Credit Scoring and Applications

Srikanth K. Iyer
Department of Mathematics
Indian Institute of Science, Bangalore, India.
skiyer@math.iisc.ernet.in

Ref.: Thomas, Edelman and Crook, SIAM, 2002

To Lend or Not to Lend

Discriminant Analysis: Decision Theory Approach

Application: $X = (X_1, X_2, \ldots, X_p) \in A$, finite

$$A = A_G + A_B \qquad L: \text{ Profit} \qquad D: \text{ Loss}$$

p_G: Proportion of Goods $\qquad$ p_B: Proportion of Bads

Expected Cost per Application

$$L \sum_{x \in A_B} p(x|G)p_G + D \sum_{x \in A_G} p(x|B)p_B = L \sum_{x \in A_B} q(G|x)p(x) + D \sum_{x \in A_G} q(B|x)p(x)$$

Minimum Expected Cost

$$\begin{aligned} A_G &= \{x : Dp(x|B)p_B \leq Lp(x|G)p_G\} \\ &= \left\{x : \frac{D}{L} \leq \frac{q(G|x)}{q(B|x)}\right\} \end{aligned}$$

Minimize

$$\sum_{x \in A_G} p(x|B)p_B = \sum_{x \in A_G} \left(\frac{p(x|B)p_B}{p(x)} p(x) \right)$$

Subject to

$$\sum_{x \in A_G} p(x) = a$$

$$\begin{aligned} A_G &= \{x : \frac{p(x|B)p_B}{p(x)} \leq c\} \\ &= \left\{ x : \frac{1-c}{c} \leq \frac{p(x|G)p_G}{p(x|B)p_B} \right\} \end{aligned}$$

X: Continuous Variables

$$A_G = \left\{ x : \frac{Dp_B}{Lp_G} \leq \frac{f(x|G)}{f(x|B)} \right\}$$

Score $\frac{p(x|G)}{p(x|B)}$ $\frac{f(x|G)}{f(x|B)}$

$X \sim$ Multivariate Normal

$$A_G = \left\{ x : \frac{Dp_B}{Lp_G} \leq \frac{f(x|G)}{f(x|B)} \right\}$$

$$f_X(x|G) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_G)\Sigma^{-1}(x-\mu_G)^T \right)$$

$$(x-\mu_B)\Sigma^{-1}(x-\mu_B)^T - (x-\mu_G)\Sigma^{-1}(x-\mu_G)^T \geq \log\left(\frac{Dp_B}{Lp_G}\right)$$

$$x\Sigma^{-1}(\mu_G-\mu_B)^T \geq \frac{\mu_G\Sigma^{-1}\mu_G^T + \mu_B\Sigma^{-1}\mu_B^T}{2} + \log\left(\frac{Dp_B}{Lp_G}\right)$$

Score $\quad x\,\Sigma^{-1}(\mu_G-\mu_B)^T$ $\quad$ **Equivalent to Linear Regression**

More than two groups

$$\sum_j c(i,j)p_j p(x|j) < \sum_j c(k,j)p_j p(x|j) \qquad \forall\ k \neq i$$

Logistic Regression

$$\log\left(\frac{p_i}{1-p_i}\right) = w_0 + \sum_{j=1}^{p} w_j x_j = w \cdot x_i'$$

$$p_i = \frac{e^{w \cdot x_i'}}{1 + e^{w . x_i'}}$$

Estimate w

Score $w \cdot x'$

Advantages: Can handle discrete, continuous, categorical (non-numeric) variables

Dummy Variables

- X_i : Categorical Variable with k attributes
- $X_{i1}, \ldots X_{ik}$ binary variables
- $X_{ij} = 1$ if individual i has attribute j

Estimation

- $P(Y_i = 1) = p_i = \frac{e^{w \cdot x_i'}}{1+e^{w.x_i'}}$
- $f(y_1, \ldots, y_n) = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i}$
- $\log\left(\frac{p_i}{1-p_i}\right) = w \cdot x_i'$
- **Maximize**

$$\begin{aligned} \log\ f(y_1, \ldots, y_n) &= \sum_{i=1}^{n}\left[y_i \log\left(\frac{p_i}{1-p_i}\right) + \log(1-p_i)\right] \\ &= \sum_{i=1}^{n}\left[y_i\ w \cdot x_i' - \log\left(1+e^{w \cdot x_i'}\right)\right] \end{aligned}$$

Classification Trees

1. Recursive Partition Algorithm
2. Splitting Rule: KS, BII, GI, EI

$$KS = \max_{s \in R} |F(s|B) - F(s|G)| = max_{\ell}|p(\ell|B) - p(\ell|G)|$$

3. Stopping-Pruning Rule
   - Stop when number of samples falling in a node is small
   - Holdout sample

$$r(T) = \sum_{t \in T_G} Dr(t, B) + \sum_{t \in T_B} Lr(t, G)$$

   - $c(T) = r(T) + d\, n(T)$
   - Choose T^* that minimizes $c(T)$.
4. Good-Bad Nodes
   - Good – Majority of Sample Cases are Good
   - Minimize Cost of Misclassification: $G/B \geq D/L$

Behavior Scoring Models of Repayment and Usage Behavior

Objectives:

- Detect danger of default in near/medium term
- Adjust credit limits, decide on marketing and operational policies

Two Methods:

- Classification Approach: Lot more variables
- Credit Scoring: Markov Chain Models

Behavior Scoring: Classification Approach

- Obtain sample history for each customer
  - Observation point
  - Performance period: 12–18 months preceding OP; Add performance characteristics to application and credit bureau information
  - Outcome point: 6–12 months after OP; Classify as good or bad (more than 3 consecutive months of missed payments).
- For any current customer carry out a classification based on past performance.
- Segment population (classification trees) and develop separate scorecards for each segment
- Ensure continuity in scores when customer moves from one segment to another

Markov Chain Approach

Problem with previous method: Using the score to alter credit limit policy invalidates the effectiveness of the score.

Simple Model

- $S = \{NC, 0, 1, \ldots, M\}$, $NC-$ No credit, $0-$ Credit but payments up-to-date, $M-$ Default
-
$$P = \begin{bmatrix} .79 & .21 & \cdot & \cdot & 0 \\ .09 & .73 & \cdot & \cdot & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0.06 & .32 & \cdot & \cdot & 0.02 \end{bmatrix}$$
- $\alpha_0 = (1, 0, \ldots, 0) \rightarrow \alpha_1 = (.64, .32, \ldots, 0) \rightarrow \cdots \rightarrow \alpha_{10} = (.315, .512, \ldots, .046)$
- $\pi = P\pi$: limiting probabilities.

Markov Chain Approach

- State $s = (b, n, i)$, where b =balance outstanding, n =number of periods of non-payment, i =other charcteristics
- $t(s, a)$ = probability an account in s repays a next period
- $w(s, i')$ = probability an account in s changes its other characteristics to i'
- $r(s)$ = expected value of orders placed per period in state s

Objective: Find credit limit policy to

– maximize expected one period reward subject to a given level of bad debt

– minimize expected bad debt in one period subject to a given level of reward

$D(s)$ = probability of default in state s

$$D(b, N, i) = 1, \qquad \forall\ i, b > 0, \qquad D(0, n, i) = 0, \qquad \forall\ i, n \leq N$$

$$D(b, n, i) = \sum_{i', a \neq 0} t(s, a) w(s, i') D(b - a, 0, i') + \sum_{i'} t(s, 0) w(s, i') D(b, n + 1, i')$$

Credit Policy

- Solve for $D(b, n, i)$ for each state
-

State	s_7	s_1	s_3	s_2	s_9	s_8	s_4	s_6	s_5
Default probability	.001	.004	.008	.02	.06	.09	.12	.15	.2
Value of orders	50	150	250	200	120	80	60	40	50

- Total expected value of orders 1000
- Acceptable level of default $= 0.1 \quad \Rightarrow \quad 15\%$ orders lost
- Choose acceptable level D^* of default. Determine credit limit $L(n, i)$ using

$$D(L(n, i), n, i) = D^*$$

Markov Decision Process Approach

- Previous Model
  - States – Repayment performance
  - Transition Matrix – Dynamics of repayment behavior
  - Reward function – value to the organization
- Decisions lenders can make that impact on the rewards and the transition between states

Definition

- $X_t, \; t = 1, 2, \ldots \in S$
- K_i = set of actions of which one is chosen each time the process enters state i
- $r_k(i)$ = reward for choosing action k in state i
- $p^k(i, j)$ = probability system moves from i to j when action k is chosen

$v_n(i)$ Optimal reward over n periods starting with state i

Objective: Choose $k \in K_i$ so as to maximize some function of the reward

Finite Horizon:

$$v_n(i) = \max_{k \in K_i} \left\{ r^k(i) + \sum_j \beta \, p^k(i,j) v_{n-1}(j) \right\}, \qquad \forall \, i \in S, \qquad n = 1, 2, \ldots$$

$0 \equiv v_0(\cdot) \rightarrow v_1(\cdot) \rightarrow v_2(\cdot) \rightarrow \cdots \rightarrow v_n(\cdot)$

Infinite Horizon:

$$v(i) = \max_{k \in K_i} \left\{ r^k(i) + \sum_j \beta \, p^k(i,j) v(j) \right\}, \qquad \forall \, i \in S$$

For some $n \geq n^*$, $v_n(i) = v(i)$ for all $i \in S$.

Markov Decision Process Behavioral Scoring System

- Actions – Credit Limit L
- Specify – $p^L(s, s')$ and $r^L(s)$
- $t^L(s, a), q^L(s, e), w^L(s, i')$– probabilities of repayment, order and transition
- Define $p^L(s, s')$ by estimating

$$
\begin{aligned}
p^L(b, n, i; b + e - a, 0, i') &= t^L(s, a) q^L(s, e) w^L(s, i'), \qquad b + e - a \leq L \\
p^L(b, n, i; b - a, 0, i') &= t^L(s, a) \left(q^L(s, 0) + \sum_{e \geq L - b + a} q^L(s, e) \right) w^L(s, i') \\
p^L(b, n, i; b + e, n + 1, i') &= t^L(s, 0) q^L(s, e) w^L(s, i'), \qquad b + e \leq L \\
p^L(b, n, i; b, n + 1, i') &= t^L(s, 0) \left(q^L(s, 0) + \sum_{e \geq L - b} q^L(s, e) \right) w^L(s, i')
\end{aligned}
$$

$$r^L(b,n,i) = f \sum_e e\, q^L(s,e) - b\, t^L(s,0)\, \delta(n-(N-1))$$

Optimal Total Discounted Expected Profit

$$v_n(s) = \max_L \left\{ r^L(s) + \sum_{s' \in S} p^L(s,s')\, \beta\, v_{n-1}(s') \right\}, \qquad \forall\, s \in S, \qquad n = 1, 2, \ldots$$

$$v(s) = \max_L \left\{ r^L(s) + \sum_{s' \in S} p^L(s,s')\, \beta\, v(s') \right\}, \qquad \forall\, s \in S$$

Optimal Credit Limits: Actions that maximize the above equations

Validation of Markov Chain Models

Estimation

- Stationary Markov Chain: R customer histories $s_0^r, s_1^r, \ldots, s_T^r$. $n^r(i), n^r(i,j)$ number of times state i, $i \longrightarrow j$ appears in the history of customer r.
  $n(i) = \sum_r n^r(i)$, $n(i,j) = \sum_r n^r(i,j)$.

$$\hat{p}(i,j) = \frac{n(i,j)}{n(i)}$$

- Non-Stationary Markov Chain: Seasonal Effect, Trend

$$\hat{p}_t(i,j) = \frac{n_t(i,j)}{n_t(i)}$$

Testing for Stationarity

$$\sum_{i \in S} \sum_{t=1}^{T-1} \sum_{j \in S} \frac{n_t(i)[\hat{p}_t(i,j) - \hat{p}(i,j)]^2}{\hat{p}(i,j)} \sim \chi^2_{m(m-1)(T-1)}$$

Testing for Markovity

$$\sum_{i,j,k} \frac{n(i,j)[\hat{p}(i,j,k) - \hat{p}(j,k)]}{\hat{p}(j,k)} \sim \chi^2_{m(m-1)^2}$$

Non-Markov

- Enlarge state space
- Segment population on basis of age, product, time of year etc.
- Mover-Stayer Model $\rightarrow$ Mixture Markov Models

Mixture Markov Models Moving at Different Speeds

Ref. Halina Frydman

- Labor Mobility
- Income dynamics
- Consumer Brand Preferences
- Rating Migration
- Credit behavior
- Tumor progression

Assumption Individual transitions are Markovian, but population is heterogeneous with respect to speed of transitions

Definition

- $\{X(t), t \geq 0\}$ stochastic process with state space $S = \{1, 2, \ldots, w\}$
- Conditional on the initial state, X_t is a mixture of homogenous Markov chains X_m, $1 \leq m \leq N$
- Mixture Probabilities $s_{i,m} = P[X = X_m | X(0) = i]$
- The generator of X_m is $A_m = \Gamma_m Q$

$$a_{ij,m} \geq 0, \qquad i \neq j, \qquad \sum_{j \neq i} a_{ij,m} = -a_{ii,m} = a_{i,m} \leq 0$$

- $\Gamma_m =$ diag$(\gamma_{1,m}, \gamma_{2,m}, \ldots, \gamma_{w,m})$, $1 \leq m \leq N-1$, $\Gamma_N = I$
- $S_m =$ diag$(s_{1,m}, \ldots, s_{w,m})$, $\qquad P_m(t) = \exp(tA_m)$

$$P(t) = \sum_{m=1}^{N} S_m P_m(t), \qquad t \geq 0$$

- $\gamma_{i,m} = 0$: No movement, $0 < \gamma_{i,m} < 1$: Faster, $\gamma_{i,m} > 1$: Slower

Possible generalization: Introduce two gamma parameters for each row, one for upward transitions (above diagonal) and the other for downward transitions.

Estimation: Maximum Likelihood Estmation via EM Algorithm

Key: Estimation of the probability that a particular **history** follows X_m

Estimation

- X is a mixture of N Markov processes
- n realizations from X : $X^1, \ldots, X^n$, with X^k observed on $[0, T^k]$.
- **Complete Information:** We know which Markov chain generated each realization
- Data for X^k : $(n_{ij}^k, \tau_i^k, Y_m^k, i, j \in S, j \neq i)$

  $n_{ij}^k =$ number of times X^k makes an $i \to j$ transition, $i \neq j$

  $\tau_i^k =$ total time X^k spends in state i

  $Y_m^k = 1$ if X^k is generated by X_m and 0 otherwise

$$s_{i,m} = P[X = X_m | X(0) = i]$$

- Likelihood of X^k conditional on knowing the initial state $X^k(0) = i_k$

$$L^k = s_{i_k,m} \prod_{i \neq j} (\gamma_{i,m} q_{ij})^{n_{ij}^k} \prod_i \exp(-\gamma_{i,m} q_i \tau_i^k)$$

- ... and knowing the Y_m^k's

$$\log L^k = \sum_{m=1}^{N} Y_m^k \left\{ \log(s_{i_k,m}) + \sum_{i \neq j} n_{ij}^k [\log(\gamma_{i,m}) + log(q_{ij})] - \sum_i \gamma_{i,m} q_i \tau_i^k \right\}$$

Full log-likelihood:

$$\begin{aligned}\log L &= \sum_{m=1}^{N} \sum_{i} b_{i,m} \left\{ \log(s_{i,m}) + n_{i,m} \log(\gamma_{i,m}) \right\} \\ &\quad + \sum_{i \neq j} n_{ij} \log(q_{ij}) - \sum_{i} q_i \sum_{m=1}^{N} \gamma_{i,m} \tau_{i,m}\end{aligned}$$

$$\begin{aligned}b_{i,m} &= \#\, X_m \text{ realizations starting in } i \\ n_{ij} &= \sum_{k=1}^{n} n_{ij}^{k} = \#\, i \to j \text{ transitions} \\ n_i &= \sum_{j \neq i} n_{ij} = \#\text{ transitions from } i \\ n_{i,m} &= \sum_{k=1}^{n} Y_m^k n_i^k = \#\text{ transitions from } i \text{ by } X_m \text{ realizations} \\ \tau_{i,m} &= \sum_{k=1}^{n} Y_m^k \tau_i^k = \text{time in } i \text{ for } X_m \text{ realizations}\end{aligned}$$

ML Estimates under complete information:

$$
\begin{aligned}
\hat{s}_{i,m} &= \frac{b_{i,m}}{b_i}, \qquad b_i = \sum_{m=1}^{N} b_{i,m}, \qquad i \in S \\
\hat{q}_i &= \frac{n_{i,N}}{\tau_{i,N}} \\
\hat{q}_{ij} &= \frac{n_{ij}}{n_i}\hat{q}_i \\
\hat{\gamma}_{i,m} &= \frac{n_{i,m}}{\hat{q}_i \tau_{i,m}}
\end{aligned}
$$

EM Algorithm

- Algorithm will estimate $(s_{i,m}, \gamma_{i,m}, q_i)$
- After algorithm converges compute q_{ij}
- **Step 1** Choose initial values $(s^0_{i,m}, \gamma^0_{i,m}, q^0_i)$
- **Expectation Step:** Estimate probability that X^k follows X_m

$$E^0[Y^k_m] = \frac{L^k_{0,m}}{\sum_{p=1}^N L^k_{0,p}}$$

$$E^0[n_{i,m}] = \sum_{k=1}^n n^k_i E^0[Y^k_m]$$

$$E^0[\tau_{i,m}] = \sum_{k=1}^n \tau^k_i E^0[Y^k_m]$$

$$E^0[b_{i,m}] = \sum_{k=1}^n I(X^k(0) = i) E^0[Y^k_m]$$

- **Maximization Step:** Compute new values $(s^1_{i,m}, \gamma^1_{i,m}, q^1_i)$ using ML Estimates
- Repeat until convergence

- **Empirical Study:** 848 bond issuers from Moody's data base observed continuously for different periods of time from January 1985 to December 1995.
- **Seven coarser rating categories:** Aaa, Aa, A, Baa, Ba, B, C and two additional states, D and WR.
- WR could be due to debt expiration or due to non-rating for failure to pay fee
- Likelihood ratio test strongly rejects the Markov chain model
- Fitted a mixture of two Markov chains with generators Q and $A = \Gamma Q$
- A is dominant with 78.31% histories evolving from A.
- Q moves out faster from lower ratings; Those in A stay in C for an average of 3.5 yrs. (includes all original bonds (1%) rated C) whereas those from Q move out of C after an average of 1/3 of a year (most were downgraded to C).
- Default probabilities are higher under Q (almost four times higher)
- In A, WR acts as an absorbing state, whereas in Q it represents a temporary withdrawal of ratings
- Three year default probability from C is 0.36 for the mixture model, while it is 0.45 for a Markov chain

Bayesian Markov Chain Approach

Previous Model: Estimate customer behavior from other customers with similar scores

Bayesian Approach: Modify based on individual customer's behavior

Example: $X \sim B(1, p)$, $X = 0$ no repayment

Beta Prior on p

$$f_{r,m}(p) = \frac{\Gamma(m)}{\Gamma(r)\Gamma(m-r)} p^{r-1}(1-p)^{m-r-1}, \qquad 0 \le p \le 1$$

Suppose initial prior belief $f_0(p) = f_{r_0,m_0}(p)$. $r_0/m_0 = P[X = 1]$.

If $X_1 = 1$ then

$$f_1(p) = \frac{P[X_1 = 1|p] f_0(p)}{P[X_1 = 1]} = f_{r_0+1,m_0+1}$$

Similarly, if $X_1 = 0$, then $f_1(p) = f_{r_0,m_0+1}$.

Let m be the number of periods for which there is history, of which payments have been made in r periods. Then, $f_m(p) = f_{r_0+r,m_0+m}(p)$. For large m, (r, m) dominates (r_0, m_0).

Fixed amount repayment loan

State of customer: $s = (b, n, r, m)$

$$D(b, n, r, m) = \frac{r}{m} D(b - a, 0, r + 1, m + 1) + \left(1 - \frac{r}{m}\right) D(b, n + 1, r, m + 1),$$

$n = 0, 1, \ldots, N$, with $D(0, n, r, m) = 0$ and $D(b, N, r, m) = 1$.

Credit Card Repayment Model

- $R =$ required repayment level (some percentage of b)
- Two parameters described by Bayesian beliefs: p, M.
- $p =$ probability of repayment, $M =$ maximum affordable repayment amount
- If $M < R$ customer will pay M. If $M \geq R$, customer will pay between R and M.
- Belief about p is given by $f_{r,m}(p)$.
- Split repayment into K possible levels, $1, 2, \ldots, K$. Define independent Bernoulli variables M_i,

$$p_i = P[M_i = 1] = P[M \geq i | M \geq i - 1], \qquad i = 1, 2, \ldots, K.$$

- p_i =probability of payment at level $i \sim f_{r_i, m_i}$

- State Transitions: $(b, n, r, m, r_1, m_1, \ldots, r_K, m_K) \rightarrow$
  - If no payment

$$(b, n+1, r, m+1, r_1, m_1, \ldots, r_K, m_K)$$

  - If payment $a < R$

$$(b-a, 0, r+1, m+1, r_1+1, m_1+1, \ldots, r_a+1, m_a+1, r_{a+1}, m_{a+1}+1, \ldots, r_K, m_K)$$

  - If payment $a \geq R$

$$(b-a, 0, r+1, m+1, r_1+1, m_1+1, \ldots, r_a+1, m_a+1, r_{a+1}, m_{a+1}, \ldots, r_K, m_K)$$

- Probability of repayment in next period $= \frac{r}{m}$
- Expected affordable repayment amount

$$E[M] = \sum_{k=1}^{K} P[M \geq k] = \frac{r_1}{m_1}\left(1 + \frac{r_2}{m_2}\left(1 + \cdots \left(\frac{r_K}{m_K}\right)\right) \cdots\right)$$

- Add probability $q(e)$, of taking new credit e to develop a Markov decision process model