Class Notes (experimental)

Jonathan Rosenblatt

April 7, 2015

Contents

Estimation

In this section, we present several estimation principeles. Their properties are not discussed, as the section is merely a reminder and a preparation for the Learning.

Moment matching

The fundamental idea: match empirical moments to theoretical. I.e., estimate

$$E[g(X)]$$

by

$$\mathbb{E}[g(X)]$$

where $\mathbb{E}[g(X)] := \frac{1}{n}\sum_i g(X_i)$, is the empirical mean.

Example: Exponential Rate

Estimate λ in $X_i \sim exp(\lambda)$, $i = 1, \dots, n$, i.i.d. $E[X] = 1/\lambda$. $\Rightarrow \hat{\lambda} = 1/\mathbb{E}[X]$

Example: Linear Regression

Estimate β in $Y \sim \mathcal{N}(X\beta, \sigma^2 I)$, a p dimensional random vector. $E[Y] = X\beta$ and $\mathbb{E}[Y] = y$. Clearly, moment mathing won't work because no β satistifes $X\beta = Y$. A technical workaround: Since β is p dimensional, I need to find some $g(Y) : \mathbb{R}^n \mapsto \mathbb{R}^p$. Well, $g(Y) := XY$ is such a mapping. I will use it, even though my technical justification is currently unsatisfactory. We thus have: $E[X'Y] = X'X\beta$ which I match to $\mathbb{E}[X'Y] = X'y$:

$$X'X\beta = X'y \Rightarrow \hat{\beta} = (X'X)^{-1}X'y.$$

Quantile matching

The fundamental idea: match empirical quantiles to theoretical. Denoting by $F_X(t)$ the CDF of X, then $F_X^{-1}(\alpha)$ is the α quantile of X. Also denoting by $\mathbb{F}_X(t)$ the Empirical CDF of $X_1, \ldots, X_n$, then $\mathbb{F}_X^{-1}(\alpha)$ is the α quantile of $X_1, \ldots, X_n$. The quantile matching method thus implies estimating

$$F_X^{-1}(\alpha)$$

by

$$\mathbb{F}_X^{-1}(\alpha).$$

Example: Exponential rate:

Estimate λ in $X_i \sim exp(\lambda)$, $i = 1, \ldots, n$, i.i.d.

$$F_X(t) = 1 - \exp(-\lambda t) = \alpha \Rightarrow F_X^{-1}(\alpha) = \frac{-\log(1-\alpha)}{\lambda} \Rightarrow F_X^{-1}(0.5) = \frac{-\log(0.5)}{\lambda} \Rightarrow \hat{\lambda} = \frac{-\log(0.5)}{\mathbb{F}_X^{-1}(0.5)}$$

.

Maximum Likelihood

The fundamental idea is that if the data generating proces (i.e., the **sampling distribution**) can be assumed, then the observations are probably some high probability instance of this process, and not a low probability event: Let $X_1, \ldots, X_n \sim P_\theta$, with density (or probability) $p_\theta(X_1, \ldots, X_n)$. Denote the likelihood, as a function of θ: $L(\theta) : p_\theta(X_1, \ldots, X_n)$.. Then $\hat{\theta}_{ML} := argmax_\theta\{L(\theta)\}$.

Example: Exponential rate:

Estimate λ in $X_i \sim exp(\lambda)$, $i = 1, \ldots, n$, i.i.d. Using the exponential PDF and the i.i.d. assumption

$$L(\lambda) = \lambda^n \exp(-\lambda \sum_i X_i)$$

.

M-Estimation (Empirical Risk Minimization)

From Estimation to Learning