1. Statistical estimation problem.

Data: observation: X

Possible distribution $P \in \mathcal{P}$

A priori data:

$$\mathcal{P} = \{P\} = \{P_\theta, \theta \in \Theta\}$$

A posteriori data:

$$X.$$

The problem: $F : \Theta \to \Phi$,

find (estimate)

$$F(\theta).$$

Estimators

$$\widehat{F} = \widehat{F}(X)$$

The quality of estimator $\widehat{F}$

$$E_\theta \; l(F(\theta), \widehat{F}) = R(\theta)$$

Examples:

Example 1. $X_1, \ldots, X_n$
iid $\qquad f(x;\theta)$

a) Parametric case: $\theta \in \Theta \subseteq \mathbb{R}^d$

$E_\theta|\widehat{\theta} - \theta|^2 \asymp \frac{d}{n}$

b) Nonparametric case: $\theta(x)$

Example 2.

$X: \quad dX(t) = \theta(t)dt + \epsilon dw(t)$

$\theta(t) = \theta(t,x), \quad x \in G \subseteq \mathbb{R}^k$
$w(t) = w(t,x)$

α) Parametric case

$\theta(t) = f(t; \theta_1, \ldots, \theta_d)$

$E_\theta |\widehat{\theta} - \theta|^2 \asymp \varepsilon^2 d$

β) Nonparametric case

2. How to construct estimators in the nonparametric case?

Data: X, $\{P_\theta, \theta \in \Theta\}$.

Θ is a metric space.

Estimate θ. The risk function is

$$R(\theta) = E_\theta \; \rho(\theta, \widehat{\theta})$$

Let

$$\varphi : \; \Theta \to T$$

where T is an "estimable" set (finite, finite dimensional etc)

Suppose that

$$\sup_\theta diam\{\varphi^{-1}\varphi(\theta)\} \leqslant \delta.$$

If

$$\hat{\theta} = \varphi^{-1}(\hat{\varphi}),$$

the error is determined by

$$E\ dist(\hat{\varphi}, \varphi(\theta)) \tag{$*$}$$

and

$$diam(\varphi^{-1}\varphi(\theta)) \tag{$**$}$$

($*$) is a probability problem, ($**$) is a problem of approximation theory.

Here <u>widths</u> enter.

For example,

$$\inf_{T,\varphi} \sup_{\theta\in\Theta} \ diam(\varphi^{-1}\varphi(\theta)),$$

T is a compact, $dim\,T \leqslant n$, is the n-th Aleksandroff width.

Suppose that $\Theta \subseteq B$ - Banach space. Take $T \subseteq B$. If T is finite, Kolmogorov's ε-entropy naturally enters.

If $T = L_n$ is an n-dimensional linear manifold, Kolmogorov n-width enters

$$\inf_{L_n,\varphi} \sup_{\theta} \|\theta - \varphi(\theta)\|$$

and the whole error is

$$\inf [E_\theta \|\widehat{\varphi} - \varphi(\theta)\| + \sup \|\theta - \varphi(\theta)\|]$$

Examples. Observation

$$du(t) = \theta(t)dt + \epsilon \; dw(t)$$

$$\theta(t) = \theta(t; x) = \theta(t; x_1, \ldots, x_d)$$

$$\theta \in \Theta$$

Likelihood function

$$
\begin{aligned}
l(\theta) &= \ln \frac{dP_\theta}{dP_0}(u) \\
&= \frac{1}{\varepsilon^2}\int_0^T <\theta(t), du(t)> - \frac{1}{2\,\varepsilon^2}\int_0^T |\theta(t)|^2 dt
\end{aligned}
$$

Here

$$
<\theta_1(t), \theta_2(t)> = \int_G \theta_1(t;x)\theta_2(t;x)dx
$$

$$
\int_a^b <f(t), dw(t)> = \lim \sum <f(t_i), w(t_{i+1}) - w(t_i)>
$$

Let $M \subseteq L_2\{[0,T] \times G\}$,

(M, ρ) – a metric space,

$$\Phi : \Theta \to M.$$

The problem: estimate $\Phi(\theta)$.

Define

$$\omega(\delta) = \sup\{\rho(\Phi(\theta_1), \Phi(\theta_2)) : \|\theta_1 - \theta_2\| \leqslant \delta\}$$

Φ^*_ϵ – is the maximum likelihood estimator (MLE). Then

$$\sup_\theta P_\theta\{\rho(\Phi^*_\varepsilon, \Phi(\theta)) > 2\,\omega(3\,\delta)\}$$

$$\leqslant \sqrt{\frac{\varepsilon}{\delta}} \exp\left\{\frac{1}{2} C_\delta(\theta) - \frac{\delta^2}{16\,\varepsilon^2}\right\}.$$

Example. The same problem

$$du(t) = \theta(t)dt + \varepsilon dw(t)$$

Approximation by linear manifolds L_n of dim n. Kolmogorov's n-width enter:

$$d_n(\Theta) = \inf_{L_n} \sup_{\theta} \inf_{y \in L_n} \|\theta - y\|$$

Theorem .

$$\Delta_\varepsilon(\Theta) = \inf_{\theta_\varepsilon} \sup_{\theta} E_\theta \|\theta_\varepsilon - \theta\| \leqslant c \cdot \inf_n (\varepsilon\sqrt{n} + d_n(\Theta))$$

Proof. $\psi_1, \ldots, \psi_n$ basis in L_n.

$$\left\| \theta - \sum_1^n (\theta, \psi_j)\psi_j \right\| \leqslant (1+\gamma) d_n$$

Estimator for $a_j = (\theta, \psi_j)$ is

$$\int_0^T < \psi_j, dw(t) >= a_j^*$$

$$\theta_\varepsilon^* = \sum a_j \psi_j$$

$$E_\theta \|\theta_\varepsilon^* - \theta\|^2 \leqslant \sum_1^n E|a_j - a_j^*|^2 + (1+\gamma)^2 d_n^2 \ .$$

$\square$

Lower bounds.

1. Shannon's capacity:

$$du = \theta dt + \varepsilon dw$$

but now θ is a random field independent on w

$$\mathcal{C}_\varepsilon(\Theta) = \sup_{\theta \in \Theta} I(u, \theta)$$

[Shannon's information is

$$I(\xi, \eta) = E\left\{ \ln \frac{dP_{\xi\eta}}{dP_\xi \times dP_\eta}(\xi, \eta) \right\}$$]

$$\mathcal{C}(\theta) \leqslant \frac{1}{2\,\varepsilon^2} \sup_{\theta \in \Theta} E\|\theta\|^2$$

Theorem .

$$\inf_{\Phi^*} \sup_{\theta} P_\theta\{\rho(\Phi^*, \Phi(\theta)) > \delta\}$$

$$\geqslant 1 - \frac{\mathcal{C}_\varepsilon(\Theta) + 1}{C_{2\,\delta}(\Phi(\Theta)) - 1}.$$

2. Bernstein's widths.

Let $\Sigma \subseteq B$.

$$b_n(\Sigma) = \sup_{M_{n+1}} \{r : \, rO \cap M_{n+1} \subseteq \Sigma\}$$

Theorem .

$$\inf_{\theta_\varepsilon} \sup_{\theta} E_\theta \|\theta_\varepsilon - \theta\|$$

$$\geqslant \varepsilon^2 \inf_{x \geqslant 1} \left(1 - \frac{\tanh x}{x}\right) \sup \left\{ n : \, \frac{b_n(\theta)}{\varepsilon\sqrt{n+1}} \geqslant 1 \right\}$$

Classes of smooth functions:

$$\theta(t;x) = \theta(x)$$

$$G = [0,1]^d$$

$$\theta(x_1, \ldots, x_d) - \text{periodic}$$

Let $P(z) = P(z_1, \ldots, z_d)$ be a polynomial.

$$W_2^P : \{\theta : |P(D)\theta|_2 \leqslant 1\}$$

where $\qquad D = \left(\frac{1}{i}\frac{\partial}{\partial x_1}, \cdots, \frac{1}{i}\frac{\partial}{\partial x_d}\right).$

1. $\quad P(z) = \left(\sum c_j z_j^{\beta_j}\right)^k$

$$\Delta_\varepsilon(\Theta) = \inf_{\widehat{\theta}} \sup_{\theta} E|\theta - \widehat{\theta}|_2 \asymp \varepsilon^{\frac{2k\beta}{2k\beta+1}}$$

$$\beta : \ \frac{1}{\beta} = \frac{1}{\beta_1} + \cdots + \frac{1}{\beta_d}$$

2. $\quad P(z) = z_1^{\beta_1} \cdots z_d^{\beta_d}$

Let $\beta = \beta_1 = \ldots = \beta_l < \beta_{l+1} \leqslant \ldots \leqslant \beta_d$

$$\Delta_\varepsilon(\Theta) \asymp \varepsilon^{\frac{2\beta}{2\beta+1}} \left(\ln \frac{1}{\varepsilon}\right)^{\frac{\beta(l-1)}{2\beta+1}} .$$

An example of density estimation.

$$X_1, \ldots, X_n,\ X_j \in (\mathfrak{X}, \mathfrak{A})$$

$\theta(x)$ density with respect toμ.

Let $K(x, y)$ on $\mathfrak{X} \times \mathfrak{X}$

$$\int K(x, y) K(x, z) d\mu(x) = K(y, z)$$

be a reproducing kernel.

$$H(K) = \left\{ f \in L_2 : \; f(y) = \int K(x,y) f(x) d\mu \right\}$$

$$\mathcal{K}(\mathcal{N}, \beta) = \{K : \|K(x,x)\|_\beta \leqslant \mathcal{N}\}$$

The $(\mathcal{N}, \beta)$-width

$$\delta_{\mathcal{N}}(\Theta, \beta) = \inf_H \sup_\theta \inf_{h \in H} \|\theta - h\|$$

Example 1. $\dim H = N$

$$K(x,y) = \sum_1^N \varphi_j(x)\varphi_j(y), \qquad K \in \mathcal{K}(N, 1).$$

Example 2. The Dirichlet kernels.

$$D_\nu(x,y) = \pi^{-k} \prod_1^k \frac{\sin \nu_j (x_j - y_j)}{x_j - y_j}$$

where k is the dimension of x, y.

$$\delta_N(\Theta, \infty) \leqslant \inf_\nu \sup \mathcal{E}_\nu(\theta)$$

where $\mathcal{E}_\nu$ is the best approximation error by entire functions of order ν.

Theorem . *Let* $\delta(N) = \inf_{\beta \geqslant 2} \delta_N(\Theta, \beta)$*. Then*

$$\inf_{\theta_\varepsilon} \sup_\theta E_\theta \|\theta - \theta_\varepsilon\|^2 \leqslant \inf_N \left(4\,\delta_N^2 + \frac{N}{n} \left(1 + \sup_\theta \|\theta\|^2 \right) \right)$$

Example: let

$$\theta(x) = \frac{1}{(2\,\pi)^d} \int_K e^{-i(\lambda,x)} \tilde{\theta}(\lambda) d\lambda.$$

Then

$$\inf_{\widehat{\theta}_n} \sup_\theta E_\theta \|\theta - \theta_\varepsilon\|^2 \sim \frac{mesK}{(2\,\pi)^d} \cdot \frac{1}{n} \; .$$