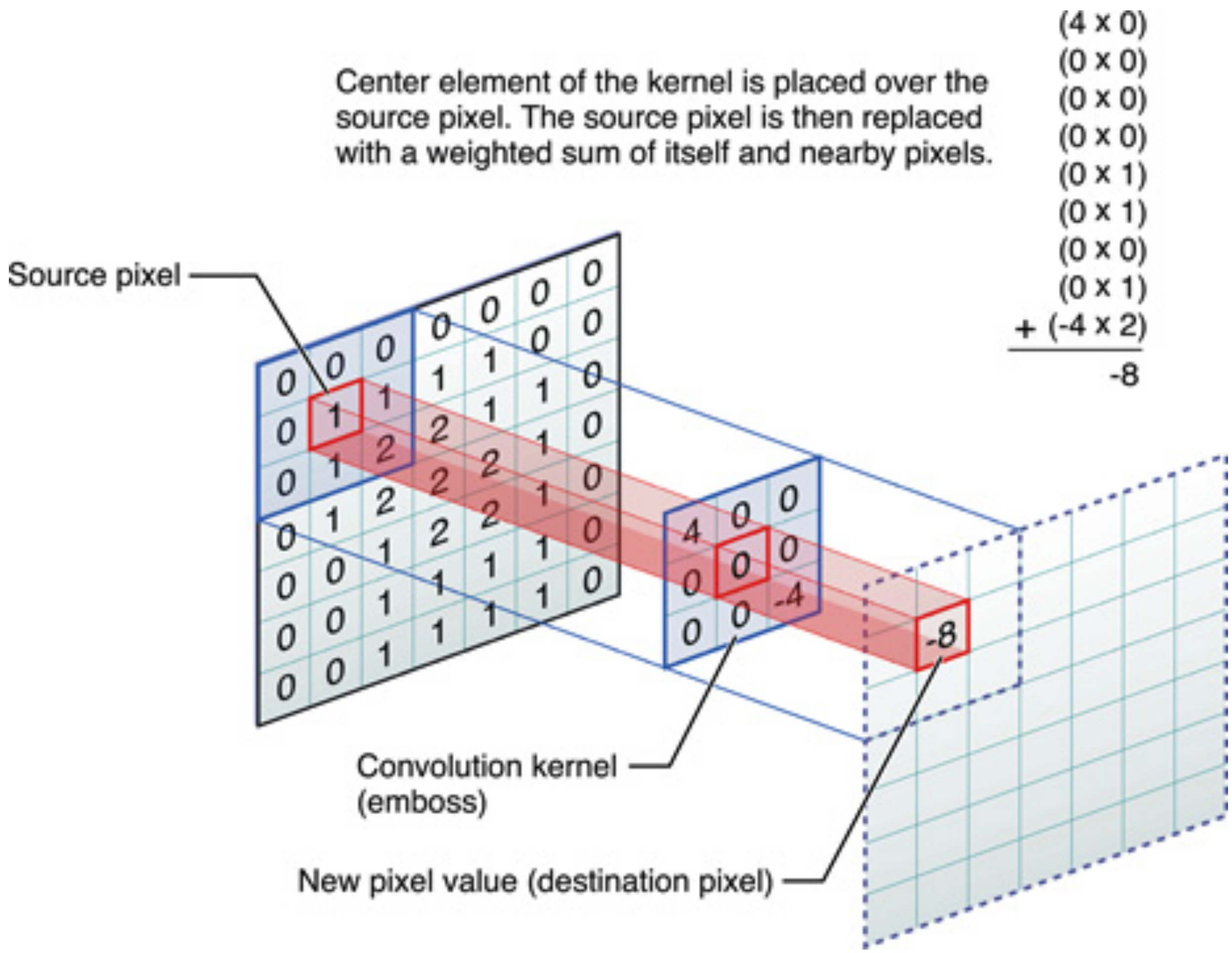
Convolution and Non-parametric Density Estimation

Correlation vs Convolution

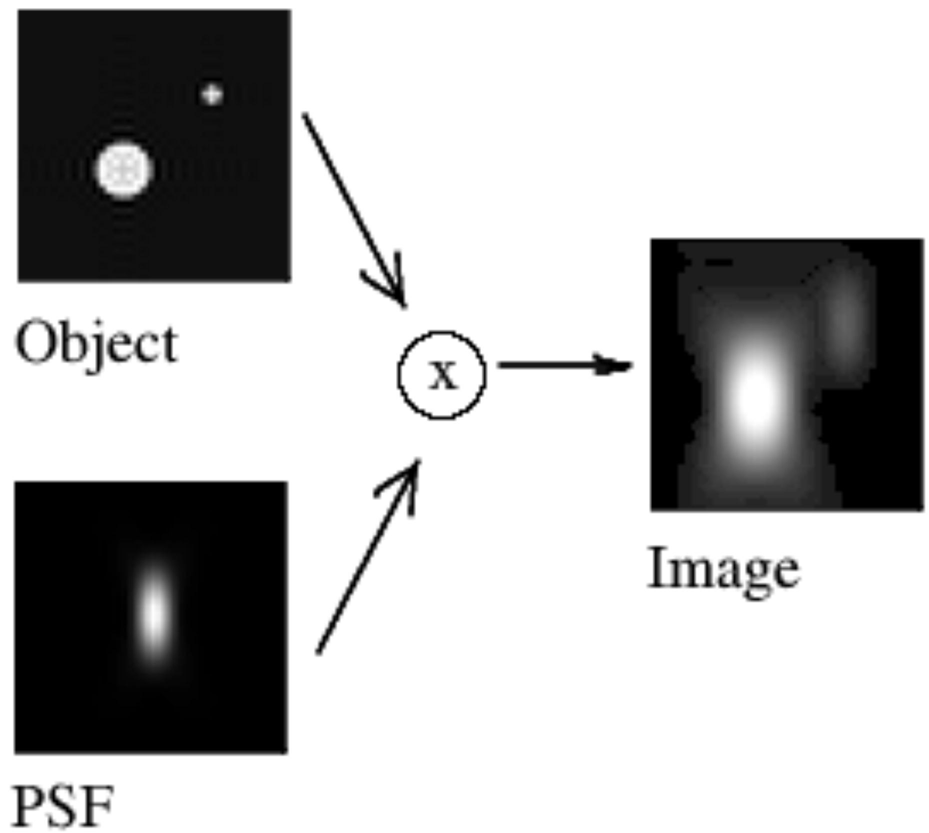
- Correlation:
  - Primary use: matching
- Convolution:
  - Models perturbations in the image generation process.

Visualizing Image Filtering



https://developer.apple.com/library/ios/DOCUMENTATION/Performance/Conceptual/vImage/ConvolutionOperations/ConvolutionOperations.html

Point Spread Functions



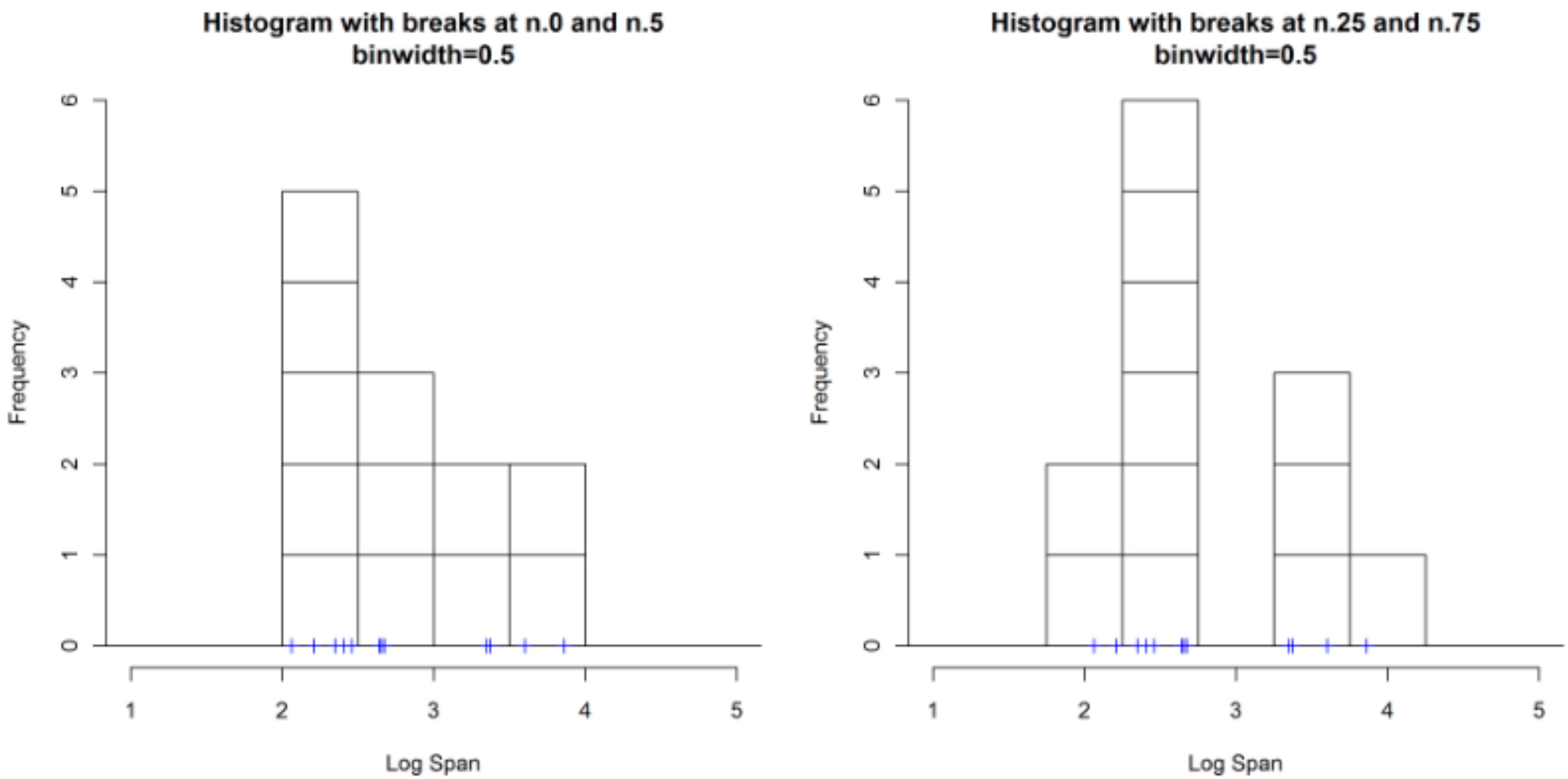
A Pachinko Machine

There goes the family fortune

Maximum Likelihood Parameter Estimation

- To the board!
- Worked through how to find the maximum likelihood mean for a sample from a Gaussian distribution with unit variance and unknown mean.
  - Write down expression for likelihood of all the data points.
  - Take derivative (of log likelihood) with respect to unknown mean.
  - Set to 0 and solve.
- Result: maximum likelihood mean is the empirical mean of the sample!

Histogram dependence on bin position



http://www.mvstat.net/tduong/research/seminars/seminar-2001-05/

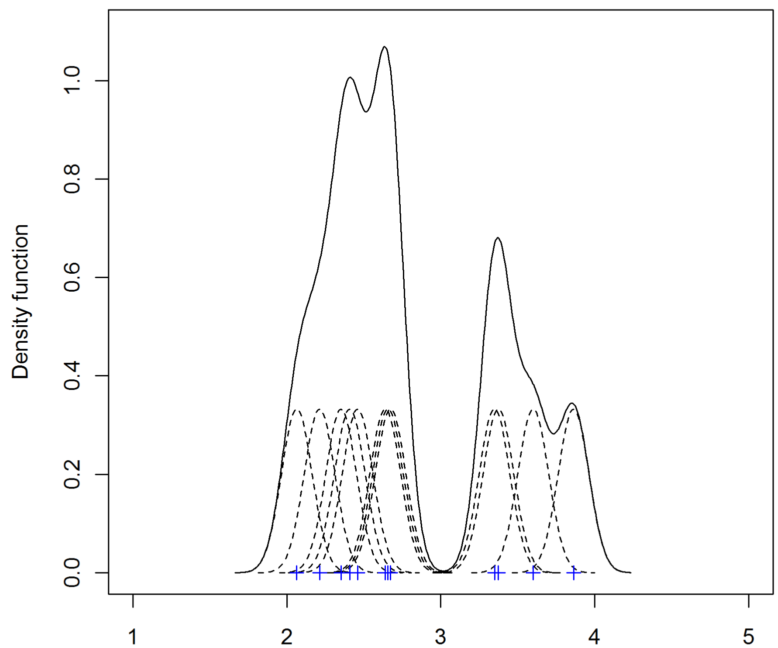
Gaussian Density Function (you might as well learn it now!)

$$p(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp - \frac{(x-\mu)^2}{2\sigma^2}$$

Likelihood of a new point x under a particular sample's Gaussian

$$p(x; x_i, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp - \frac{(x - x_i)^2}{2\sigma^2}$$

Kernel Density Estimate built from 12 samples.



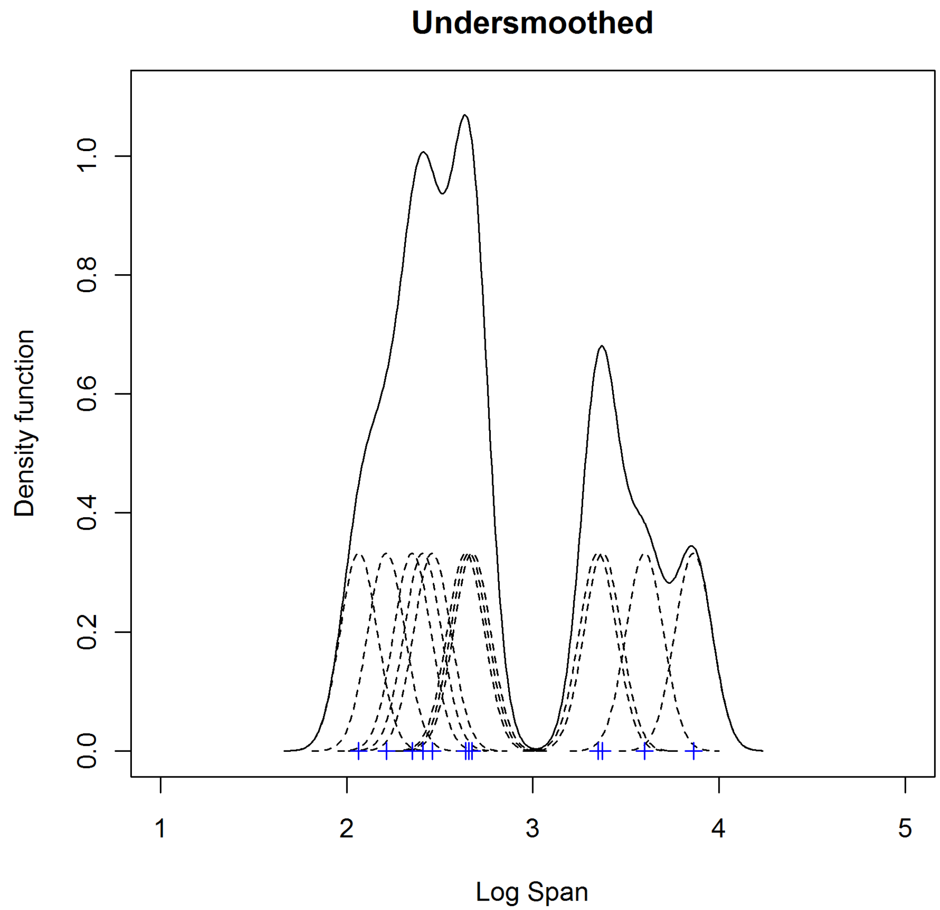
Kernel Density Estimate

$$p(x; x_1, x_2, ..., x_N, \sigma^2) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp - \frac{(x - x_i)^2}{2\sigma^2}$$

KDE as convolution

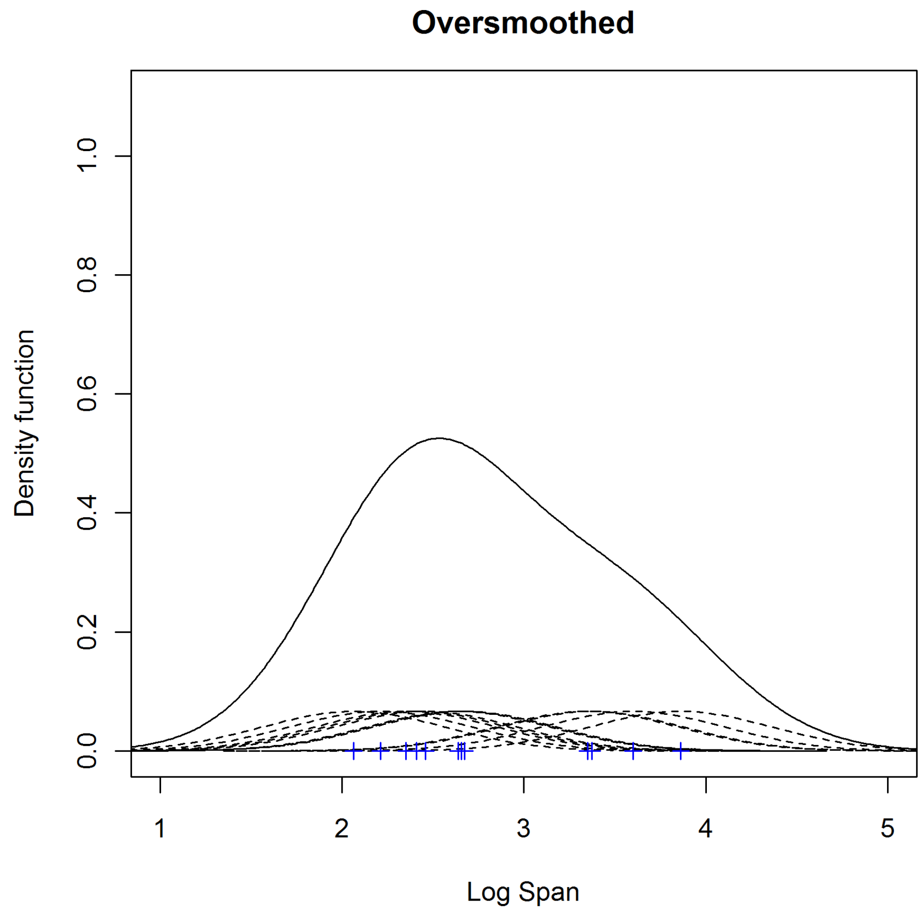
- The density is the convolution of the sample with the kernel:
- p(x) = conv(S,kernel).

Kernel Density Estimation



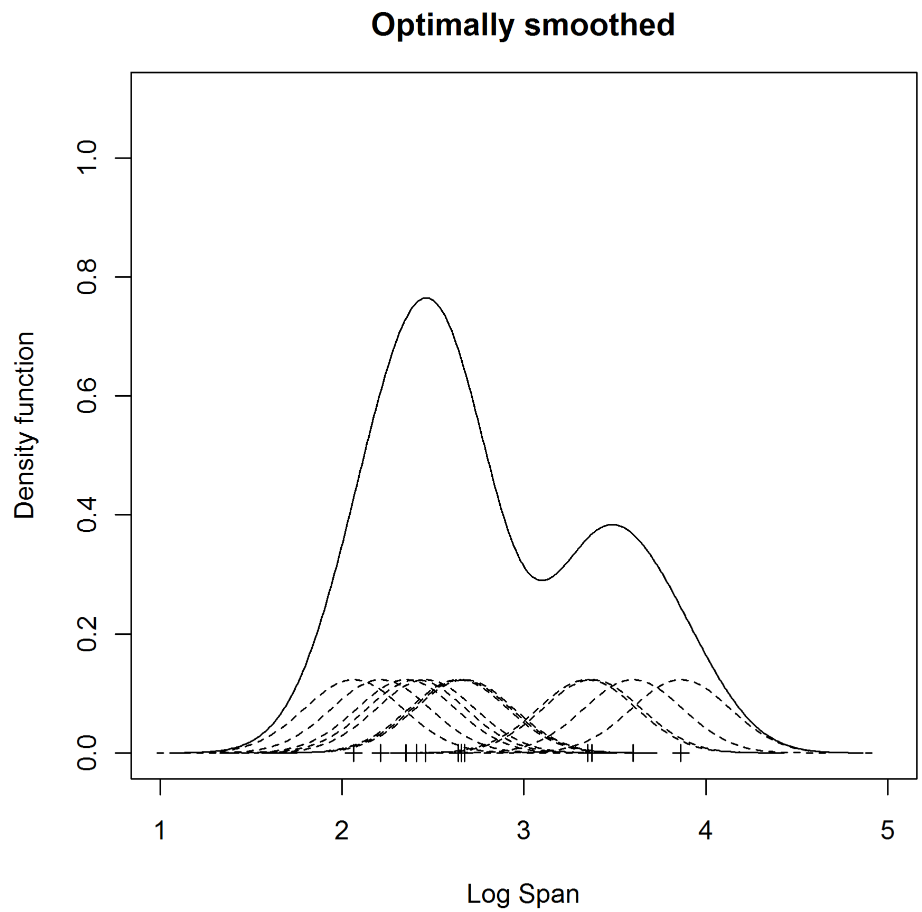
http://www.mvstat.net/tduong/research/seminars/seminar-2001-05/

Kernel Density Estimation

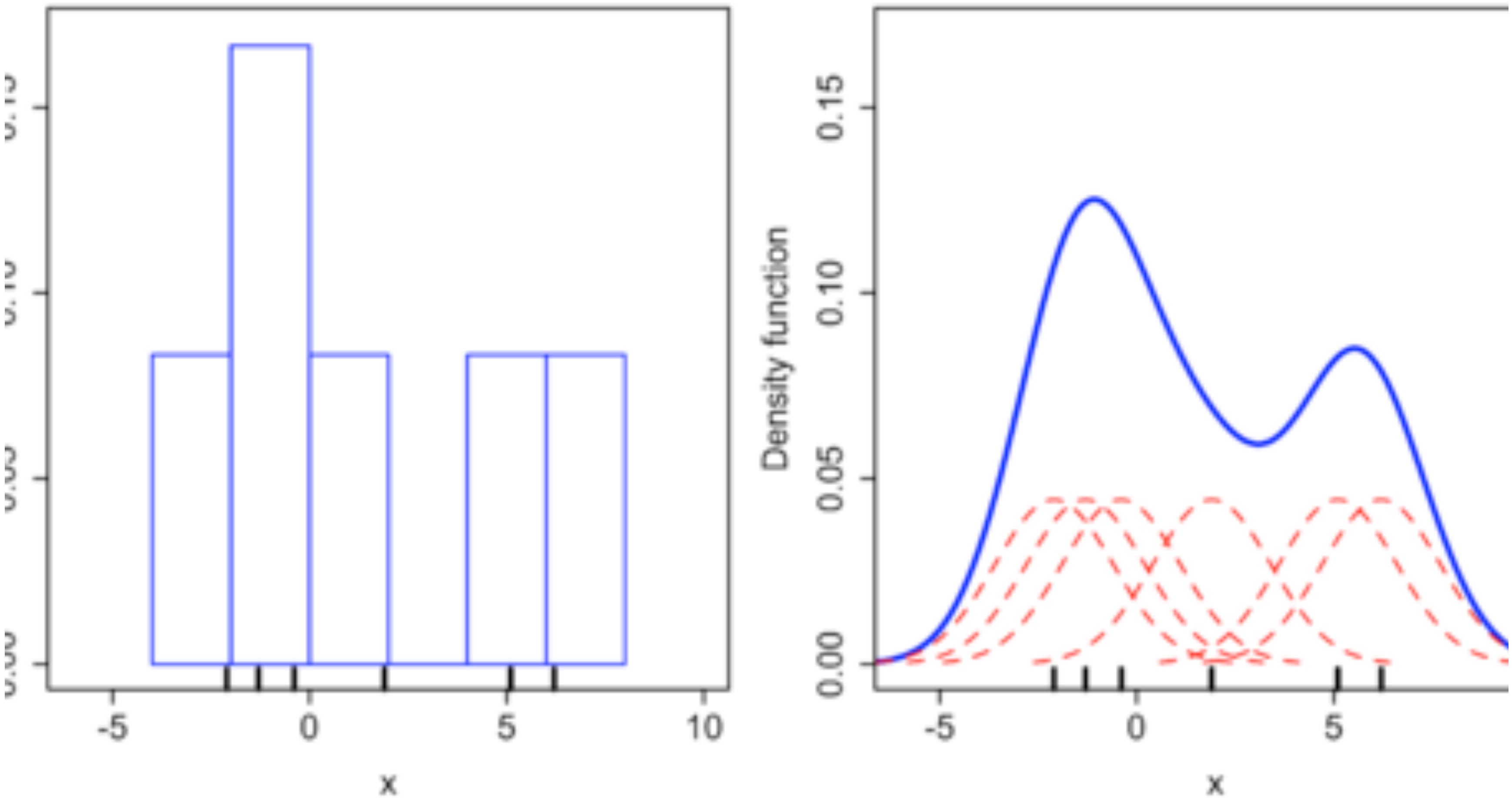


http://www.mvstat.net/tduong/research/seminars/seminar-2001-05/

Kernel Density Estimation

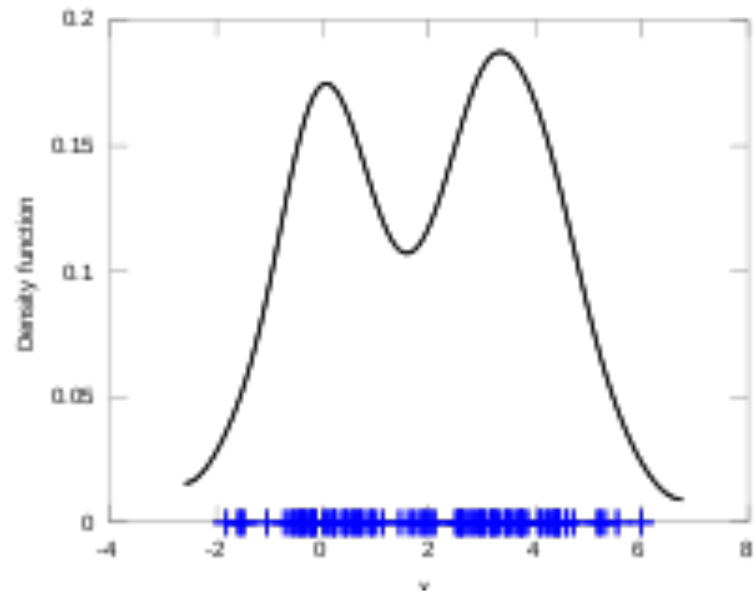


http://www.mvstat.net/tduong/research/seminars/seminar-2001-05/



Comparison of the histogram (left) and kernel density estimate (right) constructed using the same data. The 6 individual kernels are the red dashed curves, the kernel density estimate the blue curves. The data points are the rug plot on the horizontal axis. (WIKIPEDIA)

```
randn('seed',8192);
x = [randn(50,1); randn(50,1)+3.5];
[h, fhat, xgrid] = kde(x, 401);
figure;
hold on;
plot(xgrid, fhat, 'linewidth', 2, 'color', 'black');
plot(x, zeros(100,1), 'b+');
xlabel('x')
ylabel('Density function')
hold off;
```



Estimating the kernel "bandwidth"

- Use leave-one-out cross validation.
- For each point, calculate probability under density estimate with all the other points. This is the "leave-one-out estimate" of that point.
- Now consider the product of the probabilities of each point under its leave-one-out estimate.
- Find the variance of the Gaussian kernel which maximizes the leave-one-out product.