Universal inference using the **split** likelihood ratio test

Aaditya Ramdas

Dept. of Statistics and Data Science
Machine Learning Dept.
Carnegie Mellon University

Larry Wasserman

Sivaraman Balakrishnan
Summary

1. For irregular models, use split LRT instead of LRT

2. Can invert to get a universal confidence set

3. LRT and split LRT are similar for regular models

4. Model selection using sieves

5. Can derandomize by averaging

6. Sequential running-MLE LRT (+ confidence sequence)

7. Use RIPL split LRT (when computable) for convex set of densities
For regular models, LRT is easy

Consider $Y_1, \ldots, Y_n \sim p_{\theta^*}$ for some $\theta^* \in \Theta$.

Test $H_0 : \theta^* \in \Theta_0$ vs. $H_1 : \theta^* \in \Theta_1 \supset \Theta_0$
For regular models, LRT is easy

Consider $Y_1, \ldots, Y_n \sim p_{\theta^*}$ for some $\theta^* \in \Theta$.

Test $H_0: \theta^* \in \Theta_0$ vs. $H_1: \theta^* \in \Theta_1 \supset \Theta_0$

Wilk's Thm (regular models): $2 \log \frac{\mathcal{L}(\hat{\theta}_1)}{\mathcal{L}(\hat{\theta}_0)} \rightarrow \chi^2_d$ under $H_0$.

$\mathcal{L}(\theta) := \prod_{i=1}^n p_\theta(Y_i)$ is the likelihood function. $\hat{\theta}_{0/1}$ is MLE under $\Theta_{0/1}$.

d is difference in dimensionality between $\Theta_0, \Theta_1$. 

For regular models, LRT is easy

Consider $Y_1, \ldots, Y_n \sim p_{\theta^*}$ for some $\theta^* \in \Theta$.

Test $H_0 : \theta^* \in \Theta_0$ vs. $H_1 : \theta^* \in \Theta_1 \supset \Theta_0$

Wilk's Thm (regular models): $2 \log \frac{\mathcal{L}(\hat{\theta}_1)}{\mathcal{L}(\hat{\theta}_0)} \to \chi^2_d$ under $H_0$.

$\mathcal{L}(\theta) := \prod_{i=1}^{n} p_{\theta}(Y_i)$ is the likelihood function. $\hat{\theta}_{0/1}$ is MLE under $\Theta_{0/1}$.

$d$ is difference in dimensionality between $\Theta_0, \Theta_1$.

LRT rejects if $2 \log \frac{\mathcal{L}(\hat{\theta}_1)}{\mathcal{L}(\hat{\theta}_0)} \geq c_{\alpha,d}$.

$(1 - \alpha)$ quantile of $\chi^2_d$

Under regularity conditions, $\Pr_{H_0}(\text{rejection}) \leq \alpha + o_p(1)$.
Irregular composite testing problems are common

In all these cases, limiting distribution and a level-$\alpha$ test are unknown. In all these cases, we can (approximately) calculate MLE under null.
Irregular composite testing problems are common

1. (Mixtures) $H_0 : p_{\theta^*}$ is a mixture of $k$ Gaussians

In all these cases, limiting distribution and a level-$\alpha$ test are unknown. In all these cases, we can (approximately) calculate MLE under null.
Irregular composite testing problems are common

1. (Mixtures) $H_0 : p_{\theta^*}$ is a mixture of $k$ Gaussians

2. (Nonparametric shape constraints) $H_0 : p$ is log-concave

In all these cases, limiting distribution and a level-$\alpha$ test are unknown. In all these cases, we can (approximately) calculate MLE under null.
Irregular composite testing problems are common

1. (Mixtures) $H_0 : p_{\theta^*}$ is a mixture of $k$ Gaussians

2. (Nonparametric shape constraints) $H_0 : p$ is log-concave

3. (Dependence) $H_0 : p_{\theta^*}$ is Gaussian $MTP_2$ or Ising $MTP_2$

In all these cases, limiting distribution and a level-$\alpha$ test are unknown. In all these cases, we can (approximately) calculate MLE under null.
Irregular composite testing problems are common

1. (Mixtures) \( H_0 : p_{\theta^*} \) is a mixture of \( k \) Gaussians

2. (Nonparametric shape constraints) \( H_0 : p \) is log-concave

3. (Dependence) \( H_0 : p_{\theta^*} \) is Gaussian \( MTP_2 \) or Ising \( MTP_2 \)

4. (Linear model) \( H_0 : \theta^* \) is \( k \)-sparse

In all these cases, limiting distribution and a level-\( \alpha \) test are unknown. In all these cases, we can (approximately) calculate MLE under null.
Irregular composite testing problems are common

1. (Mixtures) $H_0 : p_{\theta^*}$ is a mixture of $k$ Gaussians

2. (Nonparametric shape constraints) $H_0 : p$ is log-concave

3. (Dependence) $H_0 : p_{\theta^*}$ is Gaussian $MTP_2$ or Ising $MTP_2$

4. (Linear model) $H_0 : \theta^*$ is $k$-sparse

5. (Factor models or HMMs) $H_0 : \theta^*$ has $k$ hidden factors/states

In all these cases, limiting distribution and a level-$\alpha$ test are unknown. In all these cases, we can (approximately) calculate MLE under null.
Irregular composite testing problems are common

1. (Mixtures) $H_0 : p_{θ^*}$ is a mixture of $k$ Gaussians

2. (Nonparametric shape constraints) $H_0 : p$ is log-concave

3. (Dependence) $H_0 : p_{θ^*}$ is Gaussian $MTP_2$ or Ising $MTP_2$

4. (Linear model) $H_0 : θ^*$ is $k$-sparse

5. (Factor models or HMMs) $H_0 : θ^*$ has $k$ hidden factors/states

6. (Gaussian CI testing) $H_0 : X_1 ⊥ X_2, \quad H_1 : X_1 ⊥ X_2 \mid X_3$

In all these cases, limiting distribution and a level-$α$ test are unknown. In all these cases, we can (approximately) calculate MLE under null.
Our proposal: split LRT

(regular models) LRT rejects if

\[
2 \log \frac{\mathcal{L}(\hat{\theta}_1)}{\mathcal{L}(\hat{\theta}_0)} \geq c_{\alpha,d} \cdot (1 - \alpha) \text{ quantile of } \chi^2_d
\]

(any model) split data into two parts \( D_0, D_1 \).
Our proposal: split LRT

(regular models) LRT rejects if \( 2 \log \frac{\mathcal{L}(\hat{\theta}_1)}{\mathcal{L}(\hat{\theta}_0)} \geq c_{\alpha,d} \cdot \chi^2_{d} \) 

\( (1 - \alpha) \) quantile of \( \chi^2_d \)

(any model) split data into two parts \( D_0, D_1 \).

split LRT rejects if \( 2 \log \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_0)} \geq 2 \log(1/\alpha) \).
Our proposal: split LRT

(regular models) LRT rejects if
\[ 2 \log \frac{\mathcal{L}(\hat{\theta}_1)}{\mathcal{L}(\hat{\theta}_0)} \geq c_{\alpha,d} \cdot (1 - \alpha) \text{ quantile of } \chi^2_d \]

(any model) split data into two parts $D_0, D_1$.

split LRT rejects if
\[ 2 \log \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_0)} \geq 2 \log(1/\alpha) \cdot \]

$\mathcal{L}_0(\theta) := \prod_{i \in D_0} p_{\theta}(Y_i)$ is the likelihood on $D_0$.

$\hat{\theta}_1$ is any estimator (MLE/Bayes/robust) under $\Theta_1$ on $D_1$.

$\hat{\theta}_0$ is MLE under $\Theta_0$ on $D_0$. 
Our proposal: split LRT

(regular models) LRT rejects if \( 2 \log \frac{\mathcal{L}(\hat{\theta}_1)}{\mathcal{L}(\hat{\theta}_0)} \geq c_{\alpha,d} \cdot (1 - \alpha) \) quantile of \( \chi^2_d \)

(any model) split data into two parts \( D_0, D_1 \).

split LRT rejects if \( 2 \log \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_0)} \geq 2 \log(1/\alpha) \).

\( \mathcal{L}_0(\theta) := \prod_{i \in D_0} p_{\theta}(Y_i) \) is the likelihood on \( D_0 \).

\( \hat{\theta}_1 \) is any estimator (MLE/Bayes/robust) under \( \Theta_1 \) on \( D_1 \).

\( \hat{\theta}_0 \) is MLE under \( \Theta_0 \) on \( D_0 \).

Under no regularity conditions \( \Pr(\text{rejection}) \leq \alpha \).
Split LRT is valid without assumptions

(irregular models) split data into two parts $D_0, D_1$.

Split LRT rejects if

$$2 \log \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_0)} \geq 2 \log(1/\alpha).$$

Under no regularity conditions, $\Pr_{H_0}(\text{rejection}) \leq \alpha$. 
Split LRT is valid without assumptions

(irregular models) split data into two parts $D_0, D_1$.

Split LRT rejects if

\[ 2 \log \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_0)} \geq 2 \log(1/\alpha). \]

Under no regularity conditions,

\[ \Pr_{H_0}(\text{rejection}) \leq \alpha. \]

\[
\mathbb{E}_{\theta^* \in H_0} \left[ \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_0)} \right] \leq \mathbb{E}_{\theta^* \in H_0} \left[ \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\theta^*)} \right] = \prod_{i \in D_0} \int \frac{p_{\hat{\theta}_1}(y)}{p_{\theta^*}(y)} \frac{p_{\theta^*}(y)}{p_{\theta^*}(y)} dy = 1.
\]
Split LRT is valid without assumptions

(irregular models) split data into two parts $D_0, D_1$.

split LRT rejects if $2 \log \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_0)} \geq 2 \log(1/\alpha)$.

Under no regularity conditions, $\Pr_{H_0}(\text{rejection}) \leq \alpha$.

\[
\mathbb{E}_{\theta^* \in H_0} \left[ \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_0)} \right] \leq \mathbb{E}_{\theta^* \in H_0} \left[ \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\theta^*)} \right] = \prod_{i \in D_0} \int p_{\hat{\theta}_1}(y) \frac{p_{\theta^*}(y)}{p_{\theta^*}(y)} dy = 1.
\]

By Markov's, $\Pr_{H_0} \left( \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_0)} > 1/\alpha \right) \leq \alpha$. 

\[
\text{error}
\]
Split LRT is valid without assumptions

(irregular models) split data into two parts $D_0, D_1$.

split LRT rejects if

$$2 \log \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_0)} \geq 2 \log(1/\alpha).$$

Under no regularity conditions, $\Pr_{H_0}(\text{rejection}) \leq \alpha$.

$$\mathbb{E}_{\theta^* \in H_0} \left[ \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_0)} \right] \leq \mathbb{E}_{\theta^* \in H_0} \left[ \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\theta^*)} \right] = \prod_{i \in D_0} \int \frac{p_{\hat{\theta}_1}(y)}{p_{\theta^*}(y)} p_{\theta^*}(y) \, dy = 1.$$

By Markov's, $\Pr_{H_0} \left( \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_0)} > 1/\alpha \right) \leq \alpha$.

“poor man’s Chernoff bound”
Universal confidence set for $\theta^*$

$$A_n := \left\{ \theta \in \Theta : 2 \log \frac{\mathcal{L}(\hat{\theta}_1)}{\mathcal{L}(\theta)} \leq c_{\alpha,d} \right\}$$

Under regularity conditions $\Pr_{\theta^*}(\theta^* \in A_n) \geq 1 - \alpha + o_P(1)$. 
Universal confidence set for $\theta^*$

$$A_n := \left\{ \theta \in \Theta : 2 \log \frac{\mathcal{L}(\hat{\theta}_1)}{\mathcal{L}(\theta)} \leq c_{\alpha,d} \right\}$$

Under regularity conditions $$\Pr_{\theta^*}(\theta^* \in A_n) \geq 1 - \alpha + o_P(1).$$

$$C_n := \left\{ \theta \in \Theta : 2 \log \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\theta)} \leq 2 \log(1/\alpha) \right\}$$

Under no regularity conditions, $$\Pr_{\theta^*}(\theta^* \in C_n) \geq 1 - \alpha.$$
Universal confidence set for $\theta^*$

$$A_n := \left\{ \theta \in \Theta : 2 \log \frac{\mathcal{L}(\hat{\theta}_1)}{\mathcal{L}(\theta)} \leq c_{\alpha,d} \right\}$$

Under regularity conditions, $\Pr_{\theta^*}(\theta^* \in A_n) \geq 1 - \alpha + o_P(1)$.

$$C_n := \left\{ \theta \in \Theta : 2 \log \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\theta)} \leq 2 \log(1/\alpha) \right\}$$

Under no regularity conditions, $\Pr_{\theta^*}(\theta^* \in C_n) \geq 1 - \alpha$.

For regular models (fixed $d, \alpha$), both diameters $O(1/\sqrt{n})$.

For high-dimensional Gaussians (fixed $\alpha$), $C_n$ is 4 times wider.
Can use relaxations of MLE

Sometimes, we can find (often convex) relaxation $F(\theta)$ s.t.

$$\max_{\theta \in \Theta_0, \mathcal{F}} \mathcal{F}(\theta) \geq \max_{\theta \in \Theta_0} \mathcal{L}(\theta)$$
Can use relaxations of MLE

Sometimes, we can find (often convex) relaxation $F(\theta)$ s.t.

$$\max_{\theta \in \Theta_0, F} F(\theta) \geq \max_{\theta \in \Theta_0} \mathcal{L}(\theta)$$

Define $\hat{\theta}_{0,F} := \arg \max_{\theta \in \Theta_0, F} F(\theta)$
Can use relaxations of MLE

Sometimes, we can find (often convex) relaxation $F(\theta)$ s.t.

$$\max_{\theta \in \Theta_0, \mathcal{F}} \mathcal{F}(\theta) \geq \max_{\theta \in \Theta_0} \mathcal{L}(\theta)$$

Define $\hat{\theta}_{0, \mathcal{F}} := \arg \max_{\theta \in \Theta_0, \mathcal{F}} \mathcal{F}(\theta)$

(relaxed) split LRT rejects if

$$\frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{F}_0(\hat{\theta}_{0, \mathcal{F}})} \geq 1/\alpha.$$
Can use relaxations of MLE

Sometimes, we can find (often convex) relaxation $F(\theta)$ s.t.

$$\max_{\theta \in \Theta_0, F} F(\theta) \geq \max_{\theta \in \Theta_0} L(\theta)$$

Define $\hat{\theta}_{0,F} := \arg \max_{\theta \in \Theta_0, F} F(\theta)$

(relaxed) split LRT rejects if $\frac{L_0(\hat{\theta}_1)}{F_0(\hat{\theta}_{0,F})} \geq 1/\alpha$.

Under no regularity conditions, $\Pr(\text{rejection}) \leq \alpha$. 
From testing to model selection using sieves

Nested models $\mathcal{P}_0 \subset \mathcal{P}_1 \subset \mathcal{P}_2 \ldots$

Assume $p^* \in \mathcal{P}_{j^*}$ for some (smallest) $j^*$. 
From testing to model selection using sieves

Nested models $\mathcal{P}_0 \subset \mathcal{P}_1 \subset \mathcal{P}_2 \ldots$

Assume $p^* \in \mathcal{P}_{j^*}$ for some (smallest) $j^*$.

Sieves: Test $H_{0,j} : p^* \in \mathcal{P}_j$ one by one for $j = 1, 2, \ldots$

Reject $H_{0,j}$ if
$$\prod_{i \in D_0} \frac{\hat{p}_{j+1}(Y_i)}{\hat{p}_j(Y_i)} > 1/\alpha,$$

Stop at step $\hat{j}$, when we fail to reject the null.
From testing to model selection using sieves

Nested models $\mathcal{P}_0 \subset \mathcal{P}_1 \subset \mathcal{P}_2 \ldots$

Assume $p^* \in \mathcal{P}_{j^*}$ for some (smallest) $j^*$.

Sieves: Test $H_{0,j} : p^* \in \mathcal{P}_j$ one by one for $j = 1, 2, \ldots$

Reject $H_{0,j}$ if

$$\prod_{i \in D_0} \frac{\hat{p}_{j+1}(Y_i)}{\hat{p}_j(Y_i)} > 1/\alpha,$$

Stop at step $\hat{j}$, when we fail to reject the null.

Then $j^* \geq \hat{j}$ w.p. $\geq 1 - \alpha$, i.e. $\Pr(p^* \in \mathcal{P}_{\hat{j} - 1}) \leq \alpha$.

(no multiple testing correction needed)
Derandomization by averaging

Under no regularity conditions, $\Pr_{H_0}(\text{rejection}) \leq \alpha$. 
Derandomization by averaging

1. (Crossfit LRT) Swap roles of $D_0, D_1$ and average test statistics

Under no regularity conditions, $\Pr_{H_0}(\text{rejection}) \leq \alpha$. 
Derandomization by averaging

1. (Crossfit LRT) Swap roles of $D_0, D_1$ and average test statistics

2. (K-fold LRT) Split into $K$ parts, use $K - 1$ to calculate $\hat{\theta}_1$, calculate $\hat{\theta}_0$ on last fold, evaluate test statistic on last fold, average across all folds. Alternately, we can calculate $\hat{\theta}_0$ on $K - 1$ splits, $\hat{\theta}_1$ on last split, evaluate test statistic on $K - 1$ folds, and average across all folds.

Under no regularity conditions, $\Pr_{H_0}(\text{rejection}) \leq \alpha$. 
Derandomization by averaging

1. (Crossfit LRT) Swap roles of $D_0, D_1$ and average test statistics

2. (K-fold LRT) Split into $K$ parts, use $K - 1$ to calculate $\hat{\theta}_1$, calculate $\hat{\theta}_0$ on last fold, evaluate test statistic on last fold, average across all folds. Alternately, we can calculate $\hat{\theta}_0$ on $K - 1$ splits, $\hat{\theta}_1$ on last split, evaluate test statistic on $K - 1$ folds, and average across all folds.

3. (All splits) Remove all randomization by considering all possible splits/permutations. Is there a statistics/comp. tradeoff?

Under no regularity conditions, $\Pr_{H_0}(\text{rejection}) \leq \alpha$. 
Sequential (running-MLE) split LRT

If you fail to reject, collect more data $D_{\text{new}}$, we are allowed to update the test statistic (in a particular way) and check again if it is larger than $1/\alpha$. We can repeat this indefinitely.

(no multiple testing correction needed)
Sequential (running-MLE) split LRT

If you fail to reject, collect more data $D_{\text{new}}$, we are allowed to update the test statistic (in a particular way) and check again if it is larger than $1/\alpha$. We can repeat this indefinitely.

(no multiple testing correction needed)

(Special case: one datapoint at each step) “Running-MLE LRT”

Reject when $M_t := \frac{\prod_{i=1}^{t} p_{\hat{\theta}_1(i-1)}(X_i)}{\prod_{i=1}^{t} p_{\hat{\theta}_0(t)}(X_i)} > 1/\alpha$. 
Sequential (running-MLE) split LRT

If you fail to reject, collect more data $D_{\text{new}}$, we are allowed to update the test statistic (in a particular way) and check again if it is larger than $1/\alpha$. We can repeat this indefinitely.

(no multiple testing correction needed)

(Special case: one datapoint at each step) “Running-MLE LRT”

Reject when $M_t := \frac{\prod_{i=1}^{t} p_{\hat{\theta}_1(i-1)}(X_i)}{\prod_{i=1}^{t} p_{\hat{\theta}_0(t)}(X_i)} > 1/\alpha$.

Under no regularity conditions, $\Pr_{H_0}(\exists t : \text{rejection at step } t) \leq \alpha$.

Reason: $M_t \leq L_t := \frac{\prod_{i=1}^{t} p_{\hat{\theta}_1(i-1)}(X_i)}{\prod_{i=1}^{t} p_{\theta^*}(X_i)}$, a nonnegative martingale.
The case of a convex class of null densities
The case of a convex class of null densities

RIPR split LRT rejects if
\[ 2 \log \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_{1\to\Theta_0})} \geq 2 \log(1/\alpha). \]

\[ \hat{\theta}_{1\to\Theta_0} \approx \text{reverse KL projection of } \hat{\theta}_1 \text{ onto } \Theta_0. \]

If \( \{p_\theta\}_{\theta \in \Theta_0} \) is convex, \( \Pr(\text{rejection}) \leq \alpha. \)
The case of a convex class of null densities

\[ \Pr_{H_0} \text{(rejection)} \leq \alpha. \]

RIPR split LRT rejects if
\[ 2 \log \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_{1 \rightarrow \Theta_0})} \geq 2 \log(1/\alpha). \]

\( \hat{\theta}_{1 \rightarrow \Theta_0} \approx \) reverse KL projection of \( \hat{\theta}_1 \) onto \( \Theta_0 \).

If \( \{p_\theta\}_{\theta \in \Theta_0} \) is convex, \( \Pr_{H_0}(\text{rejection}) \leq \alpha. \)

Reason (Li'99): \( \forall \theta^* \in \Theta_0, \mathbb{E}_{\theta^*} \left[ \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_{1 \rightarrow \Theta_0})} \right] \leq 1. \)
The case of a convex class of null densities

RIPR split LRT rejects if
\[ 2 \log \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_1 \rightarrow \Theta_0)} \geq 2 \log(1/\alpha). \]

\( \hat{\theta}_1 \rightarrow \Theta_0 \approx \) reverse KL projection of \( \hat{\theta}_1 \) onto \( \Theta_0 \).

If \( \{p_\theta\}_{\theta \in \Theta_0} \) is convex, \( \Pr(\text{rejection}) \leq \alpha \).

Reason (Li'99): \( \forall \theta^* \in \Theta_0, \mathbb{E}_{\theta^*} \left[ \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_1 \rightarrow \Theta_0)} \right] \leq 1. \)

RIPR split LRT dominates MLE split LRT in the convex case, but requires calculation of reverse KL projection.
The case of a convex class of null densities

RIPR split LRT rejects if $2 \log \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_{1\to\Theta_0})} \geq 2 \log(1/\alpha)$.

$\hat{\theta}_{1\to\Theta_0} \approx$ reverse KL projection of $\hat{\theta}_1$ onto $\Theta_0$.

If $\{p_\theta\}_{\theta \in \Theta_0}$ is convex, $\Pr_{H_0}(\text{rejection}) \leq \alpha$.

Reason (Li'99): $\forall \theta^* \in \Theta_0, \mathbb{E}_{\theta^*} \left[ \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_{1\to\Theta_0})} \right] \leq 1$.

RIPR split LRT dominates MLE split LRT in the convex case, but requires calculation of reverse KL projection.

Model misspecification: suppose the data comes from $q \notin \Theta$. If RIPR($q$) $\in \Theta$, then the universal set contains the RIPR whp.
Summary

1. For irregular models, use split LRT instead of LRT

2. Can invert to get a universal confidence set

3. LRT and split LRT are similar for regular models

4. Model selection using sieves

5. Can derandomize by averaging

6. Sequential running-MLE LRT (+ confidence sequence)

7. Use RIPR split LRT (when computable) for convex set of densities
Universal inference using the **split** likelihood ratio test

Aaditya Ramdas

Dept. of Statistics and Data Science
Machine Learning Dept.
Carnegie Mellon University

Larry Wasserman

Sivaraman Balakrishnan