Nonparametric and Semiparametric Data Modeling

MA750 Lecture Notes 1

Ashis Gangopadhyay

Boston University

Introduction

Nonparametric and Semiparametric approachs to modeling data has several advantages over their parametric counterparts.

- Provides a flexible method of making inferences without reference to a fixed parametric model.
- Local nature of the modeling allows easy identification of hidden features of the data.
- Allows the data to "speak for itself".

Example: Human Growth vs Age

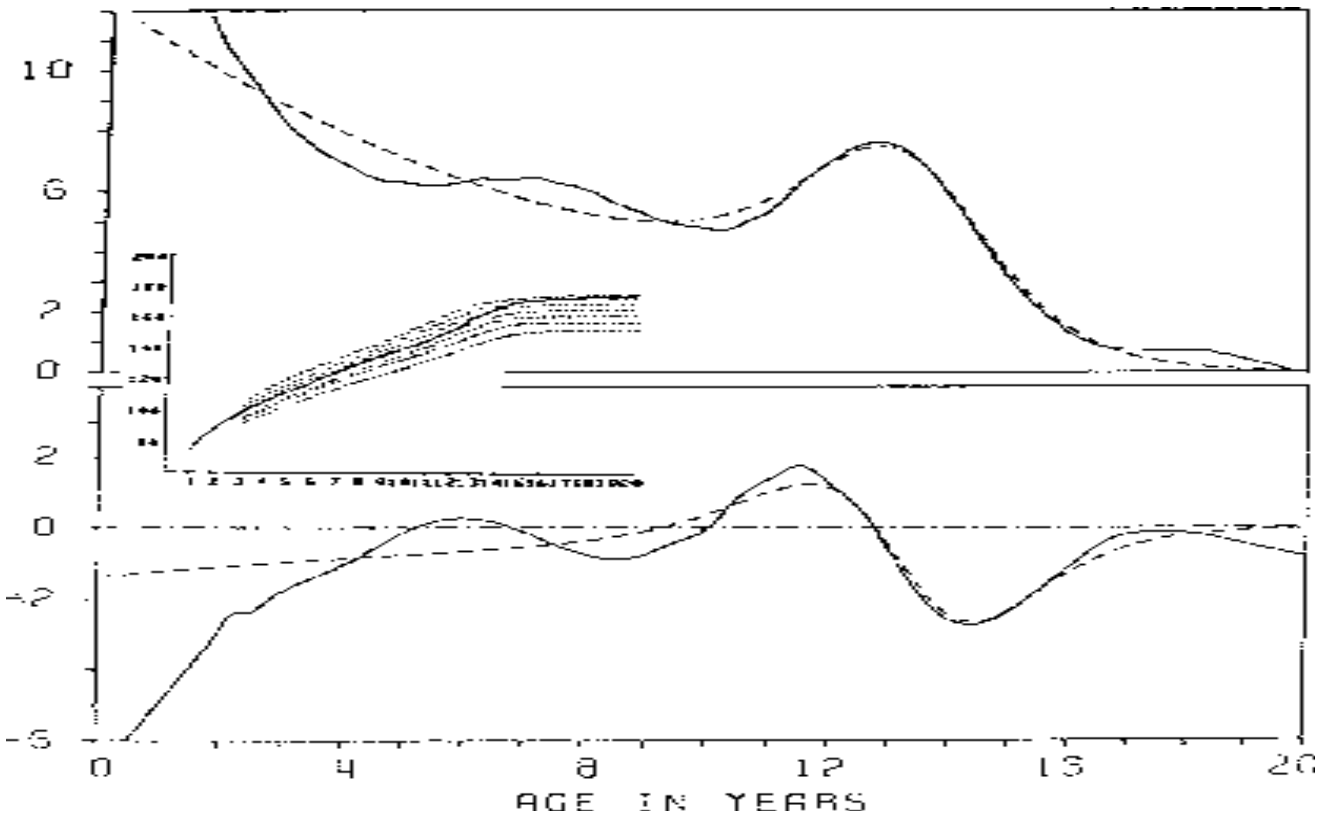

Figure 1: The small graph gives raw data of height connected by straight lines (solid line) with cross-sectional sample quantiles (dashed lines). Velocity of height growth of a girl (above) and acceleration (below) modeled by a nonparametric smoother (solid line) and a parametric fit (dashed line). From Gasser and Müller (1984)

Example: Income Distribution

Parametric Fit

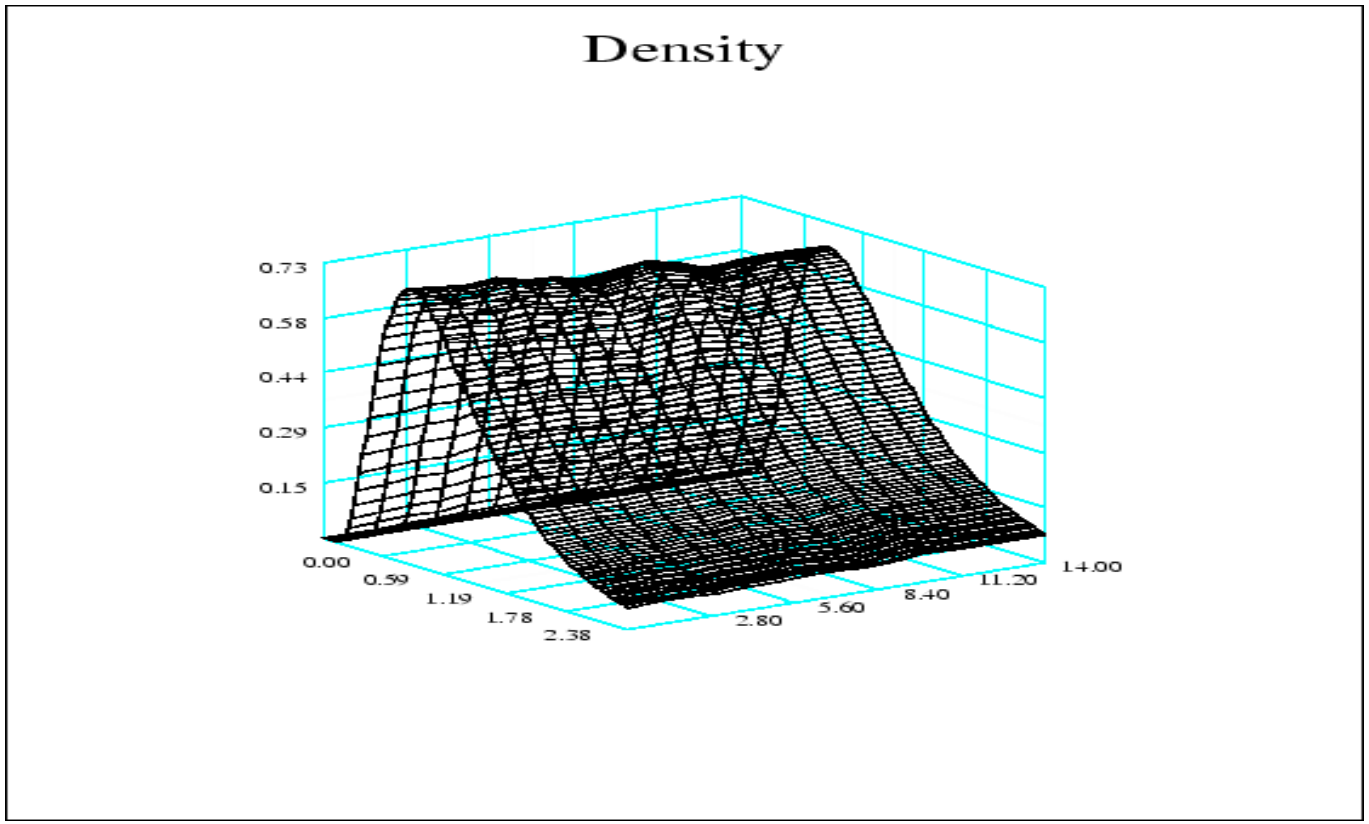

Figure 2: Net income densities over time. A parametric Singh-Madalla fit to the densities of net income from 1969 to 1983. (Hildenbrand (1986), Hardle (1992)).

Example: Income Distribution

Nonparametric Fit

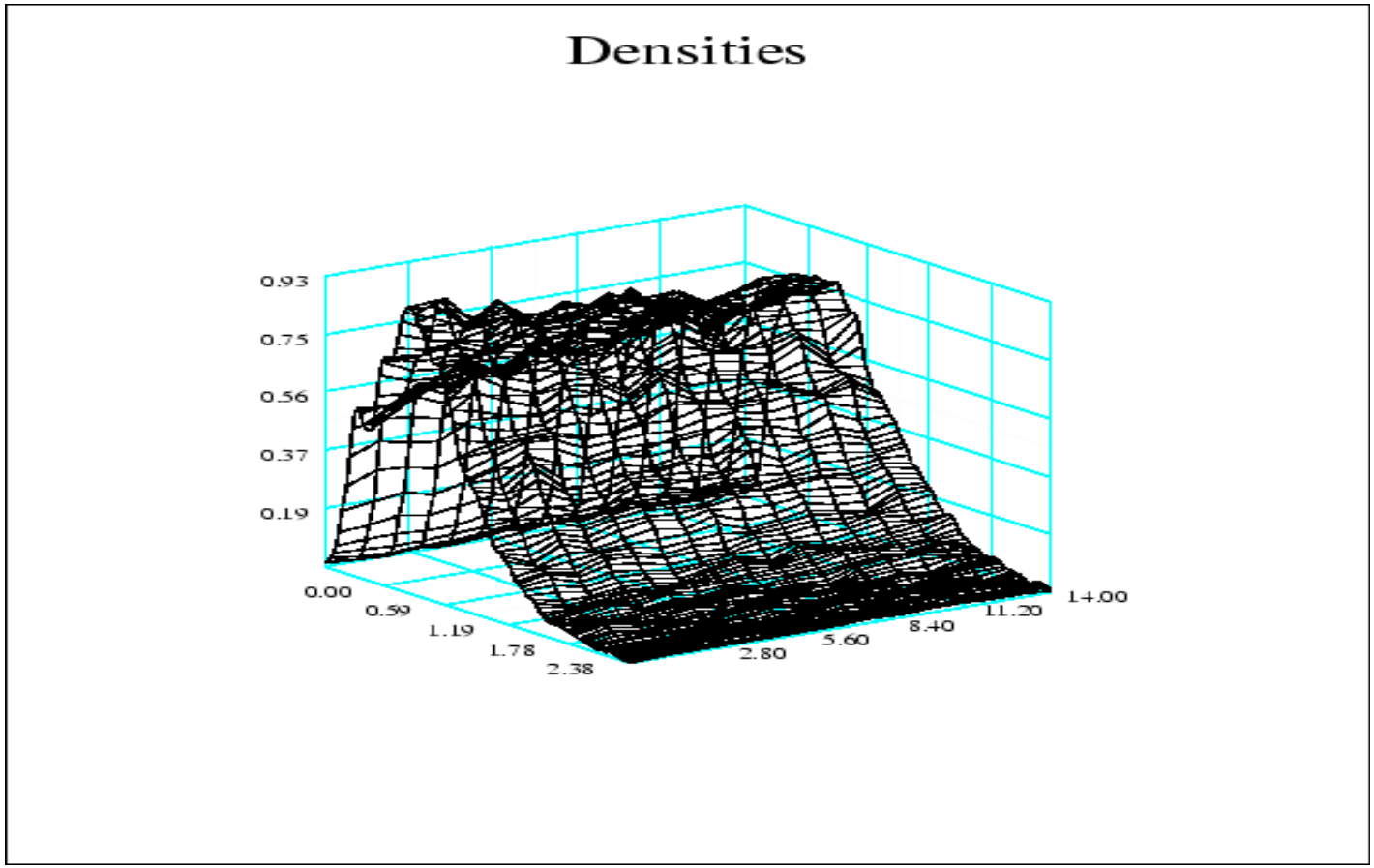

Figure 3: Net income densities over time. A nonparametric fit to the densities of net income from 1969 to 1981.

Example: Electricity Sales vs Temperature

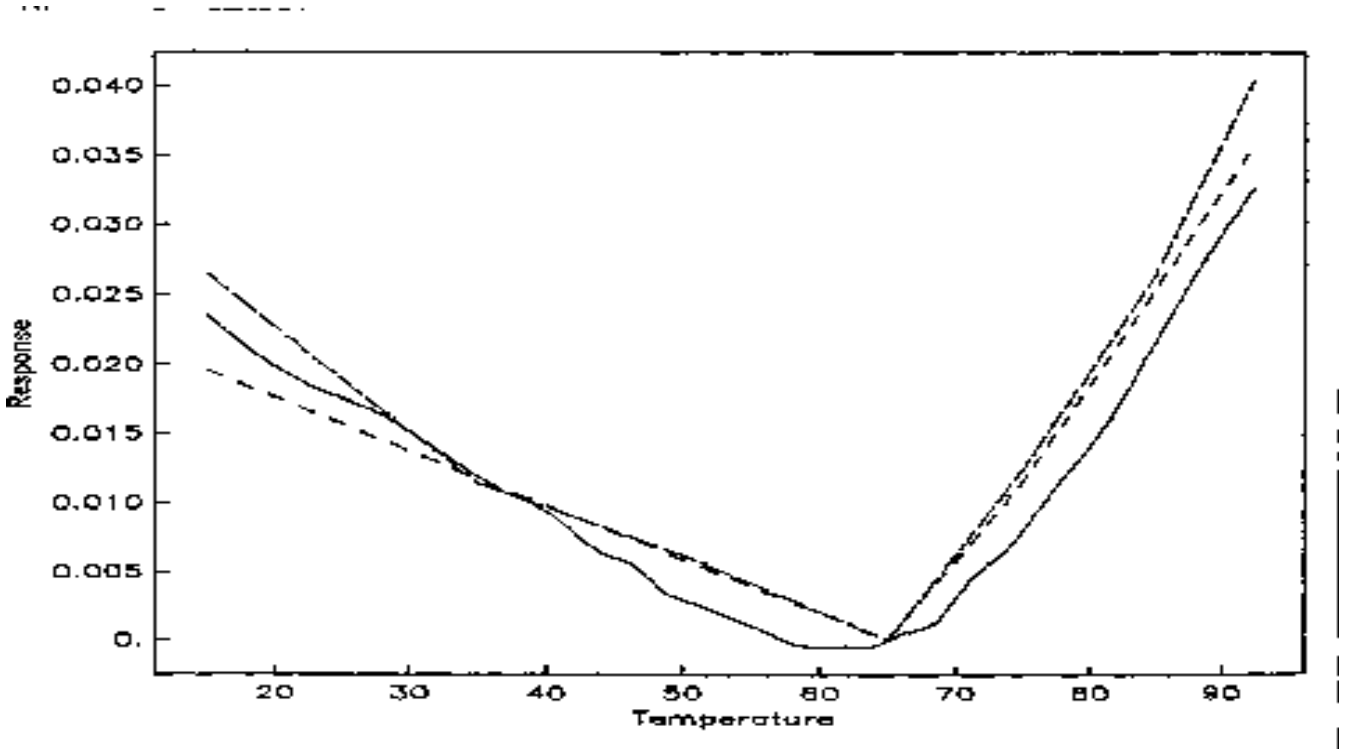

Figure 4: Temperature response function for Georgia. The nonparametric estimate is given by the solid curve and two parametric estimates by the dashed curves. Engle et al. (1986).

Example: One-day-ahead prediction of river flooding

Estmated flood probability for the St. Mary's river

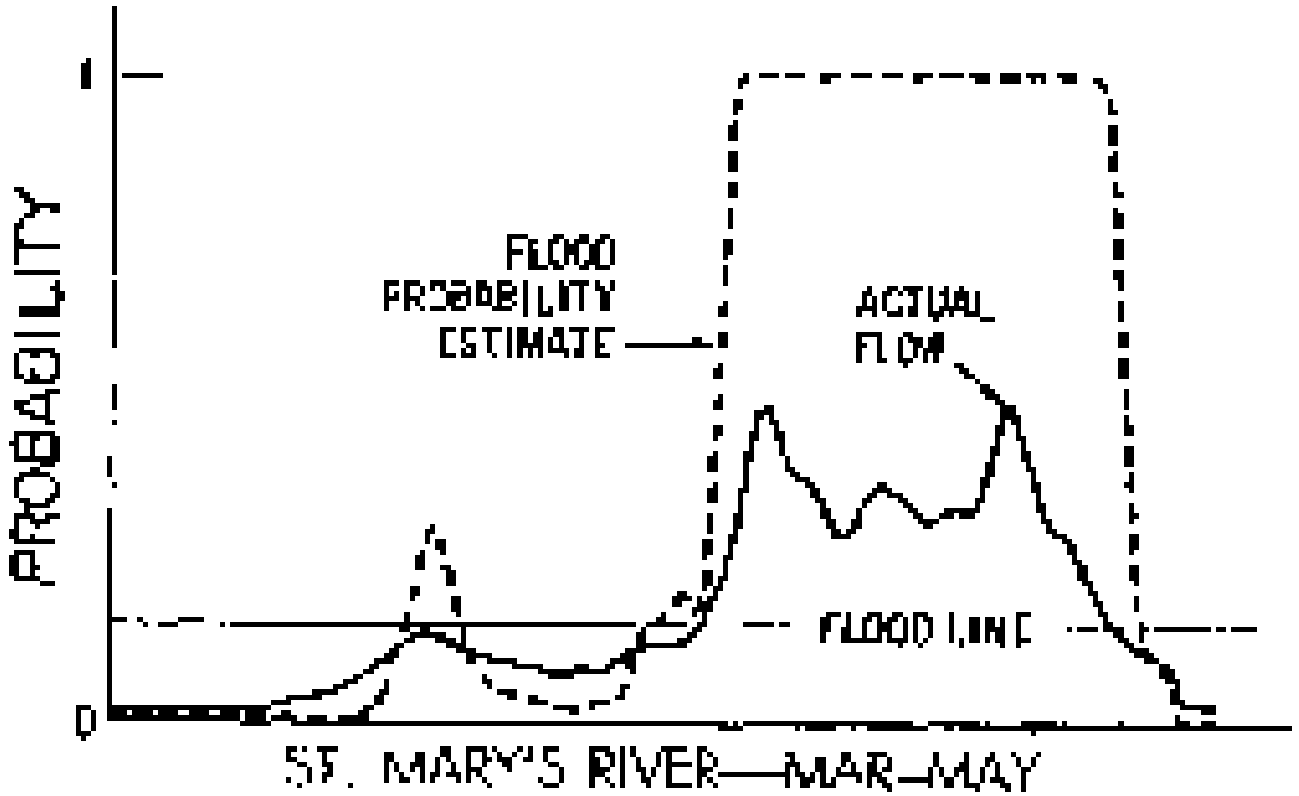

Figure 5: Nonparametric estimate of one-day-ahead flood probability (Yakowitz (1985))

Example: Binary response regression model

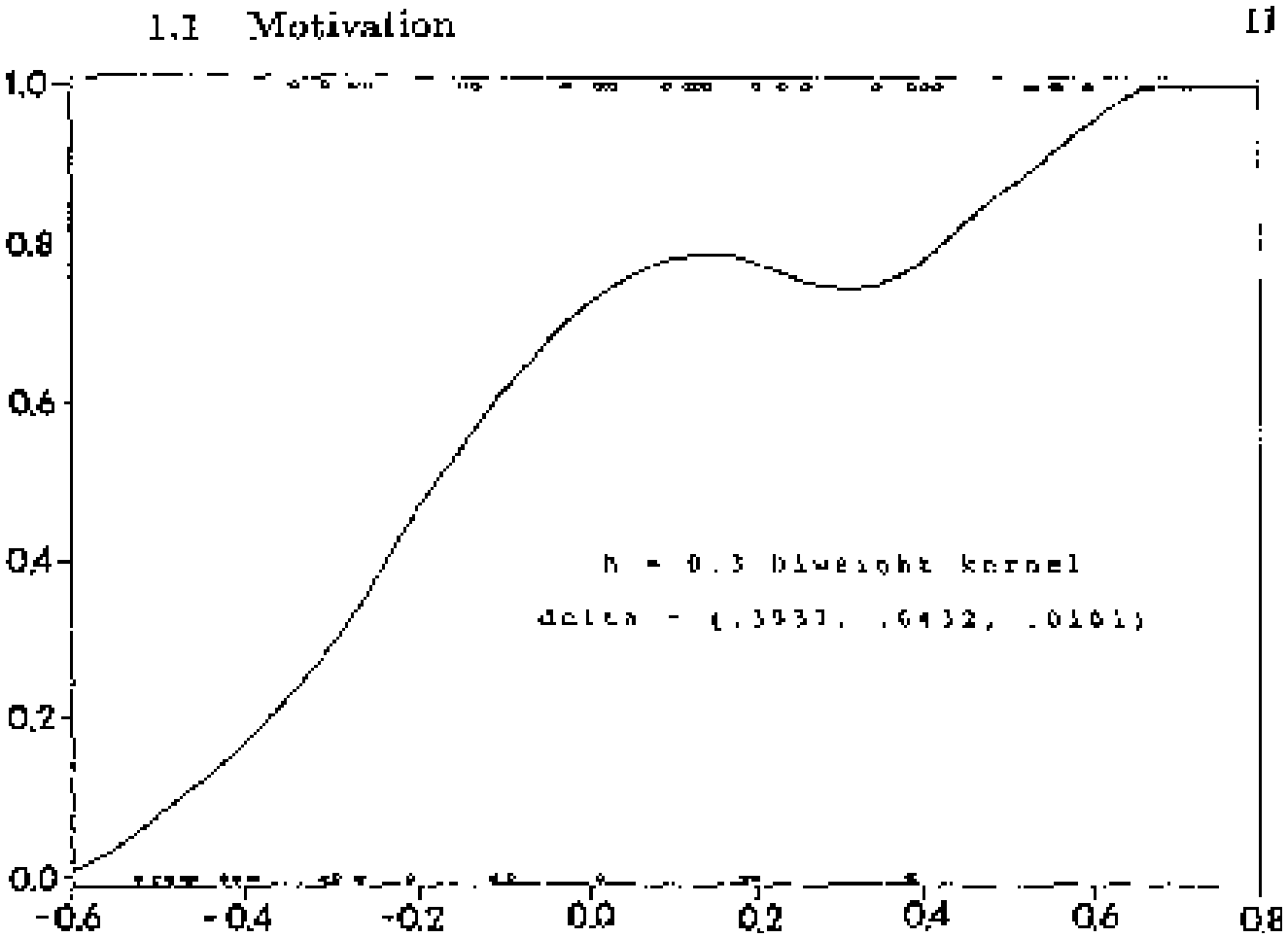

Figure 6: Indicators of fatal injury as a function of an injury stress index (X). From Härdle and Scott (1992).

Mathematical and Statistical Preliminaries

- Let $X_1, X_2, ..., X_n$ be a random sample from pdf $f(., \theta)$.
- Let $\widehat{\theta} = \widehat{\theta}(X_1, ..., X_n)$ be an estimator of θ.We will follow the standard practice and omit (in the notation only) the dependency of the estimator on the samples, i.e. we write $\widehat{\theta}$. However, it is crucial to remember that $\widehat{\theta}$ is a random variable since it is a function of n random variables.
- A desirable property of an estimator is that it is correct on average. That is, if there are repeated samplings of n samples $X_1, ..., X_n$, the estimator $\widehat{\theta}(X_1, ..., X_n)$ will have, on average, the correct value. Such estimators are called unbiased.

Definition

The bias of an estimator $\widehat{\theta}$ is $Bias(\widehat{\theta}) = E(\widehat{\theta}) - \theta$. If it is 0, the estimator is said to be unbiased.

Mathematical and Statistical Preliminaries

- There are, however, more important performance characterizations for an estimator than unbiasedness. The mean squared error (MSE) is one of the most important measure of performance of an estimator as it captures the estimation error. However, since the estimator is a random variable, we need to average over its distribution, thus capturing the average performance if there are many repeated samplings of $X_1, ..., X_n$.

Definition

The mean squared error (MSE) of an estimator is $MSE(\widehat{\theta}) = E(\widehat{\theta} - \theta)^2$.

Theorem

The mean squared error of an estimator equals
$MSE(\widehat{\theta}) = Var(\widehat{\theta}) + Bias^2(\widehat{\theta})$.

Mathematical and Statistical Preliminaries

- Since the MSE decomposes into a sum of the bias and variance of the estimator, both quantities are important and need to be as small as possible to achieve good estimation performance. It is common to trade-off some increase in bias for a larger decrease in the variance and vice-verse.

Example

Let $X_1, ..., X_n$ i.i.d. from a normal population with mean θ and variance σ^2. Consider the problem of estimating θ and θ^2

Mathematical and Statistical Preliminaries

Notation: Big O() and small o()

- Consider a sequence of real numbers a_n. We say that $a_n = O(1)$ if, as $n \to \infty$, the sequence a_n remains bounded, *i.e.*, $|a_n| \leq C$ for some constant C and for all large values of n.
- We write $a_n = o(1)$ if $a_n \to 0$ as $n \to \infty$.
- Similarly, if we write $a_n = O(b_n)$, then $a_n / b_n = O(1)$, *i.e.*, $|a_n / b_n| \leq C$ for some constant C and for all n sufficiently large.
- We write $a_n = o(b_n)$ if $a_n / b_n \to 0$ as $n \to \infty$

Example

a. If $a_n = n/(n+1)$, then $a_n = O(1)$ as $a_n \leq 1$ for all n.
b. if $a_n = 10/(n+1)$, then $a_n = o(1)$
c. If $a_n = n+5$ and $b_n = n$, then $a_n = O(b_n)$, as $a_n \leq 2b_n$ for $n \geq 5$, or $a_n \leq 5b_n$ for all n.
d. If $a_n = 1/n^2$ and $b_n = 1/n$ then $a_n = o(b_n)$.

Definition (Convergence in Probability)

Let $\{X_n\}_{n\geq 1}$ be a sequence of real random variables, and X be a random variable. We say that X_n converges to X in probability ($X_n \xrightarrow{p} X$) if for every $\varepsilon > 0$, $\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$.

Definition (Consistency of an Estimator)

An estimator $\widehat{\theta}$ ($\widehat{\theta}(X_1, ..., X_n)$) of a population parameter θ is said to be consistent if $\widehat{\theta} \xrightarrow{p} \theta$ as $n \to \infty$.

Example (Weak Law of Large Numbers)

Let $X_1.., X_n$ be a random sample from a population with mean θ and variance $\sigma^2 < \infty$. Then $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ is a consistent estimator of θ.

Definition (Convergence in Distribution)

We say that X_n converges in distribution to X ($X_n \xrightarrow{d} X$) if $\lim_{n\to\infty} F_n(x) = F(x)$ for all points of continuity of F, where $F_n(x)$ and $F(x)$ are the CDFs of X_n and X respectively.

Example (Central Limit Theorem)

Let $X_1, ..., X_n$ be i.i.d. random variables from a population with mean θ and variance $\sigma^2 < \infty$. Then $Z_n = \frac{\sqrt{n}(\overline{X}-\theta)}{\sigma} \xrightarrow{d} Z$ where Z is the standard normal random variable.

Definition

A sequence of random variables X_n is said to be bounded in probability $(O_p(1))$ if for every $\varepsilon > 0$ there exists a constant M and a positive integer n such that $P(|X_n| > M) \leq \varepsilon$ for all $n \geq N$.

That is, X_n is bounded in probability if for any arbitrarily small positive number ε, we can always find a positive constant M, such that the probability of the absolute value of X_n being larger than M is less than ϵ.

Remark

Obviously, if $X_n = O(1)$ (bounded), then $X_n = O_p(1)$. However, the converse is not true. Letting X_n denote i.i.d random variables from $N(0,1)$ distribution, then $X_n \neq O(1)$, but $X_n = O_p(1)$. In fact any random variable with well-defined cdf is an $O_p(1)$ random variable.

- Thus, we write $X_n = O_p(1)$ to indicate that X_n is bounded in probability.
- We write $X_n = o_p(1)$ to indicate that $X_n \xrightarrow{p} 0$.
- Similarly, we write $X_n = O_p(Y_n)$, if $X_n / Y_n = O_p(1)$, and $X_n = o_p(Y_n)$, if $X_n / Y_n = o_p(1)$.

Example

Let $\{X_n\}$ be such that X_i i.i.d $N(0,1)$. Let $Z_n = \sum_{i=1}^{n} X_i$.

- Then $X_n = O_p(1)$. To see this, note that for every $\varepsilon > 0$, there exists a M such that
$$P(|X_n| \leq M) = 2 \int_0^M N(x; 0, 1) dx \geq 1 - \varepsilon$$
- However, $Z_n \sim N(0, n)$. Thus, $n^{-1/2} Z_n \sim N(0,1) \Rightarrow n^{-1/2} Z_n = O_p(1)$
- Then, $Z_n = O_p(n^{1/2})$

Example (Contd.)

- It is always true that $n^{\alpha} Z_n = O_p(n^{\beta}) \Leftrightarrow Z_n = O_p(n^{\beta-\alpha})$
- It also follows that $X_n = o_p(n^{\delta})$ and $Z_n = o_p(n^{\frac{1}{2}+\delta})$ for $\delta > 0$.
- Let $Y_n = n^{-1/2}(X_n + Z_n)$. Since $n^{-1/2} X_n = O_p(n^{-1/2})$ and therefore $o_p(1)$, and $n^{-1/2} Z_n = O_p(1)$, only $n^{-1/2} Z_n$ matters in the asymptotic behavior of Y_n.