Stylized Normal Example for Dry Erase Board

Example Exam Question

Suppose we collect N random samples $y = \{y_1, y_2, ..., y_N\}$ and model these data with stylized normal distribution with pdf $f(y_i; \mu) = \frac{1}{\sqrt{2\pi}} \exp\bigl(-\frac{1}{2}(y_i - \mu)^2\bigr)$ for $y_i \in \mathbb{R}$ and $\mu \in \mathbb{R}$. Find the ML estimator of μ.

Step 1. Write down the likelihood.

Suppose a **stylized normal model** for each observation $y_i \sim \mathcal{N}(\mu, 1)$.

$$f(y_i; \mu) = \frac{1}{\sqrt{2\pi}} \exp\bigl(-\tfrac{1}{2}(y_i - \mu)^2\bigr) .$$

The joint likelihood of the data is the product of the densities (by independence).

$$L(\mu) = \prod_{i=1}^{N} f(y_i; \mu).$$

Step 2. Take log of the likelihood.

$$\ell(\mu) = \log L(\mu) = \log \prod_{i=1}^{N} f(y_i; \mu).$$

Step 3. Simplify the log-likelihood.

$$\begin{aligned}
\ell(\mu) &= \sum_{i=1}^{N} \log\left[\tfrac{1}{\sqrt{2\pi}} \exp\!\left(-\tfrac{1}{2}(y_i-\mu)^2\right)\right] \\
&= \sum_{i=1}^{N} \left\{ \log\!\left(\tfrac{1}{\sqrt{2\pi}}\right) + \log\exp\!\left(-\tfrac{1}{2}(y_i-\mu)^2\right) \right\} \\
&= \sum_{i=1}^{N} \left\{-\tfrac{1}{2}\log(2\pi) - \tfrac{1}{2}(y_i-\mu)^2\right\} \\
&= -\tfrac{N}{2}\log(2\pi) - \underbrace{\tfrac{1}{2}\sum_{i=1}^{N}(y_i-\mu)^2}_{\text{expand for easier } \frac{d\ell}{d\mu}} .
\end{aligned}$$

Step 4. Take derivative of the log-likelihood.

$$\begin{aligned}
\frac{\partial \ell}{\partial \mu} &= \frac{\partial}{\partial \mu}\left[-\tfrac{1}{2}\sum_{i=1}^{N}(y_i-\mu)^2\right] \\
&= -\tfrac{1}{2}\sum_{i=1}^{N}\frac{\partial}{\partial \mu}(y_i-\mu)^2 \\
&= -\tfrac{1}{2}\sum_{i=1}^{N}\Big[2(y_i-\mu)(-1)\Big] && \text{chain rule} \\
&= -\tfrac{1}{2}\sum_{i=1}^{N}\left[-2(y_i-\mu)\right] \\
&= \sum_{i=1}^{N}(y_i-\mu).
\end{aligned}$$

Step 5. Set derivative equal to zero.

$$\sum_{i=1}^{N}(y_i-\hat{\mu}) = 0.$$

Step 6: Solve.

$$\sum_{i=1}^{N} y_i - N\hat{\mu} = 0$$

$$N\hat{\mu} = \sum_{i=1}^{N} y_i$$

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} y_i = \mathrm{avg}(y).$$

Curvature at ML estimate

Our intuition that curvature (or the peakedness of the log-likelihood function) is somehow related to the standard error.

Let's explore this by finding the second derivative of the log-likelihood function.

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \mu^2} &= \frac{\partial}{\partial \mu}\left[\sum_{i=1}^{N}(y_i - \mu)\right] \\
&= \frac{\partial}{\partial \mu}\left[\left(\sum_{i=1}^{N} y_i\right) - N\mu)\right] \\
&= \frac{\partial}{\partial \mu}\left(\sum_{i=1}^{N} y_i\right) - \frac{\partial}{\partial \mu}(N\mu) \\
&= 0 - N \\
&= -N.
\end{aligned}$$

Observations

Facts:

- As N increases, $\dfrac{\partial^2 \ell(\mu \mid y)}{\partial \mu^2} = -N$ becomes more negative.
- As N increases, the curvature increases (i.e., the log-likelihood becomes more sharply peaked).
- As N increases, the uncertainty decreases (i.e., the SE should be smaller).

This matches our intuition exactly, so there is definitely a connection here.

But recall that the classic SE estimate is $\frac{\sigma}{\sqrt{N}}$.

For $\sigma = 1$ (as is the case in the stylized normal model), this reduces to $\frac{1}{\sqrt{N}}$.

Converting this SE to a variance, we have $\frac{1}{N}$. This is kinda close to $-N$.

At least in this case, the variance estimator is just the inverse of the negative of the second derivative. Is there something here? Can we say something more general about the inverse of the negative of the second derivative?