Brief Notes on
Association Rules
Prof. Tom Fomby
Department of Economics
Southern Methodist University
Dallas, Texas 75275
April, 2011

- Unsupervised Learning Method. Neither prediction nor classification.
- Applications. E.g. Amazon.com sidebars; Grocery store shelving.
- $\Pr(C|A)$ C = consequent items, A = antecedent items
- Transaction File: Binary Format versus Item List Format
- Confidence of a rule = $\frac{support(A \cap C)}{support\ of\ A} = \Pr(C|A) = \frac{\Pr(C \cap A)}{\Pr(A)}$
- Lift = $\frac{\Pr(C|A)}{Pr(C)}$ $: 1 \leq Lift < \infty$; Lift = 1 when C and A are independent
- $Pr(C|A)$ = absolute measure ; Lift = relative measure
- The **Apriori Algorithm** (Agrawal and Srikant (1994))
  In the **first phase** build up the item sets that meet the minimum support requirement. First, collect together all of the one-item sets that have minimum support. Then form two-item sets built from the previous minimum support one-item sets and then retain the two-item sets that have minimum support. Then form three-item sets built from the retained one-item and two-item sets and then choose those three-item sets that have minimum support. One continues this process building up larger and larger minimum support item sets until at say the K-item set level none of the K-items sets meet the minimum support requirement. Then the first phase is complete. Then in the second phase one examines all of the retained one, two, … (K-1) item sets for possible Association Rules that have minimum confidence. In the making of the Association Rules it must be remember that the consequent (C) and antecedent(A) item sets must be **disjoint**.
  Two tuning parameters:
  a) minimum support (t) of k-item sets
  b) minimum confidence of Association Rule
- The choice of "top" Rules (Associations) is sensitive to the choices of the above tuning parameters. The larger (smaller) the choices of the tuning parameters, the more simple (complex) the resulting Rules. You should examine the robustness of the "top" Rules over various choices of tuning parameter values. Rules need to make sense given our domain-specific knowledge, otherwise you shouldn't use them.