SD 675 Pattern Recognition

Winter 2009

(Assignments are to be done individually. Do not write a formal report.)

Purpose

Time to look at some pattern recognition basics ...

1. Given $p_x(x), p_y(y)$ the PDFs for independent random variables x, y.
   - Derive $p_z(z)$ for $z = x + y$
   - Show explicitly where you use the independence assumption.
2. Given independent random variables x, y with variances σ_x^2, σ_y^2.
   - Prove that the variance of a sum of independent random variables is the sum of the variances.
   - Show explicitly where you use the independence assumption.
3. Create a large vector number of uniform random values in Matlab, e.g, `rand(1, 10000)`
   - Sort the values, and then plot the histogram of the *interpoint* spacing.
   - What distribution is this?
4. Again using the uniform random values from before, create bins

   $$[0.001 \cdot i, 0.001 \cdot (i + 1)) \text{ for } i = 0, \ldots, 999$$

   We want to count the number of points in non-overlapping bins:
   - Let N_j equal the number of times we see j values in a bin. Plot N_j as a function of j
   - What distribution is this?
5. A conversion of PDFs ... suppose that x is uniformly distributed,

   $$p(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & x < 0, x > 1 \end{cases}$$

   Suppose I let $y = -\ln(1 - x)$, where "ln" is the natural logarithm.
   - What is the PDF for y (PDF conversion is discussed in Chapter 2 of the SYDE 372 Notes)?
   - Any idea why this result is useful?