U U

ECE 3530

Statistical Data Plots

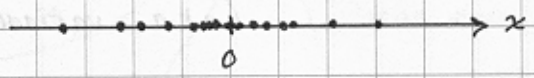
7 Apr 03
Neil E Cotter

Histogram: If we sample from a continuously distributed random variable, we just get points.

ex: Suppose we sample repeatedly from a standard gaussian density function: $X_i \sim n(0,1)$

Note: We always assume independent samples when dealing with statistics unless otherwise specified.

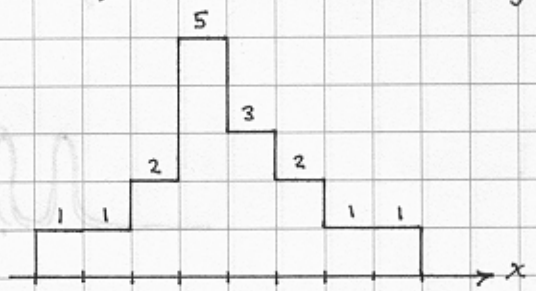
Our samples plotted as points might look like the following:



With many points, it will be hard to discern the shape of the density function, $p(x)$.

A Histogram is a bar plot showing the number of points in each interval (where we choose the intervals).

ex: For the above example, using the squares as intervals, we have the following:



The histogram allows us to roughly visualize $p(x)$.

U U

ECE 3530

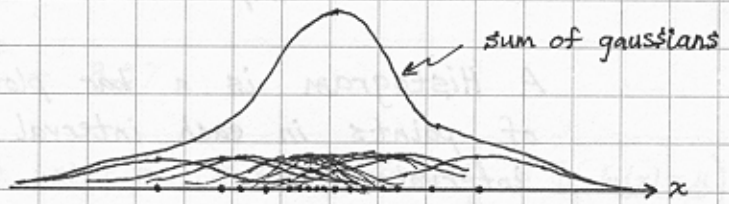
7 Ap
Neil &

Parzen Windows: By placing a small "bump" where each sample data point is located and summing them up, a smooth approximation of $p(x)$ results. We normalize so area under $p(x)$ equals one.

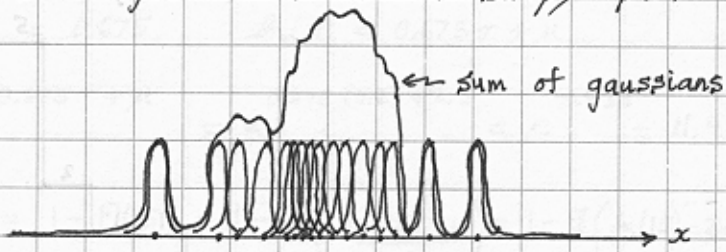
ex: We may use a gaussian-shaped bump to obtain a smooth curve approximating $p(x)$. We also must choose the width of the gaussian appropriately.

Using data from histogram example, we see the effect of the gaussian's width (a.k.a. variance σ^2).

Wide gaussian (large σ^2) blurry $p(x)$:



Narrow gaussian (small σ^2) bumpy $p(x)$:



The advantage of the Parzen Windows approach is that it may approach $p(x)$ as the number of data points approaches infinity.

ref: Pattern Classification and Scene Analysis, Richard Duda and Peter Hart, John-Wiley, NY, 1973. ISBN 0-471-22361-1.

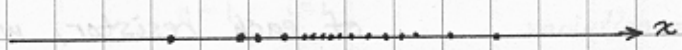
U U

ECE 3530

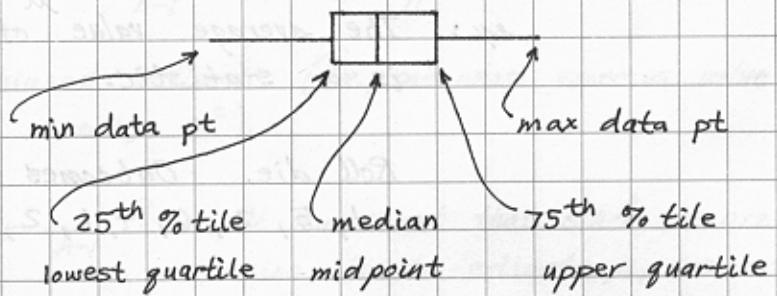
8 Apr 03
Neil E Cotter

Box Plots: Often we are most interested in knowing where the quartiles of data lie. The box plot shows quartiles.

Consider the following data points:



The box plot for the data:



Note: The box plot shows the <u>median</u> value, <u>not the mean</u> value.