情報認識
「カーネル密度推定法(第12章)」

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

これまでの復習

- 最大事後確率則: $\underset{y}{\arg\max}\ p(y|x)$
- ベイズの定理より $p(y|x) \propto \underbrace{p(x|y)}_{\text{条件付き確率}}\underbrace{p(y)}_{\text{事前確率}}$
- 事前確率は各カテゴリの標本の割合で推定

$$\hat{p}(y) = n_y / n$$

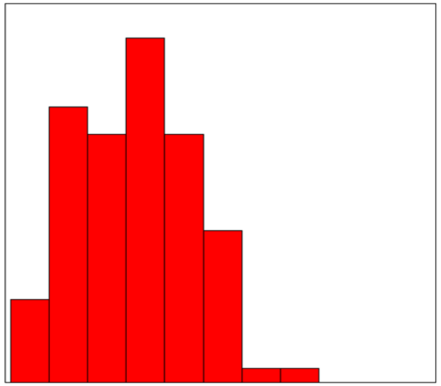
- 本講義の主題:条件付き確率をうまく推定したい
- 簡単のため，条件付きでない確率密度関数 $p(x)$ を $\{x_i\}_{i=1}^{n}$ から推定する問題を考える.

確率密度関数の推定

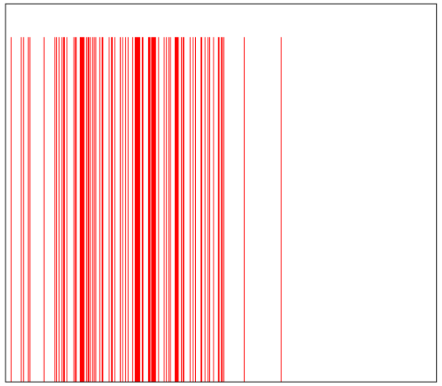
- パラメトリック法: モデルのパラメータを推定
  - 最尤推定法
  - ベイズ推定法, 最大事後確率法
- ノンパラメトリック法(non-parametric method): モデルを使わず直接確率密度関数を推定
  - カーネル密度推定法
  - 最近傍密度推定法

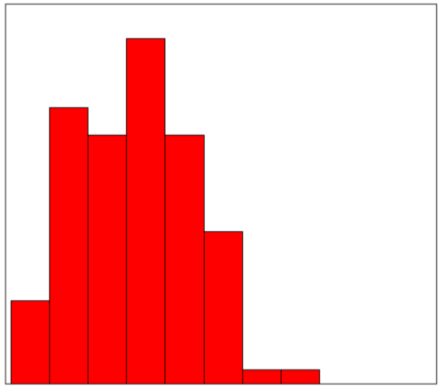
導入:ヒストグラム法

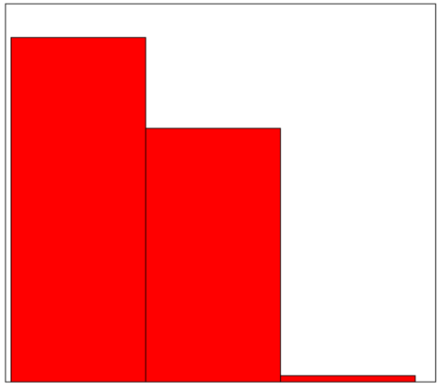
- ヒストグラム法(histogram method):単純にヒストグラムを用いて確率密度関数を推定する方法
  - パターン空間を適当に分割する
    （必ずしも同じ形に分割する必要はない）.
  - 各分割内に入る訓練標本を数える.
  - 積分が1になるように正規化する.
- 非常に単純な方法なので
  便利であるが・・・



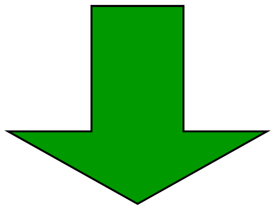
ヒストグラム法の問題点







- 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

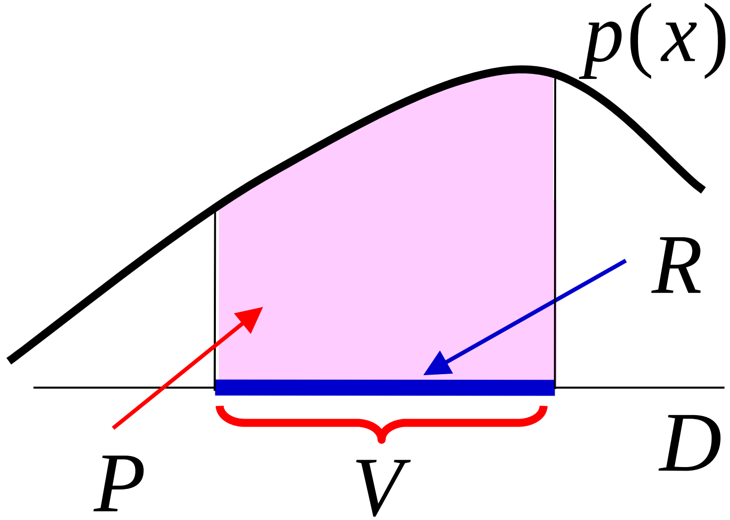
- ある注目点 x' での確率密度 $p(x')$ を推定する
- R:x'を含むパターン空間 D 内のある領域(region)
- V: R の体積(volume)

$$V = \int_R dx$$

- P:あるパターン x が R に入る確率

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

- k:n個の訓練標本のうち R に入っている個数



ノンパラメトリック法の基礎

- 確率 P を二つの方法で近似する.

A) k, n を用いれば,

$$P \approx k/n$$

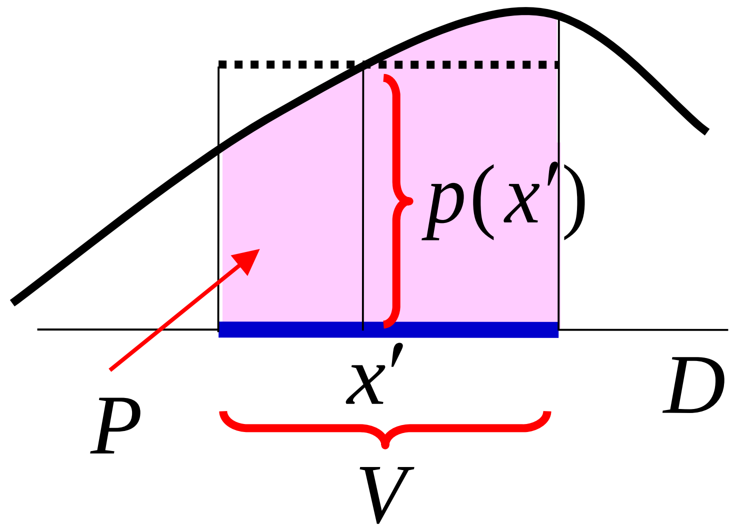
B) 注目点 x' を用いれば,

$$P \approx Vp(x')$$

- これらより

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

積分の長方形近似



近似Aの良さ

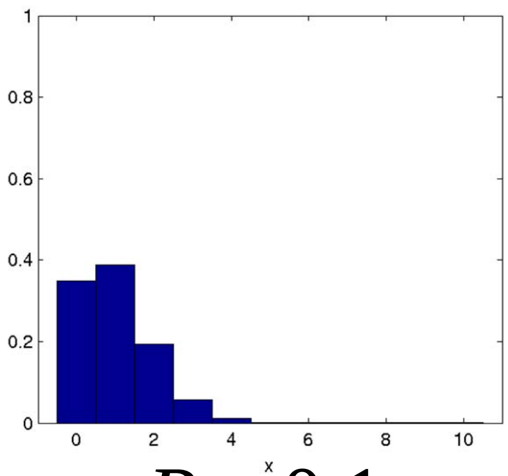
$$P \approx k/n$$

- n 個の訓練標本のうち k 個が R に入る確率は 二項分布に従う.

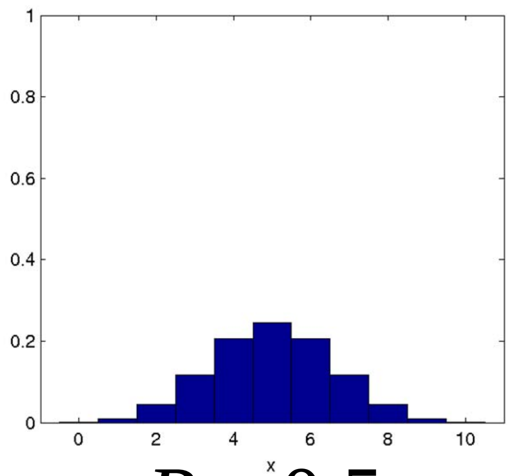
確率: ${}_nC_k P^k (1-P)^{n-k}$

期待値: nP

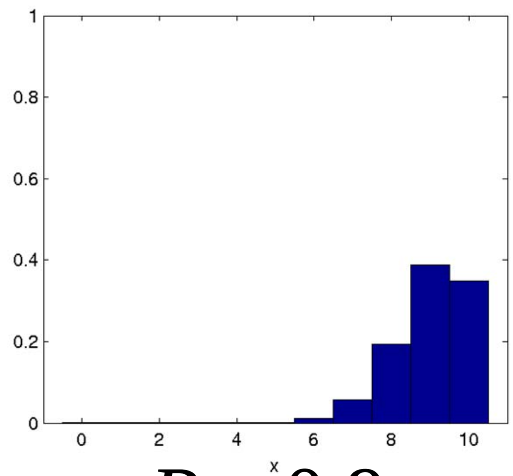
分散: $nP(1-P)$



$P = 0.1$



$P = 0.5$



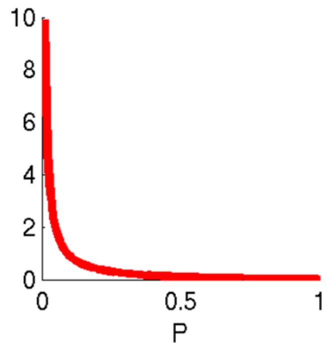
$P = 0.9$

近似Aの良さ(続き)

$$P \approx k/n$$

- 平均的にはぴったり $P = k/n$．分散は？
- 相対的な分散が重要なので期待値を1に揃える．

$$z = \frac{k}{nP} \Rightarrow E[z] = 1,\ V[z] = \frac{1-P}{nP}$$



- P が大きい方が分散が小さくなり，近似の精度が良い

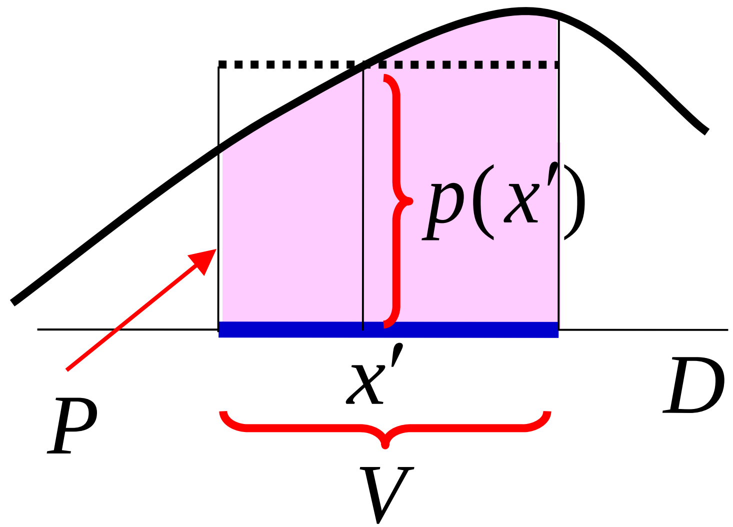
⇒ 領域 R は大きい方が良い！

近似Bの良さ

$$P \approx Vp(x')$$

- 積分の長方形近似は, $p(x)$ が R 内でほぼ一定値をとるとき, 精度が良い.

→ 領域 R は小さい方が良い！



領域の決め方(続き)

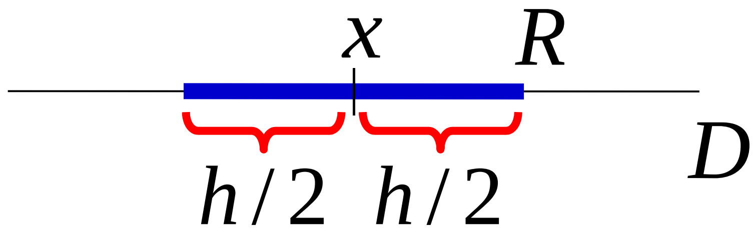
$$p(x) \approx \frac{k}{nV}$$

- 全体の近似精度を上げるためには，領域 R を程よい大きさに決める必要がある.
- 訓練標本を用いて領域 R を決める.
  - パーゼン窓法，カーネル密度推定法:
    R の形を決め V を固定したもとで k を標本から決定
  - 最近傍密度推定法:
    R の形を決め k を固定したもとで V を標本から決定

パーゼン窓法

- 領域 R として，ある点 x を中心とする一辺の長さが h の超立方体(hypercube)を用いる.
  - 体積 $V = h^d$
  - R に入る標本数

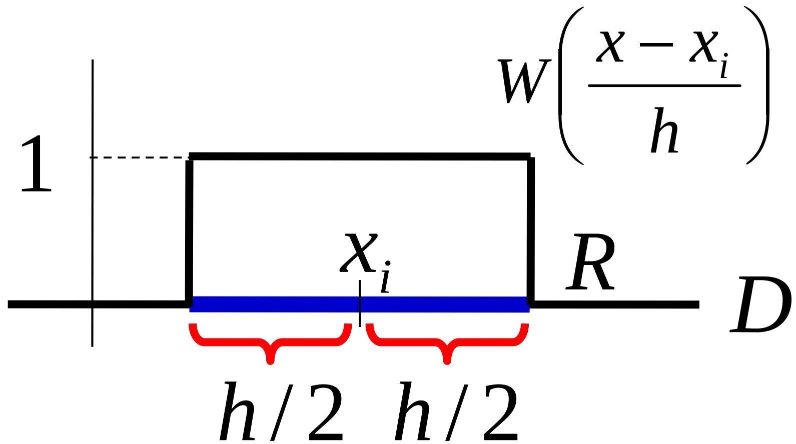
$$k = \sum_{i=1}^{n} W\left(\frac{x - x_i}{h}\right)$$



$$W(x) = \begin{cases} 1 & \max |x^{(i)}| < \frac{1}{2} \\ 0 & \text{それ以外} \end{cases}$$

パーゼン窓関数
(Parzen window function)

$$x = (x^{(1)}, x^{(2)}, \ldots x^{(d)})^T$$

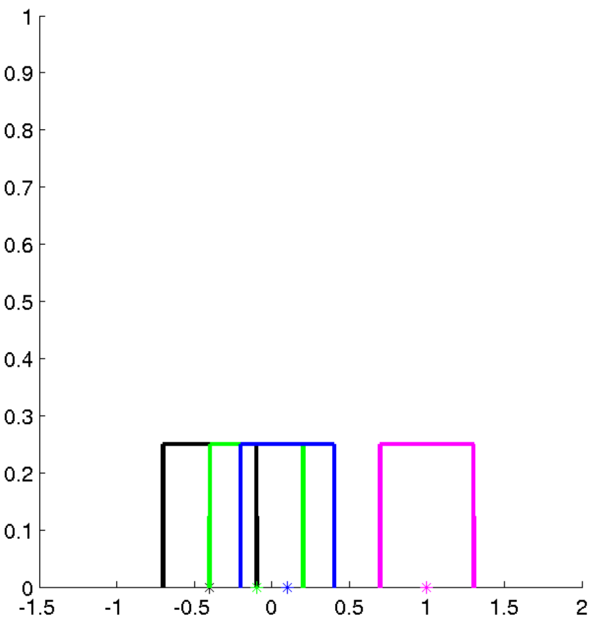


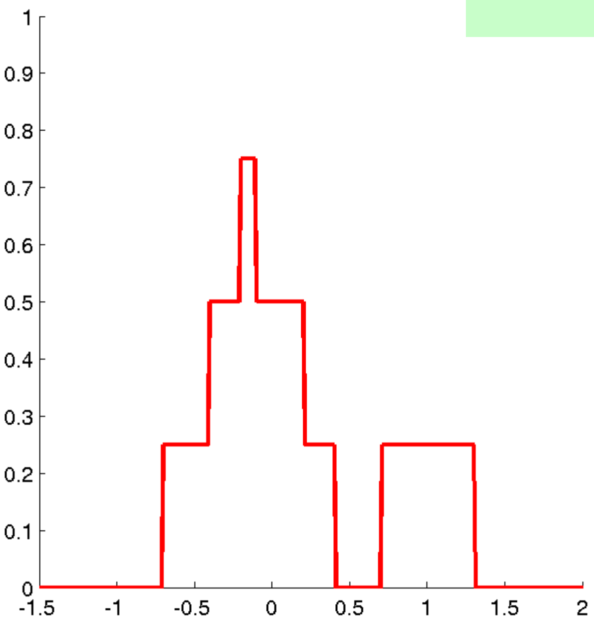
パーゼン窓法

■パーゼン窓法(Parzen window method):

$$\hat{p}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} W\left(\frac{x - x_i}{h}\right)$$

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





パーゼン窓法の特徴

- 与えられた標本からパターン領域の分割を適応的に決定できるため，ヒストグラム法のようにあらかじめ領域の分割の仕方を決定しておく必要が無い．
- 領域間での不連続性は未解決．
- 領域の分割の仕方を決める必要は無いが，パーゼン窓関数のバンド幅 h を適切に決める必要がある．

カーネル密度推定法

- パーゼン窓法の不連続性を解決したい
- カーネル関数(kernel function) $K(x)$:

$$\int_D K(x)dx = 1 \qquad K(x) \geq 0 \text{ for all } x \in D$$

- カーネル密度推定法(kernel density estimation):

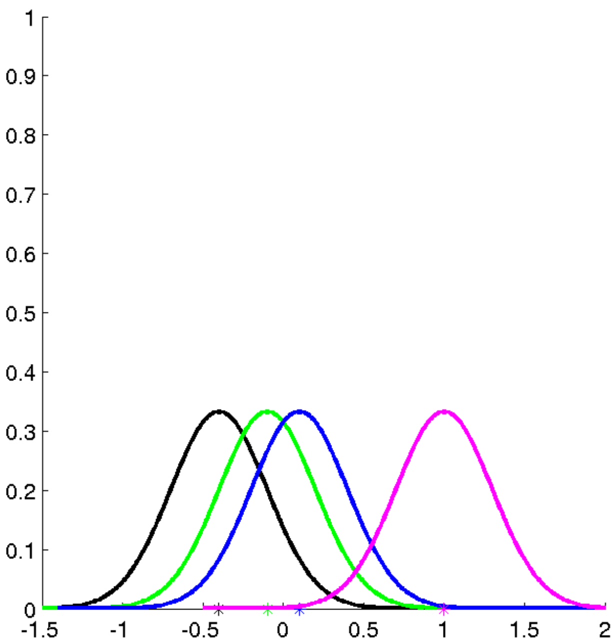
$$\hat{p}(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

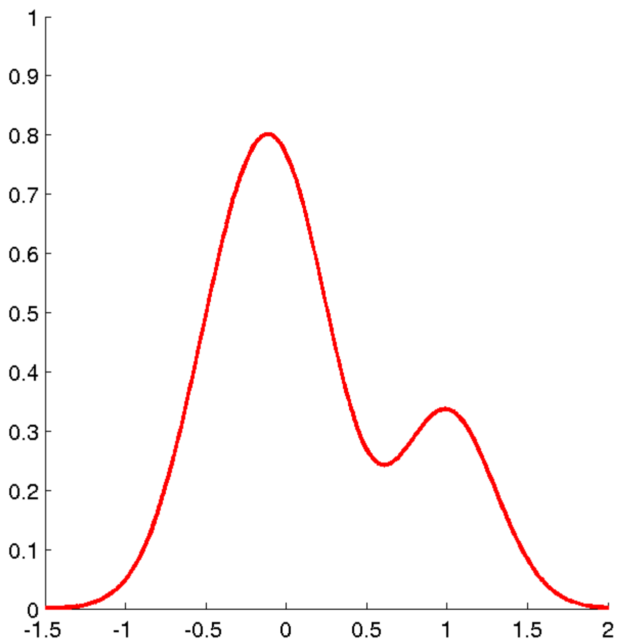
h :バンド幅(bandwidth)

カーネル密度推定法の例

■ガウスカーネル関数:

$$K(x) = \frac{1}{(2\pi)^{d/2}} \exp(-\tfrac{1}{2} x^T x)$$





一般化

$$\hat{p}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

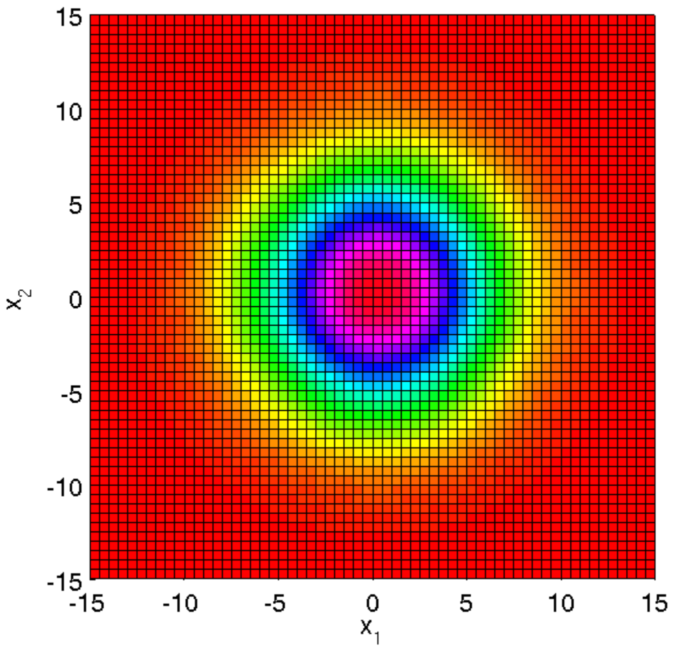
- バンド幅を一つのパラメータ h で制御すると自由度が小さい.

- バンド幅行列(bandwidth matrix) H :
  正値対称行列

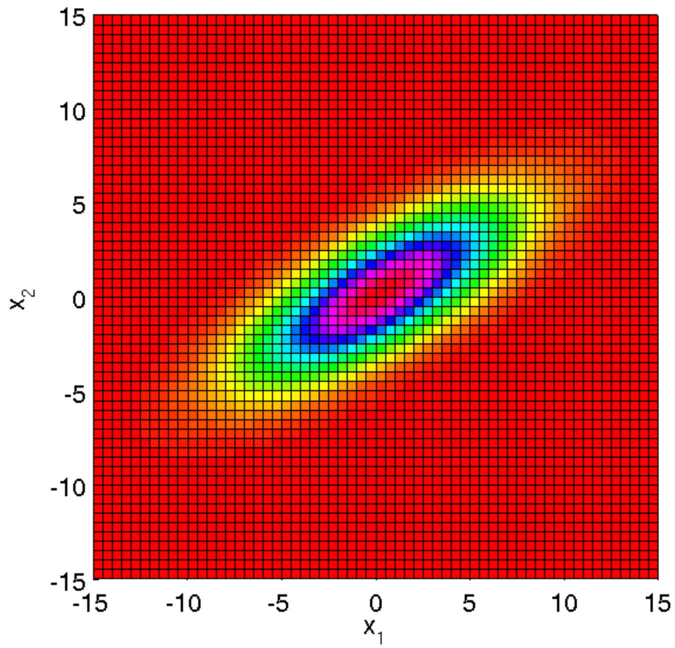
$$\hat{p}(x) = \frac{1}{n|H|} \sum_{i=1}^{n} K\left(H^{-1}(x - x_i)\right)$$

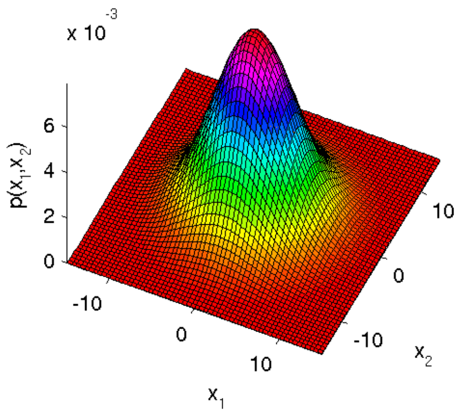
例

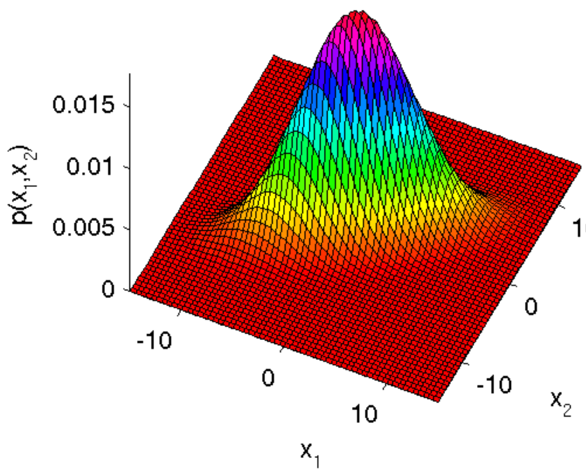
h

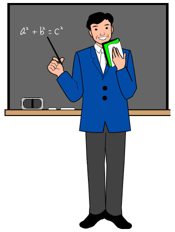

15
10
5
0
-5
-10
-15
-15 -10 -5 0 5 10 15
x_1
x_2

H


15
10
5
0
-5
-10
-15
-15 -10 -5 0 5 10 15
x_1
x_2


x 10^-3
6
4
2
0
$p(x_1,x_2)$
-10
0
10
x_1
10
0
-10
x_2


0.015
0.01
0.005
0
$p(x_1,x_2)$
-10
0
10
x_1
10
0
-10
x_2



まとめ

- ノンパラメトリック法: パラメトリックモデルを用いない確率密度関数の推定法
- ノンパラメトリック法では，領域の大きさを適当に決める必要がある．
- パーゼン窓法: 領域の形と体積を固定し，標本数をデータから決定する
- カーネル密度推定法: パーゼン窓を滑らかなカーネル関数に置き換えた方法

小レポート（第6回）

- 1次元の入力に対するカーネル密度推定法を実装し，適当なデータを用いて確率密度関数を推定せよ（サンプルプログラム参照）．
- データ標本数，真の確率分布，カーネル関数，バンド幅などの条件を変化させたとき，どのように推定結果が変わるかを考察せよ．

- 余力のある学生は，入力が2次元の場合に対しても同様の実験を行え．また，次元が増えたことによりどのような変化が生じたかを考察せよ．

Octaveのサンプルプログラム

kde.m

```
clear all
n=5000; x=myrand(n);
h=0.1;

xx=0:0.01:5;
pxh=zeros(size(xx));
for i=1:n
  pxh=pxh+normpdf(xx,x(i),h^2)/n;
end

figure(1); clf; hold on;
hist(x,0:0.1:5,10)
plot(xx,pxh,'b-')
legend('true','estimated')
print -depsc kde.eps
```

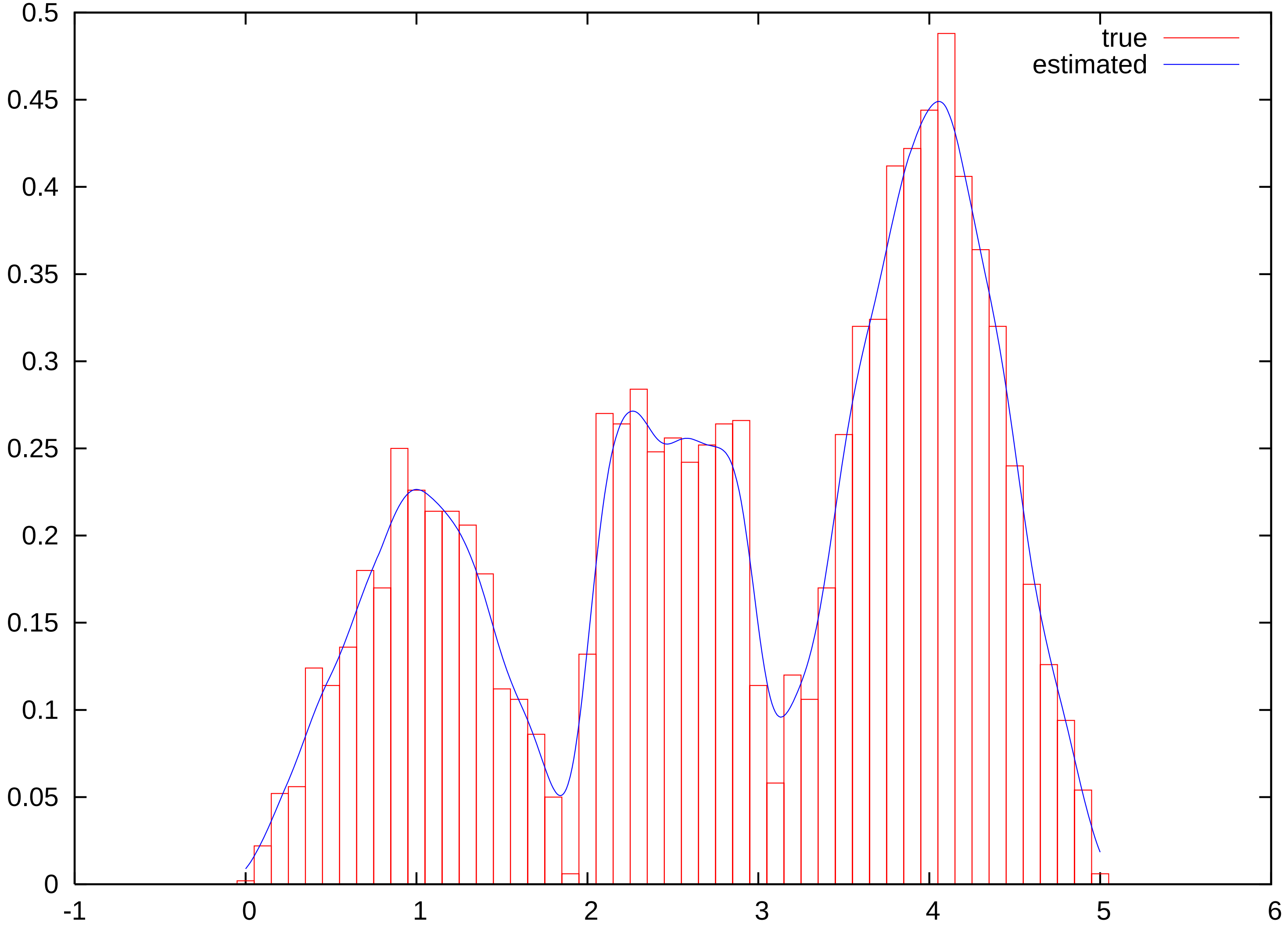
myrand.m

```
function x=myrand(n)

x=zeros(1,n);
u=rand(1,n);

flag=(0<=u & u<1/8);
x(flag)=sqrt(8*u(flag));
flag=(1/8<=u & u<1/4);
x(flag)=2-sqrt(2-8*u(flag));
flag=(1/4<=u & u<1/2);
x(flag)=1+4*u(flag);
flag=(1/2<=u & u<3/4);
x(flag)=3+sqrt(4*u(flag)-2);
flag=(3/4<=u & u<=1);
x(flag)=5-sqrt(4-4*u(flag));
```

実行例



今後の予定

- 12月2日：通常の講義
  - 最近傍密度推定法（第13章）
- 12月9日：計算機演習
- 12月16日：通常の講義
  - ガウス混合モデル（第8章）
- 12月23日：計算機演習

計算機演習は出欠を取らないので，
各自好きな時間に演習を行なって良い．
課題はOCWで事前に公開する．