Foundations of machine learning

Local asymptotic Normality

Maximilian Kasy

Department of Economics, University of Oxford

Winter 2025

The Normal means model as asymptotic approximation

- The Normal means model might seem quite special.
- But asymptotically, any sufficiently smooth parametric model is equivalent.
- Formally: The likelihood ratio process of n i.i.d. draws Y_i from the distribution

$$P^n_{\theta_0+h/\sqrt{n}},$$

  converges to the likelihood ratio process of one draw X from

$$N\left(h, I_{\theta_0}^{-1}\right)$$

- Here h is a local parameter for the model around θ_0, and I_{θ_0} is the Fisher information matrix.

- Suppose that P_θ has a density f_θ relative to some measure.
- Recall the following definitions:
  - Log-likelihood: $\ell_\theta(Y) = \log f_\theta(Y)$
  - Score: $\dot{\ell}_\theta(Y) = \partial_\theta \log f_\theta(Y)$
  - Hessian $\ddot{\ell}_\theta(Y) = \partial_\theta^2 \log f_\theta(Y)$
  - Information matrix: $I_\theta = \mathrm{Var}_\theta(\dot{\ell}_\theta(Y)) = -E_\theta[\ddot{\ell}_\theta(Y)]$
- Likelihood ratio process:

$$\prod_i \frac{f_{\theta_0+h/\sqrt{n}}(Y_i)}{f_{\theta_0}(Y_i)},$$

where $Y_1, \ldots, Y_n$ are i.i.d. $P_{\theta_0+h/\sqrt{n}}$ distributed.

Practice problem (Taylor expansion)

- Using this notation, provide a second order Taylor expansion for the log-likelihood $\ell_{\theta_0+h}(Y)$ with respect to h.
- Provide a corresponding Taylor expansion for the log-likelihood of n i.i.d. draws Y_i from the distribution $P_{\theta_0+h/\sqrt{n}}$.
- Assuming that the remainder is negligible, describe the limiting behavior (as $n \to \infty$) of the log-likelihood ratio process

$$\log \prod_i \frac{f_{\theta_0+h/\sqrt{n}}(Y_i)}{f_{\theta_0}(Y_i)}.$$

Solution

- Expansion for $\ell_{\theta_0+h}(Y)$:

$$\ell_{\theta_0+h}(Y) = \ell_{\theta_0}(Y) + h' \cdot \dot{\ell}_{\theta_0}(Y) + \tfrac{1}{2} \cdot h \cdot \ddot{\ell}_{\theta_0}(Y) \cdot h + \textit{remainder}.$$

- Expansion for the log-likelihood ratio of n i.i.d. draws:

$$\log \prod_i \frac{f_{\theta_0+h'/\sqrt{n}}(Y_i)}{f_{\theta_0}(Y_i)} = \tfrac{1}{\sqrt{n}} h' \cdot \sum_i \dot{\ell}_{\theta_0}(Y_i) + \tfrac{1}{2n} h' \cdot \sum_i \ddot{\ell}_{\theta_0}(Y_i) \cdot h + \textit{remainder}.$$

- Asymptotic behavior (by CLT, LLN):

$$\Delta_n := \tfrac{1}{\sqrt{n}} \sum_i \dot{\ell}_{\theta_0}(Y_i) \to^d N(0, I_{\theta_0}),$$

$$\tfrac{1}{2n} \cdot \sum_i \ddot{\ell}_{\theta_0}(Y_i) \to^p -\tfrac{1}{2} I_{\theta_0}.$$

- Suppose the remainder is negligible.
- Then the previous slide suggests

$$\log \prod_i \frac{f_{\theta_0+h/\sqrt{n}}(Y_i)}{f_{\theta_0}(Y_i)} =^A h' \cdot \Delta - \frac{1}{2} h' I_{\theta_0} h,$$

  where

$$\Delta \sim N\left(0, I_{\theta_0}\right).$$

- Theorem 7.2 in van der Vaart (2000), chapter 7 states sufficient conditions for this to hold.
- We show next that this is the same likelihood ratio process as for the model

$$N\left(h, I_{\theta_0}^{-1}\right).$$

Practice problem

- Suppose $X \sim N\left(h, I_{\theta_0}^{-1}\right)$
- Write out the log likelihood ratio

$$\log \frac{\varphi_{I_{\theta_0}^{-1}}(X-h)}{\varphi_{I_{\theta_0}^{-1}}(X)}.$$

Solution

- The Normal density is given by

$$\varphi_{I_{\theta_0}^{-1}}(x) = \frac{1}{\sqrt{(2\pi)^k |\det(I_{\theta_0}^{-1})|}} \cdot \exp\left(-\tfrac{1}{2} x' \cdot I_{\theta_0} \cdot x\right)$$

- Taking ratios and logs yields

$$\log \frac{\varphi_{I_{\theta_0}^{-1}}(X-h)}{\varphi_{I_{\theta_0}^{-1}}(X)} = h' \cdot I_{\theta_0} \cdot x - \tfrac{1}{2} h' \cdot I_{\theta_0} \cdot h.$$

- This is exactly the same process we obtained before, with $I_{\theta_0} \cdot X$ taking the role of Δ.

Why care

- Suppose that $Y_i \sim^{iid} P_{\theta+h/\sqrt{n}}$, and $T_n(Y_1, \ldots, Y_n)$ is an arbitrary statistic that satisfies

$$T_n \to^d L_{\theta,h}$$

for some limiting distribution $L_{\theta,h}$ and all h.

- Then $L_{\theta,h}$ is the distribution of some (possibly randomized) statistic $T(X)$!
- This is a (non-obvious) consequence of the convergence of the likelihood ratio process.
- cf. Theorem 7.10 in van der Vaart (2000).

Maximum likelihood and shrinkage

- This result applies in particular to $T =$ estimators of θ.
- Suppose that $\widehat{\theta}^{ML}$ is the maximum likelihood estimator.
- Then $\widehat{\theta}^{ML} \to^d X$, and any shrinkage estimator based on $\widehat{\theta}^{ML}$ converges in distribution to a corresponding shrinkage estimator in the limit experiment.

References

van der Vaart, A. W. (2000). Asymptotic statistics. *Cambridge University Press, chapter 7.*

Hansen, B. E. (2016). Efficient shrinkage in parametric models. Journal of Econometrics, *190(1):115–132.*