Asymptotic Variance for MLEs with Parameter Misspecification

Anish Lakkapragada

April 13, 2025

Given a set $X_1, \ldots, X_n \overset{\text{IID}}{\sim} f(x|\theta)$, we have established that under regularity conditions for $f(x|\theta)$ as $n \to \infty$, MLE $\hat{\theta}$ is asymptotically normal and $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \frac{1}{I(\theta)})$ where Fisher Information $I(\theta) = -\mathbb{E}_\theta[\frac{\partial^2}{\partial\theta^2} \log f(x|\theta)]$.

Let us now suppose that $X_1, \ldots, X_n \overset{\text{IID}}{\sim} g(x)$ for an unknown distribution $g(x)$ and we fit parameters $\theta \in \Omega$ for model $f(x|\theta)$ with MLE. *What is the asymptotic variance of MLE $\hat{\theta}$ in this case?*

Before proceeding, we assume standard regularity conditions and that $\theta \mapsto D_{\text{KL}}(g(x) \parallel f(x|\theta))$ has a unique minimizer $\theta^* \in \Omega$ where $D_{\text{KL}}(g(x) \parallel f(x|\theta))$ is the Kullback-Leibler (KL) Divergence from $f(x|\theta)$ to $g(x)$. We are now ready to begin, with a very similar derivation of asymptotic variance of the MLE under correct parameter specification. By definition of MLE, $\hat{\theta}$ maximizes $\ell_n(\theta) = \Sigma_{i=1}^n \log f(X_i|\theta) \implies \ell_n'(\hat{\theta}) = 0$. Furthermore, as we have established, $\hat{\theta} - \theta^* \to 0$ in probability as $n \to \infty$. We use a Taylor Expansion of $\ell_n'(\hat{\theta})$ around θ^*:

$$0 = \ell_n'(\hat{\theta}) \approx \ell_n'(\theta^*) + \ell_n''(\theta^*)(\hat{\theta} - \theta^*) \implies \sqrt{n}(\hat{\theta} - \theta^*) \approx \frac{\sqrt{n}\ell_n'(\theta^*)}{-\ell_n''(\theta^*)} = \frac{\ell_n'(\theta^*)/\sqrt{n}}{-\ell_n''(\theta^*)/n}$$

We first understand the denominator term $-\ell_n''(\theta^*)/n$:

$$\frac{-\ell_n''(\theta^*)}{n} = \frac{1}{n}\sum_{i=1}^{n} -\frac{\partial^2}{\partial\theta^2} \log f(X_i|\ \theta^*) \xrightarrow{p} \mathbb{E}_g[-\frac{\partial^2}{\partial\theta^2} \log f(X|\theta^*)]$$

For clarification the above shows a simple LLN argument, where $\mathbb{E}_g$ represents an expectation taken over $X \sim g$. We now tackle the numerator $\ell_n'(\theta^*)/\sqrt{n}$:

$$\frac{\ell_n'(\theta^*)}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i|\theta^*)$$

By the CLT, $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i|\theta^*) \xrightarrow{d} \mathcal{N}(\sqrt{n} \cdot \mathbb{E}_g[\frac{\partial}{\partial\theta} \log f(X|\theta^*)], \text{Var}_g[\frac{\partial}{\partial\theta} \log f(X|\theta^*)])$. Note that $\mathbb{E}_g[\frac{\partial}{\partial\theta} \log f(X|\theta^*)]$ is the derivative of the expected log-likelihood function $L(\theta) =$

$\mathbb{E}_g[\log f(X|\theta)]$ evaluated at $\theta = \theta^*$, or $L'(\theta^*) = \mathbb{E}_g[\frac{\partial}{\partial\theta}\log f(X|\theta^*)]$. By definition θ^* minimizes $D_{\text{KL}}(g(x) \,\|\, f(x|\theta)) \implies \theta^*$ maximizes $L(\theta) \implies L'(\theta^*) = 0$. This allows us to simplify $\text{Var}_g[\frac{\partial}{\partial\theta}\log f(X|\theta^*)]$:

$$\text{Var}_g[\frac{\partial}{\partial\theta}\log f(X|\theta^*)] = \mathbb{E}_g[(\frac{\partial}{\partial\theta}\log f(X|\theta^*))^2] - \mathbb{E}_g[\frac{\partial}{\partial\theta}\log f(X|\theta^*)]^2 = \mathbb{E}_g[(\frac{\partial}{\partial\theta}\log f(X|\theta^*))^2]$$

Putting these two results together, we have that:

$$\frac{\ell_n'(\theta^*)}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(X_i|\theta^*) \xrightarrow{d} \mathcal{N}(0, \mathbb{E}_g[(\frac{\partial}{\partial\theta}\log f(X|\theta^*))^2])$$

Recall our approximation $\sqrt{n}(\hat{\theta} - \theta^*) \approx \frac{\sqrt{n}\ell_n'(\theta^*)}{-\ell_n''(\theta^*)} = \frac{\ell_n'(\theta^*)/\sqrt{n}}{-\ell_n''(\theta^*)/n}$. We apply Slutsky's Lemma to analyze the convergence (in distribution) of this approximation (recall $-\ell_n''(\theta^*)/n \xrightarrow{p} \mathbb{E}_g[-\frac{\partial^2}{\partial\theta^2}\log f(X|\theta^*)]$):

$$\sqrt{n}(\hat{\theta} - \theta^*) \approx \frac{\ell_n'(\theta^*)/\sqrt{n}}{-\ell_n''(\theta^*)/n} = \frac{\mathcal{N}(0, \mathbb{E}_g[(\frac{\partial}{\partial\theta}\log f(X|\theta^*))^2])}{-\ell_n''(\theta^*)/n} \xrightarrow{d} \mathcal{N}(0, \frac{\mathbb{E}_g[(\frac{\partial}{\partial\theta}\log f(X|\theta^*))^2]}{\mathbb{E}_g[\frac{\partial^2}{\partial\theta^2}\log f(X|\theta^*)]^2})$$

This asymptotic variance for $\hat{\theta}$ is also known as the *sandwich variance.*