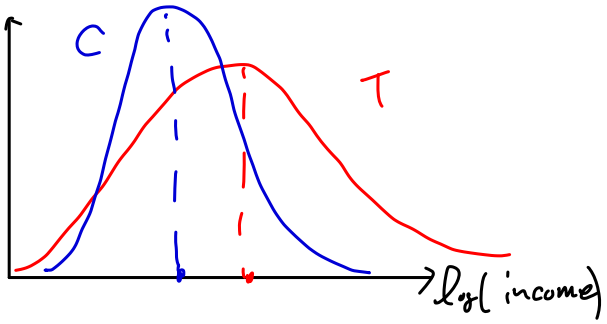
中級ミクロデータサイエンス受講ノート
分位回帰

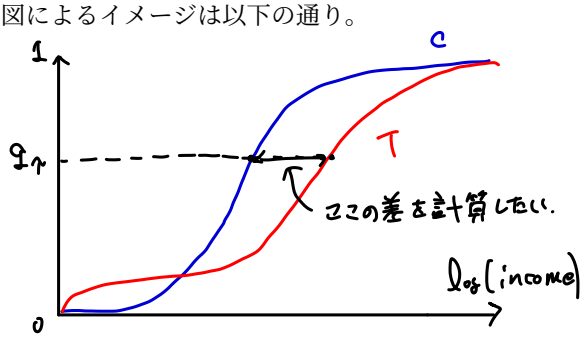
花澤 楓
学籍番号：2125242

2023 年 12 月 19 日

1　分位回帰（quantile regression）

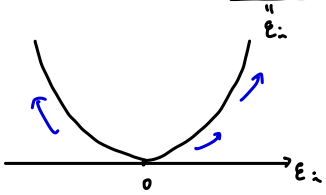
分位回帰とは、X_i を条件とした時の Y_i の条件付分位点を分析するもので、特に、格差を分析するために重要である。また、特定の個人についての推定値ではないことに注意する。図によるイメージは以下の通り。

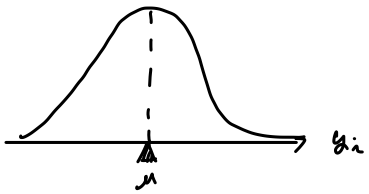




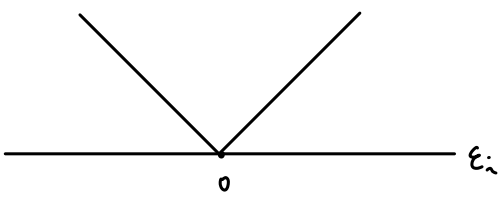
1.1　最適化問題

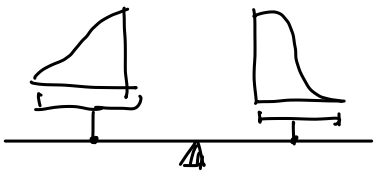
- 平均値 μ：$\min_\mu \ \sum_i^N (y_i - \mu)^2$ を考えたものが、最小二乗法（OLS）。特に、$\mu_i = \beta_0 + \beta_1 X_i$ とする。





- 中央値 $q_{1/2}$：$\min_{q_{1/2}} \ \sum_i^N |y_i - q_{1/2}|$ の解が中央値となる。これは、Least Absolute Deviation(LAD) と呼ばれる。

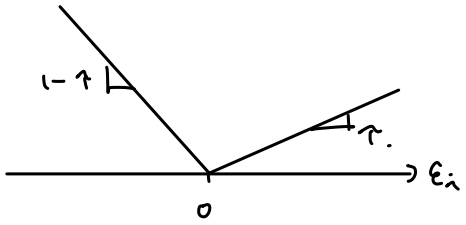




- τ 分位点 q_τ：$\min_{q_\tau} \sum_i^N \rho(y_i - q_\tau)$ の解が τ 分位点 q_τ である。これは、ある τ の元での OLS をするイメージで、$q_\tau = \beta_{0\tau} + \beta_{1\tau} X$ で分析される。ここで、

$$\rho(\varepsilon_i) = \begin{cases} \tau\varepsilon_i & \text{if } \varepsilon_i \geq 0 \\ -(1-\tau)\varepsilon_i & \text{if } \varepsilon_i < 0 \end{cases} \tag{1}$$

で定義されるチェック関数である。形状は以下の通り。



2 カーネル密度推定（kernel density estimation）

密度関数を推定したい状況を考える。このとき、データの分布の情報を見ることになり、そこから密度関数の推定量を計算する必要があるが、その際に密度関数の関数形に仮定を置くアプローチをパラメトリック推定、関数形に仮定を置かないアプローチをノンパラメトリック推定という。パラメトリック推定では、関数形の仮定がある程度良い近似になっている場合には強力だが、仮定が不適切（misspecification）の場合には分布の重要な特徴を見逃してしまい、誤った推定量を計算してしまう可能性が高い。そのため、結果が信用できないものとなる。また、密度関数の推定量はなるべく滑らかであってほしい。そのような場合、カーネル密度推定量を使用することが推奨される。

今、データ:$X_1, \ldots, X_n$ から密度関数を推定することを考える。カーネル密度推定量は以下の式で定義される。

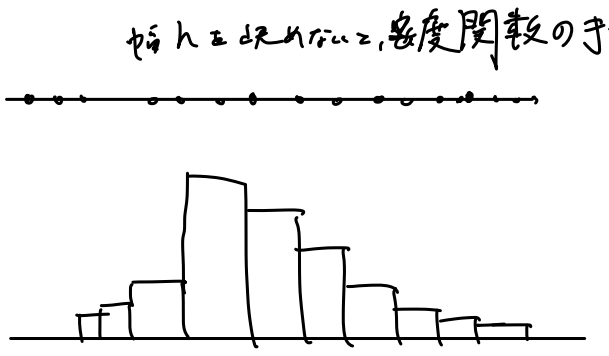
$$\hat{f}(x) = \frac{1}{nh} \sum_i^n K\left(\frac{X_i - x}{h}\right) \tag{2}$$

ここで、$K(\cdot)$ は

$$\int K(u)du = 1 \tag{3}$$

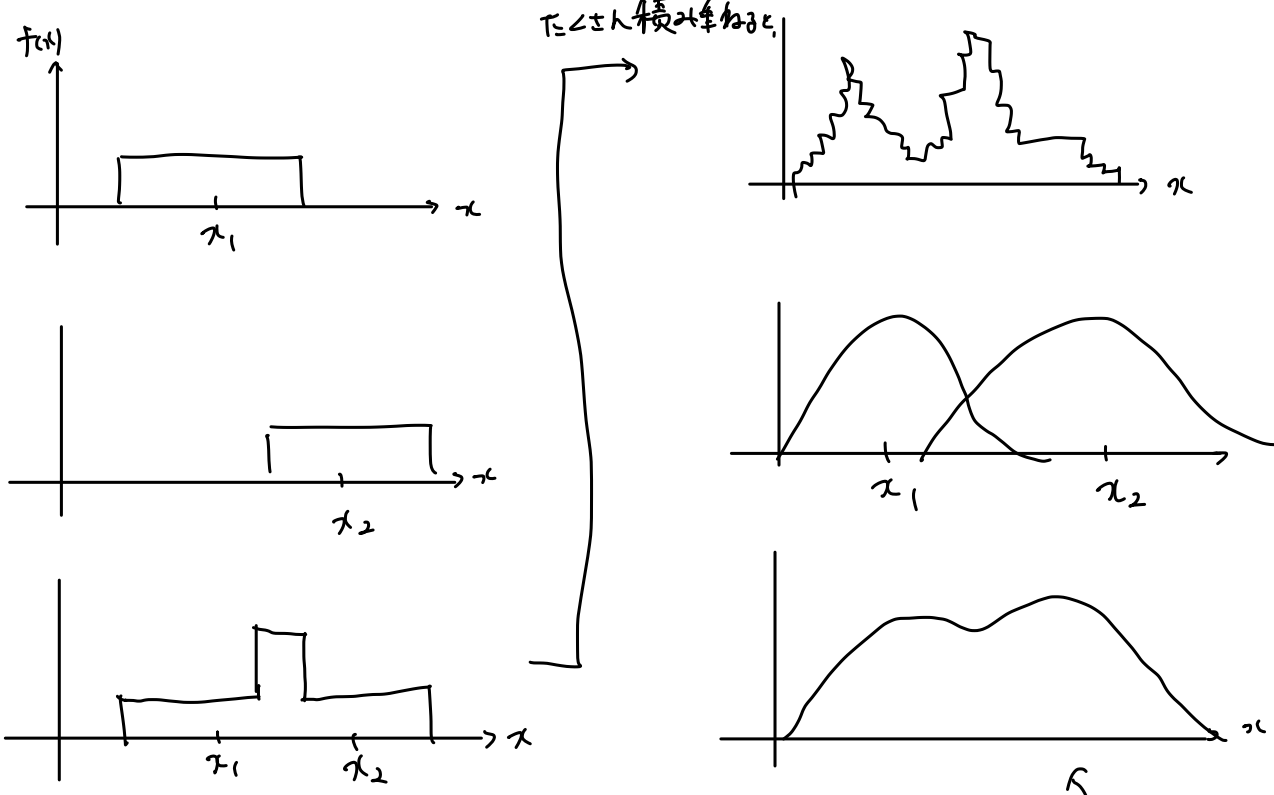
を満たす関数（カーネル関数）である。また、h は平滑化パラメータ（or バンド幅）であり、分析者が設定する必要がある（もしくは CV で最適なものを設定)。

(1) 密度：幅が必要



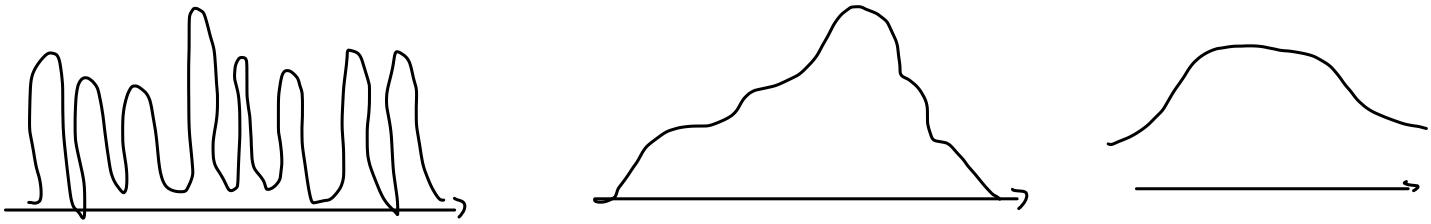
$$\hat{f}(x) = \begin{cases} \frac{1}{n} & \text{if } |x_i - x| \le \frac{h}{2} \\ 0 & \text{otherwise} \end{cases}$$ を仮定

(2) カーネル関数を積み重ねる：



(3) 最適な平滑化パラメータ：バイアスと分散はバンド幅についてトレードオフの関係

正規分布なら
より滑らかになる。
(混合正規分布)



small h.

- low bias
- high variance

medium h.

$\min_h$ bias^2 + var

large h

- high bias
- low variance