Using Statistical Tools to Improve Actuarial Model Efficiency

John Hegstrom FSA, MAAA

Need for Modeling Efficiency

- PBR/Solvency II/RBC C3 Phase III/IFRS
- Product designs and resulting pricing
- Risk Management / Economic Capital
- Stochastic on Stochastic models
- Limits of Moore's Law
- Near real-time analysis desired for some applications

PBR/Solvency II/RBC C3 Phase III/IFRS

- PBR
  - 2015?
- Solvency II
  - 2013
- RBC C3 Phase III
  - Soon?
- International Financial Reporting Standards
  - ?

3

Moore's Law

- The number of transistors that can be placed on an integrated circuit doubles about every two years
- This trend has continued for 50 years and is expected to continue until 2015 at earliest or 2600 at latest

4

How to Get Faster Actuarial Models

- Spend lots of money on hardware
- Tweak and optimize software
- Manually map smaller models/cells to larger ones
- Compress the whole model
- Compress parts of the model
- Compress some of the inputs

5

Some Statistical Tool Groupings

- Predictive Modeling
  - Customer relationship management and data mining
- Predictive Analytics
  - actuarial science, financial services, insurance, telecommunications, retail, travel, healthcare, pharmaceuticals

6

Some Statistical Tool Groupings

- Machine Learning
  - machine perception, computer vision, natural language processing, syntactic pattern recognition, search engines, medical diagnosis, bioinformatics, brain-machine interfaces, cheminformatics, detecting credit card fraud, stock market analysis, classifying DNA sequences, speech and handwriting recognition, object recognition in computer vision, game playing, software engineering, adaptive websites, robot locomotion, structural health monitoring.

7

Predictive Modeling - Wikipedia

- "Predictive modeling is the process by which a model is created or chosen to try to best predict the probability of an outcome."
- "Models can use one or more classifiers in trying to determine the probability of a set of data belonging to another set."

8

Predictive Modeling - sample application

- Probability that a term insurance contract belongs to the group of contracts that will lapse during a given year
  - Tool: logistic regression

$$f(z) = \frac{1}{1+e^{-z}} \quad z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k$$

  - f(z): probability of lapse from 0 to 1
  - Possible dependent variables: age, sex, duration of contract, etc.
  - Advantage: provides functional structure to raw data

9

Predictive Analytics - Wikipedia

- "Predictive analytics is an area of statistical analysis that deals with extracting information from data and using it to predict future trends and behavior patterns."
- "The core of predictive analytics relies on capturing relationships between explanatory variables and the predicted variables from past occurrences, and exploiting it to predict future outcomes."

10

Predictive Analytics – sample application

- Assigning underwriting class to a life insurance application
- Tool: multinomial logistic regression

$$P(y_i = j) = \frac{e^{X_i \beta_j}}{1 + \sum_{j=1}^{J} e^{X_i \beta_j}}$$

where j is the underwriting class

- Predictive variables could be age, tobacco use, health measures, and even use publicly available data such as purchase history

11

Compressing Entire Model

- Start with a stochastic financial model
- Find an adaptive statistical tool that reproduces the stochastic model in a compressed manner
- Run the adaptive model over a set of scenarios

12

Neural Network Learning of a Stochastic Financial Planning Model

- Outlined in my paper "Predictive Model Learning of Stochastic Simulations"
- http://www.soa.org/research/research-projects/life-insurance/research-pred-mod-life-insurers.aspx
- The idea: Run a large slow model several times in order to train a small fast model. The small fast model needs to be "adaptive" in order to learn.

13

Personal Financial Planning Model

- Built in Microsoft Excel
- Calculates whether or not assets will last to a given age.
- Inputs
  - Starting asset level
  - Annual Income and Expenses, tax rate
  - Interest, Inflation and Equity market returns
  - Allocation between cash, bonds and small and large company equities

14

Interest, Inflation and Equity market return scenarios

- Economic scenario generator developed under direction of the SOA's Committee on Finance Research in 2004
- Outputs used were inflation, long and short interest rates and large and small cap equity returns
- Generated 500 random scenarios for each of 2000 different deterministic investment strategies (one million total scenarios)

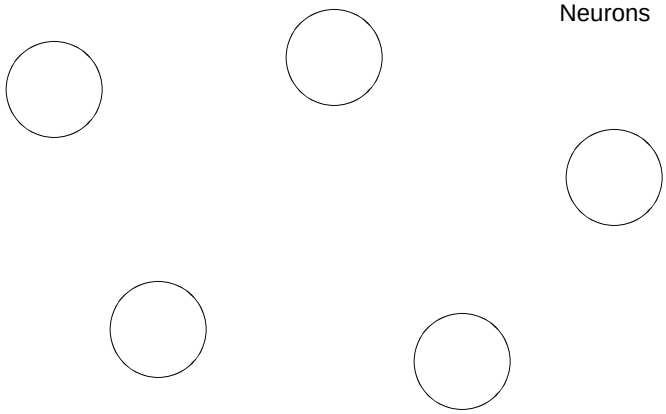
15

Choosing a Statistical Tool

- Need a powerful tool to capture the nuances of the financial planning model.
- Choose to construct and use a Neural Network.
- Neural Networks are "universal function approximators"

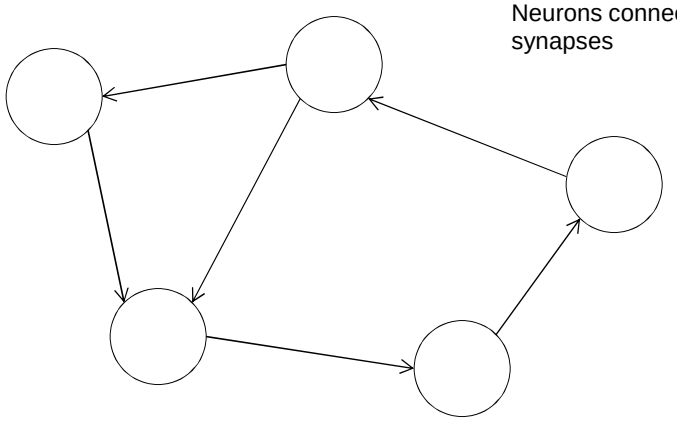
16

Neural Network - Neurons



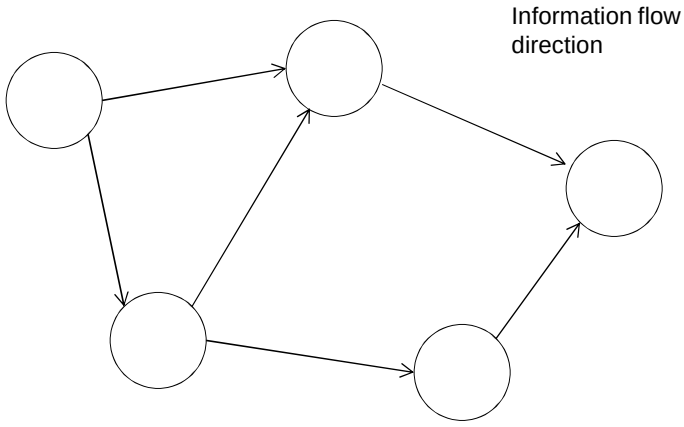
17

Neural Network - Synapses



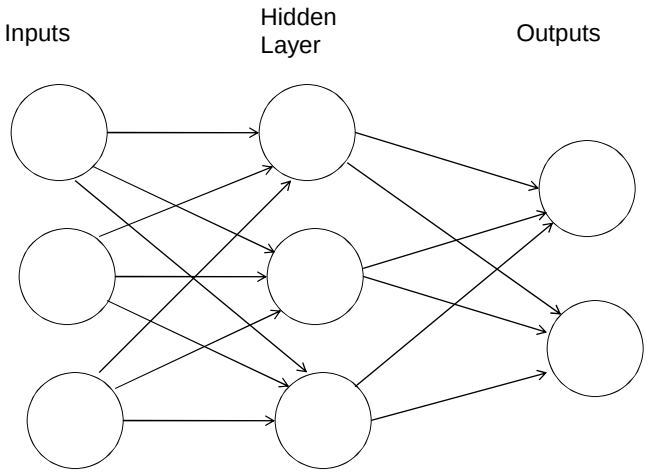
18

Neural Network – Feedforward Network



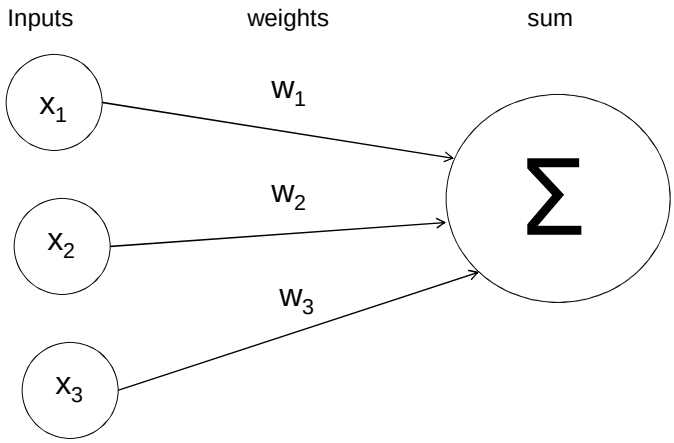
19

Neural Network – Multi-Layer Perceptron



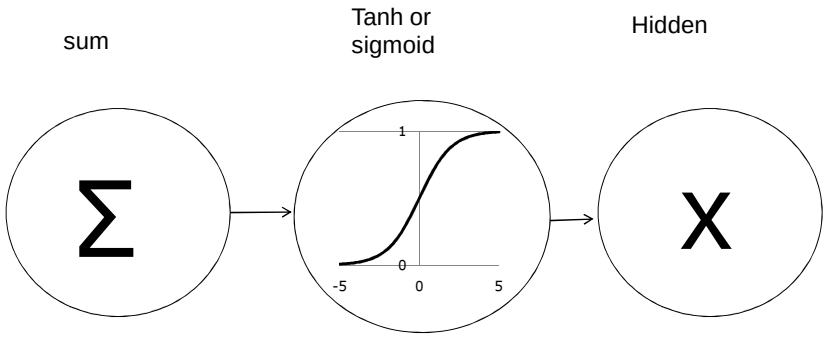
20

Neural Network – Calculations step 1



21

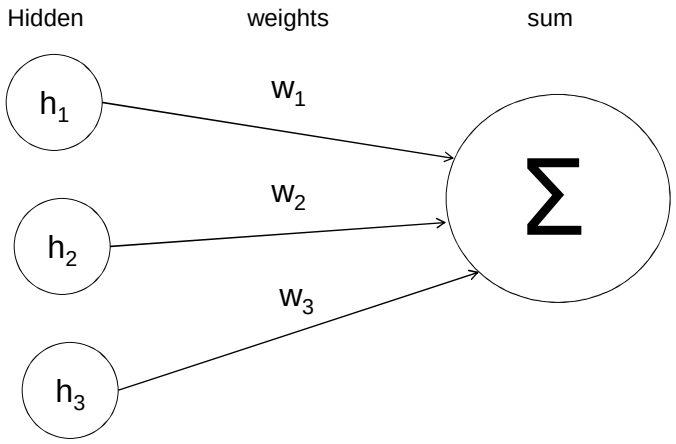
Neural Network – Calculations step 2



$$y = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \quad \text{or} \quad f(x) = \frac{1}{1 + e^{x}}$$

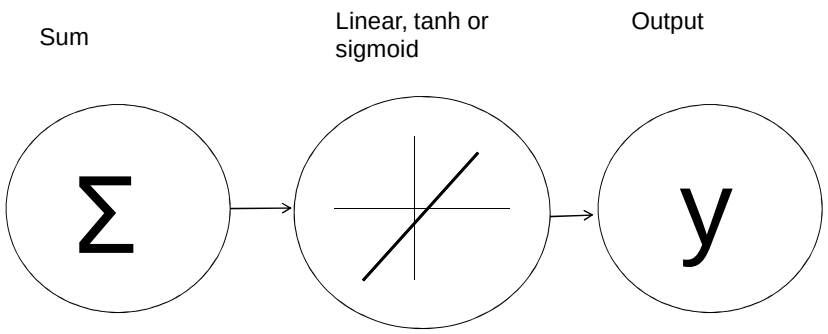
22

Neural Network – Calculations step 3



23

Neural Network – Calculations step 4



24

Neural Network – Calculation

- Random initialization of weights
- Run the network through using training cases
- Use the backpropagation algorithm to distribute the error back through the network
- Change the weights by a small amount
- Repeat process and find minimum mean squared error solution by using a numerical optimization algorithm

25

Specifying the Neural Network for Learning the Financial Planning Model

- Architecture – Multi Level Perceptron
- Inputs for a given current age, final age and starting asset level
  - Planned retirement age, expense ratio, current and retirement asset allocation percentages for cash, bond, small company stocks, and large company stocks.
- Output is the probability of assets lasting until final age

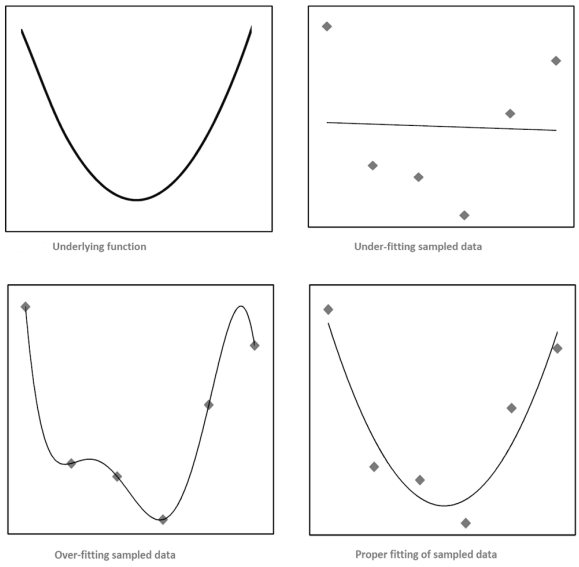
26

Specifying the Neural Network for Learning the Financial Planning Model

- Activation functions - tanh
- How many hidden layers - 1
- How many nodes in hidden layer – 5
- Error (or cost) minimization algorithm – Levenberg – Marquardt
- Want to avoid over-fitting
  - Use enough training cases compared to weights
  - Use cross validation set to measure fit

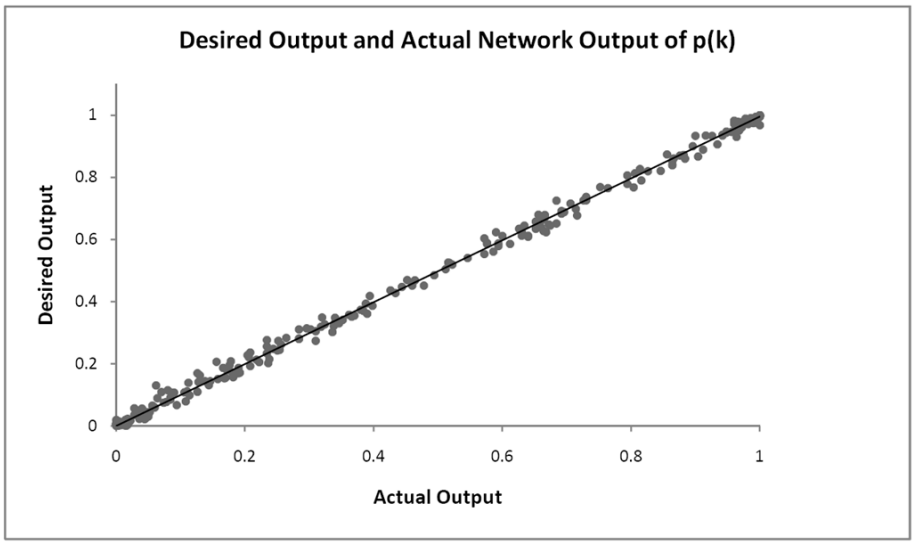
27

Danger of over or under fitting



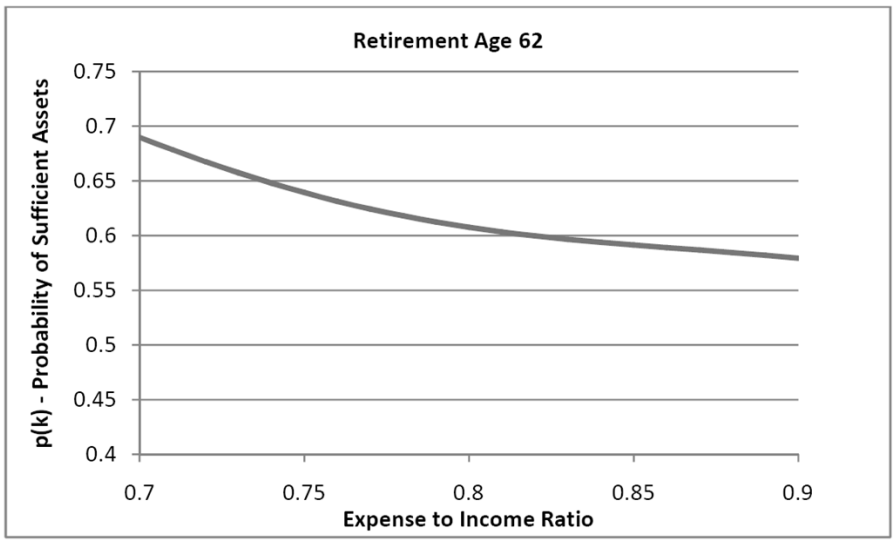
28

Neural Network Results Compared to Stochastic Simulation



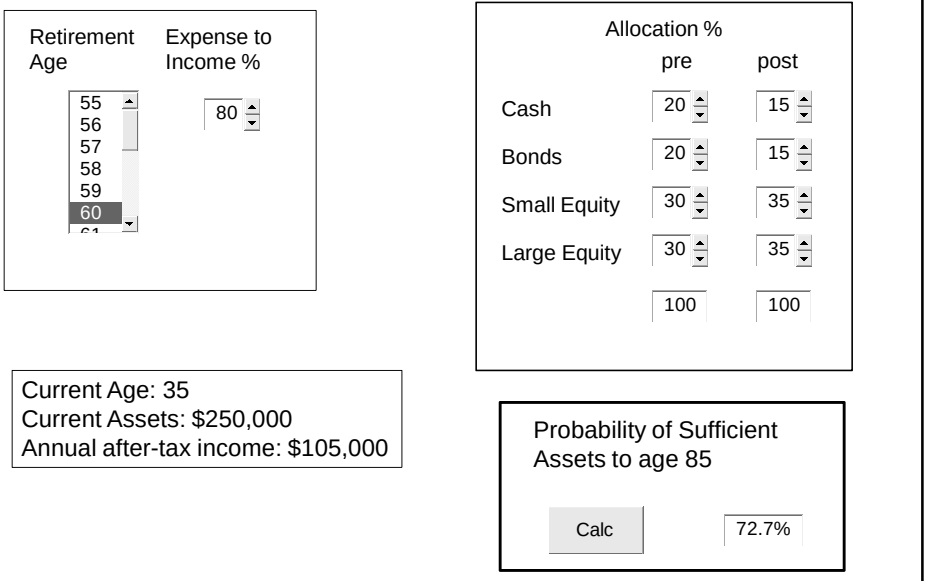
29

What can we do with this NN Model?



30

Neural Network Model – VBA PowerPoint



31

What can we do with this NN Model?

- Instantly evaluate alternate investment and retirement strategies for this individual.
- Quickly solve for the optimal investment strategies given retirement parameters.
- More efficient and effective than just creating a lookup table of source model results
- Caveats – must be trained over enough varied scenarios to capture behavior – cannot use outside of trained range

32

Corporate Planning NN Model

- Train over a range of product sales mixes and economic scenarios
- Output variables could be Expected Earnings, Capital or Risk Measures
- Allows real-time interactive modeling to occur

33

What could be done differently

- Use representative scenarios to train the neural nets more efficiently.
  - Cluster the economic scenarios and pick representative ones first before training.
  - Approach used by Steve Craighead in his article in CompAct (10/09).
- Alternative tools
  - Projection Pursuit Regression (used by Craighead)
  - Support Vector Machine Regression

34

Other Approaches for Compressing Models

- First need to define necessary model outputs for clues
- Replicating Portfolios.
- Unhook liabilities from assets.

35

Replicating Portfolios

- Fit a portfolio of price-able assets to replicate the cash flows of liabilities given sets of economic scenarios
- Commercial Implementations, straightforward

36

Replicating Portfolios

- Example: for a fixed liability cash flow of 5, 5, 105, 5, 105 you could substitute
  - A five year $100 par bond with a coupon of 5%, and
  - A three year zero coupon bond of 100
- Much more elaborate cash flows can be replicated with various asset combinations
- Advantages – arbitrary discount rate not required, term structure of rates reflected
- Disadvantages – no explicit impact of taxes, accounting, required capital

37

Compressing Inputs

- Model Point Compression
  - Traditional
  - Introducing Randomness
  - Clustering Algorithm
  - Also applies to Assets
- Economic Scenario Compression
  - Sampling
  - Statistical Algorithms

38

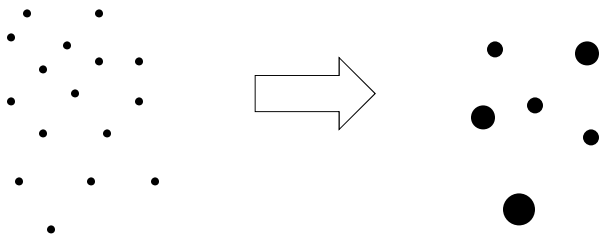
Model Point Compression

- Traditional Approach for In Force Life Business
  - Issue Year, Issue Month, Contract, Issue Age, Gender, Band/Amount, Mode, Underwriting Groupings
  - Mapping small groupings on to large groupings
  - Different groupings based on relative size and similarities - manual and tedious
  - Could randomize mode and issue month
  - End up with a lot of model points in order to accurately model all aspects of the business

39

Model Point Compression

- Cluster Modeling – algorithm developed by Freedman/Reynolds at Milliman
- Start with seriatim listing and group into clusters based on important characteristics using distance between points



40

Cluster Modeling

- Start with "Location Variables" – could be:
  - Starting Reserves
  - Liability cash flows
  - Book/Par ratio for Assets
  - Present Value of Profits
- Distance Calculation

$$\sqrt{(Var1_1 - Var1_2)^2 + (Var2_1 - Var2_2)^2 + (Var3_1 - Var3_2)^2}$$

- Importance = Size x Distance to nearest neighbor

41

Cluster Modeling

- Policy with least importance is mapped to nearest neighbor
- Repeat the last step until desire size is reached
- Review model fit

42

Economic Scenario Compression

- Many methods, many implementations
- Want accuracy in risk measures such as CTE, VaR
- Academy 10,000 Pre-Packaged Scenarios
  - Representative Scenarios using Scenario Picking tool
  - Caveats potential sampling error in CTE calculation

43

Economic Scenario Compression

- Sampling
  - Chueh – various algorithms using definitions distance of one interest rate scenario from another
  - Longley-Cook – expanded to equity scenarios
  - Christiansen – un-weighted representative subset of interest rate scenarios
  - Craighead – CLARA algorithm
- Quasi Monte Carlo and Variance Reduction methods – many mathematical papers
  - Low Discrepancy Sequences

44

Combinations

- Scenario and Model Point Compression
- Scenario Compression and Replicating Portfolios
- Scenario Compression and Whole Model Compression