1 Maximum likelihood estimator is efficient

Proof of the claim:

Maximum likelihood estimator is efficient - if an efficient estimator exists.

From the proof of the Cramer-Rao bound it follows, that
$\Rightarrow$ the equality sign in the Cauchy-Schwarz inequality (*) holds, if

$$g_{(\underline{\mathbf{D}})} \sim h_{(\underline{\mathbf{D}})}$$

$\Rightarrow$ using the definitions below eq. (*) one obtains

$$\begin{aligned}
\underbrace{\underline{\mathbf{a}}^T \overbrace{(\underline{\hat{\mathbf{w}}} - \underline{\mathbf{w}})}^{\text{efficient estimator}}}_{g_{(\underline{\mathbf{w}})}} &= \gamma_{(\underline{\mathbf{w}}^*)} \underbrace{\left(\frac{\partial \ln P}{\partial \underline{\mathbf{w}}}\right)^T \underline{\mathbf{b}}}_{h_{(\underline{\mathbf{D}})}} \\
&= \gamma \left(\tfrac{\partial \ln P}{\partial \underline{\mathbf{w}}}\right)^T \underbrace{\underline{\mathbf{M}}^{-1}\underline{\mathbf{a}}}_{\text{particular cl??? of } \underline{\mathbf{b}}}
\end{aligned} \tag{1}$$

since $\underline{\mathbf{a}}$ is an arbitrary vector, we obtain

$$\frac{\partial \ln P}{\partial \underline{\mathbf{w}}} = \frac{1}{\gamma} \underline{\mathbf{M}}(\underline{\hat{\mathbf{w}}} - \underline{\mathbf{w}}) \tag{2}$$

calculation of γ:

$$\frac{\partial \ln P}{\partial \mathrm{w}_j} = \sum_k \frac{\mathrm{M}_{jk}}{\gamma}(\hat{\mathrm{w}}_k - \mathrm{w}_k) \tag{3}$$

$$\frac{\partial^2 \ln P}{\partial \mathrm{w}_i \partial \mathrm{w}_j} = \sum_k \left\{ -\frac{\mathrm{M}_{jk}}{\gamma} \underbrace{\delta_{ik}}_{\mathrm{w}_k} + (\hat{\mathrm{w}}_k - \mathrm{w}_k) \overbrace{\frac{\partial}{\partial \mathrm{w}_i}}^{\gamma} \left(\frac{\mathrm{M}_{jk}}{\gamma}\right) \right\} \tag{4}$$

$$\begin{aligned}
\mathrm{M}_{ij} &= \left\langle \tfrac{\partial^2 \ln P}{\partial \mathrm{w}_i \partial \mathrm{w}_j} \right\rangle_p \Big|_{\underline{\mathbf{w}}^*} \\
&= \tfrac{\mathrm{M}_{ji}}{\gamma} \text{ because } < \hat{\mathrm{w}}_k >_p = \mathrm{w}_k^*
\end{aligned} \tag{5}$$

We obtain:

$$\frac{\partial \ln P}{\partial \underline{\mathbf{w}}} = \underline{\mathbf{M}}(\underline{\hat{\mathbf{w}}} - \underline{\mathbf{w}}) \text{ for all vectors } \underline{\mathbf{w}} \tag{6}$$

For the maximum likelihood estimator we get

$$\underbrace{\frac{\partial \ln P}{\partial \underline{\mathbf{w}}} \stackrel{!}{=} 0}_{\text{max. of likelihood}} \quad \Rightarrow \underline{\mathbf{w}} = \underline{\hat{\mathbf{w}}} \tag{7}$$

Since $\underline{\hat{\mathbf{w}}}$ is efficient, this also holds for the maximum likelihood estimator.