情報認識
「ノンパラメトリック法」

- 講師:　杉山　将（計算工学専攻）
- 居室:　W8E-505
- 電子メール:　sugi@cs.titech.ac.jp

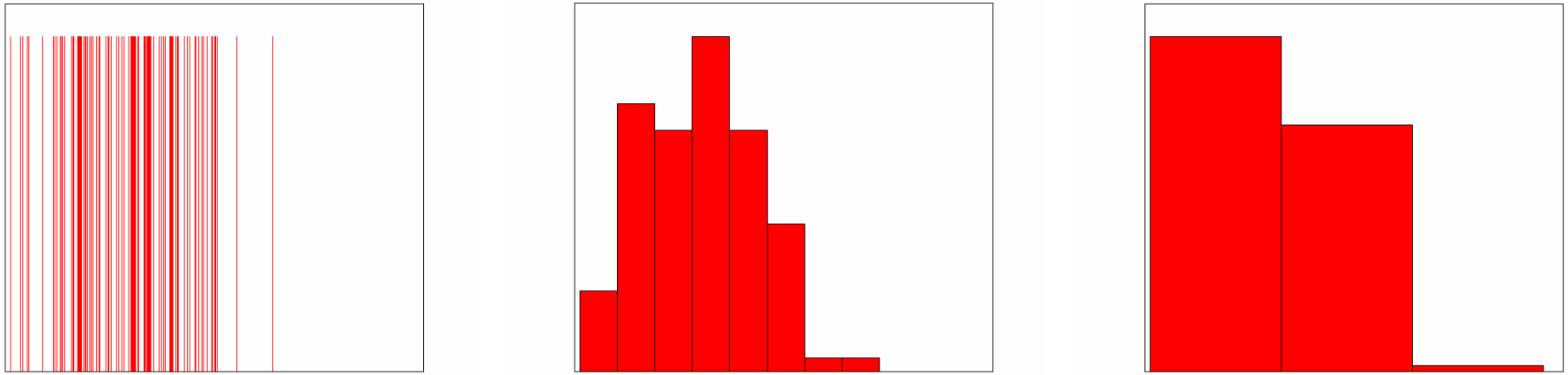
ノンパラメトリック法

- パラメトリック法：確率密度関数のモデルを仮定してその中からよいものを探す.
- ノンパラメトリック法(non-parametric method)：モデルを用いないで，より直接的に確率密度関数を推定.

ヒストグラム法

- ヒストグラム法(histogram method):単純にヒストグラムを用いて確率密度関数を推定する方法
  - パターン空間を適当に分割する
    （必ずしも同じ形に分割する必要はない）.
  - 各分割内に入る訓練標本を数える.
  - 積分が1になるように正規化する.
- 非常に単純な方法なので便利であるが・・・

ヒストグラム法の問題点



- 領域間で不連続
- 領域の分割の仕方を決めるのが難しい

もう少し工夫した方法が必要

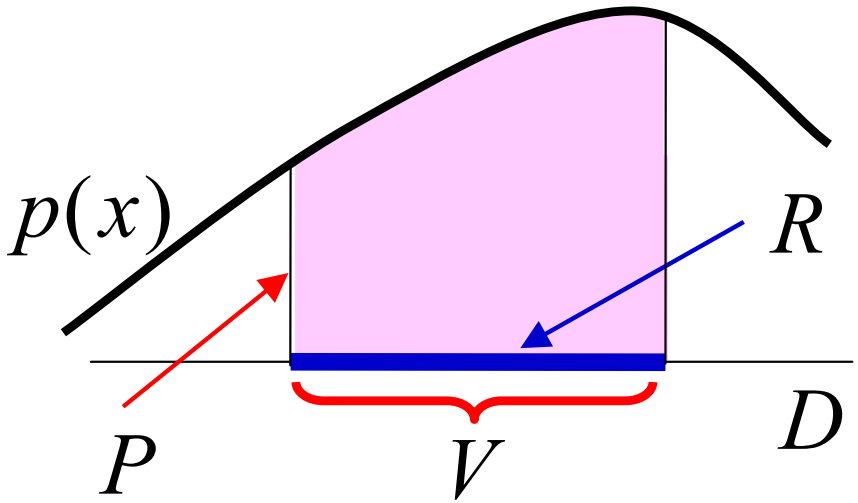
ノンパラメトリック法の表記

- R :パターン空間 D 内のある領域(region)
- V : R の体積(volume)

$$V = \int_R dx$$

- P :あるパターン x が R に入る確率

$$P = \int_R p(x)dx$$



ノンパラメトリック法の基礎

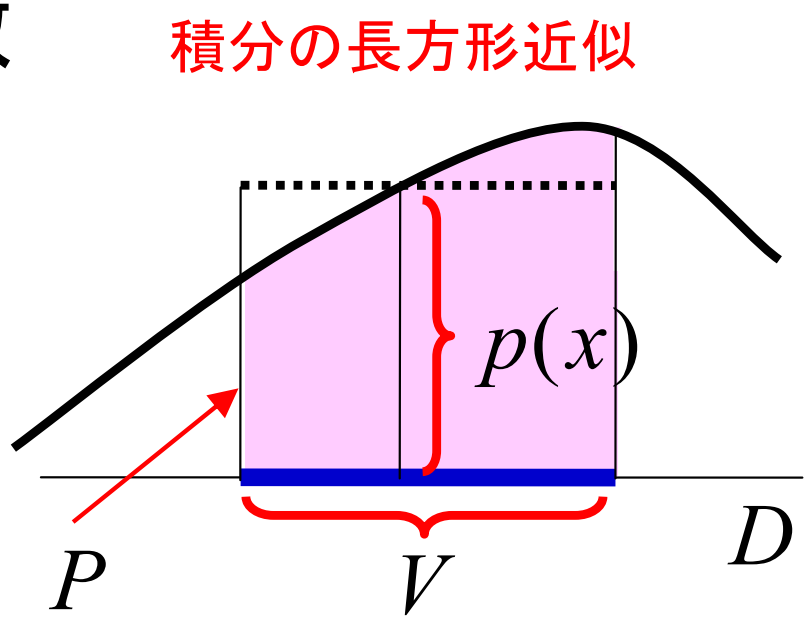
- 推定したい確率密度関数 $p(x)$ が R 内でほぼ一定値を取るとき，確率 P は領域 R 内のある点 x を用いて

$$P \approx Vp(x)$$

- k ：n個の訓練標本のうち R に入っている個数

$$P \approx k/n$$

$$\Rightarrow \quad p(x) \approx \frac{k}{nV}$$



領域の決め方

$$p(x) \approx \frac{k}{nV}$$

- 近似のよさは領域 R の決め方に依存.
- $p(x)$ が R 内でほぼ一定値をとる

  ➡ R をできるだけを小さくする
- しかし, R を小さくしすぎると $P \approx k/n$ がほとんど常に0になってしまい, 精度が悪い.

⬇

程よい大きさに R を決める必要がある

領域の決め方（続き）

$$p(x) \approx \frac{k}{nV}$$

- 訓練標本を用いて領域 R を決める.
  - 核密度推定法: V を固定したもとで k を決定
  - 最近傍密度推定法: k を固定したもとで V を決定