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 \( \Delta_n(X^n) \overset{d}{\underset{P_{\theta,n}}{\rightarrow}} N(0, K_\theta) \).

Le Cam’s third lemma implies that, with \( Z_n = K_\theta^{-1} \Delta_n \), \( Z_n \overset{d}{\underset{P_{\theta,n}}{\rightarrow}} N(h, K_\theta^{-1}) \)

The goal is to make rigorous that in LAN families, estimating \( \theta \) 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 \( (P_{\theta+\frac{h}{\sqrt{n}},n}) \) whenever \( ||h|| \leq C < +\infty \).

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 \( c, b < +\infty \):

\[
\lim_{n \to \infty} \sup_{||h|| \leq c} \int \mathbf{1}_{||z|| \leq b} |dQ_{h,n}(z) - dP_{\frac{h}{\sqrt{n}},n}(z)| = 0
\]

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 \( \Gamma > 0 \) and \( Z|h \sim \mathcal{N}(Ah, \Sigma) \) with \( \Sigma > 0 \). 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 $K \geq 0, \Gamma > 0$, the Gaussian distribution

$$G_{K,\Gamma}(\cdot | z) = N((K + \Gamma^{-1})^{-1}Kz, (K + \Gamma^{-1})^{-1})$$

**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 $N(0, \Gamma)$ 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 $X^n|h \sim P_{\sqrt{n}h}$ (LAN family), $P_n(\cdot) := \int P_{\sqrt{n}h}(\cdot)d\pi_{\Gamma,c}(h)$ the marginal distribution of $X^n$. Then, for all $\epsilon > 0$, there exist $C, N < +\infty$ such that for all $n \geq N$, $c \geq C$,

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

**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 $\alpha$-sublevel set $\{x : L(x) \leq \alpha\}$ is convex.

**Example 2.1.** $L(x) = \frac{1}{2}||x||^2 \wedge B$ 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 $X \sim N(\mu, \Sigma)$. Then:

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

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

**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 $\theta_0$ with precision $K_{\theta_0} \geq 0$. Then, with $W \sim N(0, I_k)$,

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

**Remark** Consider a quadratic mean differentiable family $\{P_{\theta}\}_{\theta \in \Theta}$ with Fisher information $I_{\theta_0}$ at parameter $\theta_0$. Then the theorem implies that:

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

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

**Proof of Theorem 2.1.**
Without loss of generality, assume that $L$ takes values in $[0, 1]$ and $\theta_0 = 0$.

Observe that

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

where $\theta = \frac{h}{\sqrt{n}}$, for any $\pi$ 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 $X^n$:

$$\check{P}_n(x) = \int P_{\frac{h}{\sqrt{n}} \sim h \sim \sqrt{n}}(\cdot) d\pi^{\Gamma,c}(h)$$

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

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

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

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

Observe that:

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

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

Moreover, by Anderson’s lemma, we have:

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

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

$$(\ast) \geq \mathbb{E} [L(Z)] - \epsilon$$