Bayes Multiple Binary Classifier - How to Make Decisions Like a Bayesian.

Wensong Wu
Division of Statistics
Department of Mathematics and Statistics
Florida International University

Mathematics Colloquium Series
Nova Southeastern University
Nov. 10, 2015

Agenda

- Introduction to Bayesian statistics.
- Bayesian Multiple Binary Classifier.

Bayesian Statistics

What is Bayesian Statistics?

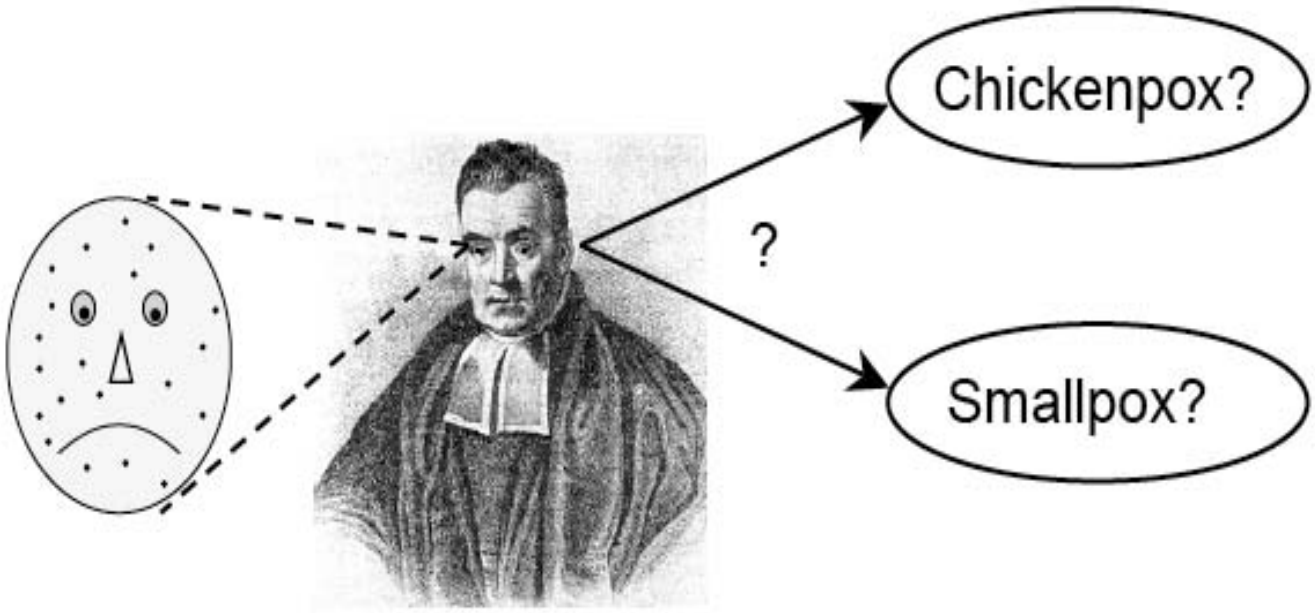
- To most statisticians: An approach to the philosophy of science, different from classical (Frequentist) statistics.
- To most practitioners: A convenient methodology to analyze data, usually highly computational.

An introduction today:

- Bayes' Rule: A Probability rule that makes all happen.
- Bayesian inference: What Bayesian can do.
- Why Bayesian has become popular.
- Bayesian decision theory: Making decision as a Bayesian!

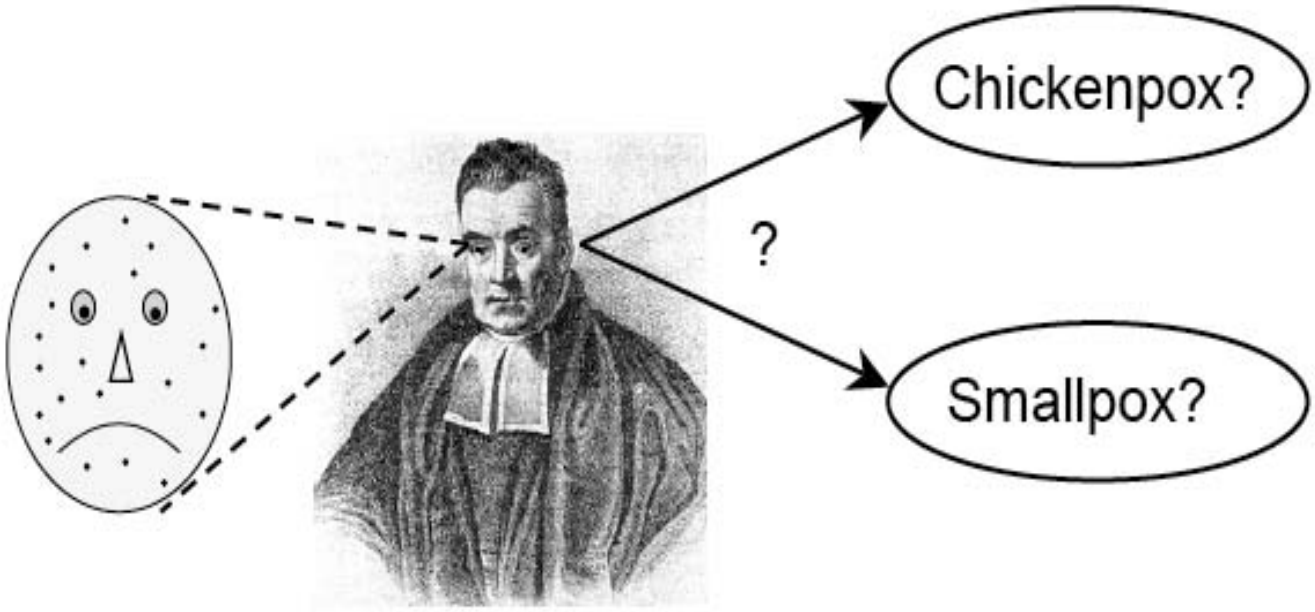
Bayes' Rule: Poxy Disease

Suppose you are a doctor, confronted with a patient who is covered in spots all over his face. The patient's symptoms are consistent with chickenpox, but they are also consistent with another, more dangerous, disease, smallpox. Decision to make: A diagnosis.



Bayes' Rule: Poxy Disease

- Observed data x= spots
- Unknown truth of disease, θ, could be θ_c=chickenpox or θ_s=smallpox.



Bayes' Rule: Poxy Disease

Likelihood: You know that 80% of people with chickenpox have spots, but also that 90% of people with smallpox have spots.

- $Likelihood(\theta_c) = p(x|\theta = \theta_c) = P(spots|chickenpox) = 0.8$
- $Likelihood(\theta_s) = p(x|\theta = \theta_s) = P(spots|smallpox) = 0.9$

Principle of classical statistical inference: the estimated unknown truth would maximize the likelihood function (MLE).

Based on MLE principal, what's your diagnosis?

Bayes' Rule: Poxy Disease

Prior probabilities: As a knowledgeable doctor, you know that chickenpox is common, whereas smallpox is rare, and is therefore intrinsically improbable. The prevalence of both diseases are 100 and 1 in every 1,000 individuals, respectively.

- $p(\theta_c) = P(chickenpox) = 0.1$
- $p(\theta_s) = P(smallpox) = 0.001$

Need to combine the prior information and the data likelihood.

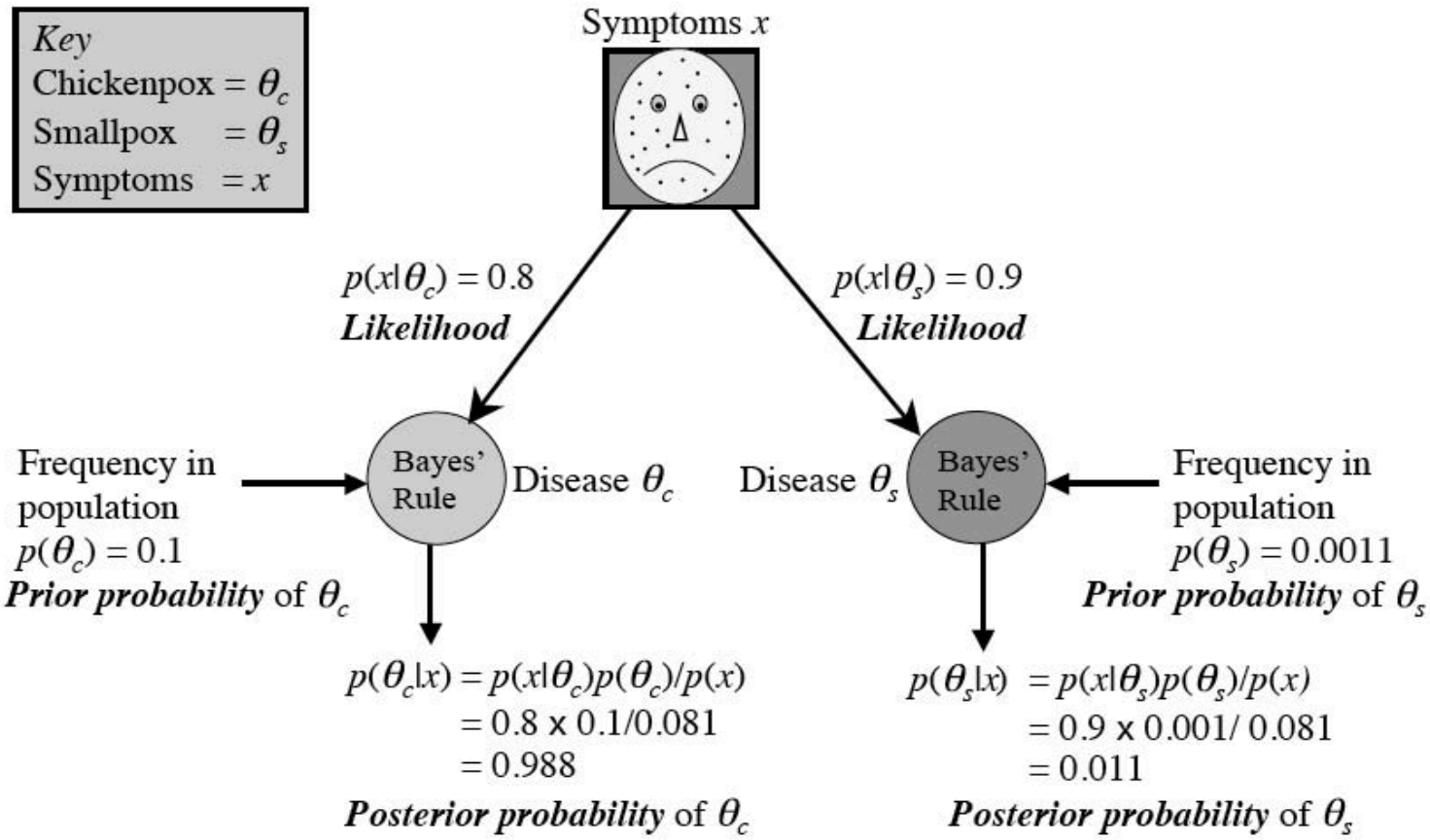
Bayes' Rule: Poxy Disease

Posterior probabilities: Use Bayes' rule to find probability of disease given data, "Bayesian update", "weighted likelihood".

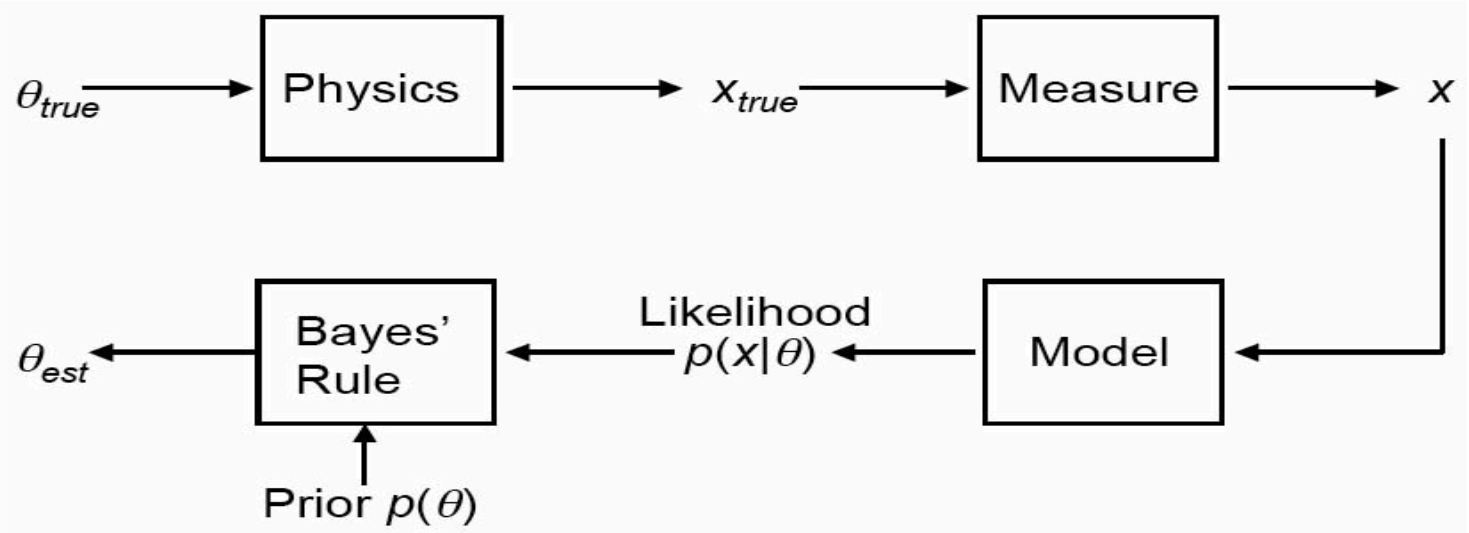
$$
\begin{aligned}
p(\theta_c|x) &= P(chickenpox|spots) = \frac{P(\text{chickenpox and spots})}{P(spots)} \\
&= \boxed{\frac{p(x|\theta_c)p(\theta_c)}{p(x|\theta_c)p(\theta_c) + p(x|\theta_s)p(\theta_s)}} \\
&= \frac{0.8 * 0.1}{0.8 * 0.1 + 0.9 * 0.001} = 0.989. \\
p(\theta_s|x) &= \boxed{\frac{p(x|\theta_s)p(\theta_s)}{p(x|\theta_c)p(\theta_c) + p(x|\theta_s)p(\theta_s)}} \\
&= \frac{0.9 * 0.001}{0.8 * 0.1 + 0.9 * 0.001} = 0.011.
\end{aligned}
$$

Bayes' Rule: Poxy Disease

One Bayesian principle: the estimated unknown truth would maximize a posterior (MAP). Diagnosis based on MAP principle?



Bayesian Inference



- Frequentist: fixed θ. Bayesian: random θ according to the prior distribution $p(\theta)$, a belief before gaining the current data.
- Choices of prior distribution: subjective, objective, Empirical Bayes, hierarchical prior, nonparametric Bayes,...
- Bayesian version of almost all inference: estimation, "confidence" interval, hypothesis "testing", prediction, model assessment and selection...

Why Bayesian Statistics Becomes Popular?

The theory is straightforward and with good properties.

- Credible interval: Mid 95% under posterior distribution.
- Hypothesis test based on post. probability of alternative $P(H_a|x)$ or Bayes factor.
- Select model based on post. probability of models $P(Model|x)$.
- Automatic shrinkage and parsimony.

Computation is feasible, now.

- Not all models have a conjugate prior, so that the posterior distribution may not be in a closed form.
- Luckily, we have Monte Carlo methods to approximate.
- In the case of high dimension of parameters MLE fails, but Bayesian works!

Bayesian Decision Theory

- Unknown truth of parameter: θ.
- Data X has a distribution given the value of θ.
- In order to discover the truth we make an action a.
- Loss function $L(a, \theta)$ represents the loss of an action vector a when the parameter is θ.
- The decision making process is represented by a decision function $\delta : \mathit{data} \rightarrow \mathit{action}$.
- The risk of decision is the average loss over all possible sample data: $R(\delta, \theta) = E_X(L(\delta(X), \theta))$.
- Classical decision theory has many different criteria to select the best decision function with the smallest risk.

Bayesian Decision Theory

- Bayesian decision theory assigns a prior distribution of $\theta \sim p(\theta)$.
- Bayes risk is the average of risk over all possible values of θ: $r(\delta) = E_\theta(R(\delta, \theta))$.
- The optimal decision making procedure is the δ^* that minimize the Bayes risk.
- As a Bayesian, you would go for the Bayes optimal action $a^* = \delta^*(data)$!

Agenda

- Introduction to Bayesian statistics.
- Bayesian Multiple Binary Classifier.

An Example: Diagnosis of Disease

Disease	Symptom 1	Symptom 2	...	Symptom 20
Yes	Yes	No	...	Medium
No	No	Yes	...	Low
⋮	⋮	⋮	⋮	⋮
Yes	Yes	Yes	...	High
?	No	No	...	Medium
?	Yes	Yes	...	Low
?	⋮	⋮	⋮	⋮
?	No	Yes	...	Low

An Example: Diagnosis of Disease

Disease	Sym1	Sym2	...
Y_1	X_{11}	X_{12}	...
Y_2	X_{21}	X_{22}	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$
Y_N	X_{N1}	X_{N2}	...
W_1	V_{11}	V_{12}	...
W_2	V_{21}	V_{22}	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$
W_M	V_{M1}	V_{M2}	...

Multiple Binary Classification

- For N subjects on file (training set), observe Y_i=I(Disease), and X_{i1}, X_{i2}, $\ldots, X_{ip}$, P categorical traits, $i = 1, 2 \ldots, N$.
- For M new subjects, observe $V_{m1}, V_{m2}, \ldots, V_{mp}$, $m = 1, 2 \ldots, M$
- Want to predict who have the disease by predicting $W_m, m = 1, 2, \ldots, M$ simultaneously.

An Example: Diagnosis of Disease

Disease	Sym1	Sym2	...
Y_1	X_{11}	X_{12}	...
Y_2	X_{21}	X_{22}	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$
Y_N	X_{N1}	X_{N2}	...
W_1	V_{11}	V_{12}	...
W_2	V_{21}	V_{22}	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$
W_M	V_{M1}	V_{M2}	...

Sparse Logic Models

- Sparsity: Only a few of symptoms are relevant.
- Want to discover the logic rule of symptoms, for example, higher risk if Symptom 1 **and** Symptom 2 are present, **or** Symptom 3 is at the low level.

Importance and Framework

- High-Dimensional Classification.
  - Screening for ADHD students in a new school based on the records of previously studied students in other schools;
  - Admission of future students based their records and on the records of previous admitted students;
  - Issue of credit cards among many applicants;
  - ...
- Theoretical framework.
  - Decision-theoretic.
  - Bayesian.
- Computational issues.
  - Make optimal multiple classifications.
  - Search for the best Logic rules.
  - Bayesian implementations.

Boolean Functions

- $B_q : \{0,1\}^q \to \{0,1\}$ is a Boolean function of dimension q if $B_q(u_1, u_2, \ldots, u_q)$ is a logic expression built from binary inputs $u_1, u_2, \ldots, u_q$ by using the operators $\wedge$ ("and"), $\vee$ ("or"), and brackets.
- Example: $B_3(u_1, u_2, u_3) = (u_1 \wedge u_2) \vee u_3 =$ I(Symptom 1 and Symptom 2 are present, or Symptom 3 is low), where
  - $u_1 =$I(Symptom 1 is present),
  - $u_2 =$I(Symptom 2 is present),
  - $u_3 =$I(Symptom 3 = Low).

Sparse Logic Regression Models

- $\mathbf{X}_i \stackrel{d}{=} \mathbf{V}_m \stackrel{iid}{\sim} G(\cdot;\eta), \eta \in \mathcal{E}$ unknown.
- $Y_i|\mathbf{X}_i = \mathbf{x}_i \stackrel{ind}{\sim} Bernoulli(\psi(\mathbf{x}_i;\boldsymbol{\beta},(A,\mathcal{C},B)))$,
  - $\psi(\mathbf{x}_i;\beta,(A,\mathcal{C},B)) = h[\beta_0 + \beta_1 \boxed{B(\mathbf{x}_i;(A,\mathcal{C}))}]$,
  - $h(\cdot)$ is a known link function,
  - $\boldsymbol{\beta} = (\beta_0,\beta_1) \in \mathcal{B}$ is unknown regression coefficients,
  - $\boxed{B(\mathbf{x}_i;(A,\mathcal{C}))} \equiv B_{\sum_{j\in A}|\mathcal{C}_j|}(\{I(x_{ij} = C) : C \in \mathcal{C}_j, j \in A\})$ is a Boolean expression involving a subset of traits $\{x_{ij} : j \in A \subset \{1,2,\ldots,P\}\}$ at levels $\mathcal{C} = \{\mathcal{C}_j \subset \mathcal{X}_{0j} : j \in A\}$.
  - Sparsity: $|A| \ll P$.
- $W_m|\mathbf{V}_m = \mathbf{v}_m \stackrel{ind}{\sim} Bernoulli(\psi(\mathbf{v}_m;\beta,(A,\mathcal{C},B)))$
- $\boldsymbol{\theta} = (\eta,\boldsymbol{\beta}) \in \Theta = \mathcal{E}\times\mathcal{B}$ is the unknown parameter vector.
- Assume $(A,\mathcal{C},B)$ uniquely determines the latent logic model.
- This defines a joint distribution of $(\mathbf{S},\mathbf{W}) = (\mathbf{X},\mathbf{Y},\mathbf{V},\mathbf{W})$ given parameter $\boldsymbol{\theta}$ and logic model $\mathcal{M} = (A,\mathcal{C},B)$.

Bayes Multiple Binary Classifier (BaMBiC)

- Action $\mathbf{a} = (a_1, a_2, \ldots, a_M) \in \mathcal{A} = \{0, 1\}^M$
  - Interpretation: $a_m = I(\text{Classify } W_m = 1)$.
- Multiple Decision Function: $\delta = (\delta_1, \ldots, \delta_M) : \mathcal{S} \to \mathcal{A}$.
- Loss function between an action $\mathbf{a}$ and an unknown truth $\mathbf{w}$: $L(\mathbf{a}, \mathbf{w})$.
- Prior $\Pi = \Pi_{\mathcal{M}} \Pi_{\boldsymbol{\theta}}$ with independent prior on the model space $\mathcal{M} = (A, \mathcal{C}, B) \sim \Pi_{\mathcal{M}}$ and parameter space $\boldsymbol{\theta} \sim \Pi_{\boldsymbol{\theta}}(\cdot)$.
- Bayes risk of MDF: $r_\Pi(\delta) = E_{\mathcal{M}} E_{\boldsymbol{\theta}} E_{(\mathbf{S},\mathbf{W})|(\mathcal{M},\boldsymbol{\theta})} L(\delta(\mathbf{S}), \mathbf{W})$.
- Bayes Multiple Binary Classifier (BaMBiC) is the δ^* such that

$$r_\Pi(\delta^*) = \inf_{\delta \in \mathcal{D}} r_\Pi(\delta).$$

Loss Function Specification

General Form: $L(\mathbf{a}, \mathbf{w}) = \lambda L_0(\mathbf{a}, \mathbf{w}) + L_1(\mathbf{a}, \mathbf{w})$,

- L_0 is a Type I-type loss: about false postives.
  - $L_0(\mathbf{a}, \mathbf{w}) = \frac{1}{M} \sum_{m=1}^{M} a_m(1 - w_m) = \text{FP}$
  - $L_0(\mathbf{a}, \mathbf{w}) = \frac{\sum_{m=1}^{M} a_m(1 - w_m)}{[\sum_{m=1}^{M}(1 - w_m)] \vee 1} = \text{FPR=1-Specificity}$
- L_1 is a Type II-type loss: about false negatives.
  - $L_1(\mathbf{a}, \mathbf{w}) = \frac{1}{M} \sum_{m=1}^{M}(1 - a_m)w_m = \text{FN}$
  - $L_1(\mathbf{a}, \mathbf{w}) = \frac{\sum_{m=1}^{M}(1 - a_m)w_m}{[\sum_{m=1}^{M} w_m] \vee 1} = \text{FNP=1-Sensitivity}$
- Many choices of (L_0, L_1) loss functions pairs.
- λ is a pre-determined cost ratio.

Bayes Multiple Binary Classifier (BaMBiC)

- Bayes risk $r_\Pi(\delta) = E_{\mathbf{S}} E_{(\mathbf{W},\mathcal{M},\theta)|\mathbf{S}} L(\delta(\mathbf{S}), \mathbf{W})$.
- Posterior average loss function of an action $\mathbf{a} \in \mathcal{A}$ and data $\mathbf{s}$

$$\tilde{L}(\mathbf{a}, \mathbf{s}) = E_{(\mathbf{W},\mathcal{M},\theta)|\mathbf{S}=\mathbf{s}}(L(\mathbf{a}, \mathbf{W})).$$

- If the prior Π specifies the independence of η and β,

$$\tilde{L}(\mathbf{a}, \mathbf{s}) = E_{\mathcal{M}|(\mathbf{X}=\mathbf{x},\mathbf{Y}=\mathbf{y})} E_{\beta|(\mathbf{X}=\mathbf{x},\mathbf{Y}=\mathbf{y},\mathcal{M})} E_{\mathbf{W}|(\mathbf{V}=\mathbf{v},\mathcal{M},\beta)} L(\mathbf{a}, \mathbf{W}).$$

- Bayes optimal action of $\mathbf{s} = (\mathbf{x}, \mathbf{y}, \mathbf{v})$

$$\mathbf{a}^*(\mathbf{s}) = \arg\min_{\mathbf{a}\in\mathcal{A}} \tilde{L}(\mathbf{a}, \mathbf{s}).$$

- Form of the BaMBiC: $\delta^* : \mathbf{S} \to \mathcal{A}$ with $\delta^*(\mathbf{S}) = \mathbf{a}^*(\mathbf{S})$.
- Difficulty 1: $\mathcal{A} = \{0, 1\}^M$ has 2^M elements.
- Difficulty 2: Too many possible logic models $\mathcal{M} = (A, \mathcal{C}, B)$.

Efficient Searching of the Bayes Optimal Action

A Two-Step Searching Strategy

- Step I: Find the best action in each of the sub-action space

$$\mathcal{A}_k : \{\mathbf{a} \in \mathcal{A} : \mathbf{a}^{\mathrm{T}}\mathbf{1} = k\}$$

  for $k = 0, 1, 2, \ldots, M$. Denote these optimal actions by $\mathbf{a}_k^*(\mathbf{x})$ for $k = 0, 1, 2, \ldots, M$.

- Step II: Find the best among the $\mathbf{a}_k^*(\mathbf{x}), k = 0, 1, 2, \ldots, M$.

- Remark: Searching order is no more than $O(M^2 \log M)$.

Efficient Searching: Step 1

Search for the best on the each sub-action space sliced by the number of "1"s;

Action Space When $M = 4$

$k = 0$ $(0,0,0,0)$

$k = 1$ $(1,0,0,0),(0,1,0,0),(0,0,1,0),(0,0,0,1)$

$k = 2$ $(1,1,0,0),(1,0,1,0),(1,0,0,1),$
$(0,1,1,0),(0,1,0,1),(0,0,1,1)$

$k = 3$ $(1,1,1,0),(1,0,1,1),(1,1,0,1),(0,0,0,1)$

$k = 4$ $(1,1,1,1)$

Efficient Searching: Step 1

Search for the best on the each sub-action space sliced by the number of "1"s;

Action Space When $M = 4$

$k = 0$	$(0,0,0,0)$
$k = 1$	$(1,0,0,0), (0,1,0,0), (0,0,1,0), (0,0,0,1)$
$k = 2$	$(1,1,0,0), (1,0,1,0), (1,0,0,1),$ $(0,1,1,0), (0,1,0,1), (0,0,1,1)$
$k = 3$	$(1,1,1,0), (1,0,1,1), (1,1,0,1), (0,0,0,1)$
$k = 4$	$(1,1,1,1)$

Efficient Searching: Step 1

Search for the best on the each sub-action space sliced by the number of "1"s.

Action Space When $M = 4$

$k = 0$	$(0,0,0,0)$
$k = 1$	$(1,0,0,0), (0,1,0,0), (0,0,1,0), (0,0,0,1)$
$k = 2$	$(1,1,0,0), (1,0,1,0), (1,0,0,1),$ $(0,1,1,0), (0,1,0,1), (0,0,1,1)$
$k = 3$	$(1,1,1,0), (1,0,1,1), (1,1,0,1), (0,0,0,1)$
$k = 4$	$(1,1,1,1)$

Efficient Searching: Step 2

Search for the best among the sub-bests.

Action Space When $M = 4$

$k = 0$	$(0, 0, 0, 0)$
$k = 1$	$(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1)$
$k = 2$	$(1, 1, 0, 0), (1,0,1,0), (1, 0, 0, 1),$ $(0, 1, 1, 0), (0, 1, 0, 1), (0, 0, 1, 1)$
$k = 3$	$(1, 1, 1, 0), (1, 0, 1, 1), (1, 1, 0, 1), (0, 0, 0, 1)$
$k = 4$	$(1, 1, 1, 1)$

Model Selection Procedure for Logic Regression Models

- Instead of model averaging when calculating posterior average loss, select one best model $\mathcal{M} = (A, \mathcal{C}, B)$ given data $\mathbf{S} = \mathbf{s}$ for better interpretation.
- Logic regression (Ruczinski et al., 2003) and its Bayesian versions (Fritsch and Ickstadt, 2007)
  - Consider equivalent Logic Trees.
  - Adaptive and stepwise tree-based algorithms.
- Our approach:

Step 1 APriori Algorithm for candidate "and" rules e.g. $R_1 = X_1 \wedge X_2$, and $R_2 = X_2 \wedge X_3$, ...

Step 2 Bayesian model selection to reduce the number of candidates.

Step 3 APriori Algorithm for candidate "or" rules of $R_1, R_2, \ldots$, e.g. $R_1 \vee R_2$,....

Step 4 Bayesian model selection to select the best rule.

APriori Algorithm

- Algorithm for frequent item set and association rule mining.
- A typical application: based on a list of shopping baskets, *if a customer buys an apple and bread, is she going to buy milk?*
- Search for "and" rules based on $(\mathbf{X}, \mathbf{Y})$: Find rules $B(\mathbf{X}; (A, \mathcal{C})) \Rightarrow (Y = 0)$ or $B(\mathbf{X}; (A, \mathcal{C})) \Rightarrow (Y = 1)$ with the following properties, denoting the rules as $LHS \Rightarrow RHS$:
  - Support$(LHS \Rightarrow RHS) = P(LHS \wedge RHS) \geq s$.
  - Confidence$(LHS \Rightarrow RHS) = P(RHS|LHS) \geq c$.
  - Length$(LHS \Rightarrow RHS) = \sum_{j \in A} |\mathcal{C}_j| + 1 \leq t$.
- Parameters s, c, and t determine the number of selected rules.
- Search for "or" rules based on $(\mathbf{X}, \mathbf{Y})$: Apply to "not $\mathbf{X}$" and by De Morgan's Law. But the meaning of support, confidence are different!

Bayesian Model Selection

- Given a logic model $\mathcal{M} = (A, \mathcal{C}, B)$ and parameter $\boldsymbol{\beta} = (\beta_0, \beta_1)$, the likelihood of data $(\mathbf{x}, \mathbf{y})$ is

$$p((\mathbf{x}, \mathbf{y})|\mathcal{M}, \boldsymbol{\beta}) = \prod_{i=1}^{N} \psi(\mathbf{x}_i; \boldsymbol{\beta}, \mathcal{M})^{y_i} \left(1 - \psi(\mathbf{x}_i; \boldsymbol{\beta}, \mathcal{M})\right)^{1-y_i},$$

  where $\psi(\mathbf{x}_i; \beta, \mathcal{M}) = h[\beta_0 + \beta_1 B(\mathbf{x}_i; (A, \mathcal{C}))]$.
- On a candidate model space, assign higher prior probability to shorter Boolean expressions: $|A| \sim d + Binomial(D - d, p_0)$, where $d = \min |A|$, $d = \max |A|$, and $p_0 \in (0, 1)$ close to 0.
- Given the length $|A|$, assign equal prior probabilities to the models of the same length.
- Assign prior of $\boldsymbol{\beta}$ independent of model: $\boldsymbol{\beta}|\mathcal{M} \sim \pi(\boldsymbol{\beta})$.

Bayesian Model Selection

- The posterior probability of model $\mathcal{M}$ given data $(\mathbf{x}, \mathbf{y})$ is

$$\pi(\mathcal{M}|(\mathbf{x}, \mathbf{y})) = \frac{p((\mathbf{x}, \mathbf{y})|\mathcal{M})\pi(\mathcal{M})}{\sum_{\mathcal{M}'} p((\mathbf{x}, \mathbf{y})|\mathcal{M}')\pi(\mathcal{M}')},$$

  where $p((\mathbf{x}, \mathbf{y})|\mathcal{M}) = \int p((\mathbf{x}, \mathbf{y})|\mathcal{M}, \boldsymbol{\beta})\pi(\boldsymbol{\beta})d\boldsymbol{\beta}$.
- The optimal model is selected as

$$\mathcal{M}^* = \arg\max_{\mathcal{M}} p((\mathbf{x}, \mathbf{y})|\mathcal{M})\pi(\mathcal{M}).$$

- In Step 2, select the "and" rules for the next step if

$$\pi(\mathcal{M}|(\mathbf{x}, \mathbf{y})) > 0.05\pi(\mathcal{M}^*|(\mathbf{x}, \mathbf{y})).$$

- If N is large, the (approximated) optimal model is selected as

$$\mathcal{M}^* = \arg\max_{\mathcal{M}} \left\{ log\ p((\mathbf{x}, \mathbf{y})|\mathcal{M}, \hat{\boldsymbol{\beta}}) + log\ \pi(\mathcal{M}) \right\},$$

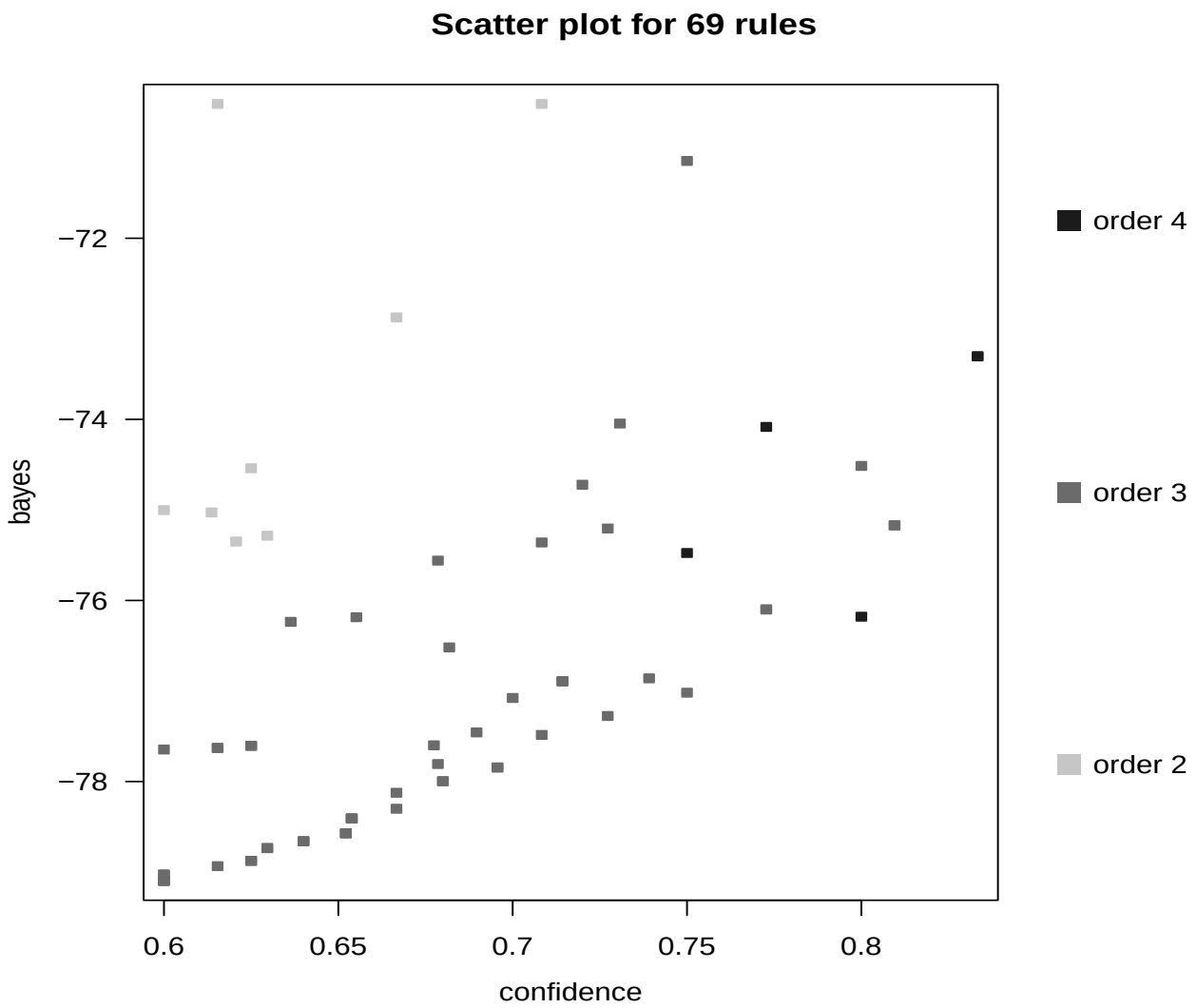
  where $\hat{\boldsymbol{\beta}}$ is the MLE under model $\mathcal{M}$.

An Illustration

- $\mathbf{X}_i, \mathbf{V}_m$, $i = 1, 2, \ldots, N = 100, m = 1, 2, \ldots, M = 100$.
- $P = 10$, the first 8 are binary, and the last 2 are of three levels with Binomial probability 0.5.
- $Y_i|\mathbf{X}_i \overset{ind}{\sim} Ber(\psi(\mathbf{X}_i; \boldsymbol{\beta}))$, $W_m|\mathbf{V}_m \overset{ind}{\sim} Ber(\psi(\mathbf{V}_i; \boldsymbol{\beta}))$.
- $\psi(\mathbf{x}; \beta) = 1/(1 + exp(\beta_0 + \beta_1 \boxed{(x_1 \wedge x_2)}))$.
- Generate data with $\beta_0 = -1, \beta_1 = 2$.
- Prior $\beta \sim N(0, 50^2)$.
- APriori Algorithm: s=0.15, c=0.6, t=5

One Simulated Data

Step 1: APriori Algorithm selected 69 "and" rules.



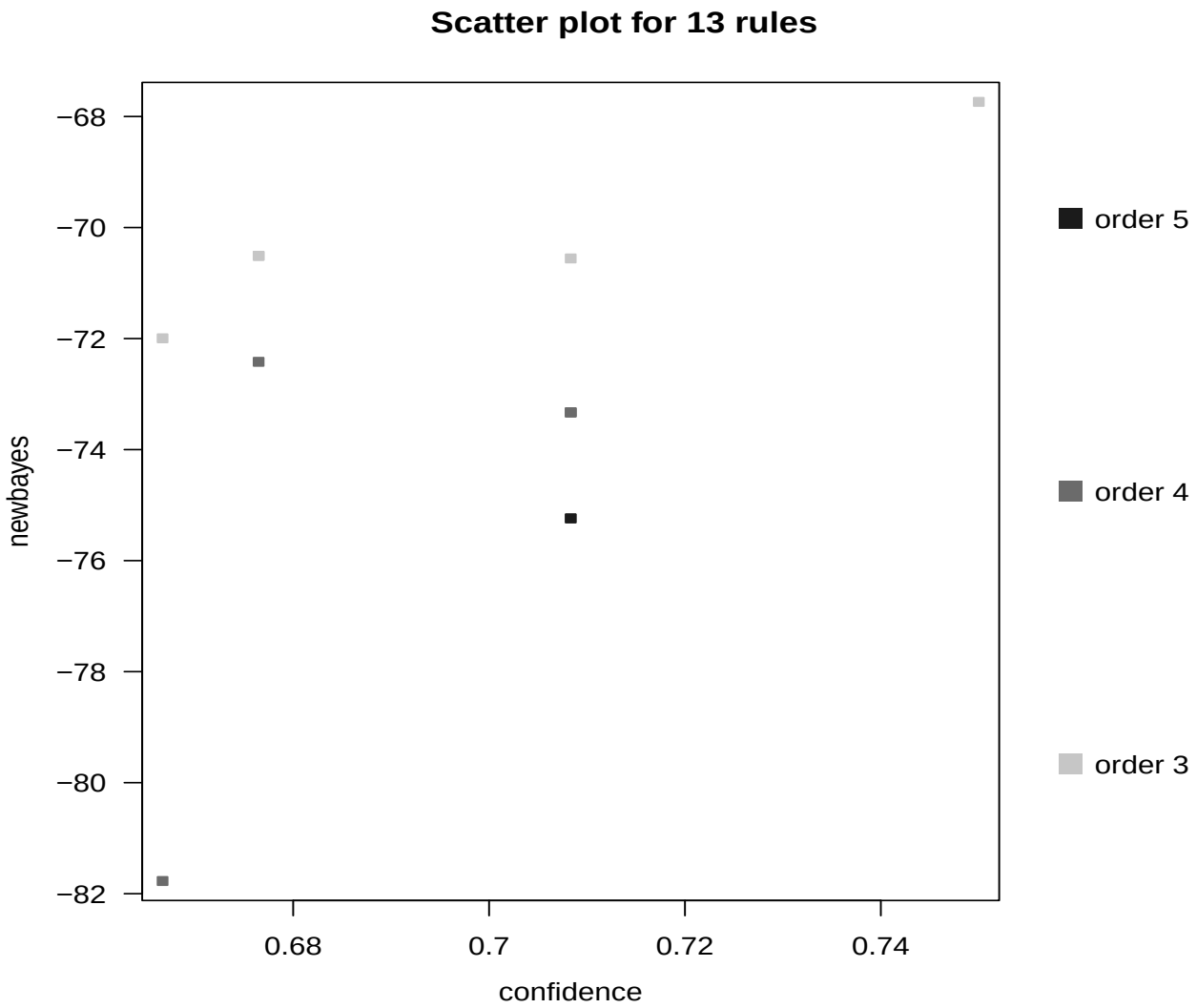
One Simulated Data

Step 2: 6 "and" rules among 69 were selected by Bayesian model selection. (Identifiability issue!)

Rule 1	x.1=1, x.2=1	$\Rightarrow$	Y=1
Rule 2	x.2=1	$\Rightarrow$	Y=1
Rule 3	x.2=0	$\Rightarrow$	Y=0
Rule 4	x.1=0	$\Rightarrow$	Y=0
Rule 5	x.1=1, x.2=1, x.6=0	$\Rightarrow$	Y=1
Rule 6	x.1=1, x.2=1, x.3=0	$\Rightarrow$	Y=1

One Simulated Data

Step 3: APriori Algorithm selected 13 "or" of previously selected 6 "and" rules.



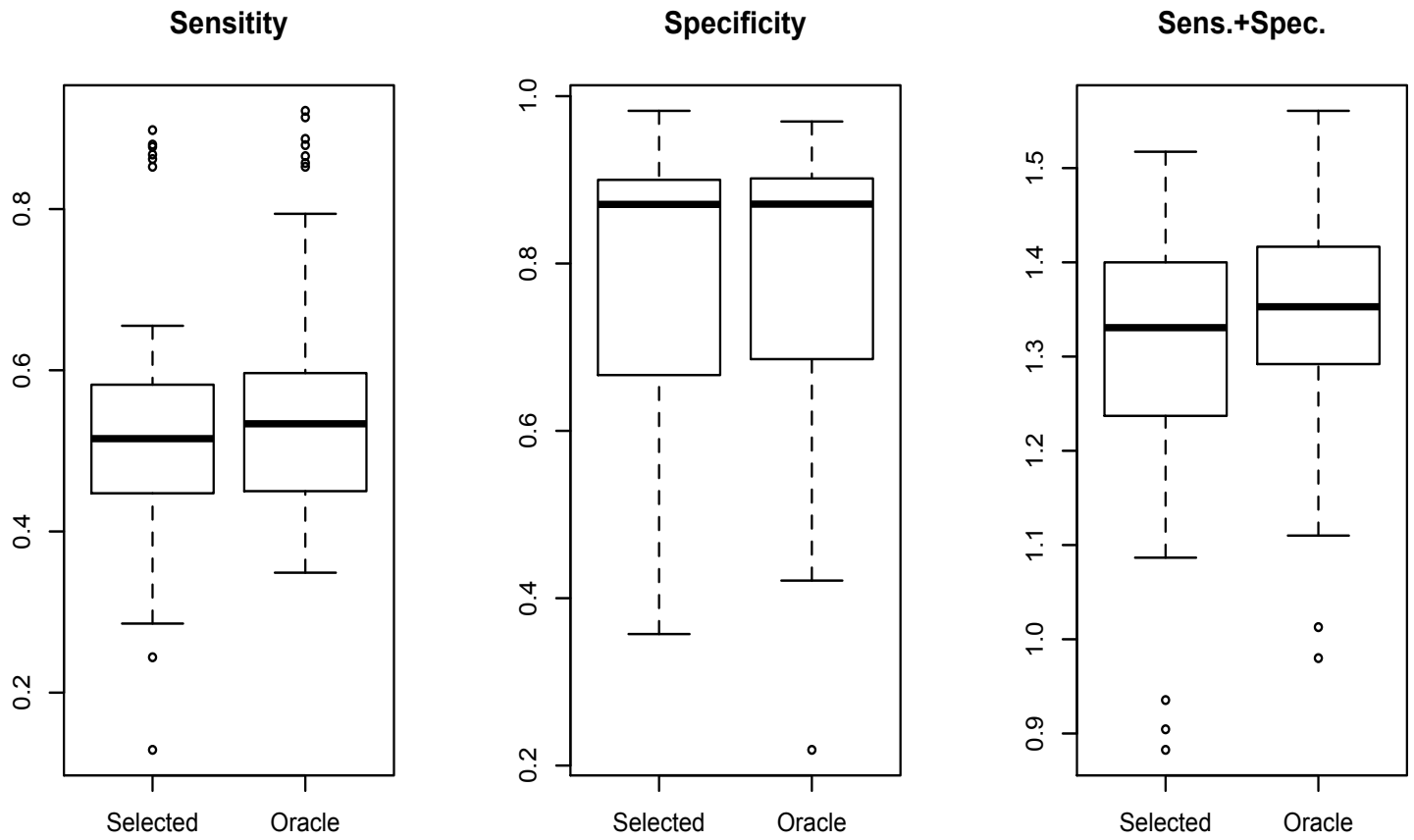
One Simulated Data

Step 4: The best model is selected among 6 "and" rules and 13 "or" rules. It is the correct rule.

Rule 1	x.1=1, x.2=1	$\Rightarrow$	Y=1
	rule 2 or rule 4	$\Rightarrow$	Y=1
Rule 2	x.2=1	$\Rightarrow$	Y=1
Rule 3	x.2=0	$\Rightarrow$	Y=0
	rule 3 or rule 5	$\Rightarrow$	Y=0
Rule 4	x.1=0	$\Rightarrow$	Y=0
Rule 5	x.1=1, x.2=1, x.6=0	$\Rightarrow$	Y=1
Rule 6	x.1=1, x.2=1, x.3=0	$\Rightarrow$	Y=1
...	...	...	...

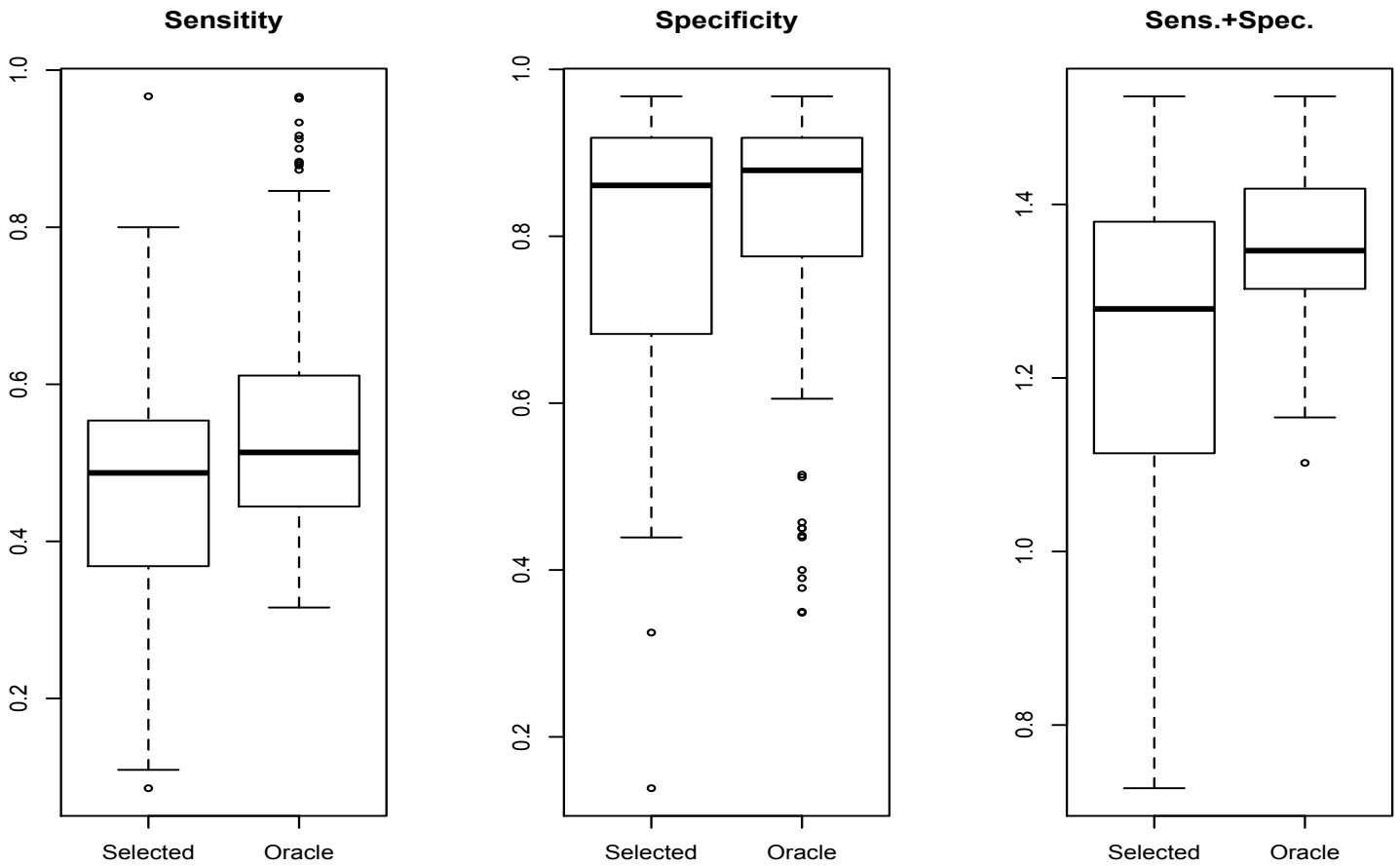
Simulations

True model $X_1 \wedge X_2$. Correct model selection 67 times out of 100 simulations.



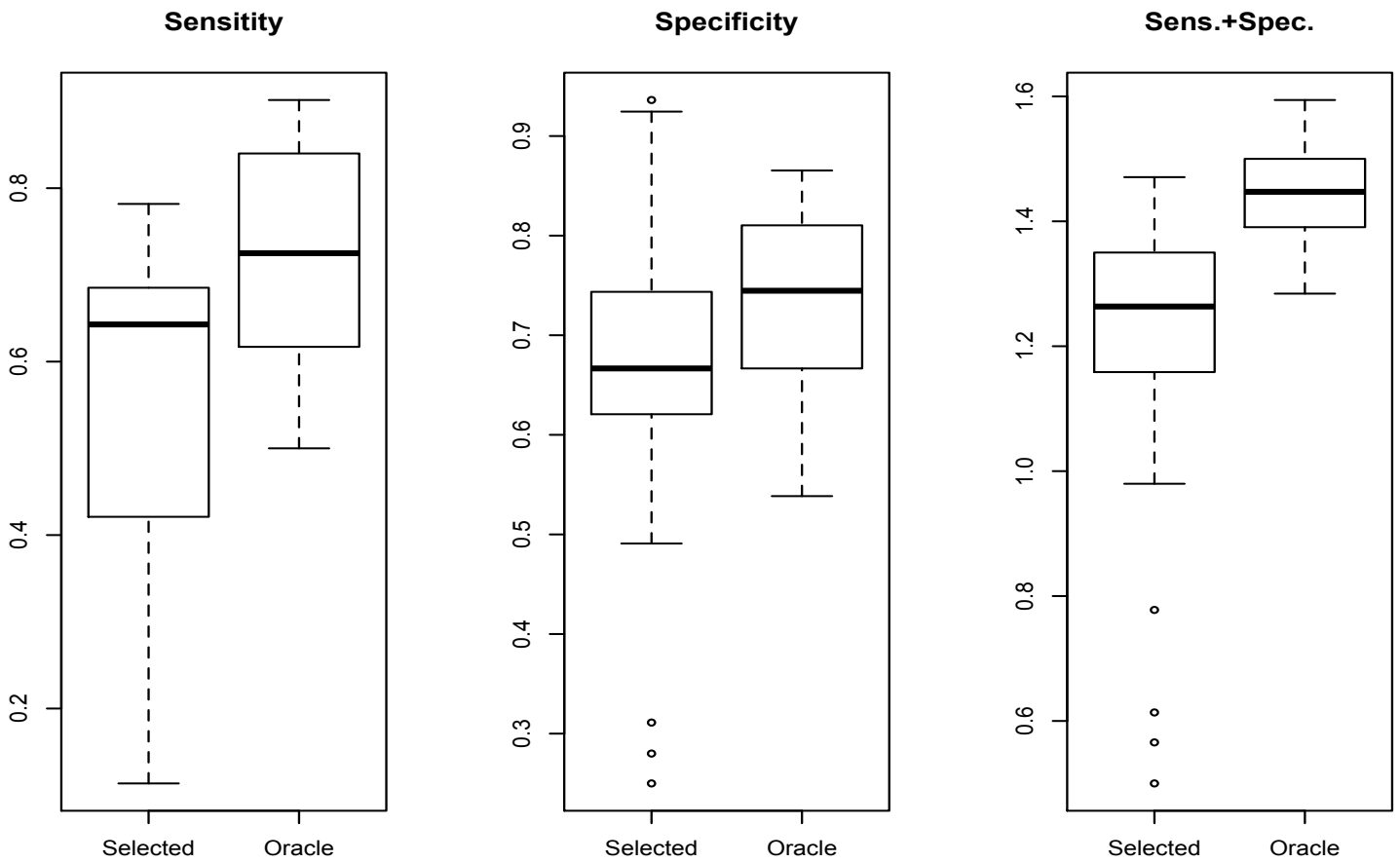
Simulations

True model $X_1 \vee X_2$. Correct model selection 70 times out of 100 simulations.



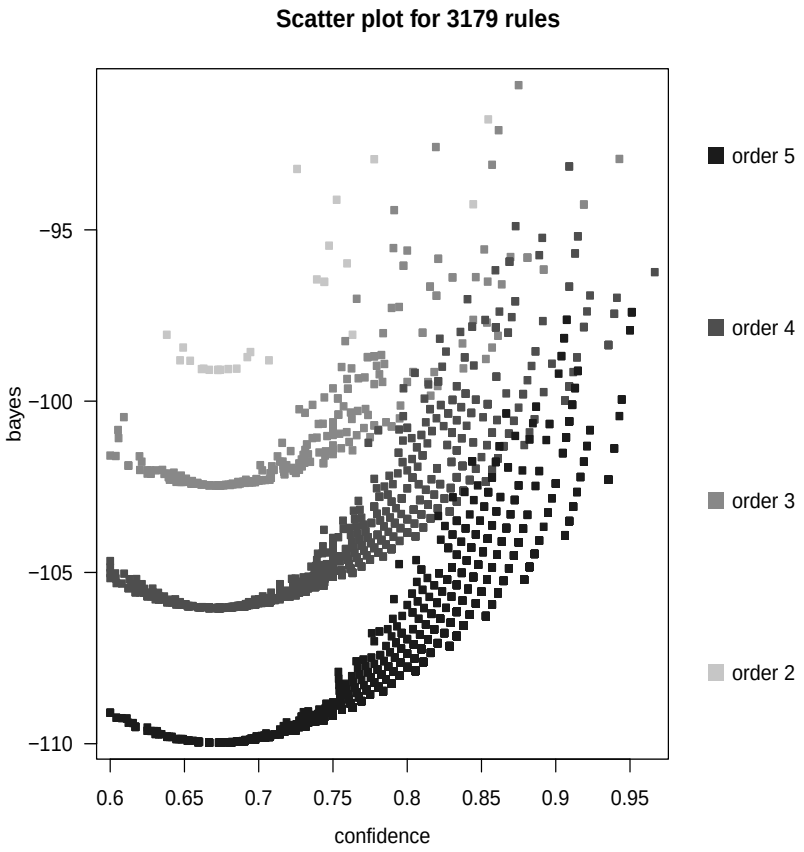
Simulations

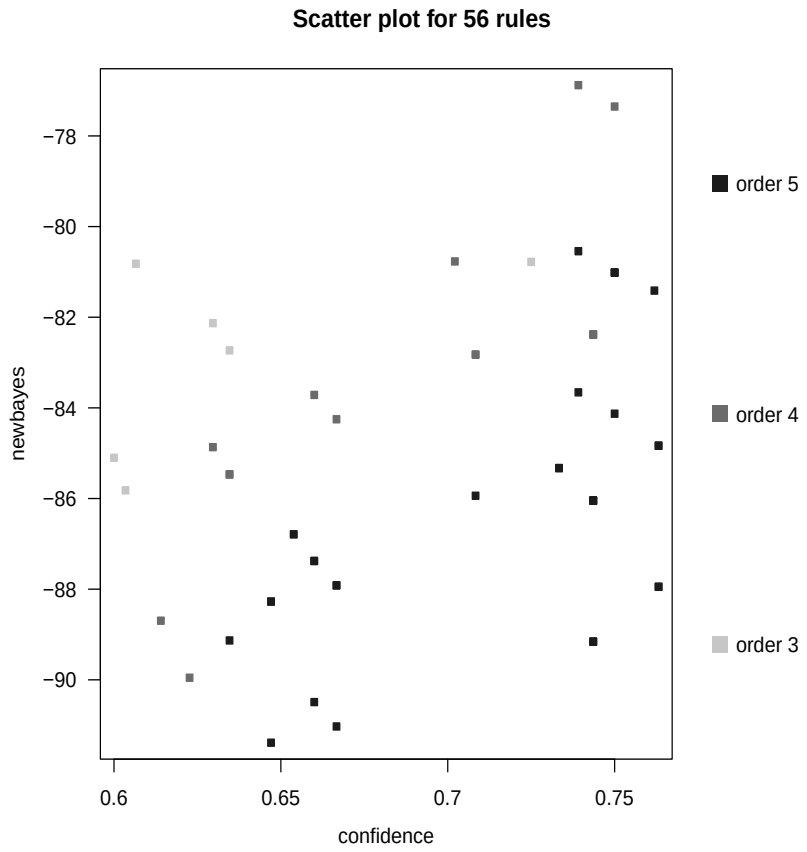
True model $X_1 \wedge (X_2 \vee X_3)$. Correct model selection 11 times out of 100 simulations.



Lupus Disease Classification

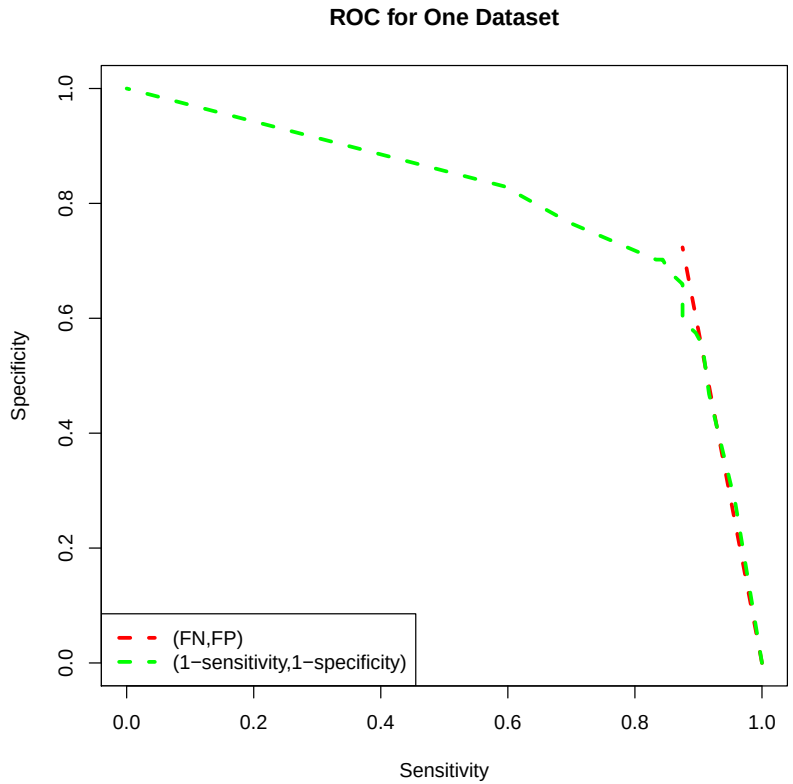
A total of 167 Lupus patients with 19 binary symptoms and 1 three-level symptom. Among them, 111 are of one type of Lupus disease called "SLE", and the other are of "MCTD".





Lupus Disease Classification

After step 2, 10 "and" rules were selected. The final selected logic model is an "or" type: History of proximal muscle weakness and (Observed joint swelling or Sclerodactyly) $\Rightarrow$ SLE.



Conclusions

On Bayesian framework

- Bayesian framework brings prior belief of unknown truth into decision making. After gaining data the posterior decision is a update of the belief.
- It becomes popular mainly because of the advent of high speed computing, especially when the unknown parameters are high dimensional.

On Bayesian multiple decision problem

- Developed a general class of Bayes multiple binary classifier.
- Considered a class of loss functions of two types of error rates.
- Developed an efficient model selection procedure.
- Illustrated in simulations and real data analysis.

Acknowledgement and Reference

Acknowledgement.

- Dr. Giri Narasimhan, School of Computing & Information Science, FIU.
- Dr. Tan Li, Dept of Biostatistics, FIU.
- Dr. Edsel Peña, Dept of Statistics, University of South Carolina.
- Dr. Annia Mesa, UM, for providing Lupus data.

Reference.

- Wensong Wu and Edsel Peña (2013), Bayes Multiple Decision Functions. *Electronic Journal of Statistics*.
- James V Stone, A Tutorial Introduction to Bayesian Analysis. 2013.