Le Cam, Birgé, Huber, and Strassen

Joonhyuk Jung

May 22, 2026

THE UNIVERSITY OF CHICAGO

Total Variation and Hellinger Distances

My first two years of studies revolve around the total variation and Hellinger distances.

$$\mathrm{TV}(p, q) := \frac{1}{2} \int |p - q|,$$

$$H(p, q) := \sqrt{\frac{1}{2} \int (\sqrt{p} - \sqrt{q})^2}.$$

- First year: $\mathrm{TV}(f_\pi, f_\eta) \overset{?}{\asymp} H(f_\pi, f_\eta)$ for Gaussian location mixtures.

$$f_\pi(x) := \int_{\mathbb{R}^d} \phi_d(x - \theta)\, d\pi(\theta).$$

Total Variation and Hellinger Distances

My first two years of studies revolve around the total variation and Hellinger distances.

$$\mathrm{TV}(p, q) := \frac{1}{2} \int |p - q|,$$

$$H(p, q) := \sqrt{\frac{1}{2} \int (\sqrt{p} - \sqrt{q})^2}.$$

- First year: $\mathrm{TV}(f_\pi, f_\eta) \overset{?}{\asymp} H(f_\pi, f_\eta)$ for Gaussian location mixtures.

$$f_\pi(x) := \int_{\mathbb{R}^d} \phi_d(x - \theta)\, d\pi(\theta).$$

- Second year (ongoing): Robust estimation of a location parameter (under Huber's contamination model).

First Year: Resolving an Open Question

- An open question [JPW23]: $\mathrm{TV}(f_\pi, f_\eta) \overset{?}{\asymp} H(f_\pi, f_\eta)$ for

$$\mathrm{supp}(\pi), \mathrm{supp}(\eta) \subseteq [-M, M]^d.$$

First Year: Resolving an Open Question

- An open question [JPW23]: $\mathrm{TV}(f_\pi, f_\eta) \overset{?}{\asymp} H(f_\pi, f_\eta)$ for
$$\mathrm{supp}(\pi), \mathrm{supp}(\eta) \subseteq [-M, M]^d.$$
- Our answer [JG26]: No. We proved that
$$H(f_\pi, f_\eta) \le \mathrm{TV}(f_\pi, f_\eta)^{1 - \frac{\Theta(1)}{\log\log(1/\mathrm{TV}(f_\pi, f_\eta))}}.$$

First Year: Resolving an Open Question

- An open question [JPW23]: $\mathrm{TV}(f_\pi, f_\eta) \overset{?}{\asymp} H(f_\pi, f_\eta)$ for
$$\mathrm{supp}(\pi), \mathrm{supp}(\eta) \subseteq [-M, M]^d.$$
- Our answer [JG26]: No. We proved that
$$H(f_\pi, f_\eta) \leq \mathrm{TV}(f_\pi, f_\eta)^{1 - \dfrac{\Theta(1)}{\log\log(1/\mathrm{TV}(f_\pi, f_\eta))}}.$$
- As a consequence, for $X_1, \ldots, X_n \overset{i.i.d.}{\sim} (1-\epsilon) P_{f_\pi} + \epsilon Q$,
$$\inf_{\widehat{f}} \sup_{\pi, Q} \mathbb{E}_{\epsilon,\pi,Q}\left[H^2\left(f_\pi, \widehat{f}\right)\right] \asymp \epsilon^{2\left(1 - \frac{\Theta(1)}{\log\log(1/\epsilon)}\right)},$$
in the "large-ϵ regime" where $\epsilon \gtrsim n^{-0.49}$.

Motivation: Gaussian Location Model (1/2)

Recall the simplest model in robust statistics.

- Classical result: For $X_1, \ldots, X_n \overset{i.i.d.}{\sim} (1-\epsilon)N(\theta, 1) + \epsilon Q$,

$$\inf_{\widehat{\theta}} \sup_{\theta, Q} \inf \left\{ \rho > 0 : \mathbb{P}_{\epsilon,\theta,Q}\left[\left(\widehat{\theta} - \theta \right)^2 \geq \rho \right] \leq 0.05 \right\} \asymp \frac{1}{n} + \epsilon^2,$$

  uniformly for all n and $\epsilon \in [0, 0.49]$.

- The Gaussian location model under Huber's contamination admits two regimes:

$$\color{blue}{\frac{1}{n} \leftarrow \text{Small-}\epsilon \text{ regime}}$$
$$\color{red}{\epsilon^2 \leftarrow \text{Large-}\epsilon \text{ regime}}$$

Motivation: Gaussian Location Model (2/2)

What determines the minimax rates in the two regimes?

$$\left(\widehat{\theta} - \theta\right)^2 \asymp \frac{1}{n} + \epsilon^2,$$

- The first term is characterized by the Hellinger distance.

$$r^2(n) := \sup\left\{\theta^2 : H^2(N(0,1), N(\theta,1)) \leq \frac{1}{n}\right\} \asymp \frac{1}{n}.$$

Motivation: Gaussian Location Model (2/2)

What determines the minimax rates in the two regimes?

$$\left(\widehat{\theta} - \theta\right)^2 \asymp \frac{1}{n} + \epsilon^2,$$

- The first term is characterized by the Hellinger distance.

$$r^2(n) := \sup\left\{\theta^2 : H^2(N(0,1), N(\theta,1)) \leq \frac{1}{n}\right\} \asymp \frac{1}{n}.$$

- The second term is given by the total variation.

$$\begin{aligned}\omega^2(\epsilon) := \sup\{\theta^2 : \exists Q_0, Q_1\ ,\\ (1-\epsilon)N(0,1) + \epsilon Q_0 = (1-\epsilon)N(\theta,1) + \epsilon Q_1\}\\ = \sup\left\{\theta^2 : \mathrm{TV}(N(0,1), N(\theta,1)) \leq \frac{\epsilon}{1-\epsilon}\right\} \asymp \epsilon^2.\end{aligned}$$

Motivation: Gaussian Location Model (2/2)

What determines the minimax rates in the two regimes?

$$\left(\widehat{\theta} - \theta\right)^2 \asymp \frac{1}{n} + \epsilon^2,$$

- The first term is characterized by the Hellinger distance.

$$r^2(n) := \sup \left\{ \theta^2 : H^2(N(0,1), N(\theta,1)) \leq \frac{1}{n} \right\} \asymp \frac{1}{n}.$$

- The second term is given by the total variation.

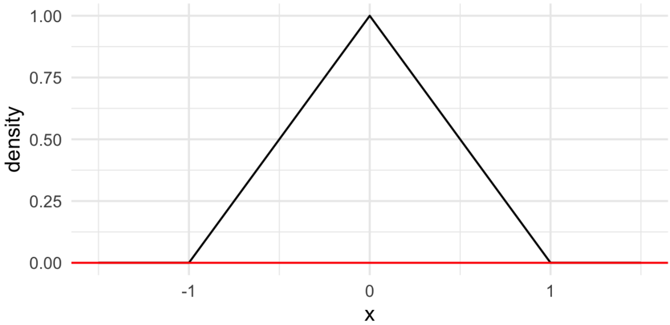
$$\begin{aligned} \omega^2(\epsilon) := \sup \{ \theta^2 : \exists Q_0, Q_1 \, , \\ (1-\epsilon)N(0,1) + \epsilon Q_0 = (1-\epsilon)N(\theta,1) + \epsilon Q_1 \} \\ = \sup \left\{ \theta^2 : \mathrm{TV}(N(0,1), N(\theta,1)) \leq \frac{\epsilon}{1-\epsilon} \right\} \asymp \epsilon^2. \end{aligned}$$

- We cannot distinguish θ from zero even when we observe $n = \infty$ sample points.

Second Year: Singular Location Model (1/4)

What happens under singular location models?

- Infinite Fisher information
- Example: Density plot of $\Delta(0, 1)$



Second Year: Singular Location Model (2/4)

First follow the same argument as the Gaussian location model.

- For $X_1, \ldots, X_n \overset{i.i.d.}{\sim} (1-\epsilon)\Delta(\theta, 1) + \epsilon Q$,

$$\left(\widehat{\theta} - \theta\right)^2 \overset{?}{\asymp} r^2(n) + \omega^2(\epsilon),$$

uniformly for all n and $\epsilon \in [0, 0.49]$, where

$$r^2(n) := \sup\left\{\theta^2 : H^2(\Delta(0,1), \Delta(\theta,1)) \leq \frac{1}{n}\right\},$$

$$\omega^2(\epsilon) := \sup\left\{\theta^2 : \mathrm{TV}(\Delta(0,1), \Delta(\theta,1)) \leq \frac{\epsilon}{1-\epsilon}\right\}.$$

Second Year: Singular Location Model (3/4)

- The first term is characterized by the Hellinger distance.

$$r^2(n) := \sup \left\{ \theta^2 : \underbrace{H^2(\Delta(0,1), \Delta(\theta,1))}_{\asymp \theta^2 \log \frac{1}{\theta}} \leq \frac{1}{n} \right\} \asymp \frac{1}{n \log n}.$$

- The second term is given by the total variation.

$$\omega^2(\epsilon) := \sup \left\{ \theta^2 : \mathrm{TV}(\Delta(0,1), \Delta(\theta,1)) \leq \frac{\epsilon}{1-\epsilon} \right\}$$
$$= 4 \left(1 - \sqrt{\frac{1-2\epsilon}{1-\epsilon}} \right)^2 \asymp \epsilon^2.$$

Second Year: Singular Location Model (4/4)

- For $X_1, \ldots, X_n \overset{i.i.d.}{\sim} (1-\epsilon)\Delta(\theta, 1) + \epsilon Q$,

$$\left(\widehat{\theta} - \theta\right)^2 \overset{?}{\asymp} \frac{1}{n \log n} + \epsilon^2,$$

Second Year: Singular Location Model (4/4)

- For $X_1, \ldots, X_n \overset{i.i.d.}{\sim} (1-\epsilon)\Delta(\theta, 1) + \epsilon Q$,

$$\left(\widehat{\theta} - \theta\right)^2 \overset{?}{\asymp} \frac{1}{n \log n} + \epsilon^2,$$

- Our answer: No. We proved that

$$\left(\widehat{\theta} - \theta\right)^2 \asymp \frac{1}{n \log n} + \underbrace{\frac{1}{n \log\left(\frac{1}{n\epsilon^2} \vee e\right)}}_{\text{Intermediate-}\epsilon} + \epsilon^2,$$

uniformly for all n and $\epsilon \in [0, 0.49]$.

Second Year: Singular Location Model (4/4)

- For $X_1, \ldots, X_n \overset{i.i.d.}{\sim} (1-\epsilon)\Delta(\theta, 1) + \epsilon Q$,

$$\left(\widehat{\theta} - \theta\right)^2 \overset{?}{\asymp} \frac{1}{n \log n} + \epsilon^2,$$

- Our answer: No. We proved that

$$\left(\widehat{\theta} - \theta\right)^2 \asymp \frac{1}{n \log n} + \underbrace{\frac{1}{n \log\left(\frac{1}{n\epsilon^2} \vee e\right)}}_{\text{Intermediate-}\epsilon} + \epsilon^2,$$

uniformly for all n and $\epsilon \in [0, 0.49]$.

- What is going on?

The Intermediate-ϵ Regime (1/3)

We recall the classical results for composite-composite hypothesis testing: $\mathfrak{H}_0^\epsilon$ v.s. $\mathfrak{H}_\theta^\epsilon$, where we define $\mathfrak{H}_\theta^\epsilon := \{(1-\epsilon)F_\theta + \epsilon Q : Q\}$.

- Le Cam, Lucien & Birgé, Lucien

$$
\begin{aligned}
1 - \sqrt{2nH^2(\mathfrak{H}_0^\epsilon, \mathfrak{H}_\theta^\epsilon)} &\leq \inf_\phi \sup_{(P_0,P_1)\in\mathfrak{H}_0^\epsilon\times\mathfrak{H}_\theta^\epsilon} P_0^{\otimes n}\phi + P_1^{\otimes n}(1-\phi) \\
&\leq 2\exp\left(-nH^2(\mathfrak{H}_0^\epsilon, \mathfrak{H}_\theta^\epsilon)\right), \\
H^2(\mathfrak{H}_0^\epsilon, \mathfrak{H}_\theta^\epsilon) &:= \inf_{(P_0,P_1)\in\mathfrak{H}_0^\epsilon\times\mathfrak{H}_\theta^\epsilon} H^2(P_0, P_1).
\end{aligned}
$$

The Intermediate-ϵ Regime (1/3)

We recall the classical results for composite-composite hypothesis testing: $\mathfrak{H}_0^\epsilon$ v.s. $\mathfrak{H}_\theta^\epsilon$, where we define $\mathfrak{H}_\theta^\epsilon := \{(1-\epsilon)F_\theta + \epsilon Q : Q\}$.

- Le Cam, Lucien & Birgé, Lucien

$$
\begin{aligned}
1 - \sqrt{2nH^2(\mathfrak{H}_0^\epsilon, \mathfrak{H}_\theta^\epsilon)} &\le \inf_\phi \sup_{(P_0,P_1)\in\mathfrak{H}_0^\epsilon\times\mathfrak{H}_\theta^\epsilon} P_0^{\otimes n}\phi + P_1^{\otimes n}(1-\phi) \\
&\le 2\exp\left(-nH^2(\mathfrak{H}_0^\epsilon, \mathfrak{H}_\theta^\epsilon)\right), \\
H^2(\mathfrak{H}_0^\epsilon, \mathfrak{H}_\theta^\epsilon) &:= \inf_{(P_0,P_1)\in\mathfrak{H}_0^\epsilon\times\mathfrak{H}_\theta^\epsilon} H^2(P_0, P_1).
\end{aligned}
$$

- Huber, Peter & Strassen, Volker

$$
\begin{aligned}
\inf_\phi \sup_{(P_0,P_1)\in\mathfrak{H}_0^\epsilon\times\mathfrak{H}_\theta^\epsilon} P_0^{\otimes n}\phi + P_1^{\otimes n}(1-\phi) &= 1 - \mathrm{TV}\left((P_0^\star)^{\otimes n}, (P_1^\star)^{\otimes n}\right), \\
H^2(P_0^\star, P_1^\star) &= H^2(\mathfrak{H}_0^\epsilon, \mathfrak{H}_\theta^\epsilon), \\
dP_1^\star/dP_0^\star &= \max(c', \min(c'', dF_\theta/dF_0)).
\end{aligned}
$$

The Intermediate-ϵ Regime (2/3)

- According to Le Cam, Birgé, Huber, and Strassen,

$$\delta^2(n,\epsilon) := \sup\left\{\theta^2 : H^2\left(\mathfrak{H}_0^\epsilon, \mathfrak{H}_\theta^\epsilon\right) \le \frac{1}{n}\right\},$$
$$\mathfrak{H}_\theta^\epsilon := \left\{(1-\epsilon)\Delta(\theta,1) + \epsilon Q : Q\right\}.$$

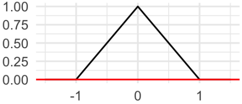
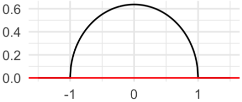
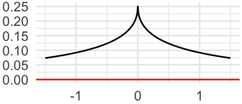
- We proved that

$$H^2\left(\mathfrak{H}_0^\epsilon, \mathfrak{H}_\theta^\epsilon\right) \asymp \begin{cases} \theta^2 \log\left(\dfrac{\theta}{\theta^2 \vee \epsilon} \vee e\right), & \omega(\epsilon) < \theta, \\ 0, & \omega(\epsilon) \ge \theta. \end{cases}$$

- Thus, solving $nH^2 \asymp 1$ gives

$$\delta^2(n,\epsilon) \asymp \frac{1}{n\log n} + \frac{1}{n\log\left(\frac{1}{n\epsilon^2} \vee e\right)} + \epsilon^2.$$

The Intermediate-ϵ Regime (3/3)

Density	Shape	$\delta^2(n,\epsilon)$
Triangle		$\frac{1}{n\log n} + \frac{1}{n\log\left(\frac{1}{n\epsilon^2}\vee e\right)} + \epsilon^2$
Semicircle		$\frac{1}{n^{4/3}} + \left(\frac{\epsilon}{n}\right)^{2/3} + \epsilon^2$
GG(1/2)		$\frac{1}{n\log n} + \frac{1}{n\log\left(\frac{1}{n\epsilon^2}\vee e\right)} + \epsilon^2$

Conclusion

- Such intermediate-ϵ regimes exist only under singular models.
- Q: What if we have more than one singularities?
- Q: Unified framework?
  - Le Cam, Lucien & Birgé, Lucien

$$nH^2(P_0^\star, P_1^\star) \asymp 1.$$

  - Huber, Peter & Strassen, Volker

$$dP_1^\star / dP_0^\star = \max(c', \min(c'', dF_\theta / dF_0)).$$

Acknowledgements

- I thank you all for listening.

Acknowledgements

- I thank you all for listening.
- I thank Prof. Chao Gao for continued guidance.

Acknowledgements

- I thank you all for listening.
- I thank Prof. Chao Gao for continued guidance.
- I thank my wife.

References I

Joonhyuk Jung and Chao Gao, *Sharp inequalities between total variation and Hellinger distances for Gaussian mixtures*, arXiv preprint arXiv:2602.03202 (2026).

Zeyu Jia, Yury Polyanskiy, and Yihong Wu, *Entropic characterization of optimal rates for learning Gaussian mixtures*, The Thirty Sixth Annual Conference on Learning Theory, PMLR, 2023, pp. 4296–4335.