36-709: Advanced Statistical Theory I **Spring 2019**

Lecture 1: January 15

Lecturer: Alessandro Rinaldo *Scribe: Elan Rosenfeld*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

1.1 Recap of Parametric Statistical Models

Definition 1.1

$$P = \{P_\theta : \theta \in \Theta\} \tag{1.1}$$

where $\Theta \subseteq \mathbb{R}^d$ *and* P_θ *is a probability distribution on* $\mathbb{R}^s$*.*

1.1.1 Example: Normal

$\Theta = \{(\mu, \Sigma) : \mu \in \mathbb{R}^k, \Sigma \in S_+^k\}$ where S_+^k is the cone of PD $k \times k$ matrices.

(Recall that a matrix M is PD $\iff x^T M x > 0$ for all $x \in R^k \neq 0$)

Then $P_\theta \sim \mathcal{N}(\mu, \Sigma)$ and $dim(\Theta) = k + \frac{k(k+1)}{2} = \frac{k^2}{2} + \frac{3}{2}k$.

1.1.2 Example: Linear Regression

$Y \sim \mathcal{N}(X\beta, \sigma^2 I_n)$ where $Y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n\times d}$, $\beta \in \mathbb{R}^{d\times 1}$, and $\sigma > 0$.

Model: $Y = X\beta + \epsilon$ where $\epsilon = (\epsilon_1, ..., \epsilon_n)$ i.i.d. from $\mathcal{N}(0, \sigma^2)$.

Observe $\underset{\sim}{X} = (x_1, ..., x_n)$ i.i.d. from P_{θ_0}

Goal: Draw inference on θ_0.

Important Assumption: P, θ_0 are fixed as $n \to \infty$. In high dimensional statistics we assume $d \to \infty$ as $n \to \infty$. In non-parametric statistics we assume P grows as $n \to \infty$.

1.1.3 Master Theorem for Parametric Models (from Jon Wellner's Notes)

Found at

`http://www.stat.cmu.edu/~arinaldo/Teaching/36709/S19/Wellner_Notes.pdf`

There is a mistake in the notes, find it for extra credit on HW1!

1.1.3.1 Assumptions

Let

1. p_θ be the density for the distribution P_θ
2. $L_n(\theta|\underset{\sim}{X}_n) = \Pi p_\theta(x_i)$ be the likelihood function, $\ell_n(\theta|\underset{\sim}{X}_n) = \log L_n(\theta|\underset{\sim}{X}_n) = \sum \log p_\theta(x_i)$
3. $\nabla_\theta \ell_n(\theta)$ be the gradient of $\ell_n(\theta)$, $H\ell_n(\theta)$ be the Hessian of $\ell_n(\theta)$
4. $I(\theta) = -\mathbb{E}_x[H\ell_n(\theta|x)]$ be the Fisher Information

Then under certain regularity conditions (smoothness, identifiability, ...) on P, let $\widetilde{\Theta}_n$ be a solution to the score equation $\nabla \ell_n(\theta) = 0$ (i.e., the MLE).

We have that:

- $\widetilde{\Theta}_n$ exists and $\widetilde{\Theta}_n \xrightarrow{p} \Theta_0$ (WLLN)
- $\sqrt{n}(\widetilde{\Theta}_n - \Theta_0) \xrightarrow{d} \mathcal{N}_d(0, I^{-1}(\Theta_0))$ (CLT)
- $2\log \widetilde{\lambda}_n \xrightarrow{d} \chi^2_d$ where $\widetilde{\lambda}_n = \frac{\ell_n(\widetilde{\Theta}_n|\underset{\sim}{X}_n)}{\ell_n(\Theta_0|\underset{\sim}{X}_n)}$ (Wilk's Theorem)
- $\sqrt{n}(\widetilde{\Theta}_n - \Theta_0)^T \hat{I}_n(\widetilde{\Theta}_n)\sqrt{n}(\widetilde{\Theta}_n - \Theta_0) \xrightarrow{d} \chi^2_d$ (Wald Test)

Some points of note:

- These results are asymptotic!
- *Critical:* Require P and number of parameters to be fixed as $n \to \infty$

1.2 High-Dimensional Statistical Models

Definition 1.2 *A high-dimensional parametric statistical model is a sequence of parametric statistical models $\{P_n\}_{n=1}^{\infty}$ where for each n, the sample space has size S_n and the parameter space has dimension d_n, where S_n, d_n are allowed to grow with n.*

1.2.1 Example: Linear Regression

$(Y_1, X_1), ..., (Y_n, X_n)$ are n R.V.s in $\mathbb{R} \times \mathbb{R}^{d_n}$ such that $Y_i = X_i^T\beta + \epsilon_i$ where $(\epsilon_1, ..., \epsilon_n) \sim \mathcal{N}(0, \sigma^2 I_n)$ and $\beta \in \mathbb{R}^{d_n}$.

1.3 Different Types of Parametric Models

1. Fixed-d models (what we've worked with before)

2. High-dimensional models

(1a) d_n is allowed to change but $d_n \in o(n)$

($x_n \in o(y_n) \iff \forall \epsilon > 0.\ \exists n_0$ s.t. $\forall n > n_0.\ |\frac{x_n}{y_n}| < \epsilon$ [i.e., $|\frac{x_n}{y_n}| \to 0$])

See work by Portnoy on these models

(1b) $d_n \gg n$

Not generally possible without additional structural assumptions (sparsity, data near a low-dimensional manifold, etc.)

Recommended Reading / Class Sources

[] M.J. Wainwright, "High-Dimensional Statistics: A Non-Asymptotic Viewpoint," *Cambridge Series in Statistical and Probabilistic Mathematics*, 2019.

Note: This text has yet to be published! The author has generously provided Prof. Rinaldo with advance copies of some chapters, so *do not* **distribute this material outside of the class**

[] R. Vershynin, "High-Dimensional Probability: An Introduction with Applications in Data Science," *Cambridge Series in Statistical and Probabilistic Mathematics*, 2018.