Optimal Confidence Sets for Parameters in Discrete Distributions

Presentation by:
Joshua D. Habiger (OSU)

Coauthors:
Melinda McCann (OSU) and Joshua Tebbs (USC)

April 6, 2012

Outline

- Background
- A Class of Estimators
- The Optimal Estimator
- Assessment of the Estimator
- Some Remarks

Background

Standard Inference

- Data: $X \sim f_\theta(x)$
- Goal: Construct $100(1-\alpha)\%$ confidence interval for θ
- Example: $X \sim N(\mu, \sigma^2) \rightarrow X \pm 1.96\sigma$
  - $\Pr(X - 1.96\sigma < \mu < X + 1.96\sigma) = 0.95$ FOR EVERY μ!

Challenge

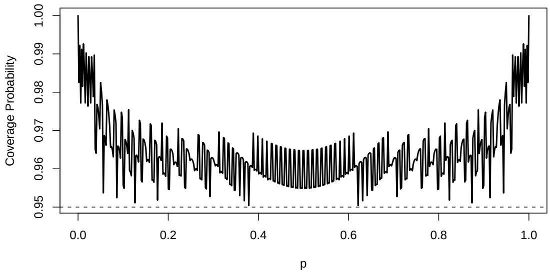
- Data: $X \sim Bin(n, p)$
- Extra Credit: Construct an estimator for p so that the coverage probability is 0.95?
- Hint: Could you define a test for testing $H_0 : p = p_0$ with type 1 error rate 0.05?

Clopper Pearson (1934)

- Idea: Lower/upper limits found by inverting lower/upper tailed **level** $\alpha/2$ test
  - Upper tailed test: reject $H_0 : p \leq p_0$ in favor of $H_1 : p > p_0$ if $X > k$
    - $\Pr_{p_0}(X > k) \leq \alpha/2$
- Interval: $\{p_0 : \text{neither } H_0 \text{ not rejected}\}$

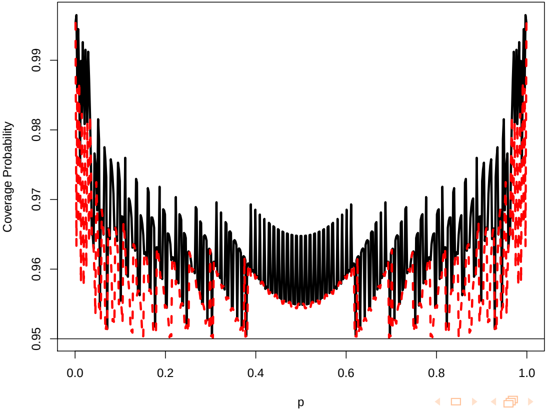
Clopper Pearson Coverage

$$Coverage(p) = \Pr(p \in Interval) \geq 0.95$$

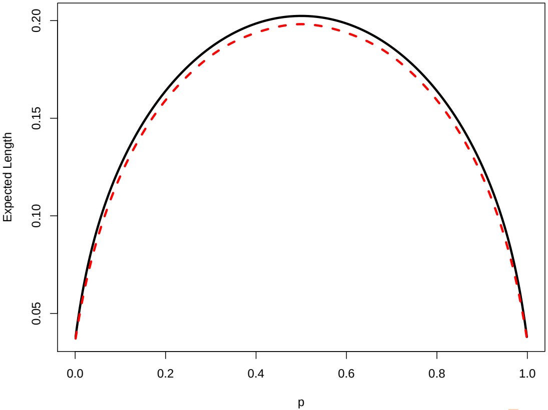


Improvement: Blaker (2000)

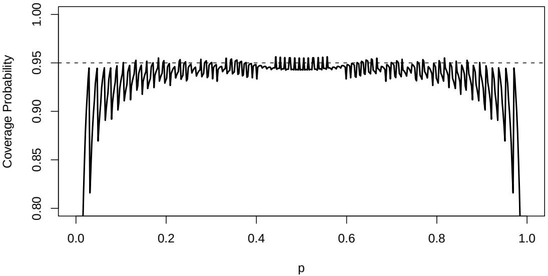
lower tailed level + upper tailed level ≤ 0.05



Expected Lengths

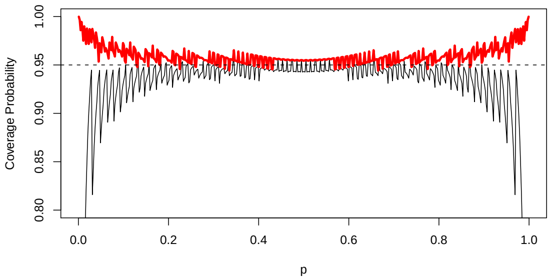


Standard Nonconservative Interval



$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \text{with} \quad \hat{p} = \frac{X}{n}$$

Agresti Coul (1998)

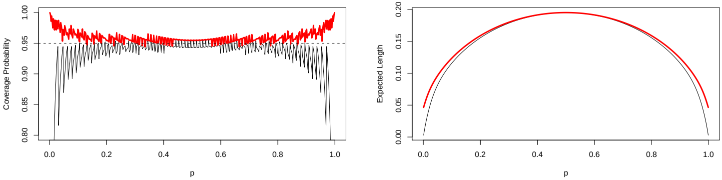


$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \text{with} \quad \hat{p} = \frac{X+2}{n+4}$$

Other Intervals

- Many Intervals - See Agresti, Gottard (2007) or Brown, Cai, DasGupta (2002) or Newcombe (1998)
  - Wilson interval, Jefferys interval, mid-p interval, . . .
- **Which is best?**
  - Considerations: Expected Length and Coverage Probability

Wald vs. Agresti Coul



- Is one interval ***always*** better? - unlucky p
- Is one interval ***generally*** better?

Good Intervals

Intervals are "good" if (cf. Newcomb(1998), Brown et. al (2002), ...)

1. *mean* coverage "near" $1 - \alpha$
2. *mean* expected length small

Definitions:

$$\text{Mean Coverage} = \int_0^1 \text{Coverage}(p)dp$$

$$\text{Mean Expected Length} = \int_0^1 \text{Expected Length}(p)dp$$

Example

- Brown et. al (2001) (n = 40).

Interval	Mean Coverage	Mean Expected Length
Standard	0.900	0.230
Wilson	0.952	0.240
Agresti Coull	0.960	0.245
Jefferys Prior	0.951	.0239

Note: Actual numerical values are approximations based on figures

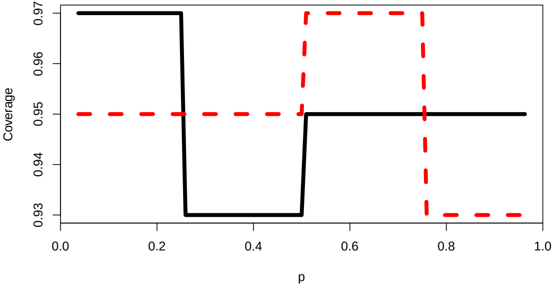
Critical Thinking

Hypothetical Question:

Interval	Mean Coverage	Mean Expected Length
Interval 1	0.96	0.25
Interval 2	0.94	0.24

- Who wins?

More Critical Thinking



- Who wins?

Main Problems

1. Mean coverage “near $1 - \alpha$”
2. Different applications may call for different intervals

Problems not limited to binomial data

Goals and Route to a Solution

Goals:

1. Provide a definition of optimality that is
   a. precise
   b. flexible
2. Find the optimal estimator

Route:

1. To accomplish 1:
   a. Among all estimators whose **mean coverage =** $1 - \alpha$ it has **smallest mean expected length**
   b. Allow for weighted mean and consider a general setting
2. Do some math

A Class of Interval Estimators

Basic Elements

- $X \in \mathcal{X}$ for $\mathcal{X}$ countable
- $X \sim p_\theta(x)$, $\theta \in \Theta \subseteq \Re$
- Estimator: $A : \mathcal{X} \to \sigma(\Theta)$
- Loss function: $L(A(x), \theta) = 1 - I(\theta \in A(x))$

Risk/Coverage

$$R(\theta, A) = \sum_{x \in \mathcal{X}} L(A(x), \theta) p_\theta(x)$$

$$C(\theta, A) = 1 - R(\theta, A)$$

- Want to choose A s.t. $R(\theta, A) = \alpha$, but . . .

Mean Risk/Coverage

User specified weight: $w : \Theta \rightarrow \Re^{+}$

- Large $w(\theta) \rightarrow$ performance of A at θ more important
- Assume w has properties of a density fxn on Θ

Mean risk: $r(w, A) = \int_{\Theta} R(\theta, A) w(\theta) d\theta$

Mean Coverage: $c(w, A) = 1 - r(w, A)$

1-α mean interval estimators

Definition:If $r(w, A) = \alpha$ then A is a $1 - \alpha$ **mean interval estimator** for θ with respect to w.

Notation:

$$A_w = A_w(\alpha)$$

Weighted Set

Definition: For a fixed w and x, we say that $A(x)$ is a $1-\alpha_x$ ***w-weighted set*** for θ if

$$\int_\Theta L(A(x),\theta)w_x(\theta)d\theta = \alpha_x,$$

where

$$w_x(\theta) = \frac{p_\theta(x)w(\theta)}{p(x)} \quad \text{and} \quad p(x) = \int_\Theta p_\theta(x)w(\theta)d\theta$$

Notation:

$$A_w(x;\alpha_x)$$

Remark: Are we Bayesian?

Bayesian: $A_w(x; \alpha_x)$ a $1 - \alpha_x$ credible interval

What would a Bayesian do?

1. Philosophically: Choose w to reflect likely values for θ
2. Mathematically: Choose $\alpha_x = .05$, say, and get $A(x; .05)$

Class of estimators

Proposition: A_w is a $1-\alpha$ mean interval estimator with respect to w iff

$$\sum_{x \in \mathcal{X}} \alpha_x p(x) = \alpha. \qquad (1)$$

- **Class of such estimators satisfying (1):** $\mathcal{A}_w(\alpha)$
  - Note the Bayesian estimator ($\alpha_x = \alpha$) is in $\mathcal{A}_w(\alpha)$

The Optimal Estimator

Which estimator?

Key Issues:

1. Many choices of α_x satisfy $\sum_{x \in \mathcal{X}} \alpha_x p(x) = \alpha$
2. Given α_x, many $A_w(x; \alpha_x)$ satisfy

$$\int_{\Theta} L(A(x; \alpha_x), \theta) w_x(\theta) d\theta = \alpha_x$$

Length

Lengths:

1. Expected length: $\Lambda(\theta, A_w) = \sum_{x \in \mathcal{X}} \lambda(A_w(x; \alpha_x))p_\theta(x)$
2. Mean expected length: $\Lambda(A_w) = \int_\Theta \Lambda(\theta, A_w)w(\theta)d\theta$

Goal: Find the $A_w^*(\alpha) \in \mathcal{A}_w(\alpha)$ with smallest mean expected length - **minimum mean expected length estimator (MMELE)**

Assumptions

Some assumptions:

(A1) w_x unimodal and continuous

(A2) $A_w(x; \alpha_x) = [l_x(\alpha_x), u_x(\alpha_x)]$ for $l_x(\alpha_x)$, $u_x(\alpha_x)$ elements of the closure of Θ

(A3) If w_x is monotone, for any $\epsilon > 0$, there exists $\theta_0 \in \Theta$ s. t. $w_x(\theta_0) < \epsilon$. Else, there exists $\theta_1 \in \Theta$ and $\theta_2 \in \Theta$ s. t. $w_x(\theta_1), w_x(\theta_2) < \epsilon$.

Optimal w-Weighted Estimate

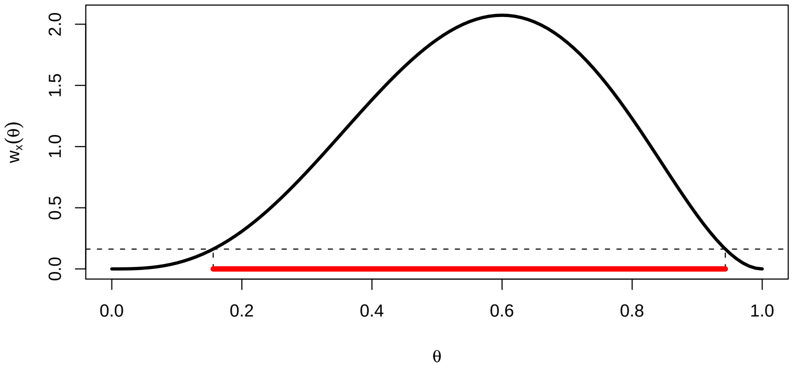
Lemma

For x and α_x fixed and under (A1), (A2), and (A3), $A_w^*(x;\alpha_x)$ satisfies

$$w_x(l_x^*(\alpha_x)) = w_x(u_x^*(\alpha_x))$$

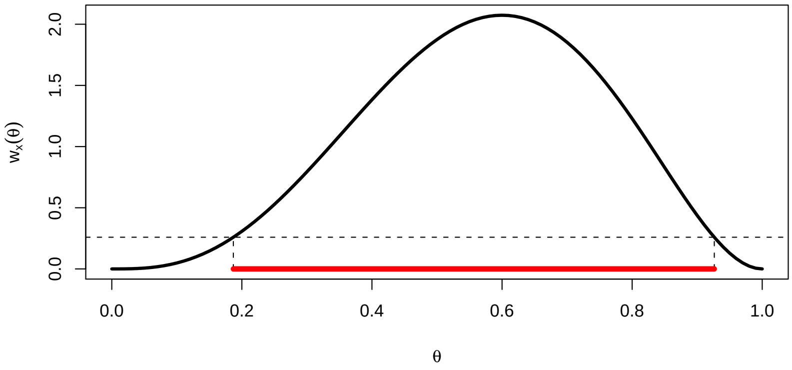
if $w_x(\cdot)$ is not monotone. Otherwise, use a one-sided interval.

Illustration



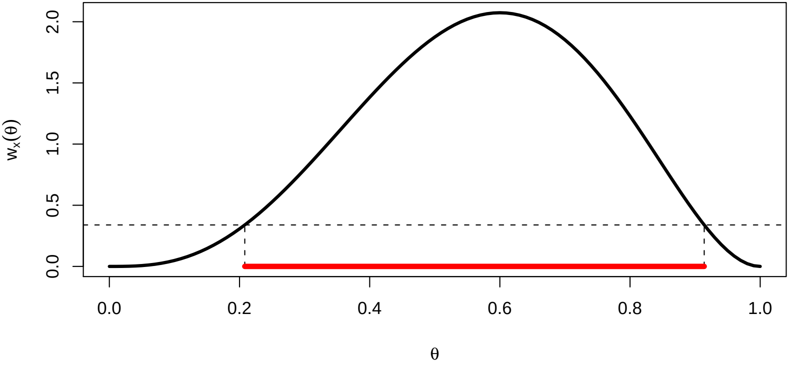
$\alpha_x = .01$

Illustration



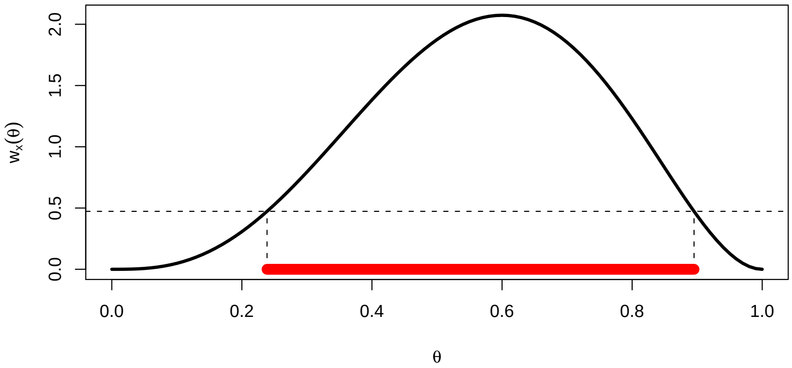
$\alpha_x = .02$

Illustration



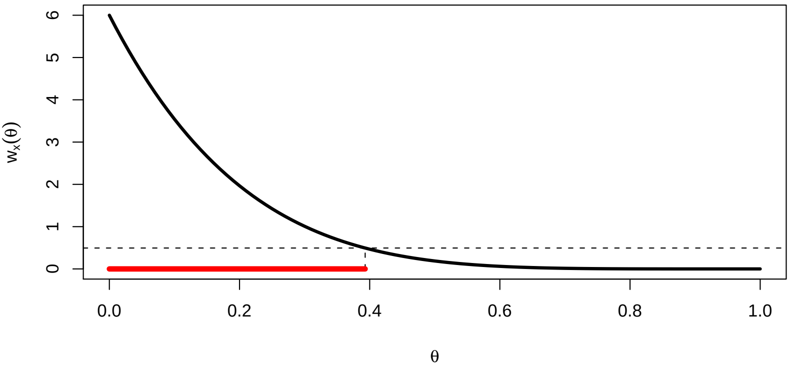
$\alpha_x = .03$

Illustration



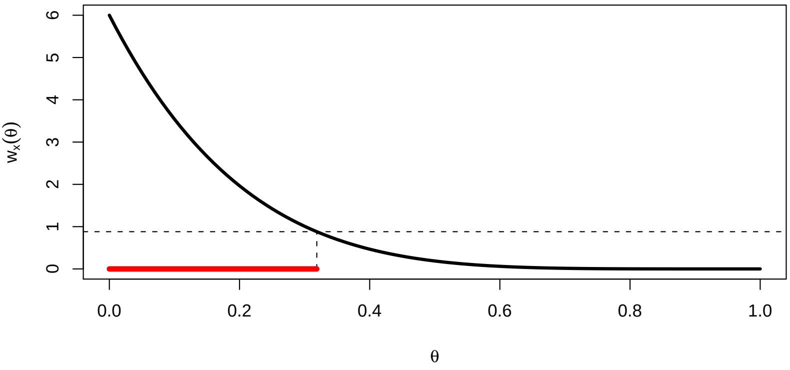
$\alpha_x = .05$

Illustration



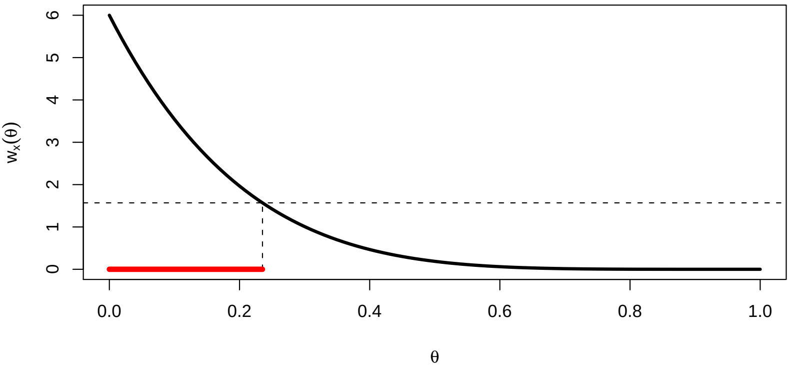
$\alpha_x = .05$

Illustration



$\alpha_x = .10$

Illustration



$\alpha_x = .20$

Towards the MMELE

Lemma

Let $\mathcal{A}_w^*(\alpha)$ be the collection of all $A_w \in \mathcal{A}_w(\alpha)$ defined by $A_w^*(x;\alpha_x) = [l_x^*(\alpha_x), u_x^*(\alpha_x)]$. Then, the MMELE $A_w^* \in \mathcal{A}_w^*(\alpha)$.

Point: We just need to find each α_x

Towards the MMELE

Any w-weighted estimate in the above smaller class can be specified in two ways

1. Specify $\alpha_x \to w_x(l_x^*(\alpha_x)) = w_x(u_x^*(\alpha_x))$
2. Specify density value $y \to$ find $l_x^*(y)$ and $u_x^*(y)$ satisfying $w_x(l_x^*) = w_x(u_x^*) = y$

Important Point:

$y \to [l_x(y), u_x(y)] \to \alpha_x(y) \to \alpha(y) = \sum_{x \in \mathcal{X}} \alpha_x(y) p(x)$

The MMELE

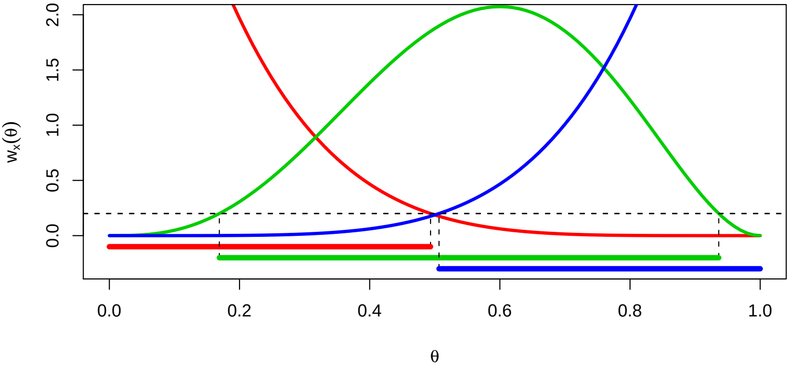
Theorem

Under (A1), (A2), and (A3) the ***unique*** *MMELE is*

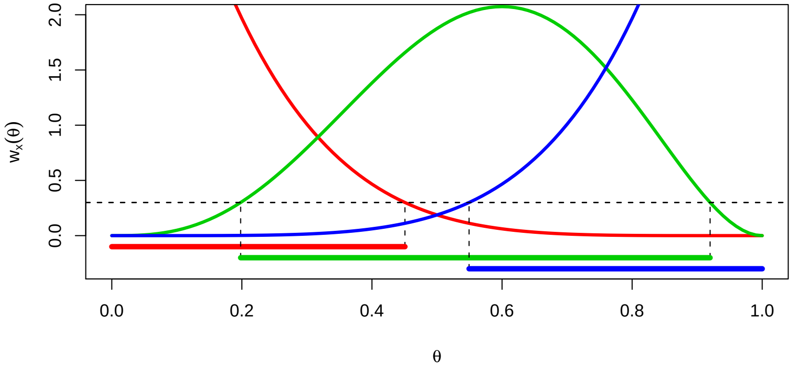
$$A_w^*(x; \alpha_x^*) = [l_x(y^*), u_x(y^*)]$$

for each x where y^* is the solution to $\alpha(y) = \alpha$.

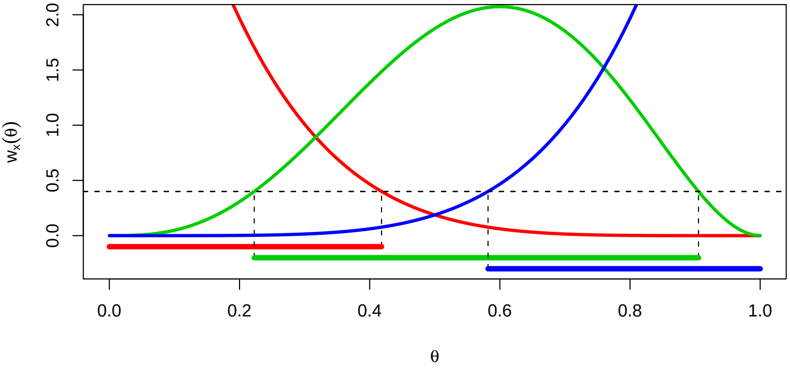
Illustration



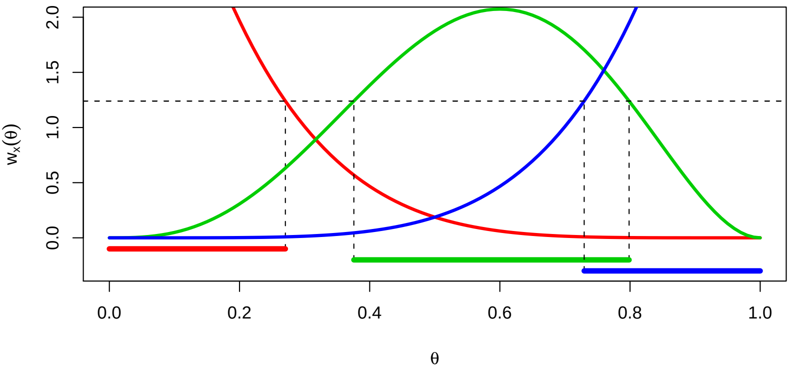
Illustration



Illustration



Illustration



Key Point

EACH STEP REQUIRES $\rightarrow$ SINGLE ROOT FINDING ALGORITHM

1. For fixed y and each x get $l_x(y)$ and $u_x(y)$
2. Increase y until $\alpha(y^*) = \alpha$

Extention

The previous theorem assumes a solution to $w_x(u) = y^*$ exists.

Theorem

If no solution to $w_x(u) = y$ exists, take $l_x = u_x = \theta_0$. This yields the unique MMELE.

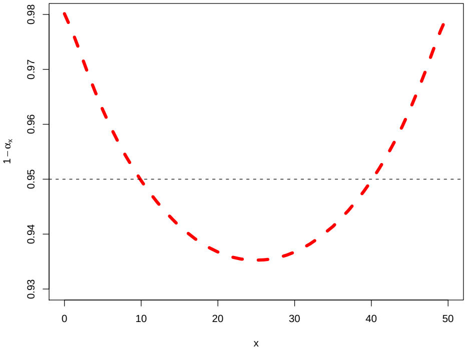
- Scenario arises when w_x is "flat"

Operating Characteristics

Set up

- $X \sim Bin(50, \theta)$
- We will consider three scenarios
    1. **Minimum length Bayes Interval (α_x = .05)**
    2. **MMELE with Uniform Weights**
    3. **MMELE with beta(2,1) weights - triangle shape**

α allocation: MMELE vs. Bayes



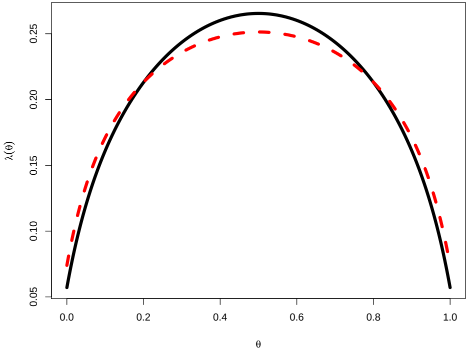
α borrowing

Expression in proof: $\frac{d}{d\alpha_x}\lambda(A_w(x;\alpha_x)) = constant$

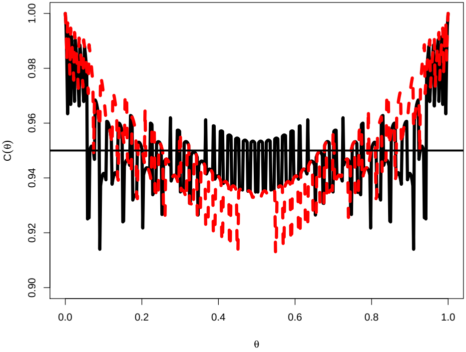
- When $X/n \approx 0$ or 1 small standard error
  - unit decrease in $1-\alpha_x$ **small decrease in length**
- When $X/n \approx .5$
  - unit decrease in $1-\alpha_x \rightarrow$ **large decrease in length**

α **borrowed** from inefficient intervals and spent on more efficient intervals

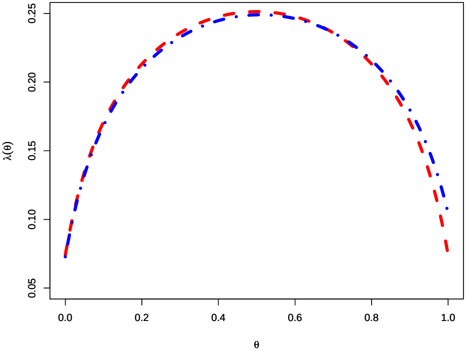
Length: MMELE vs. Bayes



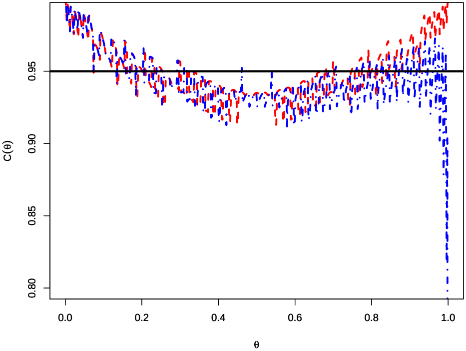
Coverage: MMELE vs. Bayes



Length: Weighted vs. Unweighted



Coverage: Weighted vs. Unweighted



Women in Irish Prisons

Allwright et. al (2000)

- $n = 57$ women in Irish Prisons
- $x = 24$ with hepatitis C infections

Weights/prior

- $w^{(1)} = beta(\cdot; 1, 1)$
- $w^{(2)} = beta(\cdot; 1, 2)$
- $w^{(3)} = beta(\cdot; 2, 4)$

Results

Table: Irish prisoner data. MMELE and Bayesian intervals for three weight functions.

w	Procedure	Interval	Length
$w^{(1)}$	MMELE	(0.306, 0.542)	0.236
$w^{(1)}$	Bayes	(0.301, 0.551)	0.249
$w^{(2)}$	MMELE	(0.300, 0.534)	0.233
$w^{(2)}$	Bayes	(0.296, 0.543)	0.246
$w^{(3)}$	MMELE	(0.297, 0.530)	0.233
$w^{(3)}$	Bayes	(0.295, 0.536)	0.240

Some Remarks

Recap

Coverage and expected length not computable

$\downarrow$

mean coverage and **mean** expected length are computable

$\downarrow$

precise definition of "nonconservative" and "optimal".

Properties of the MMELE

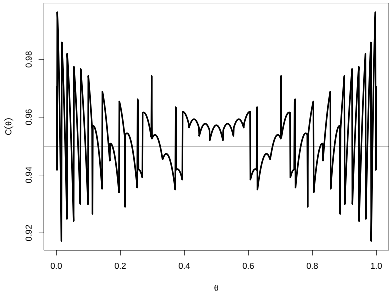
- User specified weight function $\rightarrow$ flexible
- Easy to compute
  - Single root finding algorithm
- α allocation

Extensions Needed

- We consider single parameter setting
  - Other Applications: odds ratio, relative risk, poisson mean
- Nuisance parameters?
- Asymptotics?
- Other merit criterion?

Example

What merit criterion is considered here?



On nonconservative inference

- $X \sim N(\mu, 1)$
- $H_0 : \mu \in [-1, 1]$ vs. H_1 : not H_0
- Decision: Reject H_0 if $|X| > k$
- Choice of k?
  - Conservative: $\sup_{\mu \in [-1,1]} \Pr_\mu(|X| > k) = \alpha$
  - Nonconservative: $\int_{[-1,1]} \Pr_\mu(|X| > k) d\mu = \alpha$

QUESTIONS