Adaptive Confidence Intervals for the Test Error in Classification

Eric B. Laber & Susan A. Murphy

University of Michigan

August 5, 2009
Setup

Simple beginnings:

1. Observe iid training data \( \mathcal{D} = \{(x_i, y_i)\}_{i=1}^n \)
   - inputs \( X \in \mathbb{R}^p \)
   - outputs \( Y \in \{-1, 1\} \)

2. Construct classifier \( \hat{c}_D(X) : \mathbb{R}^p \rightarrow \{-1, 1\} \)

3. Use classifier for prediction at new inputs

Questions:

• How well will my classifier perform its task?
  
  Point estimate for test error
  \[ \tau(\hat{c}_D) = P(Y \neq \hat{c}_D(X)) \]

• How confident am I in the above estimate?
  
  Confidence interval for test error
  \[ \tau(\hat{c}_D) \]
Setup

Simple beginnings:

1. Observe iid training data $D = \{(x_i, y_i)\}_{i=1}^{n}$
   - inputs $X \in \mathbb{R}^p$
   - outputs $Y \in \{-1, 1\}$

2. Construct classifier $\hat{c}_D(X) : \mathbb{R}^p \mapsto \{-1, 1\}$

3. Use classifier for prediction at new inputs

Questions:

- How well will my classifier perform its task?
  
  **Point estimate for test error** $\tau(\hat{c}_D) \triangleq P1_{Y \neq \hat{c}_D(X)}$

- How confident am I in the above estimate?
  
  **Confidence interval** for test error $\tau(\hat{c}_D)$
Background

• Long standing problem
• Primary focus has been point estimation
  • CV methods: Krzanowski and Hand [1986], Langford [2005], Yang [2006]
  • Hybrid methods: Fu [2005], Kim [2009]
• More than 200 references!
Background

- Long standing problem
- Primary focus has been point estimation
  - CV methods: Krzanowski and Hand [1986], Langford [2005], Yang [2006]
  - Hybrid methods: Fu [2005], Kim [2009]
  - More than 200 references!
- Confidence intervals secondary interest
  Point estimate paradigm
  **Step 1:** Develop best point estimate
  **Step 2:** Estimate standard error
Background

Why is this problem still open?

- Non-regular
- Previous methods either
  - ad-hoc
  - assume regularity
- Point estimation paradigm
Background

Why is this problem still open?
- Non-regular
- Previous methods either
  - ad-hoc
  - assume regularity
- Point estimation paradigm

Path to a solution
- Validity in non-regular setting
- Confidence interval primary focus
The problem

- Construct a linear classifier using surrogate loss $L(X, Y, \beta)$
  1. $\hat{\beta}_n \triangleq \arg \min_{\beta \in \mathbb{R}^p} \mathbb{P}_n L(X, Y, \beta)$
  2. $\hat{c}_D(X) = \text{sign} \left( X^t \hat{\beta}_n \right)$
The problem

- Construct a linear classifier using surrogate loss $L(X, Y, \beta)$
  1. $\hat{\beta}_n \triangleq \arg \min_{\beta \in \mathbb{R}^p} \mathbb{P}_n L(X, Y, \beta)$
  2. $\hat{c}_D(X) = \text{sign} \left(X^t \hat{\beta}_n\right)$

- Review: surrogate loss function $L(X, Y, \beta)$
  - like to minimize empirical test error $\mathbb{P}_n 1_{Y \neq \text{sign}(X^t \beta)}$
  - non-smoothness $\Rightarrow$ computational difficulty
  - replace $1_{Y \neq \text{sign}(X^t \beta)} = 1_{YX^t \beta < 0}$ with smooth surrogate
    - Support Vector Machines:
      $L(X, Y, \beta) = (1 - YX^t \beta)_+ + \gamma \|\beta\|^2$
    - Binomial Deviance:
      $L(X, Y, \beta) = \log(1 + e^{-YX^t \beta})$
    - Squared Error:
      $L(X, Y, \beta) = (1 - YX^t \beta)^2$
The problem cont’d

• Test error $\tau(\hat{\beta}_n) \triangleq P_{Y|X^t} \hat{\beta}_n < 0$
• Test error $\tau(\hat{\beta}_n) \triangleq P_{Y|X^t\hat{\beta}_n<0}$

• Goal: given $\gamma \in (0, 1)$ construct $\hat{u}$ and $\hat{l}$ so that

$$P_D \left\{ \hat{l} \leq \tau(\hat{\beta}_n) \leq \hat{u} \right\} \geq 1 - \gamma$$
Assumptions

Some technical assumptions:

(A1) \( L(X, Y, \beta) \) is convex with respect to \( \beta \) for each \( (x, y) \in \mathbb{R}^p \times \{-1, 1\} \)

(A2) \( Q(\beta) \triangleq PL(X, Y, \beta) \) exists and is finite for all \( \beta \in \mathbb{R}^p \)

(A3) \( \beta^* \triangleq \arg \min_{\beta \in \mathbb{R}^p} Q(\beta) \) exists and is unique

(A4) Let \( g(X, Y, \beta) \) be a sub-gradient of \( L(X, Y, \beta) \). Then \( P||g(X, Y, \beta)||^2 < \infty \) for all \( \beta \) in a neighborhood of \( \beta^* \).

(A5) \( Q(\beta) \) is twice continuously differentiable at \( \beta^* \) and \( H \triangleq \nabla^2 Q(\beta^*) \) is positive definite.
Assumptions

Some technical assumptions:

(A1) \( L(X, Y, \beta) \) is convex with respect to \( \beta \) for each \( (x, y) \in \mathbb{R}^p \times \{-1, 1\} \)

(A2) \( Q(\beta) \triangleq PL(X, Y, \beta) \) exists and is finite for all \( \beta \in \mathbb{R}^p \)

(A3) \( \beta^* \triangleq \arg \min_{\beta \in \mathbb{R}^p} Q(\beta) \) exists and is unique

(A4) Let \( g(X, Y, \beta) \) be a sub-gradient of \( L(X, Y, \beta) \). Then \( P\|g(X, Y, \beta)\|^2 < \infty \) for all \( \beta \) in a neighborhood of \( \beta^* \).

(A5) \( Q(\beta) \) is twice continuously differentiable at \( \beta^* \) and \( H \triangleq \nabla^2 Q(\beta^*) \) is positive definite.

Note: hereafter these are regarded as standing assumptions.
Some things to note:

- Model space may not be correct
- $p << n$
- Cannot afford a test set
Non-regularity

• Simple estimate of $\tau(\hat{\beta}_n)$ is $\hat{\tau}(\hat{\beta}_n) \triangleq \mathbb{P}_n 1_{YX^t\hat{\beta}_n < 0}$

• Natural starting point for inference:

$$
\sqrt{n}(\hat{\tau}(\hat{\beta}_n) - \tau(\hat{\beta}_n)) \triangleq \sqrt{n}(\mathbb{P}_n - P) 1_{YX^t\hat{\beta}_n < 0}
= \sqrt{n}(\mathbb{P}_n - P) 1_{X^t\beta^*=0} 1_{YX^t\sqrt{n}(\hat{\beta}_n - \beta^*) < 0}
+ \sqrt{n}(\mathbb{P}_n - P) 1_{X^t\beta^* \neq 0} 1_{YX^t\hat{\beta}_n < 0}
$$

• $P1_{X^t\beta^*=0} > 0$ implies $\sqrt{n}(\hat{\tau}(\hat{\beta}_n) - \tau(\hat{\beta}_n))$ has non-regular limit

  • points near the boundary cause jittering
  • bootstrap and normal approximations are invalid
Non-regularity

So what?

\[ x^\beta \neq 0 \]

- Small sample performance "contaminated" by non-regular setting [Cheng 2009], [Xie 2009]
- Points near boundary indistinguishable from points on boundary
- Indicator function unstable
Non-regularity

So what? $X^t \beta^*$ is never equal to zero.
Non-regularity

So what? $X^t \beta^*$ is never equal to zero.

- Small sample performance “contaminated” by non-regular setting [Cheng 2009], [Xie 2009]
- Points near boundary indistinguishable from points on boundary
- Indicator function unstable
Simple example

Suppose

- \((X_1, X_2) \sim Unif[0, 5]^2\)
- \(\epsilon \sim N(0, 1/4)\)
- \(Y = \text{sign} \left( X_2 - (4/25)X_1^2 - 1 + \epsilon \right)\)

Properties of this example

- \(P1_{x^t \beta^* = 0} = 0\)
- Linear classifier is a good fit
- E.g. if \(n = 30\)
  - \(\mathbb{E} \left( \tau(\hat{\beta}_n) \right) \approx .11\)
  - Bayes error \(\approx .09\)
Simple example cont’d

Under “regular” framework

- Centered bootstrap $\sqrt{n}(\hat{P}_n^{(b)} - P_n)1_{YX^t\hat{\beta}_n^{(b)} < 0}$

- Normal approximation $\hat{\tau}(\hat{\beta}_n) \pm z_{1-\gamma/2} \sqrt{\frac{\hat{\tau}(\hat{\beta}_n)(1-\hat{\tau}(\hat{\beta}_n))}{n}}$

are both asymptotically valid
Simple example cont’d

Under “regular” framework

• Centered bootstrap \( \sqrt{n}(\hat{P}_n^{(b)} - P_n)1_{YXt\hat{\beta}_n^{(b)}<0} \)

• Normal approximation \( \hat{\tau}(\hat{\beta}_n) \pm z_{1-\gamma/2} \sqrt{\frac{\hat{\tau}(\hat{\beta}_n)(1-\hat{\tau}(\hat{\beta}_n))}{n}} \)

are both asymptotically valid

- Coverage estimated using 1000 Monte Carlo iterations
- Below nominal coverage even for \( n = 250 \)
- Coverage especially poor for small samples
Adaptive confidence interval

Basic idea: data adaptive bound on $\sqrt{n}(\hat{\tau}(\hat{\beta}_n) - \tau(\hat{\beta}_n))$
Adaptive confidence interval

Basic idea: data adaptive bound on $\sqrt{n}(\hat{\tau}(\hat{\beta}_n) - \tau(\hat{\beta}_n))$

- Recall $\sqrt{n}(\hat{\tau}(\hat{\beta}_n) - \tau(\hat{\beta}_n))$ is equal to

$$\sqrt{n}(\mathbb{P}_n - P)1_{X^t\beta^* = 0}1_{YX^t\hat{\beta}_n < 0} + \sqrt{n}(\mathbb{P}_n - P)1_{X^t\beta^* \neq 0}1_{YX^t\hat{\beta}_n < 0}$$
Adaptive confidence interval

Basic idea: data adaptive bound on $\sqrt{n}(\hat{\tau}(\hat{\beta}_n) - \tau(\hat{\beta}_n))$

- Recall $\sqrt{n}(\hat{\tau}(\hat{\beta}_n) - \tau(\hat{\beta}_n))$ is equal to
  
  $$\sqrt{n}(\mathbb{P}_n - P)1_{X^t\beta^*=0}1_{YX^t\beta_n<0} + \sqrt{n}(\mathbb{P}_n - P)1_{X^t\beta^*\neq 0}1_{YX^t\beta_n<0}$$

- Natural bound

$$\sup_{u \in \mathbb{R}^p} \sqrt{n}(\mathbb{P}_n - P)1_{X^t\beta^*=0}1_{YX^t\beta_n<0} + \sqrt{n}(\mathbb{P}_n - P)1_{X^t\beta^*\neq 0}1_{YX^t\beta_n<0}$$
Adaptive confidence interval

Basic idea: data adaptive bound on $\sqrt{n}(\hat{\tau}(\hat{\beta}_n) - \tau(\hat{\beta}_n))$

- Recall $\sqrt{n}(\hat{\tau}(\hat{\beta}_n) - \tau(\hat{\beta}_n))$ is equal to

$$\sqrt{n}(\mathbb{P}_n - P)1_{X^t\beta^* = 0}YX^t\hat{\beta}_n < 0 + \sqrt{n}(\mathbb{P}_n - P)1_{X^t\beta^* \neq 0}YX^t\hat{\beta}_n < 0$$

- Natural bound

$$\sup_{u \in \mathbb{R}^p} \sqrt{n}(\mathbb{P}_n - P)1_{X^t\beta^* = 0}YX^tu < 0 + \sqrt{n}(\mathbb{P}_n - P)1_{X^t\beta^* \neq 0}YX^t\hat{\beta}_n < 0$$

- Operationalize decision $X^t\beta^* = 0$ using

$$C_n \triangleq \sup_{u \in \mathbb{R}^p} \sqrt{n}(\mathbb{P}_n - P)1_{\frac{(X^t\beta^*_n)^2}{X^t\Sigma X} \leq a_n}YX^tu < 0 + \sqrt{n}(\mathbb{P}_n - P)1_{\frac{(X^t\beta^*_n)^2}{X^t\Sigma X} > a_n}YX^t\hat{\beta}_n < 0$$

where $a_n \to \infty$, $a_n, = o(n)$ and $\Sigma = n\text{Cov}(\hat{\beta}_n)$.
Adaptive confidence interval

Basic idea: data adaptive bound on $\sqrt{n}(\hat{\tau}((\hat{\beta}_n) - \tau(\hat{\beta}_n))$

- Recall $\sqrt{n}(\hat{\tau}((\hat{\beta}_n) - \tau(\hat{\beta}_n))$ is equal to

\[
\sqrt{n}(\mathbb{P}_n - P)1_{X^t\beta^* = 0} 1_{YX^t\hat{\beta}_n < 0} + \sqrt{n}(\mathbb{P}_n - P)1_{X^t\beta^* \neq 0} 1_{YX^t\hat{\beta}_n < 0}
\]

- Natural bound

\[
\sup_{u \in \mathbb{R}^p} \sqrt{n}(\mathbb{P}_n - P)1_{X^t\beta^* = 0} 1_{YX^t u < 0} + \sqrt{n}(\mathbb{P}_n - P)1_{X^t\beta^* \neq 0} 1_{YX^t\hat{\beta}_n < 0}
\]

- Operationalize decision $X^t\beta^* = 0$ using

\[
C_n \triangleq \sup_{u \in \mathbb{R}^p} \sqrt{n}(\mathbb{P}_n - P)1_{(X^t\hat{\beta}_n)^2 \leq a_n^{-1}} 1_{YX^t u < 0} + \sqrt{n}(\mathbb{P}_n - P)1_{(X^t\hat{\beta}_n)^2 > a_n^{-1}} 1_{YX^t\hat{\beta}_n < 0}
\]

where $a_n \to \infty$, $a_n = o(n)$ and $\Sigma = n\text{Cov}(\hat{\beta}_n)$
Adaptive confidence interval

Some observations:

**Theorem (Convergence)**

If (A1)-(A5) hold:

1. \( \sqrt{n} \left( \hat{\tau}(\hat{\beta}_n) - \tau(\hat{\beta}_n) \right) \rightsquigarrow V(z_\infty) + B(\beta^*) \)

2. \( \sqrt{n} \left( \hat{\tau}(\hat{\beta}_n) - \tau(\hat{\beta}_n) \right) \leq C_n \) for all \( n \)

3. \( C_n \rightsquigarrow \sup_{u \in \mathbb{R}^p} V(u) + B(\beta^*) \).

where \( V \) and \( B \) are Gaussian processes and \( z_\infty \) is a \( p \)-dim normal with covariance \( \Sigma \).
Adaptive confidence interval

Some observations:

Theorem (Convergence)
If (A1)-(A5) hold:

1. \( \sqrt{n} \left( \hat{\tau}(\hat{\beta}_n) - \tau(\hat{\beta}_n) \right) \overset{d}{\rightarrow} \mathbb{V}(z_\infty) + \mathbb{B}(\beta^*) \)

2. \( \sqrt{n} \left( \hat{\tau}(\hat{\beta}_n) - \tau(\hat{\beta}_n) \right) \leq C_n \text{ for all } n \)

3. \( C_n \overset{d}{\rightarrow} \sup_{u \in \mathbb{R}^p} \mathbb{V}(u) + \mathbb{B}(\beta^*) \).

where \( \mathbb{V} \) and \( \mathbb{B} \) are Gaussian processes and \( z_\infty \) is a \( p \)-dim normal with covariance \( \Sigma \).

Theorem (Adaptation)
Assuming (A1)-(A5) hold then if either the Bayes decision boundary is linear or \( P(X^t \beta^* = 0) = 0 \) then \( C_n \) and \( \sqrt{n}(\hat{\tau}(\hat{\beta}_n) - \tau(\hat{\beta}_n)) \) have the same limiting distribution.
Adaptive confidence interval

Q: But how can I try this at home?
Adaptive confidence interval

Q: But how can I try this at home?
A: Use the bootstrap!
Adaptive confidence interval

Q: But how can I try this at home?
A: Use the bootstrap!

$$\hat{C}_n(b) \triangleq \sup_{u \in \mathbb{R}^p} \sqrt{n}(\hat{P}_n(b) - P_n) 1_{\frac{(X^t \hat{\beta}_n(b))^2}{X^t \hat{\Sigma}_n X} \leq a_n} 1_{Y X^t u < 0}$$

$$+ \sqrt{n}(\hat{P}_n(b) - P_n) 1_{\frac{(X^t \hat{\beta}_n(b))^2}{X^t \hat{\Sigma}_n X} > a_n} 1_{Y X^t \hat{\beta}_n(b) < 0}$$
Let $\hat{u}$ be the $1 - \gamma/2$ of percentile $\hat{C}_n^{(b)}$ then:

$$P \left\{ \sqrt{n}(P_n - P)1_{YX^t\hat{\beta}_n < 0} \leq \hat{u} \right\} \geq P \{ C_n \leq \hat{u} \}$$

$$\approx P \left\{ \hat{C}_n^{(b)} \leq \hat{u} \right\} = 1 - \delta/2$$

so that

$$P \left\{ P_n 1_{YX^t\hat{\beta}_n < 0} - \hat{u}/\sqrt{n} \leq P 1_{YX^t\hat{\beta}_n < 0} \right\} = P \left\{ \hat{\tau}(\hat{\beta}_n) - \hat{u}/\sqrt{n} \leq \tau(\hat{\beta}_n) \right\}$$

$$\geq 1 - \delta/2$$

and $[\hat{\tau}(\hat{\beta}_n) - \hat{u}/\sqrt{n}, \infty)$ is an approximate confidence interval
Adaptive confidence interval

Theorem

Suppose that (A1)-(A5) hold then $C_n$ and $\hat{C}_n^{(b)}$ converge to the same limiting distribution in probability.
Computation

• Computing $\hat{C}_n^{(b)}$ is a Mixed Integer Programming problem
  • requires specialized software (e.g. CPLEX)
  • computational cost excessive for large problems
• Admits convex relaxation
  • solved in polynomial time
  • negligible loss in solution quality
  • no specialized software required (can be solved in R)
Experiments

Compare performance of
- Adaptive confidence interval (ACI)
- CV-Normal approximation [Yang 2006]
- BCCVP-BR approximation [Jiang 2008]
Experiments

Compare performance of

- Adaptive confidence interval (ACI)
- CV-Normal approximation [Yang 2006]
- BCCVP-BR approximation [Jiang 2008]

Details

- 1000 Monte Carlo iterations
- 10 data sets
- Compare estimated coverage and width
- \( a_n \triangleq \sqrt{n} \max(1.0, \frac{\chi^2_{0.995}}{\sqrt{n}}) \)
Results

Target coverage .950, loss function \( L(X, Y, \beta) = (1 - YX^t\beta)^2 \), \( n = 30 \)

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Dim</th>
<th>ACI</th>
<th>CV-Normal</th>
<th>BCCVP-BR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ThreePt</td>
<td>2</td>
<td>.948</td>
<td>.930</td>
<td>.893</td>
</tr>
<tr>
<td>Donut</td>
<td>3</td>
<td>.965</td>
<td>.967</td>
<td>.908</td>
</tr>
<tr>
<td>Quad</td>
<td>3</td>
<td>.959</td>
<td>.983</td>
<td>.945</td>
</tr>
<tr>
<td>Magic</td>
<td>11</td>
<td>.926</td>
<td>.996</td>
<td>.979</td>
</tr>
<tr>
<td>Mam.</td>
<td>6</td>
<td>.957</td>
<td>.989</td>
<td>.966</td>
</tr>
<tr>
<td>Ion.</td>
<td>9</td>
<td>.954</td>
<td>.989</td>
<td>.972</td>
</tr>
<tr>
<td>Bal.</td>
<td>5</td>
<td>.976</td>
<td>.989</td>
<td>.966</td>
</tr>
<tr>
<td>Liver</td>
<td>7</td>
<td>.956</td>
<td>.997</td>
<td>.970</td>
</tr>
<tr>
<td>Spam</td>
<td>10</td>
<td>.984</td>
<td>.998</td>
<td>.975</td>
</tr>
<tr>
<td>Heart</td>
<td>9</td>
<td>.960</td>
<td>.995</td>
<td>.976</td>
</tr>
</tbody>
</table>

Table: Estimated coverage of competing confidence procedures. Coverage is highlighted if not different from .950 at the .01 level.
## Results

Target coverage .950, loss function $L(X, Y, \beta) = (1 - YX^t\beta)^2$, $n = 30$

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Dim</th>
<th>ACI</th>
<th>CV-Normal</th>
<th>BCCVP-BR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ThreePt</td>
<td>2</td>
<td>.385</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Donut</td>
<td>3</td>
<td>.424</td>
<td>.483</td>
<td></td>
</tr>
<tr>
<td>Quad</td>
<td>3</td>
<td>.246</td>
<td>.360</td>
<td>.267</td>
</tr>
<tr>
<td>Magic</td>
<td>11</td>
<td>.548</td>
<td></td>
<td>.500</td>
</tr>
<tr>
<td>Mam.</td>
<td>6</td>
<td>.374</td>
<td>.456</td>
<td>.384</td>
</tr>
<tr>
<td>Ion.</td>
<td>9</td>
<td>.344</td>
<td>.466</td>
<td>.388</td>
</tr>
<tr>
<td>Bal.</td>
<td>5</td>
<td>.271</td>
<td>.350</td>
<td>.232</td>
</tr>
<tr>
<td>Liver</td>
<td>7</td>
<td>.534</td>
<td>.527</td>
<td>.500</td>
</tr>
<tr>
<td>Spam</td>
<td>10</td>
<td>.428</td>
<td>.496</td>
<td>.418</td>
</tr>
<tr>
<td>Heart</td>
<td>9</td>
<td>.367</td>
<td>.476</td>
<td>.404</td>
</tr>
</tbody>
</table>

**Table:** Estimated width of competing confidence procedures. Width is highlighted if coverage is at least .950 and the interval is smallest.
Results

Target coverage .950, loss function $L(X, Y, \beta) = \log(1 + e^{-YX^t\beta})$, $n = 30$

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Dim</th>
<th>ACI</th>
<th>CV-Normal</th>
<th>BCCVP-BR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ThreePt</td>
<td>2</td>
<td>.976</td>
<td>.893</td>
<td>.914</td>
</tr>
<tr>
<td>Donut</td>
<td>3</td>
<td>.968</td>
<td>.966</td>
<td>.908</td>
</tr>
<tr>
<td>Quad</td>
<td>3</td>
<td>.978</td>
<td>.997</td>
<td>.945</td>
</tr>
<tr>
<td>Magic</td>
<td>11</td>
<td>.955</td>
<td>.999</td>
<td>.983</td>
</tr>
<tr>
<td>Mam.</td>
<td>6</td>
<td>.951</td>
<td>.993</td>
<td>.974</td>
</tr>
<tr>
<td>Ion.</td>
<td>9</td>
<td>.947</td>
<td>.995</td>
<td>.985</td>
</tr>
<tr>
<td>Bal.</td>
<td>5</td>
<td>.979</td>
<td>.996</td>
<td>.972</td>
</tr>
<tr>
<td>Liver</td>
<td>7</td>
<td>.946</td>
<td>.997</td>
<td>.972</td>
</tr>
<tr>
<td>Spam</td>
<td>10</td>
<td>.985</td>
<td>.999</td>
<td>.981</td>
</tr>
<tr>
<td>Heart</td>
<td>9</td>
<td>.960</td>
<td>.995</td>
<td>.976</td>
</tr>
</tbody>
</table>

Table: Estimated coverage of competing confidence procedures. Coverage is highlighted if not different from .950 at the .01 level.
## Results

Target coverage .950, loss function $L(X, Y, \beta) = \log(1 + e^{-YX^t\beta})$, $n = 30$

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Dim</th>
<th>ACI</th>
<th>CV-Normal</th>
<th>BCCVP-BR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ThreePt</td>
<td>2</td>
<td>.374</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Donut</td>
<td>3</td>
<td>.434</td>
<td>.485</td>
<td></td>
</tr>
<tr>
<td>Quad</td>
<td>3</td>
<td>.310</td>
<td>.372</td>
<td>.267</td>
</tr>
<tr>
<td>Magic</td>
<td>11</td>
<td>.466</td>
<td>.526</td>
<td>.504</td>
</tr>
<tr>
<td>Mam.</td>
<td>6</td>
<td>.373</td>
<td>.448</td>
<td>.387</td>
</tr>
<tr>
<td>Ion.</td>
<td>9</td>
<td>.305</td>
<td>.459</td>
<td>.401</td>
</tr>
<tr>
<td>Bal.</td>
<td>5</td>
<td>.262</td>
<td>.349</td>
<td>.257</td>
</tr>
<tr>
<td>Liver</td>
<td>7</td>
<td>.533</td>
<td>.526</td>
<td>.518</td>
</tr>
<tr>
<td>Spam</td>
<td>10</td>
<td>.454</td>
<td>.494</td>
<td>.423</td>
</tr>
<tr>
<td>Heart</td>
<td>9</td>
<td>.367</td>
<td>.476</td>
<td>.404</td>
</tr>
</tbody>
</table>

**Table:** Estimated width of competing confidence procedures. Width is highlighted if coverage is at least .950 and the interval is smallest.
Results

Target coverage .950, loss function \( L(X, Y, \beta) = \log(1 + e^{-YX^t\beta}) \), \( n = 250 \)

\[
\begin{array}{|c|c|c|c|c|}
\hline
\text{Data Set} & \text{Dim} & \text{ACI} & \text{CV-Normal} & \text{BCCVP-BR} \\
\hline
\text{ThreePt} & 2 & .935 & .387 & .930 \\
\text{Donut} & 3 & .974 & .895 & .988 \\
\text{Quad} & 2 & .965 & .999 & \text{.940} \text{ (Highlighted)} \\
\text{Magic} & 11 & .962 & .999 & .983 \\
\text{Mam.} & 6 & .960 & .995 & .968 \\
\text{Ion.} & 9 & .952 & .996 & .949 \\
\text{Bal.} & 5 & .946 & .991 & \text{.963} \text{ (Highlighted)} \\
\text{Liver} & 7 & .971 & .996 & .984 \\
\text{Spam} & 10 & .979 & .999 & \text{.958} \text{ (Highlighted)} \\
\text{Heart} & 9 & .958 & .989 & .976 \\
\hline
\end{array}
\]

Table: Estimated coverage of competing confidence procedures. Coverage is highlighted if not different from .950 at the .01 level.
## Results

Target coverage .950, loss function $L(X, Y, \beta) = \log(1 + e^{-YX^t\beta})$, $n = 250$

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Dim</th>
<th>ACI</th>
<th>CV-Normal</th>
<th>BCCVP-BR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ThreePt</td>
<td>2</td>
<td>.122</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Donut</td>
<td>3</td>
<td>.123</td>
<td></td>
<td>.201</td>
</tr>
<tr>
<td>Quad</td>
<td>3</td>
<td>.0811</td>
<td>.104</td>
<td>.0885</td>
</tr>
<tr>
<td>Magic</td>
<td>11</td>
<td>.125</td>
<td>.157</td>
<td>.122</td>
</tr>
<tr>
<td>Mam.</td>
<td>6</td>
<td>.112</td>
<td>.140</td>
<td>.105</td>
</tr>
<tr>
<td>Ion.</td>
<td>9</td>
<td>.103</td>
<td>.127</td>
<td>.104</td>
</tr>
<tr>
<td>Bal.</td>
<td>5</td>
<td>.0623</td>
<td>.0772</td>
<td>.0620</td>
</tr>
<tr>
<td>Liver</td>
<td>7</td>
<td>.151</td>
<td>.172</td>
<td>.143</td>
</tr>
<tr>
<td>Spam</td>
<td>10</td>
<td>.125</td>
<td>.140</td>
<td>.108</td>
</tr>
<tr>
<td>Heart</td>
<td>9</td>
<td>.102</td>
<td>.132</td>
<td>.105</td>
</tr>
</tbody>
</table>

**Table:** Estimated width of competing confidence procedures. Width is highlighted if coverage is at least .950 and the interval is smallest.
Conclusions

ACI

- Provides nominal coverage
- Non-trivial width
- Consistent under non-regular setting
- Computationally efficient
Conclusions

ACI
- Provides nominal coverage
- Non-trivial width
- Consistent under non-regular setting
- Computationally efficient

Notes
- Consistent under close alternatives
- \( p >> n \) important extension
- Choice of \( a_n \)
Thank you.