The application here is that if \( \Theta \in \mathbb{R}^k \), for a nicely behaved "MLE" \( \hat{\Theta}_n \) with

\[
I_n \xrightarrow{k \times k} \sqrt{n} (\hat{\Theta}_n - \Theta) \xrightarrow{d} \mathcal{N}_k(0, I^{-1}(\Theta))
\]

means

\[
\sqrt{n} (\hat{\Theta}_n - \Theta)' (n I^{-1}(\Theta))^{-1} \sqrt{n} (\hat{\Theta}_n - \Theta) \xrightarrow{d} \chi^2_k
\]

\((\hat{\Theta}_n - \Theta)' n I^{-1}(\Theta) (\hat{\Theta}_n - \Theta)\)

and so an (unusable) confidence set for \( \Theta \) is

\[
\{ \Theta | (\hat{\Theta}_n - \Theta)' (n I^{-1}(\Theta)) (\hat{\Theta}_n - \Theta) < c \}
\]

This involves the unknown \( I_1 \) - two practical fixes are then

1) Use of the "Expected FI" - i.e. replace \( I_1(\Theta) \) by \( I_n(\hat{\Theta}_n) \)

2) Use of the "Observed FI" - i.e. Think with

\[
H_n(\Theta) = \left( \frac{\partial^2 \ell_n(\Theta)}{\partial \Theta_i \partial \Theta_j} \right)_{k \times k}
\]

the Hessian matrix for the log likelihood

that \( H_n(\Theta) \) is a sum of iid terms and it's
plausible that
\[-\frac{1}{n} H_n(\theta) \overset{P_0}{\to} \mathcal{I}_1(\theta)\]
and then further that
\[-\frac{1}{n} H_n(\hat{\theta}_n) \overset{P_0}{\to} \mathcal{I}_1(\theta)\]
so that one might well replace \(n\mathcal{I}_1(\theta)\) with
\[n (-\frac{1}{n} H_n(\hat{\theta}_n)) = -H_n(\hat{\theta}_n)\]
to get the approximate confidence set for \(\theta\)
\[
\{\theta \mid (\hat{\theta}_n - \theta)(-H_n(\hat{\theta}_n))(\hat{\theta}_n - \theta) < c^2\}
\]

In this multi-parameter context I might have in mind inference for only some sub-vector of \(\theta\), say \(\theta_i\) of dimension \(l < n - l\) suppose
\[
\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_k \end{pmatrix}_{k \times 1} \quad \text{and} \quad \hat{\theta}_n = \begin{pmatrix} \hat{\theta}_{i1} \\ \hat{\theta}_{i2} \\ \vdots \\ \hat{\theta}_{ik} \end{pmatrix}_{k-l \times 1}
\]
and my primary interest is in \(\theta_i\) - I need to be careful as I think about using MVN stuff to do inference here.

Covariance matrix for approximating
\[\text{as for } \hat{\theta}_n\]
\[
= \mathcal{I}^{-1}(\theta) = (n\mathcal{I}_1(\theta))^{-1}
\]
\[
\approx (n\mathcal{I}_1(\hat{\theta}_n))^{-1}
\]
\[
\approx (-H_n(\hat{\theta}_n))^{-1}
\]
Then, the covariance matrix for the approximating dens for $\hat{\Theta}_n$ is

upper left $L \times L$ block of $\mathbb{I}^{-1}(\Theta) = (n \mathbb{I}_L(\Theta))^{-1}$

upper left $L \times L$ block of $(n \mathbb{I}_L(\hat{\Theta}_n))^{-1}$

upper left $L \times L$ block of $(-n \mathbb{I}_L(\hat{\Theta}_n))^{-1}$

and these blocks are NOT in general the inverses of the upper left blocks of the matrices inside the ( ) above.

The folklore is that in all this using observed rather than expected $\mathbb{I}$ does a better job of producing actual coverage probability close to nominal (based on the limiting dens).

Something even better in this regard is based on a different use of limiting dens for "MLE's" -

**Theorem 3** (of ML handout)

Under appropriate conditions in an iid model, if $\{\mathcal{S}_n(X)\}$ is consistent for $\Theta$ at $\Theta_0$, and with $\Theta_0$ a root of the likelihood equation

$$L_n(\Theta) = 0$$

Then

$$2 \left( l_n(\hat{\Theta}_n) - l_n(\Theta_0) \right) \xrightarrow{\text{d}} \chi^2_1$$
Note that the expression

\[ 2 \left( \frac{\ln(\delta_n(x)) - \ln(\theta_0)}{\Delta} \right) \]

is "sort of"

\[ 2 \log \left( \frac{\sup_\theta \ln(\theta)}{\ln(\theta_0)} \right) \]

\( \text{likelihood ratio statistic for testing } H_0: \theta = \theta_0 \)
\( \text{vs } H_1: \theta \neq \theta_0 \)

So Theorem 9 gives me a way to set critical values for LRTs of point null hypotheses for large \( n \) - For more importantly, it also gives me a way to make confidence sets for \( \theta \) by inverting tests - This amounts to doing the following - If \( c \) is the upper \( \alpha \) pt of \( \chi^2 \)

\[ P_{\theta_0} \left[ 2 \left( \ln(\delta_n(x)) - \ln(\theta_0) \right) < c \right] \approx 1 - \alpha \]

\[ P_{\theta_0} \left[ \ln(\delta_n(x)) - \frac{1}{2} c < \ln(\theta_0) \right] \]

That is, the set of \( \theta \)'s with \( \ln(\theta) \) no more than \( \frac{1}{2} c \) below the maximum of the log likelihood functions as a confidence set for \( \theta \). That is
with data $X = x$

$$\{ \theta | l_n(\theta) > h_n(\delta_n(x)) - \frac{1}{2}c^2 \}$$

can be used as a confidence set for $\theta$

Thus is an important multi-parameter extension of this, but before doing that I want to give you some idea of how this comes about.

Argument for Thm 3 of the handout:

Again write $\hat{\theta}_n$ instead of $\delta_n(x)$ (recall BTW that the argument for asymptotic normality for $\hat{\theta}_n$ is done by expanding $l_n(\cdot)$ in a Taylor series around $\theta_0$) - thus expand $l_n(\cdot)$ around $\hat{\theta}_n$

$$l_n(\theta_0) = l_n(\hat{\theta}_n) + (\theta_0 - \hat{\theta}_n) l'_n(\hat{\theta}_n) + \frac{1}{2} (\theta_0 - \hat{\theta}_n)^2 l''_n(\hat{\theta}_n)$$

$$\quad + \frac{1}{6} (\theta_0 - \hat{\theta}_n)^3 l'''_n(\theta')$$

for some $\theta'$ between $\hat{\theta}_n$ and $\theta_0$ - so

$$2(l_n(\hat{\theta}_n) - l_n(\theta_0)) = \underbrace{- (\theta_0 - \hat{\theta}_n) l'_n(\hat{\theta}_n)}_{A_n^*}$$

$$- \underbrace{2(\frac{1}{2})(\theta_0 - \hat{\theta}_n)^2 l''_n(\hat{\theta}_n)}_{B_n^*}$$

$$- \underbrace{\frac{1}{3} (\theta_0 - \hat{\theta}_n)^3 l'''(\theta')}_{C_n^*}$$
\( \hat{\Theta}_n \) an "MLE" makes \( L_n(\hat{\Theta}_n) = 0 \), makes \( A^*_n = 0 \)

\[
B^*_n = (\Theta_0 - \hat{\Theta}_n)^2 (-L''(\hat{\Theta}_n)) \\
= (\sqrt{n}(\Theta_0 - \hat{\Theta}_n))^{-2} (\frac{-1}{n} L''(\hat{\Theta}_n)) \\
\xrightarrow{\mathbb{P}} \mathcal{N}(0, I^{-1}_1(\Theta_0)) \quad \text{not too surprising if this converges to } I_1(\Theta_0)
\]

So \( B^*_n \xrightarrow{\mathbb{P}} \left( \sqrt{I_1(\Theta_0)} \cdot \mathcal{N}(0, I^{-1}_1(\Theta_0)) \right)^2 \)

(since the square of a standard normal is \( \chi^2_1 \))

And sure enough, standard regularity conditions are set up (just as for the proof of asymptotic normality of "MLEs") to produce

\[ C^*_n \xrightarrow{\mathbb{P}} 0 \]

There is an important multivariate version of this \( \chi^2 \) limit for a LRT statistic - That goes as follows -

\[
\Theta = \begin{pmatrix} \Theta_1 \\ (k-1) \times 1 \end{pmatrix} \\
\hat{\Theta}_n \text{ an "MLE" similarly partitioned}
\]

Suppose that for each \( \Theta_1 \in \mathbb{R}^d \)

\( \Theta^*_n(\Theta_1) \in \mathbb{R}^{k-1} \) is a "maximizer" of \( L_n(\Theta_1, \cdot) \)

over choices of \( \Theta_2 \)
Then
\[ l^*_n(\theta_1) = \max_{\theta_2} L_n(\theta_1, \theta_2, \theta_2^*) \]
\[ = \max_{\theta_2} L_n(\theta_1, \theta_2) \]
is called the "profile likelihood" for \( \theta_1 \), and can essentially be used like a likelihood to do inference for \( \theta_1 \). There is, e.g. The large sample result

\[ \text{``The'' under appropriate regularity conditions in an iid model, if } \Theta \subset \mathbb{R}^k \]
\[ Z \left( l_n(\hat{\theta}_n) - l^*_n(\theta_{10}) \right) \xrightarrow{\text{d}} \chi^2 \]
\[ \text{max of } l_n(\theta) \]
\[ \text{is also the max of } l^*_n(\theta_1) \]
\[ \text{any } \theta \text{ with } \theta_1 = \theta_0 \]

I can then use this to set critical values for LRT's
\[ \text{at } H_0: \theta_1 = \theta_{10} \text{ and if } c^* \text{ is the upper } \alpha \text{ pt of } \chi^2 \]
\[ \begin{align*}
P_{\theta_0} \left[ s(\hat{\theta}_n) - l^*_n(\theta_{10}) < c^* \right] & \approx 1 - \alpha \\
\text{Notes:} \quad & \\
\text{P}_{\theta_0} \left[ l_n(\hat{\theta}_n) - \frac{1}{2} c^* < l^*_n(\theta_{10}) \right] & \\
\text{So that with data } X = x \text{ and MLE } \hat{\theta}_n(\bar{x}), \text{ }
\{ \theta_1 \mid l_n^*(\theta_1) > l_n(\hat{\theta}_n(\bar{x})) - \frac{1}{2} c^* \}
\text{can be used as an approximate } (1-\alpha)x100\% \text{ confidence set for } \theta_1