only repeat the result.

$\longrightarrow$ Theorem 3.22 Skript.