Stats 300b: Theory of Statistics **Winter 2017**

Lecture 20 – March 16

Lecturer: John Duchi *Scribe: Jonathan Lacotte*

Warning: *these notes may contain factual errors*

Reading:

Outline

- limiting Gaussian experiments
- local asymptotic minimax theorem

1 Limiting Gaussianity

Definition 1.1. A collection $\{P_{\theta,n}\}_{\theta\in\Theta,\, n\in\mathbb{N}}$ is locally asymptotically normal (LAN) with precision/information $K_\theta \in \mathbb{R}^{d\times d}$ if there exists $\Delta_n \in \mathbb{R}^d$ such that:

$$\log\left(\frac{dP_{\theta+\frac{h}{\sqrt{n}},n}}{dP_{\theta,n}}(X^n)\right) = h^T\Delta_n(X^n) - \frac{1}{2}h^T K_\theta h + o_{P_{\theta,n}}(||h||)$$

where $\Delta_n(X^n) \xrightarrow[P_{\theta,n}]{d} \mathcal{N}(0, K_\theta)$.

Le Cam's third lemma implies that, with $Z_n = K_\theta^{-1}\Delta_n$, $Z_n \xrightarrow[P_{\theta+\frac{h}{\sqrt{n}},n}]{d} \mathcal{N}(h, K_\theta^{-1})$

The goal is to make rigorous that in LAN families, estimating θ is same (in limit) as estimating the mean from a Gaussian location family.

Throughout, we assume that $\theta_0 = 0$ (wlog).

Lemma 1.1. Let $Z_n = K^{-1}\Delta_n$ (in LAN family). Then, (Z_n) is uniformly tight under $(P_{\frac{h}{\sqrt{n}},n})$ whenever $||h|| \le C < +\infty$.
Moreover, if we define $dQ_{h,n}(z) = \exp\left(-\frac{1}{2}(z-h)^T K(z-h) - z^T K z\right) dP_{\theta,n}(z)$ then we get for all $c, b < +\infty$:

$$\limsup_{n\infty} \sup_{||h||\le c} \int \mathbf{1}_{\{||z||\le b\}} |dQ_{h,n}(z) - dP_{\frac{h}{\sqrt{n}},n}(z)| = 0$$

Idea: The tilted distribution 'looks' Gaussian and integrates same as $dP_{\frac{h}{\sqrt{n}},n}$ over compact sets.

We will use the following lemma to show that, with Gaussian prior, a LAN family give asymptotically Gaussian posterior.

Lemma 1.2. Let $h \sim \mathcal{N}(0,\Gamma)$ with $\Gamma \succ 0$ and $Z|h \succ \mathcal{N}(Ah, \Sigma)$ with $\Sigma \succ 0$. Then

$$h|Z = z \sim \mathcal{N}\left((\Gamma^{-1} + A^T\Sigma^{-1}A)^{-1}A^T\Sigma^{-1}z,\ (\Gamma^{-1} + A^T\Sigma^{-1}A)^{-1}\right)$$

Define, for $K \succeq 0$, $\Gamma \succ 0$, the Gaussian distribution

$$G_{K,\Gamma}(\,.\,|z) = \mathcal{N}\left((K+\Gamma^{-1})^{-1}Kz,\, (K+\Gamma^{-1})^{-1}\right)$$

Idea: In LAN family, $Z_n = K^{-1}\Delta_n$ should give "Gaussian" posterior to shifts $\frac{h}{\sqrt{n}}$.

Notation: We use $\pi^{\Gamma,c}$ to denote the truncated Gaussian $\mathcal{N}(0,\Gamma)$ to $||h|| \leq c$.

Theorem 1.1 Let $\pi^{\Gamma,c}$ be the prior distribution distribution over h and assume that data X^n verify $X^n|h \sim P_{\frac{h}{\sqrt{n}},n}$. Denote $Z_n = K^{-1}\Delta_n(X^n)$ (LAN family), $P_n(\,.\,) := \int P_{\frac{h}{\sqrt{n}},n}(\,.\,)d\pi^{\Gamma,c}(h)$ the marginal distribution of X^n. Then, for all $\epsilon > 0$, there exist $C, N < +\infty$ such that for all $n \geq N$, $c \geq C$,

$$\int ||G_{K,\Gamma}(\,.\,|z_n(x^n)) - \pi^{\Gamma,c}(\,.\,|x^n)||_{TV} dP_n(x^n) \leq \epsilon$$

Remark: The true posterior of a LAN family, under truncated Gaussian prior, is, on average, really close to a Gaussian distribution, conditioned on $Z_n = K^{-1}\Delta_n(x^n)$.

2 Local asymptotic minimax theorem

Definition 2.1. A function $L : \mathbb{R}^d \mapsto \mathbb{R}$ is quasi-convex if for all $\alpha \in \mathbb{R}$, the α-sublevel set $\{x : L(x) \leq \alpha\}$ is convex.

Example 2.1. $L(x) = \frac{1}{2}||x||_2^2 \wedge B$ is quasi-convex for any $B \in \mathbb{R}$.

Lemma 2.1. (Anderson) Let L be symmetric and quasi-convex. Let $A \in \mathbb{R}^{d\times k}$ and $X \sim \mathcal{N}(\mu, \Sigma)$. Then:

$$\inf_{v\in\mathbb{R}^k} \mathbb{E}\left[L(AX - v)\right] = \mathbb{E}\left[L(A(X-\mu))\right] = \mathbb{E}\left[L(A\Sigma^{\frac{1}{2}}W)\right]$$

where $W \sim \mathcal{N}(0, I_k)$.

Theorem 2.1. (Local asymptotic minimax)
Let $L : \mathbb{R}^d \mapsto \mathbb{R}$ be quasi-convex, symmetric and bounded. Let $\{P_{\theta,n}\}$ be LAN at θ_0 with precision $K_{\theta_0} \geq 0$. Then, with $W \sim \mathcal{N}(0, I_k)$,

$$\liminf_{c\infty} \liminf_{n\infty} \inf_{\hat{\theta}_n} \sup_{||h||\leq c,\, \theta=\theta_0+\frac{h}{\sqrt{n}}} \mathbb{E}_{P_{\theta,n}}\left[L(\sqrt{n}(\hat{\theta}_n - \theta))\right] \geq \mathbb{E}\left[L(K_{\theta_0}^{-\frac{1}{2}}W)\right]$$

Remark Consider a quadratic mean differentiable family $\{P_\theta\}_{\theta\in\Theta}$ with Fisher information I_{θ_0} at parameter θ_0. Then the theorem implies that:

$$\liminf_{c\infty} \liminf_{n\infty} \inf_{\hat{\theta}_n} \int \mathbb{E}_{P^n_{\theta_0+\frac{h}{\sqrt{n}}}}\left[L(\sqrt{n}(\hat{\theta}_n(X_1,\ldots,X_n) - \theta))d\pi_c(h)\right] \geq \mathbb{E}[L(Z)]$$

with $Z \sim \mathcal{N}(0, I_{\theta_0}^{-1})$.

Proof of Theorem 2.1.

Without loss of generality, assume that L takes values in $[0,1]$ and $\theta_0 = 0$.
Observe that

$$\sup_{||h|| \leq c} \mathbb{E}_{P_{\frac{h}{\sqrt{n}},n}} \left[L(\sqrt{n}(\hat{\theta}_n - \theta)) \right] \geq \int \mathbb{E}_{P_{\frac{h}{\sqrt{n}},n}} \left[L(\sqrt{n}\hat{\theta}_n - h) \right] d\pi(h)$$

where $\theta = \frac{h}{\sqrt{n}}$, for any π with support in $\{||h|| \leq c\}$.

Consider $\pi := \pi^{\Gamma,c}$, prior of h, to be the normal distribution $\mathcal{N}(0,\Gamma)$, truncated to $\{||h|| \leq c\}$ and denote the marginal distribution of X^n:

$$\bar{P}_n(\,.\,) = \int P_{\frac{h}{\sqrt{n}},n}(\,.\,) d\pi^{\Gamma,c}(h)$$

Then, the left hand-side $(*)$ of the last inequality satisfies:

$$(*) \geq \int \mathbb{E}\left[L(\sqrt{n}\hat{\theta}_n - h) \,|\, X^n = x^n \right] d\bar{P}_n(x^n) \geq \int \inf_{\hat{h}} \mathbb{E}\left[L(\hat{h} - h) \,|\, X^n = x^n \right] d\bar{P}_n(x^n)$$

Using the previous notation $G_{K,\Gamma}$, we get:

$$(*) \geq \int \inf_{\hat{h}} \mathbb{E}_{G_{K,\Gamma}} \left[L(\hat{h} - h) \,|\, x^n \right] d\bar{P}_n(x^n) - \int \sup_{h,\hat{h}} L(\hat{h} - h) \left(dG_{K,\Gamma}(h \,|\, x^n) - \pi(h \,|\, x^n) \right) d\bar{P}_n(x^n)$$

Observe that:

$$\int \sup_{h,\hat{h}} L(\hat{h} - h) \left(dG_{K,\Gamma}(h \,|\, x^n) - \pi(h \,|\, x^n) \right) d\bar{P}(x^n) \leq \int ||G_{K,\Gamma}(\,.\,|\, x^n) - \pi(\,.\,|x^n)||_{TV} d\bar{P}_n(x^n)$$

and that, by Theorem 1.1, the right-hand side of the last inequality is less than ϵ, for any $\epsilon > 0$, c appropriately chosen and n sufficiently large.

Moreover, by Anderson's lemma, we have:

$$\int \inf_{\hat{h}} \mathbb{E}_{G_{K,\Gamma}} \left[L(\hat{h} - h) \,|\, x^n \right] d\bar{P}_n(x^n) \geq \int \mathbb{E}\left[L(\mathcal{N}(0, (K + \Gamma^{-1})^{-1})) \right] d\bar{P}_n(x^n)$$

Taking $\Gamma \to \infty$, we get:

$$(*) \geq \mathbb{E}\left[L(Z) \right] - \epsilon$$