Part 2: ROC Curves for Cohort Data

Haben Michael
Adviser: Lu Tian

Summer 2017

Outline

Introduction: Data and ROC Curves

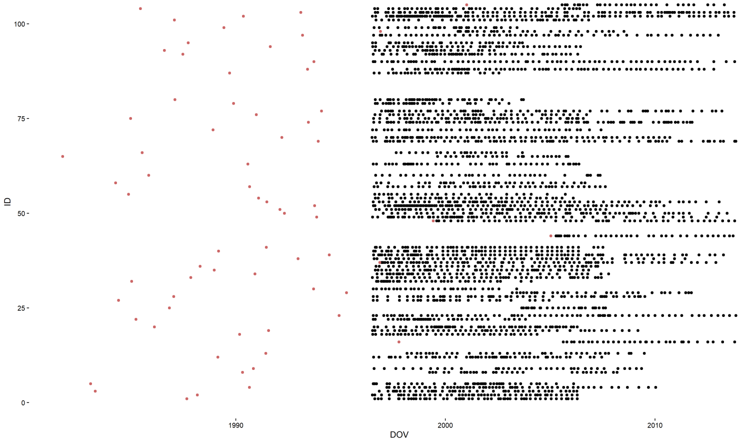
Longitudinal Dependence (ROC Curves)

Longitudinal and Pairwise Dependence (AUC)

Future Work

Yale Prospective Longitudinal Pediatric HIV Cohort (1996–2013)

- 104 patients with HIV monitored over a period of 17 years, on average 6.7 yo at enrollment
- median 32.5 visits/patient, median 1 visit/3 months



visit schedule

Among the measurements taken at each visit, we focus on:

- response: “blip” status, a binary variable representing a transient spike in viral load
- predictor #1: CD4 count, # CD4 cells/mm^3 blood
- predictor #2: CD4 percentage, the proportion of CD4 in a small sample

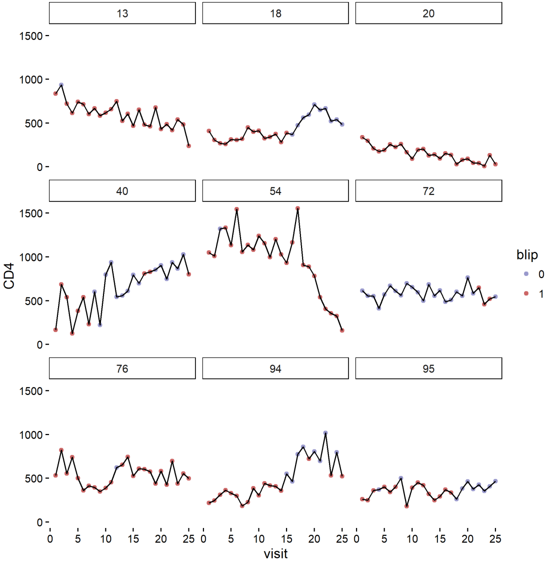
The goal of therapy is to keep CD4 count high and to suppress viral load, and both are tested regularly (every 3-6 months)

- viral load is the amount of HIV in a sample, indicating, e.g., transmission risk
- CD4 count measures immunosuppression–risk of opportunistic infections and strength of immune system ($\sim 500+$ normal, < 200 is an AIDS diagnosis)
  - “strongest predictor of HIV progression”–US DHHS
- CD4 percentage measures the proportion of white blood cells that are CD4 cells
  - usually more stable than CD4 count, but regarded as a poorer predictor of disease progression

Even when viral load is suppressed, transient spikes (“blips”) in viral load are often observed

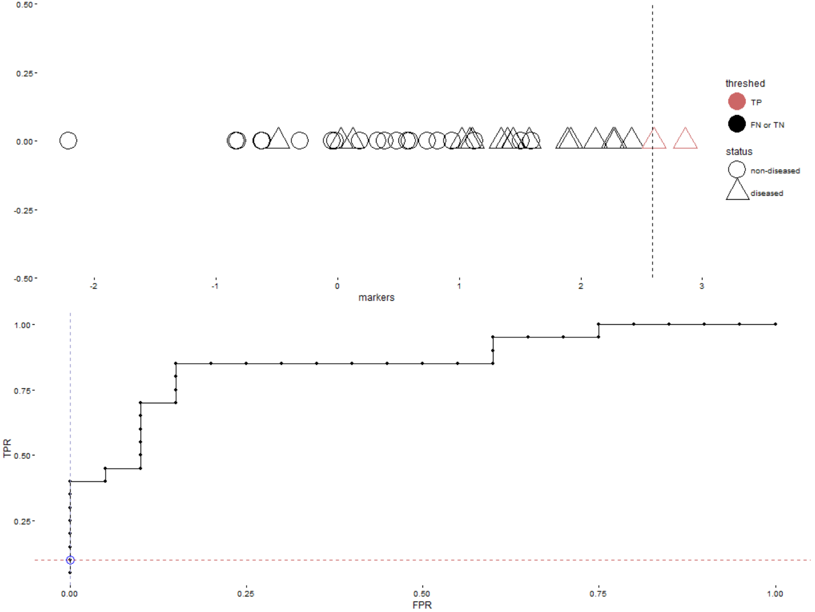
Blips are often ignored by clinicians, but recent research has established an association with the depletion of CD4 cells

CD4 count is the currently used predictor of blip status, but CD4 percentage is much cheaper to measure. How can we compare the quality of the two predictors?

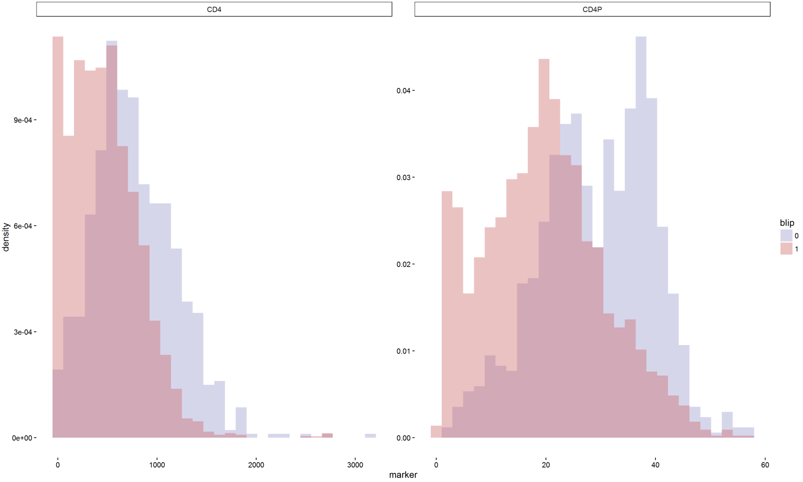

13
18
20
40
54
72
76
94
95
CD4
1500
1000
500
0
visit
0
5
10
15
20
25
blip
● 0
● 1

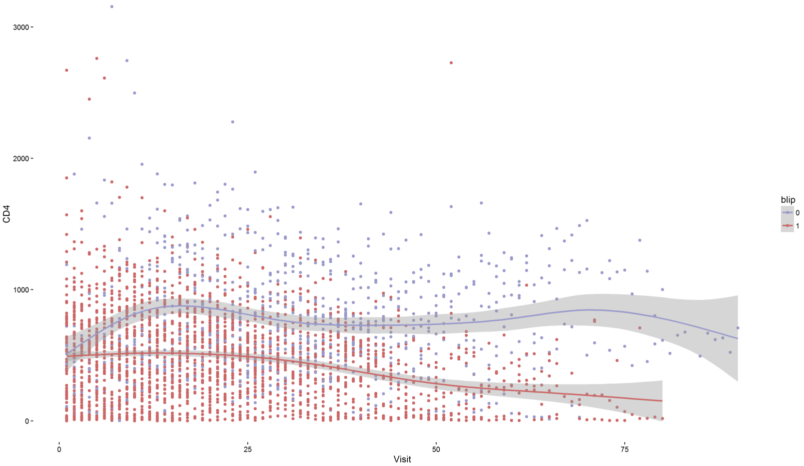
ROC curve

- graphical diagnostic of the performance of a threshold-based classifier
- plot of TPR vs FPR for a range of thresholds
- area under curve (AUC), summary statistic between 0 and 1, probability that a non-diseased marker is less than a diseased marker


0.50
0.25
0.00
-0.25
-0.50
-2
-1
0
1
2
3
markers
threshed
TP
FN or TN
status
non-diseased
diseased
1.00
0.75
TPR
0.50
0.25
0.00
0.25
0.50
0.75
1.00
FPR

Suppose we want to use an ROC curve or derived statistics to compare CD4 count and percentage as predictors of blip status.




3000
2000
1000
0
CD4
0
25
50
75
Visit
blip
0
1

Three types of dependence to contend with:

- longitudinal dependence (a) within and (b) between a given patient's diseased and non-diseased observations,
  - common variance estimators assume iid observations within each population, and between populations
- and (c) pairwise dependence between a given patient's predictors, CD4 count and percent
  - common methods of comparison of ROC or AUC data assume independent samples for each predictor

(data from different patients assumed independent)

$$
\begin{array}{c}
(S_{11}, \begin{pmatrix} X_{11} \\ X'_{11} \end{pmatrix}), (S_{12}, \begin{pmatrix} X_{12} \\ X'_{12} \end{pmatrix}), \ldots, (S_{1n_1}, \begin{pmatrix} X_{1n_1} \\ X'_{1n_1} \end{pmatrix}) \\
(S_{21}, \begin{pmatrix} X_{21} \\ X'_{21} \end{pmatrix}), (S_{22}, \begin{pmatrix} X_{22} \\ X'_{22} \end{pmatrix}), \ldots, (S_{2n_2}, \begin{pmatrix} X_{2n_2} \\ X'_{2n_2} \end{pmatrix}) \\
\vdots \\
(S_{N1}, \begin{pmatrix} X_{N1} \\ X'_{N1} \end{pmatrix}), (S_{N2}, \begin{pmatrix} X_{N2} \\ X'_{N2} \end{pmatrix}), \ldots, (S_{Nn_N}, \begin{pmatrix} X_{Nn_N} \\ X'_{Nn_N} \end{pmatrix})
\end{array}
$$

patients ↓

→

time/visits

Outline

Introduction: Data and ROC Curves

Longitudinal Dependence (ROC Curves)

Longitudinal and Pairwise Dependence (AUC)

Future Work

$$\text{patients}\ \left\downarrow\vphantom{\begin{array}{c}1\\1\\1\\1\\1\\1\\1\\1\end{array}}\right. \quad \begin{array}{c} (S_{11}, \begin{pmatrix} X_{11} \\ X'_{11} \end{pmatrix}), (S_{12}, \begin{pmatrix} X_{12} \\ X'_{12} \end{pmatrix}), \ldots, (S_{1n_1}, \begin{pmatrix} X_{1n_1} \\ X'_{1n_1} \end{pmatrix}) \\ (S_{21}, \begin{pmatrix} X_{21} \\ X'_{21} \end{pmatrix}), (S_{22}, \begin{pmatrix} X_{22} \\ X'_{22} \end{pmatrix}), \ldots, (S_{2n_2}, \begin{pmatrix} X_{2n_2} \\ X'_{2n_2} \end{pmatrix}) \\ \vdots \\ (S_{N1}, \begin{pmatrix} X_{N1} \\ X'_{N1} \end{pmatrix}), (S_{N2}, \begin{pmatrix} X_{N2} \\ X'_{N2} \end{pmatrix}), \ldots, (S_{Nn_N}, \begin{pmatrix} X_{Nn_N} \\ X'_{Nn_N} \end{pmatrix}) \end{array}$$

$$\xrightarrow{\hspace{3cm}}$$

time/visits

- With non-iid data the empirical ROC curve isn't consistent
  - consistent for what?
- A common approach is to fit a glm

  $$\text{logit}\{P(S_{ij} = 1) \mid X_{i1}, \ldots, X_{in_i}\} = \alpha_i + \beta_i X_{ij}$$

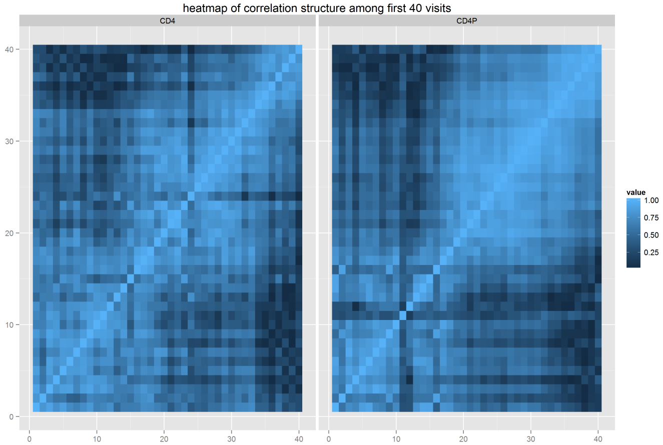
  with (α_i, β_i) multivariate normal and construct an ROC curve from the fitted values

  $$\left\{(\hat{\alpha}_i + \hat{\beta}_i X_{ij}, S_{ij}), i = 1, \cdots, n; j = 1, \cdots, n_i\right\}$$

$$\{(\alpha_i + \beta_i X_{ij}, S_{ij}), i = 1, \cdots, n; j = 1, \cdots, n_i\}$$

- Predictiveness of the biomarkers $X_{i1}, \ldots, X_{in_i}$, versus a "score" $f_i(X_{i1}, \ldots, X_{in_i})$, e.g., $\alpha_i + \beta_i X_{ij}$
- This measures predictiveness of both the biomarker and the model
  - E.g., intercept term large in magnitude relative to slope term, status is nearly constant within a cluster, nearly perfect ROC curves, regardless of quality of the biomarker X_{ij}.

- It is reasonable that the analyst should exploit predictiveness of the model if that is available
- If the model is poor, the scoring system may not do justice to the biomarker



Even assuming the model is correct, problems remain:

- In practice, the model is fit to the data,

  $$\left\{(\hat{\alpha}_i + \hat{\beta}_i X_{ij}, S_{ij}), i = 1, \cdots, n; j = 1, \cdots, n_i\right\}$$

  so the predictiveness of the model will be overestimated

  - Can't set aside a test set of patients because the subject effects are drawn independently. Could set aside along the j dimension, which is close to what we do.
- The fit uses the entire data set, including "future" observations
  - We adopt the vantage of an analyst who has a patient's history of biomarkers and disease statuses $(X_{ij}, S_{ij}), j = 1, \ldots, n_i$, is confronted with a new biomarker $X_{i,j+1}$, and must now predict a new disease status. What is the quality of the analyst's prediction?

- The procedure seems to be measuring the predictiveness of the biomarker for a given patient after adequate follow-up
    - We can generalize the notion of ROC curve to the cohort setting in other ways

- refine notion of ROC curve
- relax parametric assumptions
- account for generalization error
- use available data–model relationship between past data and current status–but no more

We model biomarker levels X_{ij} as an autoregressive process conditional on disease status $S_{ij} \in \{0, 1\}$, in turn modeled as a two-state markov process

$$\begin{aligned} \mathrm{P}\{S_{ij} = b | S_{i(j-1)} = a, \mathcal{H}_{i(j-1)}\} &= p_{abi}, a, b \in \{0,1\}, j = 2, \cdots, \\ X_{ij} | \{S_{ij} = a, \mathcal{H}_{ij}\} &= \theta_i^{(a)} + \rho_0 X_{i(j-1)} + \epsilon_{ij}, j = 1, \cdots, \\ \mathrm{P}(S_{i1} = a) = p_a \quad &\text{and} \quad X_{i0} = 0, \end{aligned}$$

where

$$\begin{aligned} \rho_0 &\in (-1, 1) \\ \epsilon_{ij} &\overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_0^2), j = 1, \cdots, n_i, \\ \xi_i = \begin{pmatrix} \theta_i^{(0)} \\ \theta_i^{(1)} \\ g(p_{01i}) \\ g(p_{11i}) \end{pmatrix} &\overset{\text{iid}}{\sim} N\left\{\begin{pmatrix} \theta_0 \\ \theta_1 \\ \mu_{01} \\ \mu_{11} \end{pmatrix}, \begin{pmatrix} \mathbf{\Sigma}_\theta & 0 \\ 0 & \mathbf{\Sigma}_p \end{pmatrix}\right\}, i = 1, \ldots, n, \end{aligned}$$

and $\eta = (p_0, \theta_0, \theta_1, \mu_{01}, \mu_{11}, \mathbf{\Sigma}_\theta, \mathbf{\Sigma}_p)$ are hyperparameters

$$\begin{tikzcd}
\cdots \arrow[r] & X_{i,j-1} \arrow[r] & X_{i,j} \arrow[r] & X_{i,j+1} \arrow[r] & \cdots \\
\cdots \arrow[r] & S_{i,j-1} \arrow[u] \arrow[r] & S_{i,j} \arrow[u] \arrow[r] & S_{i,j+1} \arrow[u] \arrow[r] & \cdots
\end{tikzcd}$$

$$\begin{tikzcd}
\cdots \arrow[r] & X_{i,j-1} \arrow[r] & X_{i,j} \arrow[r] & X_{i,j+1} \arrow[r] & \cdots \\
\cdots \arrow[r] \arrow[ur] & S_{i,j-1} \arrow[u] \arrow[ur] \arrow[r] & S_{i,j} \arrow[u] \arrow[ur] \arrow[r] & S_{i,j+1} \arrow[u] \arrow[ur] \arrow[r] & \cdots
\end{tikzcd}$$

$$\begin{tikzcd}
\cdots \arrow[r] \arrow[dr] & X_{i,j-1} \arrow[r] \arrow[dr] & X_{i,j} \arrow[r] \arrow[dr] & X_{i,j+1} \arrow[r] \arrow[dr] & \cdots \\
\cdots \arrow[r] & S_{i,j-1} \arrow[u] \arrow[r] & S_{i,j} \arrow[u] \arrow[r] & S_{i,j+1} \arrow[u] \arrow[r] & \cdots
\end{tikzcd}$$

Random effects models are submodels:

$$\begin{aligned} g\left\{\mathbb{P}\left(S_{ij}=1 \mid \mathcal{H}_{ij},\xi_i\right)\right\} &= g\left\{\mathbb{P}\left(S_{ij}=1 \mid X_{i(j-1)},X_{ij},S_{i(j-1)},\xi_i\right)\right\} \\ &= \alpha_{i0}+\alpha_{i1}S_{i(j-1)}+\alpha_{i2}X_{i(j-1)}+\alpha_{i3}X_{ij} \\ &\underset{def}{=} M_{ij}, \end{aligned}$$

where

$$\begin{aligned} \alpha_{i0} &= \frac{1}{2\sigma_0^2}\left\{(\theta_i^{(0)})^2-(\theta_i^{(1)})^2\right\}+\mu_{01i}, \\ \alpha_{i1} &= \mu_{11i}-\mu_{01i}, \\ \alpha_{i2} &= \frac{1}{\sigma_0^2}\rho_0\left(\theta_i^{(0)}-\theta_i^{(1)}\right), \\ \alpha_{i3} &= -\frac{1}{\sigma_0^2}\left(\theta_i^{(0)}-\theta_i^{(1)}\right). \end{aligned} \tag{1}$$

Individual ROC curve

- contrast an individual's diseased and non-diseased survival functions

$$\mathbb{P}(M_{ij} > m \mid S_{ij} = 1, \xi_i) \text{ vs. } \mathbb{P}(M_{ij} > m \mid S_{ij} = 0, \xi_i)$$

  with

$$M_{ij} = \alpha_{i0} + \alpha_{i1} S_{i(j-1)} + \alpha_{i2} X_{i(j-1)} + \alpha_{i3} X_{ij}$$

- Two CDFs imply an ROC curve (and derived statistics); are these the correct CDFs?
  - use all available history: under our model the current disease status S_{ij} given the score M_{ij} is conditionally independent of the remaining data
  - are functions of the available data

Individual ROC curve

- $$\mathbb{P}(M_{ij} > m \mid S_{ij} = 1, \xi_i) \text{ vs. } \mathbb{P}(M_{ij} > m \mid S_{ij} = 0, \xi_i)$$

  with

  $$M_{ij} = \alpha_{i0} + \alpha_{i1} S_{i(j-1)} + \alpha_{i2} X_{i(j-1)} + \alpha_{i3} X_{ij}$$
- The ROC curve is a function of the subject effect ξ_i through the subject specific-coefficients, which must be estimated
  - E.g., posterior means (empirical bayes)

$$\hat{\xi}_i = \mathbb{E}(\xi_i \mid \text{ available subject data, } \eta)$$

The quality of the approximation to the patient's true ROC depends on the size of the patient's history. When the patient history is too small, an alternative is the average individual ROC curve

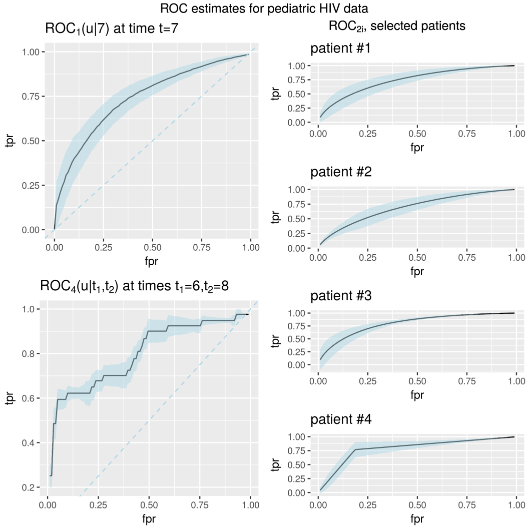
$$\mathrm{ROC}_1(u,j) = \mathbb{E}_\xi \left\{ \mathrm{ROC}(u,j|\xi,\eta) \right\}, \tag{2}$$

where the expectation is taken with respect to the random effect ξ

- in practice, sample $\xi_i \mid \eta$ under our model and average the empirical ROCs

$$\widehat{\mathrm{ROC}}_1(u,j) = B^{-1} \sum_{b=1}^{B} \widehat{\mathrm{ROC}}(u,j \mid \tilde{\xi}, \hat{\eta})$$

- $\sqrt{n}\{\widehat{\mathrm{ROC}_1}(u,j) - \mathrm{ROC}_1(u,j)\}$ converges weakly to a mean zero gaussian process (in u), provided $\sqrt{n}(\hat{\eta} - \eta) \rightsquigarrow G$



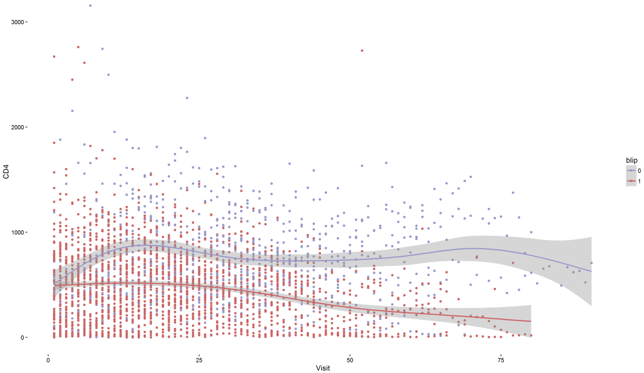
Expected individual ROC $ROC_1(u|t)$ and population ROC $ROC_4(u|t_1, t_2)$ at visits $t_1 = 6$ through $t_2 = 8$ with 95% bootstrap CI; limiting individual ROC ROC_{2i} for four patients (pediatric HIV data).

Population ROC curve

- Consider a time-indexed ROC curve based on the marginal distribution of the scores M_{ij}

$$\mathrm{ROC}_3(u \mid j) = S_{j,1}(S_{j,0}^{-1}(u))$$

$$S_{j,a}(u) = \mathbb{P}_j(M_{.j} > u \mid S_{.j} = a)$$



Population ROC curve

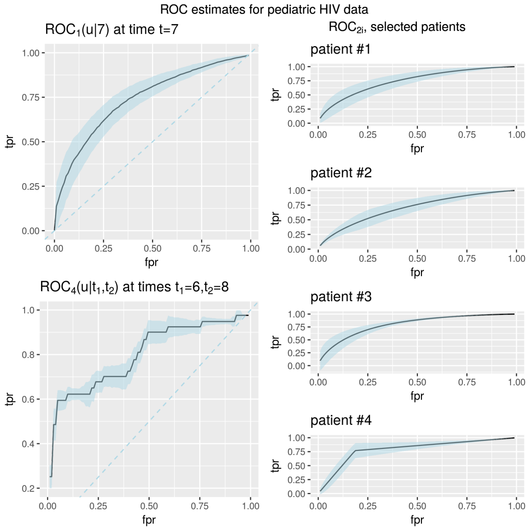
- Conditional on the population parameter, the patient data are iid, so the same holds for $M_{ij}, i = 1, \ldots, N$, j fixed
- The empirical ROC curve based on pairs $(M_{ij}, Y_{ij}), i = 1, \ldots, N$,

  $$\widehat{\mathrm{ROC}}_3(u \mid j) = \hat{S}_{j,1}(\hat{S}_{j,0}^{-1}(u))$$

  is a valid measure of the predictive performance of the scores M_{ij}, regardless of the validity of the model (beyond independence)

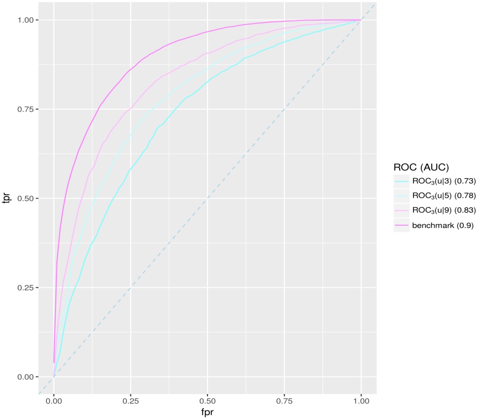
With mild regularity conditions on the population parameter,

- $\widehat{\mathrm{ROC}}_3(u \mid j)$ is consistent for $\mathrm{ROC}_3(u \mid j)$ as the number of patients grows
- $\sqrt{n}\{\widehat{\mathrm{ROC}}_3(u \mid j) - \mathrm{ROC}_3(u \mid j)\}$ converges to a mean zero gaussian process (indexed by the FPR u)
- Variance can be approximated by the bootstrap



Expected individual ROC $ROC_1(u|t)$ and population ROC $ROC_4(u|t_1, t_2)$ at visits $t_1 = 6$ through $t_2 = 8$ with 95% bootstrap CI; limiting individual ROC ROC_{2i} for four patients (pediatric HIV data).

Population ROC for misspecified model



FPR	0.10	0.25	0.50	0.75
t				
3	0.95 (0.045)	0.91 (0.065)	0.94 (0.043)	0.96 (0.027)
5	0.94 (0.052)	0.93 (0.048)	0.92 (0.042)	0.93 (0.029)
9	0.94 (0.045)	0.97 (0.042)	0.93 (0.048)	0.94 (0.027)

Outline

Introduction: Data and ROC Curves

Longitudinal Dependence (ROC Curves)

Longitudinal and Pairwise Dependence (AUC)

Future Work

$$
\text{patients}\ \Bigg\downarrow
\begin{array}{c}
(S_{11}, \begin{pmatrix} X_{11} \\ X'_{11} \end{pmatrix}), (S_{12}, \begin{pmatrix} X_{12} \\ X'_{12} \end{pmatrix}), \ldots, (S_{1n_1}, \begin{pmatrix} X_{1n_1} \\ X'_{1n_1} \end{pmatrix}) \\
(S_{21}, \begin{pmatrix} X_{21} \\ X'_{21} \end{pmatrix}), (S_{22}, \begin{pmatrix} X_{22} \\ X'_{22} \end{pmatrix}), \ldots, (S_{2n_2}, \begin{pmatrix} X_{2n_2} \\ X'_{2n_2} \end{pmatrix}) \\
\vdots \\
(S_{N1}, \begin{pmatrix} X_{N1} \\ X'_{N1} \end{pmatrix}), (S_{N2}, \begin{pmatrix} X_{N2} \\ X'_{N2} \end{pmatrix}), \ldots, (S_{Nn_N}, \begin{pmatrix} X_{Nn_N} \\ X'_{Nn_N} \end{pmatrix})
\end{array}
$$

$$\xrightarrow{\hspace{6em}}$$

time/visits

In considering both longitudinal and pairwise dependence we estimate a more tractable target, the AUC

- $X_1, \ldots, X_N \overset{iid}{\sim} F_X$ non-diseased observations
- $Y_1, \ldots, Y_N \overset{iid}{\sim} F_Y$ diseased observations
- AUC is $P(X < Y)$
- Unbiased estimator is $\sum_{i,j} \mathbb{P}(X_i < Y_j)$ (Mann-Whitney U-statistic)

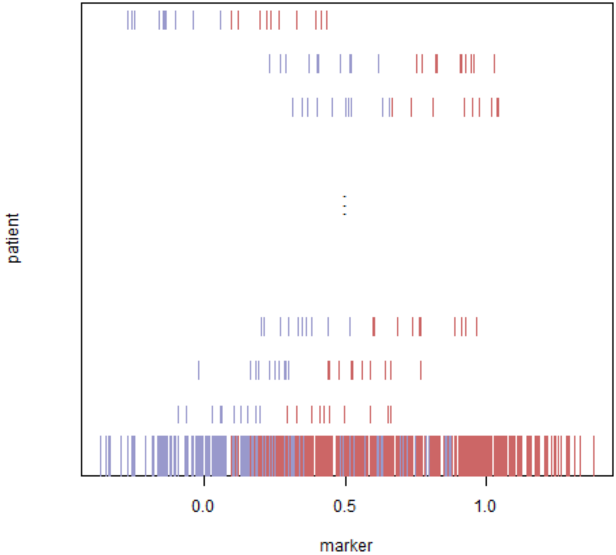
As with the ROC, we can generalize to individual and population AUCs

$$\begin{aligned} U_N &= P(X < Y \mid \text{ X and Y belong to the same cluster}) \\ &= \frac{1}{N}\sum_{i=1}^{N} P_{\mu_i}(X_{ij} < Y_{ij}) \\ V_N &= P(X < Y) \end{aligned}$$

estimators

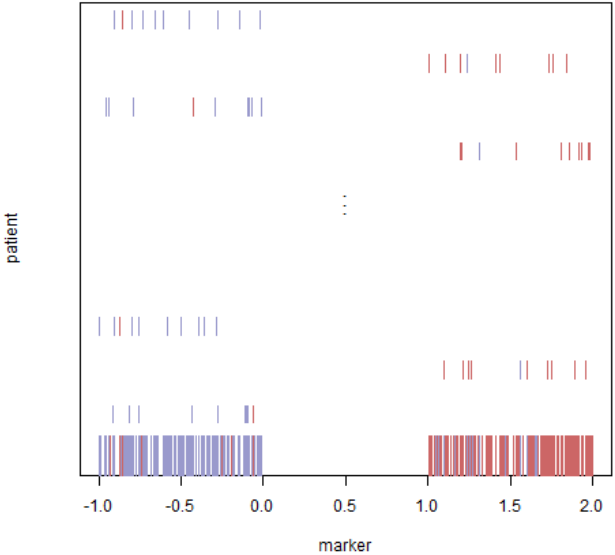
$$\hat{U}_N = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{n_i m_i}\sum_{j,k}\{X_{ij} < Y_{ik}\}$$

$$\hat{V}_N = \frac{1}{\sum m_i \sum n_i}\sum_{i,j}\sum_{r,s}\{X_{ir} < Y_{js}\}$$



$$\theta_i \sim Unif[0,1]$$
$$X_{ij} \mid \theta_i \sim Unif[\theta_i - \delta, \theta_i]$$
$$Y_{ij} \mid \theta_i \sim Unif[\theta_i, \theta_i + \delta]$$

$$\hat{U}_N = 1$$
$$\hat{V}_N \rightarrow_p 1/2 + \epsilon \ (N \rightarrow \infty)$$



$$B_i \sim \text{bernoulli}$$

$$X_{i1}, \ldots, X_{i,n-1}, Y_{i1} \mid B_i \sim Unif[2B_i - 1, 2B_i]$$

$$\hat{U}_N \rightarrow_{a.s.} 1/2$$

$$\hat{V}_N \rightarrow_{a.s.} 1 - \epsilon \ (n, N \rightarrow \infty)$$

Expected individual AUC $(\hat{U}_N, \hat{U}'_N)$

- Independent sequence of bounded RVs, apply CLT

$$N^{-1/2}(\begin{pmatrix} \hat{U}_N \\ \hat{U}'_N \end{pmatrix} - \begin{pmatrix} U_N \\ U'_N \end{pmatrix}) \rightsquigarrow \mathcal{N}(0, \Sigma)$$

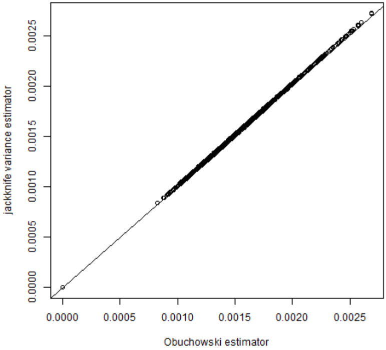
- Use usual covariance estimators to perform inference

Population AUC ($\hat{V}_N$, $\hat{V}'_N$)

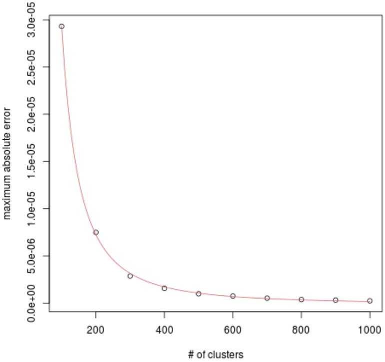
- More commonly studied than individual AUC (opposite for ROC)
- Almost a U-statistic, but (probably) not quite, because of the random vector lengths as normalization
- We want a variance estimator in order to perform inference
- We could use the bootstrap, treating the clusters as independent observations, but maybe lose too much power

- Obuchowski ’97 describes a variance estimator. Developed with radiology applications in mind. The different biomarkers correspond to different “readers.”
- nonparametric, asymptotic estimator
- little in the way of proof in the original paper
- in practice, seems very robust

Gives close to the same estimate as the jackknife variance estimator, treating the clusters as the independent observations "left out"



Replication of Obuchowski simulation with jackknife variance estimator. (Approximately 3.6e+5 observations; status and maker correlations of 0, .4, and .8; normal and nonnormal data; 100 clusters with 2 observations/cluster.)



The Obuchowski '97 simulation was replicated while varying $N = \#$ of clusters, computing the maximum absolute difference between the 2 estimators on each replicate. Also plotted is a fit to $1/N^2$.

The difference between the two variance estimators is $O(1/N^2)$

Two conclusions:

1. Consistency of the Obuchowski estimator can be established from consistency of the jackknife estimator

- Arvesen '69 establishes consistency of the jackknife variance estimator for well-behaved functions of U-statistics

$$\hat{V}_N = \frac{N^{-2}\sum_{i,j}\sum_{r,s}\{X_{ir} < Y_{js}\}}{(N^{-1}\sum m_i)(N^{-1}\sum n_i})$$

- also establishes consistency for independent, non-identically distributed data, matching our assumption on the patients
- minor modification of his proof needed to account for within-cluster dependence across diseaesed and non-diseased observations

Two conclusions:

2. might as well use bootstrap at the cluster level

Outline

Introduction: Data and ROC Curves

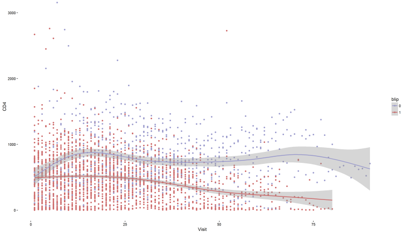
Longitudinal Dependence (ROC Curves)

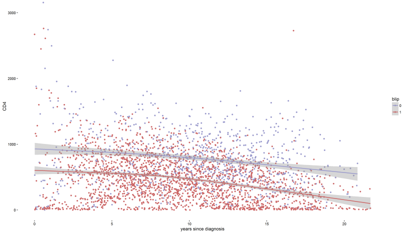
Longitudinal and Pairwise Dependence (AUC)

Future Work

- HIV data have a natural time parameter–the visits where the patients are treated and measured
- Other data may consist of irregularly measured markers
  - individual ROC curves will mean different things for different patients
  - population ROC curve loses interpretation

For the HIV data, are visits the “natural” time parameter?


CD4
Visit
blip


CD4
years since diagnosis
blip