Proof that Normal Distribution maximizes entropy $S$ over all normalized distributions, with mean $\mu$ and variance $\sigma^2$.

$$ S = \int p(m) \ln \left( \frac{p(m)}{\phi(m)} \right) dm $$

minimize/maximize $S$ with constraint using Euler/Lagrange eqns:

$$ \int p(m) dm = 1 \Rightarrow \int \phi(m) (m - \mu) dm = 0 $$
$$ \int \phi(m) (m - \mu)^2 dm = \sigma^2 \Rightarrow (\lambda - \lambda \mu)^2 = 0 $$

$$ \Phi = p(m) \ln p(m) + \lambda_1 p(m)^2 + \lambda_2 p(m) (m - \mu) + \lambda_3 p(m) (m - \mu)^2 $$
$$ \frac{\partial \Phi}{\partial m} = \ln p(m) + 1 + \lambda_1 + \lambda_2 (m - \mu) + \lambda_3 (m - \mu)^2 $$
$$ \ln p(m) = (1 - \lambda_1) = \lambda_2 (m - \mu) = \lambda_3 (m - \mu)^2 $$

Constraints require $1 - \lambda_1 = \ln \frac{\sqrt{2\pi \sigma^2}}{\phi(m)}$,
so $\phi(m) = \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left\{ -\frac{(m - \mu)^2}{2\sigma^2} \right\}$.

Now let $S = \int p(m) \ln \left( \frac{p(m)}{\phi(m)} \right) dm$ with $\phi(m)$ prescribed, and minimize with constraints:

$$ \int p(m) dm = 1 \Rightarrow \int \phi(m) (m - \mu) dm = 0 $$
$$ \int \phi(m) (m - \mu)^2 dm = \sigma^2 $$
$$ \Phi = p(m) \ln p(m) \phi(m) + \lambda_1 p(m) + \lambda_2^T p(m) (m - \mu) $$

$$ \frac{\partial \Phi}{\partial m} = \ln \phi(m) + 1 - \ln \phi(m) + 2 \lambda_1 + \lambda_2^T (m - \mu) $$

$$ \phi(m) = \exp \left\{ -\frac{(m - \mu)^2}{2\sigma^2} \right\} $$

Now suppose $\phi(m) = \exp \left\{ -\frac{1}{2} \delta m^T C m \right\}$

$$ p(m) = \exp \left\{ \frac{1}{2} \delta m^T C m \right\} $$

with $C(m) = \frac{1}{2} (m - s)^T C m (m - s) + (1 + \lambda) - 2 \lambda^T (d - \delta m)$

Now mean $\mu$ also max. likelihood point satisfies $\frac{\partial S}{\partial \mu} \bigg|_{\mu = \mu} = 0$.
\[
\frac{\partial f}{\partial m} = 0 = C_m (m^T - s) + G_T m
\]

\[m^T - s = -C_m G_T m\]

pre-multiply by \(G\) to get \(G_{m^T} - G_s = -C_m G_T m\)

and use \(G_{m^T} = m\) to get \((d - G_s) = -G_{m^T} m\)

Assume \(G_{m^T}\) has inverse and solve for \(\lambda\)

\[\lambda = -(G_{m^T} G_T)^{-1} (d - G_s)\]

and now plug back into \(\frac{\partial f}{\partial m} = 0\) to get

\[(m^T - s) = -C_m (G_T m)\]

\[(m^T - s) = C_m G_T (G_{m^T} G_T)^{-1} (d - G_s)\]

which is a "minimum length" type solution.

Thus a minimum length MLE solution can be derived from "minimum relative entropy" HRE considerations.